



universität
wien

DIPLOMARBEIT

Titel der Diplomarbeit

„Evaluating the CLIL programme
of a secondary school in Austria“

verfasst von

Fabian Szymanek

angestrebter akademischer Grad

Magister der Philosophie (Mag.phil.)

Wien, 2013

Studienkennzahl lt. Studienblatt: A 190 344 353

Studienrichtung lt. Studienblatt: UF Englisch UF Spanisch

Betreuerin: Ao. Univ.-Prof. Mag. Dr. Christiane Dalton-Puffer

Acknowledgements

Content and Language Integrated Learning is a concept that has fascinated me as a student at secondary school, at university and now as a teacher. I first came into contact with CLIL as a student when I had some History and Geography lessons in English. At the time I did not know that what was happening in class was CLIL. Moving on in my education, when I came to university I first learned about CLIL as a teaching approach and was able to connect my previous experiences at school with this concept. My most recent “CLIL-step” was when I started teaching Mathematics in English in a secondary school.

Learning foreign languages is something I have been familiar with for all my life, having spent my early childhood in Muscat, Oman and attending kindergarten as well as preschool there. My mother sometimes tells me how she brought me to kindergarten in Muscat on my first day and how for the first two weeks I came home crying because I must have had no idea what all the others were talking about. Everybody was speaking English and I had no idea initially, but after two weeks, so I am told, I came home babbling along in English. Since then, and most committedly in upper secondary I have taken up French, Spanish, Italian and Russian and enjoyed every minute of understanding how language works. Pretty soon I developed a passion for sharing that knowledge and trying to make other people see how great it was to speak different languages.

Completing this thesis has been a task that at some point seemed insurmountable. Many people, be that friends or family, had told me to not take up a teaching job next to university but I had to have it my way. The result is that this thesis is now complete exactly two years after requesting permission to work on the given topic. Good things take while, they say.

This shows very basically that it was not always easy, and that I was not always successful when trying to balance the demands of this thesis and the challenges of everyday life. However, at this point of completion I am indebted to a number of people for their support.

Firstly and foremostly, I would like to thank my thesis advisor Prof. Christiane Dalton-Puffer who has been a part of this process since the very beginning, its moment of inception. While I already knew that my thesis would have something to do with Content and Language Integrated Learning (CLIL), I was unsure what exactly the scope of my thesis would be. From that starting point, Prof. Dalton has provided guidance and support all the way, through

feedback, skype conversations and understanding when I had to move my personal deadline yet once again for whatever reason. Thank you! I always walked away from our conversations feeling positive that I would be able to do it after all.

Given the fact that this thesis examines the CLIL programme of the school I work at I am also very grateful for the help of all involved – my colleagues, students and headmaster who agreed to become a part of this study through interviews and providing information and useful conversations on the development on CLIL at that particular school. I am also grateful for the moral support that I have received and the help from colleagues when there was a little less time for school available from me, also for the understanding and care that I have felt when sometimes I could not complete some marking on time or when colleagues and students and even some of the parents alike would ask me how things were going and motivate me to keep on going and see the light at the seemingly endless tunnel.

Special thanks go to Bernhard Prentner and Erika Schneider who I am immensely thankful for transcribing all the interviews and thereby saving me a lot of time.

I am also heavily indebted to a number of good friends who mustn't go unmentioned for their undying support, their shoulders to lean on and patience with me when I couldn't always make time for them because of the work that had to be done. Also I thank them for their time they took to listen to my ranting and continuously encourage me to make sure I wouldn't just give up 😊 – thank you Irene Höllwerth, Sarah Übleis, Vanessa Walzl, Julia Skala & Miriam Soltész.

More special thanks have to go to my flatmate – Angela Kirchner – who witnessed the “growing pains” of this thesis and was exposed to most of the negative consequences this process had on me. Thank you for your patience and support throughout.

Finally I would like to thank my parents, Astrid Haiger and Marek Szymanek, as well as Debra Forsyth for their support throughout the years. It was very much through their help, continuous guidance and sacrifice that made it possible for me to be where I am now, for allowing me to be who and what I want to be and to pursue the career and passion that I wanted to and for providing the necessary moral and financial support but also exercising some pressure that would make me move forward on this sometimes seemingly endless project! Thank you!

TABLE OF CONTENTS

1. INTRODUCTION	1
2. CONTENT & LANGUAGE INTEGRATED LEARNING (CLIL).....	4
2.1. DEFINING CLIL.....	4
2.2. HOW CLIL? WHY CLIL? – PURPOSES AND BENEFITS OF CLIL.....	7
2.3. CLIL IN AUSTRIA.....	13
2.4. IMPLEMENTING CLIL IN SCHOOLS	19
2.4.1. Pre-launch essentials.....	19
2.4.2. Students.....	20
2.4.3. Schools	21
2.4.4. Teachers.....	22
2.4.5. Curriculum	23
2.4.6. Partners.....	23
2.4.7. Parents.....	24
2.5. CRITICISM	24
3. THEORY OF EVALUATION.....	27
3.1. INTRODUCTION TO THE CONCEPT OF EVALUATION – DEFINITION & BASIC TERMS	28
3.1.1. Definition(s).....	28
3.1.2. Purpose of Evaluation	30
3.1.3. Types of evaluations.....	31
3.1.4. Stakeholding in evaluation.....	31
3.2. UNDERSTANDING DIFFERENT TYPES OF EVALUATION – HISTORICAL DEVELOPMENT	33
3.2.1. The Tylerian model & first reactions	34
3.2.2. Stake's "countenance model" or "responsive model".....	37
3.2.3. Standards.....	40
3.2.4. Illuminative evaluation & Triangulation.....	41
3.2.5. Lawrence Stenhouse	43
3.2.6. 4 th generation evaluations	44
4. SETTING THE SCENE: THE SCHOOL, THE PROGRAMME, THE RESEARCHER	49
4.1. THE SCHOOL & THE CLIL-PROGRAMME.....	49

4.1.1. <i>Organisation and goals of the CLIL-programme</i>	51
4.1.2. <i>Funding for the programme</i>	53
4.1.3. <i>Admission to the EAA-programme</i>	54
4.2. APPROACHING INTERNAL EVALUATION & VALIDITY – NEGOTIATING THE TWO ROLES OF THE RESEARCHER	56
5. METHODOLOGY	65
5.1. TURNING THEORY INTO PRACTICE – DEFINING THE EVALUATION.....	65
5.2. THE INTERVIEW QUESTIONS	70
5.3. THE INTERVIEWEES.....	74
5.3.1. <i>Teachers</i>	74
5.3.2. <i>Students</i>	75
5.3.3. <i>Parents</i>	75
5.4. DATA ANALYSIS	76
6. FINDINGS.....	78
6.1. FINDINGS BY STAKEHOLDER GROUPS	78
6.1.1. <i>Students</i>	78
6.1.2. <i>Parents</i>	81
6.1.3. <i>Teachers</i>	84
6.2. RESULTS BY CATEGORY	90
6.3. AGENDA FOR NEGOTIATION & REFLECTION.....	93
7. CONCLUSION	95
8. REFERENCES	98
8.1. GRAPHS, PICTURES	98
8.2. BOOKS & ARTICLES & REFERENCE WORKS	98
8.3. ONLINE RESOURCES, WEBPAGES MENTIONED.....	102
APPENDIX.....	103

1. Introduction

The concept of Content and Language Integrated Learning (CLIL) is not a new idea anymore in the Austrian education system. The basic idea thereof has been implemented in a number of ways so far, especially as schools, educators, parents and students are continuously realizing that in a globalized world there is no serious educational path anymore that does not involve the teaching of communicative skills and therefore the teaching of communication in the world's still singular world-wide lingua franca: English. In Austria the subject English as a foreign language has been a part of compulsory education for decades now and the idea that even preceded CLIL, namely that of using English as a medium of instruction in more subjects than just the subject of English itself, has also become increasingly widely spread for many years.

Content and Language Integrated Learning is an approach in language teaching that has gained great popularity across the world and is being used in countless classrooms in all settings of education – be that primary, secondary or tertiary education, regardless of the focus a school or educational institution may have. However, it does require a closer look at the theory behind it since it does not simply comprise the idea of conducting as much teaching as possible in a foreign language, but claims to be something more powerful.

In Austria the “trend” or “movement” of introducing concepts similar to CLIL next to standard instruction of foreign languages in the regular curriculum is generally not more than two decades old and the ways these have been implemented vary greatly. Also the support and recognition these programmes have been receiving varies.

The thesis at hand examines one of these “incarnations” of CLIL in an Austrian secondary school in lower Austria. The term “incarnation” is used deliberately here for the simple reason that there are many forms of what is perceived, understood and implemented as CLIL but it remains to be seen which of them really adhere to the true definition of this teaching approach.

Chapter 2 will be devoted to defining and understanding the concept of Content and Language Integrated Learning and its motivations as well as providing a short theoretical background that positions CLIL in the study of language teaching and teaching in general.

Furthermore some light will be shed on the current situation of CLIL in Europe and particularly, as far as possible, in Austria.

The school is located just outside of the Austrian capital Vienna in an area that is considered one of the wealthiest in the country. I have been teaching at this school for a little over two years now. When stating that this thesis will “examine” the CLIL programme that is run at this school, the means of examining referred to here is an evaluation.

Evaluation is a trend, at least in Austrian education that is definitely a lot more recent than the introduction of CLIL and similar concepts in Austrian schools. This thesis will also look at the concept of evaluation more closely, will examine the history of evaluation in order to provide a sound basis for the evaluation that is to be undertaken in the course of this research project. Defining the term “evaluation” for the purpose of this thesis and explaining which factors have shaped the development of evaluations will be an integral part.

What has to be understood immediately from the beginning of this research project is that it is motivated by one sole purpose, namely the one of doing this thesis. It is not an evaluation I was asked to do by any party or am paid to carry out. The evaluation here will therefore be what is called an internal evaluation. This is due to the fact that a participant of the programme studied and evaluated is doing the evaluation “from the inside” of the programme. This perspective within an evaluation gives rise to a certain potential for criticism, namely whether or not the results of the study can still be deemed scientific. This issue will also be dealt with in the course of chapter 4 whose purpose it is to describe the scene of the evaluation: the school and the programme but as well, in a second part, the role of the researcher within the programme will be discussed in order to provide a complete picture and to address aforementioned room for criticism.

In a next step (chapter 5) the procedure of the evaluation will be explained, and reflected upon and thereby a connection will be established to the chapter on the theory of evaluation. The aforementioned purpose will be achieved by following an approach that is the result of the study of the history of evaluation. Drawing from a number of different scholars in the field a meaningful way of arriving at an evaluative process will be determined. The process of gaining evidence will be through qualitative interviews with key participants of the programme on the basis of which conclusions will be drawn. The most prominent approach in this undertaking is called “Fourth Generation Evaluation”. Its

advocates, Guba & Lincoln (1989) acknowledge the power of the individual within the construction of knowledge and do not believe in one “single, universal, objective truth” (Kiely & Rea-Dickins 2005: 40). This idea is what motivates the methodology of this study.

The main purpose is to then put the accumulation of viewpoints and opinions expressed into perspective and in relation with each other and to identify what the participants view as strengths and weaknesses, to find commonalities and differences in perception between the parties involved and analyse these.

It will therefore not be the purpose of this study to determine how effectively the CLIL programme of this school is running through quantitative measures and if the students’ language proficiency in English is really increasing and expanding.

Chapter 6 will then proceed to making sense and constructing meaning of the data gathered in the interviews. Categories will be established, different viewpoints will be assessed and juxtaposed in order to make all the voices part of a picture that intends to be as wholesome as it can. Finally, it will make careful suggestions for a possible future development of this specific programme at this specific school.

To quote from Thomas Jefferson who wrote in a letter to William Charles Jarvis in 1820:

I know of no safe depository of the ultimate powers of the society but the people themselves; and if we think them not enlightened enough to exercise their control with a wholesome discretion, the remedy is not to take it from them, but to inform their discretion.

Part of the purpose of this thesis is to inform people’s discretion – my own, as a practitioner within the programme and maybe the one of other individuals involved in the project.

2. Content & Language Integrated Learning (CLIL)

The chapter at hand will study in more detail the concept of Content and Language Integrated Learning (CLIL). In a first step the aim is to provide a definition of CLIL, point out differences between CLIL and related concepts and ideas and to briefly trace the development of this teaching approach before moving on to giving an insight into the situation of CLIL in Europe and particularly in Austria.

2.1. Defining CLIL

In the following the concept of CLIL will be explained in more detail, with some attention paid to addressing key differences between other similar language programmes. Starting out from a currently very widely used definition certain aspects that define CLIL will be discussed.

The most frequent definition that I have come across during reading on CLIL is the one given by Do Coyle, Philip Hood and David Marsh (2010: 1):

Content and Language Integrated Learning (CLIL) is a dual-focused educational approach in which an additional language is used for the learning and teaching of both content *and* language. That is, in the teaching and learning process, there is a focus not only on content, and not only on language. Each is interwoven, even if the emphasis is greater on one or the other at a given time.

Furthermore they refer to CLIL as “[...] an approach, which is neither language learning nor subject learning, but an amalgam [or fusion] of both and is linked to the processes of convergence” (2010: 4).

What this means is that in a CLIL lesson there will ideally be two goals – honing a language skill and teaching content (e.g. Geography, History, etc.) at the same time. A good example would be a History lesson that introduces and discusses the Roman Empire and fosters the learners’ correct use of the Past Tense. The important thing to point out at this stage is that a simple teaching of a content subject in an additional language is not sufficient when the aim is to achieve CLIL teaching.

Regarding the languages used in a CLIL-setting Dalton-Puffer (2011: 183) points out that in general “[...] languages tend to be recruited from a small group of prestigious languages [...]” and that English is most commonly used as a CLIL medium. She even goes as far as somewhat renaming CLIL to “CEIL – content-and-English integrated learning” (2011: 183).

In reaction to the rising trend of multilingualism a number of educational settings have been created whose aim it is to provide the necessary foreign language input but not all of them are CLIL. There have been a number of

“[...] bilingual educational approaches such as Canadian immersion programmes, content-based instruction or other programmes that involve the use of regional minority or heritage languages as medium of instruction.” (Llinares et al. 2012: 1)

Llinares, Morton and Whittaker (2012: 2) point out that there has been some difficulty in understanding the differences between the terms “immersion” and “CLIL”. In their book they make a point of carefully differentiating between these two ideas and (ibid: 2) claim that they differ in areas such as: “language of instruction, teachers, starting age, teaching materials, language objectives, inclusion of immigrant students and research”.

Regarding language of instruction Llinares, Morton and Whittaker maintain that CLIL programmes will be conducted in a foreign language that is not a part of the students’ daily lives, i.e. “[...] is not present in the students local communities” (ibid: 2), which stands in contrast to immersion settings. This also explains why very frequently CLIL teachers are not native speakers of the language used in CLIL instruction. With regard to language objectives Llinares, Morton and Whittaker point out that immersion programmes aim at a significantly higher level of foreign language competence than any CLIL programme while it is the aim of both to achieve “functional competence” (ibid: 2) in the foreign language. Furthermore it has to be understood that due to the fact that in CLIL teaching there is a dual objective (namely teaching a content element as well as a linguistic skill) most teaching materials will have to be adapted for CLIL purposes as there is only a very limited (and only slowly increasing) number of materials developed especially for a true CLIL purpose. In immersion contexts mostly genuine teaching materials from the foreign language will be used.

Other researchers such as Maljers, Marsh and Wolff (2007: 8) refer to CLIL as

[...] a generic term that covers some 20 or more educational approaches [...] [and that] [a]lthough these have differed in terminology used (immersion, languages across the curriculum, bilingual education, etc..) they share certain common methodologies. CLIL was introduced as an inclusive ‘umbrella’ term by which to capture and further develop these.

What seems to be the case is that the term CLIL has come to be understood as a title or common denominator for all educational settings that include teaching non-language subjects (content subjects) in a foreign language. This foreign language (most commonly

English) is not part of the students' communities and everyday lives. An example for that would be the situation of CLIL in Austria. In mainstream education the first language of instruction is German, the teaching of English as the first foreign language is omnipresent but English is not an official or necessary language in the system (cf. Dalton-Puffer 2011: 183). This is in sharp contrast to previously mentioned immersion models.

What remains from this short section is that CLIL is not (yet) fully defined and that this particular way of calling the many educational realities that are similar to CLIL term has not (yet) found its way into all educational settings. The definition seems to be widely used for scholarly purposes on a university level, which also creates a unified understanding for research purposes, but at individual schools the term is still relatively uncommon.

The many forms of CLIL and the lack of completely straight-forward and clear cut definitions and interpretations thereof for application in educational scenarios throughout the world is can also be attributed to the fact that every educational setting (different schools, different resources, different teaching staff, etc.) differs from another in some way. This also impacts on the form of CLIL.

However, a crucial part of the definition for me is the focus on the words “dual focus”, “integrated/integration” and “intervowen” (see definition above). Wolff (2007: 16) maintains that “[l]anguage is both content and medium in the CLIL classroom” and again this is what is important – a CLIL classroom is not about transmitting content in a foreign language. CLIL is an approach that unites, builds on and hones the faculties of a content subject learner and those of a foreign language learner. As Wolff puts it (2007: 19)

In CLIL the learner's role as a foreign language learner and as a content subject learner merge. The learner learns the concepts and schemata of the content subject in a new language, i.e. the concept (or schema) and the linguistic items or structures designating it are acquired simultaneously.

And with this I want to lead over to the next section whose purpose it is to explain the purposes of CLIL and the theory behind it in more detail. After now approaching a definition of the term it seems logical to examine the driving theoretical forces behind it and motivations for it.

2.2. How CLIL? Why CLIL? – Purposes and benefits of CLIL

The illustration on the right is one that has commonly been adopted in research on the theory of CLIL. It shows the 4Cs of CLIL (context, culture, cognition and communication) according to Coyle, Hood & Marsh

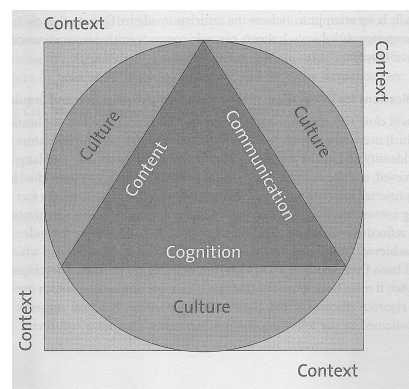


Figure 1 - The 4Cs of CLIL
Source: Coyle, Hood & Marsh (2010: 41)

(2010:41). To my mind the illustration demonstrates why one speaks of integration in CLIL, since the 4Cs are depicted next to but also within each other. This is due to the fact that CLIL is understood to be more than just the learning of specific content or learning of a particular foreign language. All CLIL learning is embedded in a certain context – context that is needed to make any kind of input meaningful. The very centre of CLIL would seem to be the link between content, cognition and communication (used here interchangeably with the term “language”). Learning a language cannot happen without meaningful context to use it for and context cannot be communicated without the necessary communicative (linguistic & cultural) skills. The importance of the word “integration” in CLIL becomes obvious.

Coyle, Hood and Marsh (2010: 41) maintain that

[i]t suggests that effective CLIL takes place as a result of this symbiosis, through:

- progression in knowledge, skills and understanding of the content;
- engagement in associated cognitive processing;
- interaction in the communicative context;
- development of appropriate language knowledge and skills;
- the acquisition of a deepening cultural awareness, which is in turn brought about by the positioning of self and ‘otherness’

Using this as a basis I would like to comment on some of the theoretical background to CLIL as presented by various researchers. CLIL – as indicated previously – has a dual focus and in order to fulfil that it takes an understanding of the faculties of both a language learner and a content subject learner and what these are grounded in. With that it is easier to understand how these can be combined in order to ensure the desired effect.

In the following I would like to name a number of benefits of CLIL that researchers have identified in recent research or which they derive from learning theories. Dale and Tanner (2012: 11-13) list quite a few, I have chosen the most significant ones to be represented

here. Wolff (2007: 21f.) also provides a list of beneficial outcomes with a slightly different focus or orientation. I do not, however, want to merely present the outcomes suggested by Dale, Tanner and Wolff but also complement these with relevant theory. The purpose here is to explain how some of these outcomes are achieved and also to present some of the theoretical background that CLIL researchers draw from when justifying the practises of CLIL.

Wolff claims that CLIL learners tend to be better foreign language learners and better content learners due to the fact that they “process” both content as well as the foreign language “more deeply” (2007: 22). If content is learnt through another language, it is cognitively more demanding and this is highly conducive to the previously mentioned idea of deep learning. CLIL learners are surrounded by the foreign language for more time and more regularly than non-CLIL learners which Wolff points out as a further benefit. It has to be mentioned at this point that researchers such as Dalton-Puffer have pointed out that CLIL students do not achieve higher language skills and competences plainly through increased exposure to the foreign language (what Dalton-Puffer refers to as the “language bath” which “[...] somehow stimulates the individual learning process” (2007: 3)) but that the idea of language and teaching language objectives is much more complex. However, what Dalton-Puffer does point out is that regarding the language skills of CLIL students the receptive ones are more affected than others (cp. 2008: 143). In other words, language skills such as listening and reading as well as vocabulary (that includes to a certain extent morphology) are those that are found to increase a lot more under CLIL influence than others (e.g. syntax, writing, pronunciation, to name a few). This of course suggests that exposure to the foreign language is beneficial but it is not what defines CLIL.

Wolff (2007: 22) sees CLIL students to be better prepared for their later professional life since CLIL presents students with more relevant academic language from contexts that are more useable, CLIL will have acquainted them with a broader range of topic-related vocabulary. Furthermore CLIL teaching promotes skills and strategies such as strategies for (autonomous) problem-solving, teamwork and presentation techniques. I do, however, feel that skills as these can be very easily acquired in a standard language classroom as well as in a good content classroom, depending on the teacher, and are therefore not unique to CLIL students.

“CLIL learners develop cognitively and their brains work harder” is a benefit brought up by Dale and Tanner (2012: 11) and this is based on cognitive learning theories. This is a topic I would like to address in some more detail. When researching the theoretical background of CLIL two learning theories play an important role and are mainly brought forward – constructivist and participatory learning theories. These obviously do not only apply to CLIL, but to content learning as well as foreign language learning separately and are the basis and foundation of many more specific theories, such as in the field of second language acquisition.

Constructivism in an educational context suggests “[...] that learning is an active process in which learners construct new ideas or concepts based upon their current knowledge state” (Dalton-Puffer 2007: 7). The key and necessary first step to successful teaching therefore is “[...] to present the material to be learned in such a way that it matches the learner’s current state of understanding and encourages students to discover principles by themselves” (ibid: 8). The idea that new knowledge can be constructed on the foundation of already existing knowledge ties in very closely and is expanded a little further by what Vygotsky (in Dalton-Puffer 2007: 9) refers to as the “zone of proximal development”. This comprises the line of thought that for every individual student it will not be possible to learn just anything at a certain time, but what can be learned is based on the current knowledge. It also “describe[s] the kind of learning which is always challenging yet potentially within reach of individual learners [...]” (Coyle et al. 2010: 29). This means, in a CLIL context, that at a certain stage it will not be possible for a student to usefully acquire and retain a certain language feature (e.g. past perfect tense) in the course of a CLIL lesson, when he/she is not sufficiently, cognitively ready for it. The same will apply to any particular content. In CLIL this means that at times the focus will have to lie on providing more content information at a certain time while reducing the difficulty of the language and vice versa. What CLIL researchers claim to have shown, however, is that students are much more cognitively engaged because they are processing new content through a foreign language. Coyle, Marsh & Hood (2010: 29) suggest that a cognitive challenge, which will lead to deeper learning, is made up of tasks that demand what Bloom’s taxonomy calls “higher-order thinking skills”. These comprise any task that requires students to analyse, evaluate or create, i.e. individually using new knowledge (linguistic or content) in a creative way. These skills are in opposition to what is called “lower-order thinking skills” such as remembering, understanding and applying (cp. Ibid: 31).

What Coyle, Marsh & Hood suggest in this context is that thinking (cognition) is a part of CLIL teaching that has to be planned well in order to achieve a high level of “cognitive engagement” through tasks that provide a balance between higher- and lower order thinking skills (“thinking curriculum” – Coyle et al. 2010: 30).

The second main theory that a substantial part of CLIL theory (and language learning theories in general) is based on is “Vygotsky’s theory of socio-cognitive development” which comprises the “[...] idea that social interaction plays a fundamental role in the development of cognition [...]” (Dalton-Puffer 2007: 8). What this means is that knowledge cannot be successfully constructed without social interaction. This in turn means that building new knowledge is virtually impossible without language, since that is an almost universally vital necessity for interaction. This shows an essential link between language and cognition – put very simply and bluntly: without language no (or hardly any) communication, without communication no (or little and less effective) construction of knowledge.

“CLIL learners’ language progresses more” (Dale & Tanner 2012: 12) – in CLIL the L2 has to be used to articulate the new content. This is cognitively more demanding than in an L1 and therefore encourages deeper learning and processing. “Learners who spend time focusing on *how* language is used (form), as well as *what* is being said (meaning) also progress faster in learning a language [...]” (ibid: 12). In CLIL, students do not only learn to use a language, they also use it to learn (content) and that can often lead to a disjuncture “between their levels of cognitive functioning and linguistic competence” (Coyle et al. 2010: 35). This means that it can happen (and in fact it is likely to) that students understand the content but lack the language skills to express and articulate said content. For a CLIL teacher that means that sometimes linguistic skills will be required in the foreign language that have yet to be covered in the standard foreign language classes. In order to remedy this situation a CLIL teacher therefore has to plan the content objectives in accordance with the language objectives (cp. Coyle et al. 2010: 36). According to Coyle, Marsh and Hood this is only possible if the CLIL teacher distinguishes between the following three concepts: “*language of learning*”, “*language for learning*” and “*language through learning*”.

- *Language of learning* is an analysis of language needed for learners to access basic concepts and skills relating to the subject theme or topic.
- *Language for learning* focuses on the kind of language needed to operate in a foreign language environment.

- Language *through* learning is based on the principle that effective learning cannot take place without the active involvement of language and thinking. When learners are encouraged to articulate their understanding, then a deeper level of learning takes place.

(Coyle, Marsh & Hood 2010: 37)

Understanding these different conceptions of language allows for effective preparation of CLIL teaching and compensates for the aforementioned disjuncture between the level of understanding of the content matter and the level of ability to express this in the foreign language. Students need language structures to voice and apply the content matter (language *of* learning) and they also need to be given linguistic material that allows for interaction in the classroom situation (language *for* learning). This highlights the interdependency of language and content.

The concept of language *through* learning is also strongly encouraged by Dale & Tanner (2012: 5) who suggest that “[...] subject teachers pay attention to both language and content in their lessons, to help learners learn both language and content as they learn a school subject.”

This leads me to briefly discuss the role of **content within CLIL**. Since in CLIL, as opposed to the Communicative Language Teaching (CLT) approach, it is an objective for students to also learn and process concrete content rather than just practising the language, content learning activities have to be tailored in order to allow the practise of a language skill too. Content therefore is what the language is used for and provides a “scaffold for the language learning process” as Wolff argues (2007: 20). He maintains that “[a]cademic content is easier to define than everyday content and therefore easier to handle with respect to language learning”. This again leads to the aforementioned deeper learning and processing of both language and content and a great part of what makes CLIL such a successful teaching approach. Wolff points out (2007: 21) that there is still very little research on measurable effects of CLIL but what he makes reference to (Lamsfuß-Schenk, 2002) indicates that

concepts and schemata seem to be more precise when learners express themselves in a foreign language, and the language items they choose are usually situated on a high academic level. The reason for this has to do with the fact that content is processed more deeply by the learners when it is in a foreign language [...] [and] [...] input and output are in the same language.

Researchers such as Vollmer (2006) argue that CLIL-students outperform non-CLIL-students on a content level due to the fact that “[...] CLIL students work more persistently on tasks

and show a higher tolerance of frustration [...]", which promotes "intensified mental construction activity [...] resulting in deeper semantic processing and better understanding of curricular concepts." (Vollmer 2006 in Dalton-Puffer 2011: 188).

This counters a very common concern, especially among parents and non-language teachers, but also among critical researchers (Airey: 2009; Airey & Lindner: 2006), that CLIL causes deficiencies with the content subject matter and that the use of a foreign language is considered an obstacle in the learning of the content subject matter. These shortcomings on the part of the content matter are attributed to the aforementioned disjuncture because it is then understood that the content has to be simplified in order for it to be processed in the foreign language.

In this context I would like to briefly mention some of the research carried out into the question of outcomes in CLIL, taking a closer look at those studies looking at the outcomes on a content level rather than a language level. Dalton-Puffer (2011: 188) maintains that conducting research whose goal it is to determine whether CLIL is beneficial not only for foreign language skills but also for content knowledge has proven to be rather difficult. This is due to the fact that "[...] relatively few countries conduct standardized testing in science and social studies subjects." (ibid: 188). This makes it hard to draw comparisons between CLIL learners and non-CLIL learners. There is a small number of studies that show varying results when it comes to content learning. Dalton-Puffer mentions three different studies (Admiraal et al: 2006, Jäppinen: 2005, Badertscher & Bieri: 2009) all of which either proving slight advantages, slight disadvantages or no difference at all between CLIL and non-CLIL students in performance checks on content subject matter. Dalton-Puffer therefore raises the question "how it is possible that learners can produce equally good results [on a content level] even if they studied the content in an imperfectly known language?" (Dalton-Puffer 2011: 189). There are two research findings that I consider relevant in answering these questions and both are derived from studying CLIL classroom interaction. The first is one found by Nikula (2010) who claims that the fact that in CLIL teaching teachers are mostly non-native speakers of the foreign language has the consequence that learners "[...] claim a larger share of the discourse space" (Dalton-Puffer 2011: 190). This means that learners in a CLIL classroom tend to be able assume a more prominent role in class as in their L1 subject teaching and therefore are able to process the content more deeply, simply because they

get to talk more. This ties in with what was previously said about the construction of knowledge through language. The second research finding that may provide an answer to the question from Dalton-Puffer is closely related to Nikula's (2010) and originates from Badertscher's and Bieri's research undertaking (2009). Badertscher and Bieri (2009) compared CLIL and mainstream teaching and discovered that not only are the number of negotiation sequences in CLIL classes is double the number of those in mainstream teaching but also that teachers in CLIL lessons spend more time providing help and assistance when they notice that students have difficulties understanding the content (cf. Dalton-Puffer 2011: 191).

"CLIL learners develop **intercultural awareness**" (Dale & Tanner 2012: 13). This brings the fourth element, the fourth "C" into the picture – culture. Culture is an aspect that cannot be ignored in language teaching. Based on the theories of cultural theorists such as Vygotsky "language, thinking and culture" are constructed through social interaction. Coyle, Marsh & Hood (2010: 39) argue that "[i]f we follow the idea that culture determines the way we interpret the world, and that we use language to express this interpretation, then CLIL opens an intercultural door [...]"

This section was intended to present some theory behind the concept of CLIL as well as to discuss its potential benefits. It has to be pointed out at this point, that as Wolff (2007) claims, the research into CLIL results that allows comparison is still rather rare and, as has already been and will also be very briefly shown at a later stage, there is also criticism expressed towards CLIL for a number of reasons.

After a short presentation of relevant theory and benefits of CLIL I will now continue with an examination of the situation of CLIL in Austria.

2.3. CLIL in Austria

This section will briefly examine the situation of CLIL in Austria. The data available on CLIL is limited and is drawn from two main sources. The first of these sources is the Austrian Centre for Language Competence in Graz (formerly known as the "Zentrum für Schulentwicklung, Bereich III (ZSE III)" (rough translation: centre for development in education) which was founded in 1991 and has been monitoring the development of CLIL-related programmes in Austria and – in a very limited manner – has been trying to be a somewhat unifying force for

the Austrian CLIL movement. The second source is the Eurydice Network, which is an institution funded by the European Commission whose major goal it is to “[...] provide[s] information on and analyses of European education systems and policies.”¹

Both theses sources maintain that the general idea of CLIL has been around and actively sought in Austria since the early 1990s. Increasingly, parents were voicing a strong interest and also demanding a solid training and education in foreign languages, especially in English.

A report issued by Eurydice in 2005 takes a closer look at the situation of bilingual (and therefore CLIL-like) programmes in a number of European countries in order to examine their implementation and mentions another reason that triggered the movement in Austria and that is the fact that due to an “expected influx of English-speaking children whose parents were employed in international organisations and companies moving to Vienna” (Eurydice 2005: 6) schooling had to be provided by the (Viennese) Board of Education to accommodate these students. In response Vienna Bilingual School (VBS) was founded whose main idea it was to educate their students in both German & English in equal amounts.

With that (VBS) as a basis and an inspiration as well as medium that raised awareness a number of forms of CLIL have emerged over time. The following list provides a brief overview of definitions and intentions of these programmes:

EAA/FAA: Englisch als Arbeitssprache (= English as a working language) which is a special form of FAA (= Fremdsprache als Arbeitssprache). The acronym denotes any form of teaching where a subject is taught in a foreign language, which is most commonly English in Austria.

EMI “English as a Medium of Instruction”: is used interchangeably with the acronym EAA

EAA seeks to achieve functional proficiency in the foreign language, enabling learners to communicate in it on topics appropriate to their age group, while also mastering subject content in accordance with the curriculum. (Eurydice 2005: 3).

Further objectives laid out for EAA/EMI by the Zentrum für Schulentwicklung (ZSE III) are optimal preparation for demands of later career choices/jobs as well as increased reflection on the usefulness of a foreign language through its application in a content subject. (cp.

¹ <http://eacea.ec.europa.eu/education/eurydice/> (accessed Mar. 22nd, 2013)

Oestreich & Grogger 1997: 5). Eurydice (2005: 3) by referring to Mewald (2004: 47) put it as follows:

EAA 'conceives of language as a tool that can be employed to teach subject-specific content, by temporarily merging content teaching and language learning. The use of the terms 'content teaching' and 'language learning' implies that the organisation of lessons should promote conscious subject tuition leading to conscious learning of content and, at the same time that the foreign language should be learnt'.

This clearly separates the teaching of content from the teaching of language or rather it defines an EAA lesson as a lesson where both these forms of learning co-occur.

EAC "English Across the Curriculum" is often understood as the idea of delivering English language skills (in varying levels of difficulty) in nearly all subjects a school offers, which often makes it seem like a whole-school approach that is "in support of a cross-curricular approach of foreign language input and intellectual networking" (Mewald: 2004, 50 in Eurydice 2004: 3).

DLP "Dual Language Programme" is described as a situation where "mainly German-speaking students [are] taught subject content in the target language, predominantly by qualified German-speaking teachers (sometimes supported by target language mother tongue teachers) for clearly defined periods of time" (Eurydice 2004: 4/24).

CLIL: this acronym tends to be used on a university or teacher-training level for the simple reason that it serves as a common denominator which is widely used and understood across Europe

Bilingual: referring back to the Vienna Bilingual Schooling programme (an initiative which has inspired 3 other bilingual schools in Austria so far) it has to be pointed out that there is a noticeable difference between bilingual schools and DLP (EAA, etc.). While in bilingual schooling there is a roughly equal amount of native German speaking and native English speaking students in every class being taught all their subjects in both German and English in relatively equal amounts, in schools with DLP (EAA, etc.) programmes in place there are classes of predominantly German native speakers or German L2 speakers being taught certain content subjects in a foreign language (mostly English, some French) for a much more limited period of time.

It has to be noted at this point that the programme evaluated in the course of this thesis obviously belongs to the group of the latter programmes.

When trying to gather information about CLIL in Austria one might be somewhat startled at how little information is provided by the Austrian Ministry of Education itself. Dalton-Puffer points out (2011: 184) that on a European level CLIL is mainly driven by two very different entities, one of which being “high-level policy-making” and the other being “grass-roots activities” which, as Dalton-Puffer (ibid: 184) puts it “[...] dovetail[...] parental and teacher choices”. In Austria, so it would seem and the following paragraphs intend to show that, it is mainly the latter of the two driving forces mentioned above that is regulating or moving CLIL teaching forward or in Dalton-Puffer’s (ibid: 185) words “[...] the impetus [is left] to the grassroots stakeholders.”

In an approach to examine the situation of CLIL in Austrian classrooms from the perspective of an applied linguist Dalton-Puffer (2002: 5) points out that “[...] there is very little information about what the ‘local conditions’ actually look like that would go beyond the anecdotal.” This is due to the fact that the boards of education of the 9 respective provinces of Austria make varying effort to track and make records of the amount and implementation of CLIL happening in classrooms since the 14th amendment (14.SchOG-Novelle) to Austrian educational law which grants every school a certain level of autonomy as regards the implementation of certain focal aspects (focus on natural sciences, on foreign languages, on sports etc.).

What a number of researchers (Dalton-Puffer, Gierlinger, etc.) point out is the lack of a nationally unifying movement or basis for all the CLIL activities in Austria and that there is surprisingly little cooperation between practitioners themselves as well as between practitioners and applied linguists. Gierlinger (2002: 89) speaks of so called “Insellösungen” by which he refers to all the activities resembling CLIL in Austria as insular (and therefore badly interwoven). Dalton-Puffer (2002: 13) voices another concern by pointing out that “[w]ith regard to EaA, [...], a clear formulation of its objective(s) is still a desideratum.” There are a couple of general goals for any EAA programme but what I believe is still missing (more than 10 years later) entirely are the language objectives – a list of linguistic skills and also the basis for any of these objectives in second language acquisition theory (hence the missing link between practitioners and applied linguistics).

It has become clear so far that there are essentially two different types of CLIL provision in Austria. The Eurydice report (2004: 5) classifies them as follows:

a) “Mainstream school provision”: DLP/EAA/EAC programmes are worked into the proceedings in a school in diverging quantities and form consisting of a spectrum spanning from only a few ‘mini-projects’ carried out in a content subject with English (or other foreign languages which will be specified below) to intensive regular teaching of a content subject in the foreign language. “CLIL lessons are usually timetabled as content lessons” (Dalton-Puffer 2011: 184) but the formal instruction of the L2 as a distinct subject is maintained. The languages used are, depending on the region, either “non-indigenous” languages such as English, French, Italian and Spanish, or “minority/regional” languages that are officially recognised in Austria such as Slovene, Croatian, Czech, Hungarian and Slovak. These are almost exclusively taught in the areas close to the borders between Austria and the respective countries. In many cases the schools or the boards of education employ native speakers of the target language to aid this process.

b) “Pilot projects”: truly bilingual teaching in bilingual settings (VBS and similar schools in two other locations)

In Austrian educational law the legal basis for the implementation of CLIL in mainstream education is §16 section 3 of the Schulunterrichtsgesetz (SchUG 1986), parts of which are represented here:

[...] kann die Schulbehörde erster Instanz auf Antrag des Schulleiters [...] die Verwendung einer lebenden Fremdsprache als Unterrichtssprache (Arbeitssprache) anordnen, wenn dies [...] zur besseren Ausbildung in Fremdsprachen zweckmäßig erscheint und dadurch die Zugänglichkeit der einzelnen Formen und Fachrichtungen der Schularten nicht beeinträchtigt wird. Diese Anordnung kann sich auch auf einzelne Klassen oder einzelne Unterrichtsgegenstände beziehen.[...]

This provides one of the goals of CLIL-like projects, namely the higher level of education in a foreign language, which is sufficient reason for the board of education to grant any head of a school permission to allow a programme such as EAA, EAC, DLP, etc. The law also reflects the reason for the fact that there are so many different implementations of such programmes since it does in no way specify the amount of time or content subjects are actually taught in a foreign language.

The last point I want to raise in this section is one that seems very straightforward, namely how many Austrian schools participate in the movement and have implemented some form of CLIL programme. The answer to that question, however, is not so easily given for reasons mentioned above. In a report written by Margarete Nezbeda, who is a member of the

Austrian Centre for Language Competence (Österreichisches Sprachenzentrum), also known as the ZSE III (see above) it is one of her main intentions to provide an overview of “FsAA-Aktivitäten” (CLIL-like activities) in Austria” in 2005. After contacting the Austrian Ministry of Education for nationwide data on CLIL activities in schools the only thing provided was the information that in upper secondary schools with vocational foci (BHS) the number of students taught by CLIL had increased by 16,226% between the academic years of 2002/03 and 2003/04. However, she does then move on to providing extensive data acquired from the respective boards of education of the 9 Austrian provinces that lists all schools of a province. Maljers, Marsh and Wolff (2007: 18) clearly demonstrate the lack of current data on the issue by maintaining that

[o]n average about 15% of all Austrian Secondary schools provide a kind of CLIL instruction. The rate for lower secondary ‘Hauptschule’ [HS] is about 7%, for secondary academic schools [AHS] about 27%, and for vocational schools [BHS] about 30%. These numbers are estimates and based on a survey carried out in 1997.

This last quote, or rather the year mentioned in it (1997) again underlines what I previously called “somewhat startling”: the fact that the most comprehensive study of EAA nationwide was carried out in 1997 reflects a rather dire situation as far as educational statistics in Austria are concerned.

On a final note, the subjects most likely to be taught in the course of any kind of CLIL are geography, history and biology. Very frequently this is due to the fact that Austrian qualified secondary school teachers have a qualification for two subjects and combinations of a foreign language and a content subject are frequent, which in turn means that these teachers are more likely to be able to offer to teach in CLIL settings. Maljers, Marsh and Wolff (2007: 18) further note that there are efforts made to introduce CLIL into subjects like mathematics, physics and chemistry by providing special training for teachers “to meet the didactic and linguistic challenges”. On a primary level all the literature previously mentioned basically states that during the first two (of the 4) years of primary education foreign language instruction foreign languages are “[...] in a cross-curricular way, integrated into the compulsory subjects [...] in [s]hort sequences amounting to a total of one hour per week”. (Maljers et al. 2007: 16). During the second two years of primary schooling this concept is maintained “with additional focus on everyday communication”.

This section has shown that as Dalton-Puffer (2008: 139) puts that “[...] CLIL is still far from being a consolidated and fully articulated educational model in any of the European countries [...] and that it there is still considerable effort to be made in order “[...] to consolidate the theoretical underpinnings of CLIL and create a conceptual framework that is both coherent and applicable in local conditions.”

I believe that from a 2013 standpoint this theoretical research has evolved greatly – works by Coyle, Marsh, Mehisto just to name a few authors. The simple fact that upon entering the search word “CLIL” into a publication search on amazon.com one is presented with literature that dates back to 2007 demonstrates this. However, I feel that the theoretical research has not yet found much practical application that really draws from current CLIL research, at least in Austrian CLIL-like activities.

2.4. Implementing CLIL in schools

In the following section I want to present some findings of research into the actual implementation of CLIL projects at schools. Drawing from two main sources I want to provide some information on what researchers have found to be important steps to be taken when contemplating the implementation of a new CLIL programme at a school. It has to be noted that the lists provided are quite extensive, which is why I will provide an abridged version discussing those points I consider relevant for this thesis and the programme evaluated.

One of the two main sources is a report by Peeter Mehisto (2007) who refers to evidence from an Estonian CLIL programme. The other is a list provided by Margarete Nezbeda (2005: 18f.) whose basis is feedback from headmasters and teachers in Austrian schools.

Mehisto (2007: 62) lays out 7 categories that have to be considered before launching a CLIL programme: (1) pre-launch essentials, (2) students, (3) schools, (4) teachers, (5) curriculum, (6) partners and (7) parents. These categories will be adopted and both sources combined and tailored to what I consider Austrian needs or likely scenarios.

2.4.1. Pre-launch essentials

Some of the pre-launch essentials include ensuring that the managers are aware of the requirements and long-term consequences: there have to be enough teachers ready to take up the extra work that designing CLIL lessons requires. One also has to bear in mind that the

number of teachers necessary to maintain the programme will most likely increase over the years as more and more year groups become involved. It also takes thorough understanding of CLIL methodology, which in many cases will require in-service training for those teachers unaccustomed to the principles of CLIL.

On another level issues such as other schools in the district and the number of students within a school and those enrolling have to be considered. Mehisto points out that it is important to consider the fact that the number of students allows a two-track system (provision of a different option next to the CLIL programme) and on a slightly more macro-level bear in mind possible consequences that the setting up of a CLIL programme might have in the district as the CLIL programme could cause decreasing enrolment in other neighbouring schools (cp. Mehisto 2007: 62) and therefore lead to imbalances and lobbying against your school.

2.4.2. Students

Regarding the students two aspects have to be considered: the recruitment procedure and making sure that students stay part of the programme throughout its running. As far as recruitment is concerned Mehisto (2007: 63) recommends it be voluntary and that admission criteria be transparent and visibly adhered to during the process. Furthermore he points out that admission is normally either based on a first come first serve basis or on a lottery, both of which are normally accompanied by an interview with prospective students and their parents to “make certain that parents understand the long-term implications of programme enrolment” (Mehisto 2007: 63). The latter is a process that requires thorough work – providing parents with all the relevant information. One aspect Mehisto includes as a criterion is one that I feel is particularly relevant to this thesis and the CLIL programme being evaluated. He strongly urges prospective managers of CLIL programmes to make an effort to “[...] dispel the misconception that the programme is not suited for the average student” since “parents and students sometimes consider CLIL programmes more cognitively demanding than they really are” which can lead to it “attract[ing] a disproportionately large number of academically bright students” and furthermore create the idea of an “elitist programme” which “is likely to undermine both the CLIL and the regular programmes [...]”. (Mehisto 2007: 63). All of the relevant information (overview, CLIL theory, etc.) including this last, highly important, aspect is suggested to be given in an extended meeting of teachers

(those teaching CLIL as well as those not), headmaster, CLIL-programme director, parents and students. This would – in an Austrian context at least – be a necessary prerequisite to another meeting that is a legal requirement according to Austrian law. A panel that is locally referred to as the SGA (Schulgemeinschaftsausschuss), which roughly translates into school partnership board or school council, consisting of 3 representatives each for the three groups involved in the setting of a single school – teachers, parents and students. According to the previously mentioned 14th amendment, Austrian educational law grants any school a certain level of autonomy in its curriculum design, etc. providing any changes are subject to a vote in the SGA. Introducing and implementing a CLIL programme is a prime example for a decision that would have to be passed in the SGA before it can be initiated at all. Hence, the extended meeting proposed by Mehisto (2007: 63) would provide room for an open discussion that should come before any decision.

When it comes to keeping the students in the programme Mehisto provides an extensive list, which, however, is not entirely relevant to most Austrian CLIL-scenarios, which is why in the following I will provide a very condensed version thereof. It is considered very helpful for the students to be given the opportunity for regular contact with people who speak the CLIL language, which lets the students see the application of what they are learning and will in turn raise motivation. Another factor that is important in raising motivation is to take care when choosing topics that will be dealt with in the course of the CLIL classes. On another level notice has to be taken of the fact that normally the students in the CLIL programme will need some kind of “support system” (Mehisto 2007: 65). Mehisto suggests establishing a homework club that allows student to collaborate and catch up on content points of a CLIL class that might not have been fully comprehended due to their attention being on the language component during the lesson.

2.4.3. Schools

Let me begin with the most important point made in this section right away – “preventing the two-schools-in-one phenomenon” (Mehisto 2007: 66). The coexistence of both the CLIL- and the regular (or non-CLIL) programme should not create a perception of them drifting apart and thereby causing feelings of difference between the students of the respective programmes and also their parents. Therefore the school’s vision and mission should incorporate both programmes and both vision and mission should be developed “involving

the entire staff” in order “to ensure that the school ‘s internal stakeholder interests and needs are taken into account, and that the implementation of the plan is supported” (Mehisto 2007: 66). If all the teachers are on the same page with both programmes and its aims, a unified force is at work that can go a long way in preventing perceived gaps between the two programmes. Mehisto therefore also suggests the establishment of workshops that can serve as a regular medium of exchange for teachers of either programme.

The headmaster of a school plays a critical role in this process too by granting both programmes an equal amount of attention and support. On whichever occasion the headmaster has to be careful to praise both programmes and their respective worth, especially when showing around visitors to the school, again in order to not “feed a sense of inferiority in regular programme students” (Mehisto 2007: 67). Another measure to achieve this sense of unity is by making sure that an equal amount of parents from both programmes are chosen as representatives in the school council (SGA).

Needless to say, there are a lot more points that Mehisto raises such as space and room situations as well as multilingual signage and creating displays in the entrance area that will introduce visitors to both programmes.

2.4.4. Teachers

When it comes to hiring new teachers Mehisto (2007: 68) makes it clear that it is vital for a CLIL teacher to have excellent command of the CLIL language next to subject qualifications, alongside a pronounced willingness to learn CLIL methodology. In an Austrian scenario one would have to point out that in general mainstream education it is not entirely the headmaster’s decision which teacher is assigned to which school. The board of education makes this decision and allows a certain level of weighing in from the headmaster. It also has to be pointed out that in an Austrian context most CLIL teaching would happen either with a subject teacher and a native speaker assistant, or a subject teacher who is also a qualified language teacher, or in rare cases a subject teacher who possesses sufficient command of the CLIL language (English in most cases). Nezbeda (2005: 19) – from an Austrian perspective – raises the issue of the teachers’ pronunciation. Many subject teachers are afraid of going into CLIL teaching because they feel their command of English and their pronunciation are not sufficient. This has led to heated discussions at CLIL schools and is an issue that needs to be addressed from the early onset of the programme, when defining its goals. Questions like

“Is it necessary to speak flawless English in order to teach CLIL effectively or not?”, “Are we aiming to teach a broad spectrum of linguistic structures on a wide range of topics and see the use of English in its definition as a lingua franca?” are only two examples.

Another issue both Nezbeda (2005: 19) and Mehisto (2007: 71) raise is providing support for non-language teachers who enter CLIL teachers also on another front. Many non-language teachers feel that the content is set back due to adding the language component to a lesson. It is therefore essential to instruct these teachers in CLIL methodology and make clear the advantages and benefits of language learning and allow enable them to see that “[a] little investment of class time into language learning will greatly facilitate the teaching of content” (Mehisto 2007: 71).

Finally, another issue brought forward by Nezbeda is the necessity to create an awareness among CLIL-teachers so they understand themselves not only as instructors of a content subject but also as facilitators or “initiators” (2005: 19) of a language learning process.

2.4.5. Curriculum

When developing the curriculum for both the CLIL- and the non-CLIL programme course outlines have to have the national curriculum as a starting point but take into account the aims that the CLIL management team intend for the programme – as previously mentioned the 14th amendment to Austrian educational law provides for some room here.

Again, both researchers point out that this process has to be carried out in a team as large as possible, regardless if teachers are CLIL or non-CLIL staff. This then also makes the process of designing and procuring funding for CLIL teaching materials a lot easier. Designing materials and making them available to the colleagues at school reduces the workload and will also create a feeling of unity among the teachers and “[...] help to develop a clear and consistent message” (Mehisto 2007: 76).

2.4.6. Partners

As will be done at a later stage it is vital to analyse the stakeholders involved in a school setting – external as well as internal. The notion of stakeholding will be dealt with at a later stage (chapter 3 on evaluation). Including stakeholders and their voice, feedback and expectations is another vital step in designing a school’s strategic plan that will help to move both programmes and their coexistence forward. Furthermore Mehisto (2007: 75)

collaboration between neighbouring schools, for example he proposes the idea of “[s]chools that are new to CLIL [...] ‘buddying-up’ with experienced schools” in order to exchange experiences.

2.4.7. Parents

Parents need and want to be able to provide support for their children and therefore require and will, according to Mehisto’s experience (2007: 75), demand information and a lot of opportunity for contact with the headmaster, the programme director and CLIL teachers in order to understand the implications of their child participating in the programme. A couple of examples are as follows: encourage the child to talk about what happened in school; allow the child to watch CLIL language TV programmes, movies, etc.; not to re-teach everything every evening in the home language; understanding the differences between a typical learning cycle as opposed to the ones of language development and content acquisition, which will allow them to arrive at realistic expectations (cp. Mehisto 2007: 76f.).

The points discussed in this section will be referred to directly at a later stage of this thesis, when presenting the school and the CLIL-programme within it and when drawing conclusions from the findings of the interviews conducted in the course of the evaluation. Thus, the ideas from both Mehisto and Nezbeda will prove useful as a general perspective and basis for comparison. Also, it has to be pointed out again, in summary, that while there are many details and aspects to be considered on a number of fronts one of the main criteria for a programme to be successfully implemented is that it takes a great deal of training, teamwork, open communication and readiness for professional development (feedback, sharing materials, etc.). The next section will briefly present a critical view of CLIL with research grounded in Spain.

2.5. Criticism

Drawing from research results in Spanish CLIL programmes Antony Bruton has assumed a rather critical position of CLIL in general. In reaction to a study carried out by Lorenzo et al. (2010) in Andalusia which concluded

[...] the foreign language level of Clil children after eighteen months of the programme was over 20 per cent higher than that of non-Clil children, and argued that this was caused by the use of Clil. (Butler 2012: 13)

What Bruton criticized about this study is the way it was conducted – there was never a pre-test that provided any results to compare the post-test results with, which makes Lorenzo's et al. (2010) deductions questionable, according to Bruton (cf. Butler 2012: 13). Simply testing both CLIL- and non-CLIL groups after 18 months of CLIL provision does not allow reliable conclusions about CLIL and its effects. Another factor Bruton pointed out is that the CLIL- and non-CLIL groups were not created randomly but the parents of the students made the decision which group their child would become a part of. For Bruton (2010: 237) this is highly problematic due to the fact that many CLIL teachers felt "[...] that it was generally the children of higher socio-economic-status parents who chose the CLIL option in their school." Consequently, motivation plays an important role and the fact that CLIL students in Spain show a high readiness to participate in extra English classes outside of their regular lessons (cp. Bruton 2010: 238). By making a number of other points (including the use of native speaker language assistants in CLIL classrooms, or the number of motivated teachers in CLIL programmes) he raises the question whether or not CLIL really is "[...] so beneficial, or just selective?" (2011: 523).

Bruton points out the CLIL cannot be seen and therefore studied completely independently of the educational system that it is embedded in (e.g. in the respective countries) and "[...] not ignore the broader educational implications of such initiatives" (Bruton 2011: 240). He maintains that for the study mentioned above (Lorenzo 2010) important factors like great deficiencies in general (non-CLIL) foreign language teaching have to be taken into consideration before carrying out such a study and even before introducing CLIL in this area in general.

Any state educational system should ensure adequate standards in the L1 medium for all students, before spreading the FL [= foreign language, FS] medium across part of the curriculum for certain students, possibly to the detriment of some of the rest, who remain in the existing seemingly deficit FL scenarios. (Bruton 2011: 531).

The reason why this short presentation of academic discussions is presented here is the fact that within the school whose CLIL programme is being studied some similar criticism is being voiced. Furthermore the previous pages of this chapter put CLIL in a, I believe deservedly, good light. However, it has to be pointed out that the research available on the effectiveness

of CLIL is limited and that there is also criticism being raised against this movement that has spread all across Europe for a number of reasons.

Chapter two was intended to present the concept of Content and Language Integrated Learning. Starting out from finding a definition of this approach, the relevant theory was discussed in connection with what researchers have listed as benefits of CLIL. In the course of this literature review a number of points of criticism was raised and research was presented into the outcomes of CLIL, especially on a content level. It was pointed out that clear-cut research in the field of CLIL outcomes is only growing slowly.

In a further section the situation of CLIL in Austria was examined and was found to reflect great parts of the situation of CLIL in most European countries, namely that it appears in many different shapes and forms, mostly under different names in a number of educational settings and that there is little national (or Europe-wide) common theory of practice that is adhered and referred to when designing a CLIL course or programme and that “[...] precise learning goals and objectives are largely missing.” (Dalton-Puffer 2011: 185).

3. Theory of Evaluation

Evaluation has become a ‘buzz word’ over the past decades, especially in education. The aim of this thesis is to evaluate a language programme, more specifically, a CLIL programme at a secondary school in Austria. In order to do that effectively and meaningfully, a thorough insight into the theory of evaluation is necessary.

This chapter will provide some definitions of the term evaluation and will then proceed to tracing the historical development of the concept of evaluation. This will be done by drawing on various theorists in the field. Starting out from how evaluation was first conceived, understood and implemented up to present-day evaluations the reader will be able to see how not only the purpose but the whole justification or motivation for carrying out evaluations has evolved over time. This is, in essence, a retracing of what Lynch (1996: 12) dubs “[t]he paradigm dialog” and covers issues arising in connection with methods used for evaluating in general and the validity of these different methods.

One has to keep in mind at this point that evaluation is obviously a concept that did not develop only in relation to education or language programmes. Many principles that evaluations in the field of language teaching apply were based on evaluations which

[...] took place in the fields of industry and engineering [...] and [were only later] extended to social, publicly funded programs as the principal means of managing these programs during the second half of the twentieth century. (Kiely & Rea-Dickins 2005: 20).

Defining evaluation is not a straight-forward task – numerous theorists have assigned different purposes to evaluations (in the realm of applied linguistics and, more specifically, language programme evaluation), thereby justifying different methods for evaluations and consequently arriving at different definitions of the term evaluation altogether.

Before embarking on the discussion of the theory of evaluation, I want to point out one more thing: while this chapter deals with the theory of evaluation in general, chapter 5 (Methodology) will then refer back to this chapter when it comes to the point of explaining how the theory described will be applied in the empirical part of this research project and which strands of theory I have chosen to adhere to and why.

3.1. Introduction to the concept of evaluation – definition & basic terms

3.1.1. Definition(s)

Evaluating is something that we do in our daily lives. Statements such as “What a beautiful day!” or “What ghastly weather this is!” are evaluations, very informal ones one might add, but nevertheless evaluations. It is therefore important to remember that the kind of evaluations discussed here differs to everyday “evaluative judgements” (Rea-Dickins & Germaine 1992: 4) in so far as that in

[...] an educational context [evaluation] should be systematic and undertaken according to certain guiding principles using carefully defined criteria.

The purpose of this section is to briefly present and discuss definitions of the term evaluation and arrive at a working definition of my own that is relevant for this thesis.

To begin with, consider the following definitions:

- Evaluation is the process of examining a program or process to determine what’s working, what’s not, and why. Evaluation determines the value of programs and acts as blueprints for judgment and improvement. (Rossett & Sheldon, 2001)²
- Evaluation is the systematic determination of merit, worth, and significance of a learning or training process by using criteria against a set of standards.³
- Evaluation is defined here as the systematic attempt to gather information in order to make judgments or decisions.
- Evaluation is about the relationships between different program components, the procedures and epistemologies developed by the people involved in programs, and the processes and outcomes which are used to show the value of a program – accountability – and enhance this value – development. (Kiely & Rea-Dickins 2005: 5).
- Evaluation is a form of enquiry, ranging from research to systematic approaches to decision-making. (Kiely & Rea-Dickins 2005: 6).
- Evaluation, as a process of determining the worth of a program, [...] (Kiely & Rea-Dickins 2005: 12).

Many of these definitions have something in common – namely the phrase “worth/merit of a program(me)”: the point of an evaluation seems to be to find out if a programme (whatever the field these originated from may be: e.g. a language programme in a school, a

² taken from: http://www.nwlink.com/~donclark/hrd/isd/types_of_evaluations.html , accessed 12 Feb 2012

³ both taken from: <http://www.nwlink.com/~donclark/hrd/sat6.html>, accessed 12 Feb 2012

new procedure within an office, etc.) is working, and that basically means finding out if it is doing what it was intended to do and therefore meet the expected outcomes and as a consequence, for example, deserve to be continued and to receive further funding, etc.

While all of the definitions listed above have similarities, they do differ as well (especially in their focus and, even more importantly, in the given aims) and this can to a large degree be explained by the historical development of the concept of evaluation. In this order they reflect, to a certain extent, how the approaches have been altered and influenced by diverging theories of knowledge building over the past decades. The historical development of evaluation will be dealt with in greater detail in section 3.2. and makes up for the main component of this chapter, the study of evaluations in general, because it explains not only the idea of evaluation but different concepts of evaluations and their vital components.

With these definitions in mind I briefly want to present some of the relevant literature providing ideas and thoughts on what evaluation is used for today and what purposes are served by carrying out an evaluation in general.

The first question at hand is what evaluation is, and what it is not. Especially in the context of education it has to be clarified that evaluation is neither testing nor assessment. These concepts are commonly confused, especially in education. Testing is a vital part of evaluation, as is assessment but evaluation can be considered an umbrella term that contains the other two (cf. Rea-Dickins & Germaine 1992: 3 ff.) since a test carried out within an evaluation provides results that can lead to an assessment (a judgment) about a programme or a certain feature thereof. These assessments will, if put into context with a number of other factors, make up for the complete evaluation (results). These factors include among others aspects such as the environment a programme is set within or the person or institution carrying out the evaluation.

Before moving on to further aspects of the study of evaluation I would like to close with the definition that seems most fitting for the purpose of this thesis. It is a combination of the definitions listed above and is as follows:

“Evaluation is a systematic process whose purpose it is to determine the worth of a programme, its strengths and weaknesses, its potentials for improvement. This is achieved through a thorough study of the components of a programme and the relationship between these.”

3.1.2. Purpose of Evaluation

The aforementioned criteria have to be clearly defined beforehand and the definition of these depend on the motivation and the purpose of an evaluation, which is the second question that needs to be dealt with briefly – the “WHY” of evaluation. According to Rea-Dickins & Germaine (1992: 7) there are two different motivations: (a) to “[...] *explain*[...] and *confirm*[...] existing procedures” and (b) to gain information [in order] to bring about *innovation* or *change*. In other words evaluation is motivated by the need to find out if a programme is working the way it was intended to and if not, find out how it can be changed and improved as well as gaining an understanding of unanticipated procedures which might have developed within a programme and see what effects these might have on the outcomes of the programme.

To explain this in a bit more detail, Rea-Dickins & Germaine (1992: 23) group the purposes for carrying out an evaluation in two different sections: (a) general purposes – these include aspects such as *accountability*, *curriculum development* and *betterment* as well as *self-development*. The first – *accountability* – applies if an evaluation is carried out mainly in order to “[...] report on a product and give an evaluative judgment, whether something is intrinsically a ‘good thing’ or not.” (Rea-Dickins & Germaine 1992: 24). Mostly the findings of such evaluations are used as a report for somebody, in many cases the person or institution who provides the funding for the programme or decides if a programme is to be pursued. Very rarely are such evaluations intended for the other two reasons given above – *development* and *betterment*. The results of an evaluation with this purpose will more likely go directly to the people participating in the programme or its directors in order to incur innovation and change (see above). In the case of educational evaluation concepts such as *betterment* and *(self-) development* would imply the reconsideration of certain components of a programme, or its objectives and the search for possibilities for improvement.

The other section Rea-Dickins & Germaine group purposes into is (b) specific, topic-related purposes. Very much simplified, one can say that evaluations with this purpose in mind will look at a very specific aspect within the programme and try to answer a very specific question (e.g. “Is the school textbook used for teaching CLIL in this school appropriate?”).

3.1.3. Types of evaluations

Theorists like Scriven (1967: 42) group evaluations into two general kinds: *formative* and *summative*.

Formative evaluations aim at determining the value of a programme *during* the formation process of the activities within an ongoing programme. It therefore aids the process by identifying strengths and weaknesses within a programme before all its components are fully established and allows for changes to be made along the way. Thereby the final outcomes can be modified beforehand.

Summative evaluations on the other hand place the focus solely on the outcomes of a programme. They are “[...] used to assess whether the results of the object being evaluated [...] met the stated goals.”⁴

It is therefore logical to conclude that summative evaluations very often have the aforementioned purpose of accountability whereas formative evaluations will most likely serve the purpose of betterment and (self-) development or specific, topic-related questions as previously stated.

3.1.4. Stakeholding in evaluation

Keeping in mind what has been said so far, I want to briefly introduce and discuss one further relevant term in the context of evaluation: using the example of a language programme applied in an educational context (e.g.: a bilingual language teaching programme in a school), one can identify a number of parties involved: on an internal level, the teachers and students who are directly affected by the work with the programme. Furthermore there are parents who obviously take a more external view and who, for example, may be asked to make (mostly financial) contributions to this language programme. Moreover, certain external sponsors might contribute part of the funding and furthermore the regional or national boards of education will be interested to know if the resources used for the implementation of a language programme are well spent or might be put to better use elsewhere. All these parties are referred to as *stakeholders*, who in the

⁴ source: http://www.nwlink.com/~donclark/hrd/isd/types_of_evaluations.html , accessed December 30th, 2012

situation of an evaluation take different viewpoints and these viewpoints are what influences the design of an evaluation. Also they will have different goals in mind for the programme in question, which in turn might lead to different questions for an evaluation. This is what essentially affects and defines the criteria that Rea-Dickins & Germaine mention as the factors that distinguish everyday evaluative judgments from evaluations that are more formal and serve a more scientific purpose.

To summarize it can be said that the endeavour of evaluating a language programme can be motivated by a number of different ideas, depending on the stakeholder. Members of a board of education will want to know if the programme is showing the effects that were predicted and desired and will want to compare the effects and results of different programmes (which are an investment, or rather an allocation of resources, most commonly made by the board of education) – so their purpose is “comparison and decision-making between alternatives” (Kiely & Rea-Dickins 2006: 26). The teachers and students, however, might be less concerned with the actual outcomes but much more with the internal organisation of a programme and how to best continue its development. This illustrates the important position stakeholders assume within an evaluation and how they define the character of such an undertaking.

In this context I want to briefly explain my stakeholder position very briefly – as a teacher who is part of the programme that is being evaluated, but as well as the person carrying out the evaluation, there are a number of stakeholder goals and positions that I assume. I assume the role of the evaluator because I am the university student writing this thesis and the evaluation forms a necessary part of its completion and eventually my degree. In this sense, my goal is not related to the school and the programme but is a personal one. As a teacher working within the programme I will want to use this thesis and evaluation as an opportunity to contribute to the development of the programme at the school and maybe be able to inform future decisions made by the programme directors.

This section has already shown that there will almost always be differing viewpoints in any kind of evaluation process. In order to understand these and their development better the following section will trace the historical development of the concept of evaluation.

3.2. Understanding different types of evaluation – historical development

In their book *Program Evaluation in Language Education* Kiely and Rea-Dickins (2005) make a distinction between two different general kinds of evaluation: they describe evaluation with a “focus on design and method” (2005: 17-36) and evaluation with a “focus on context and use” (ibid: 37-55). The first group therefore concentrate more on how the evaluation is set up, how it is designed and what methods are used whereas the second one pays a lot more attention to the context the evaluation is set in and in how far its results can be put into meaningful use.

The former focus is chronologically situated before the latter and stems from the fact that evaluations were seen much more similar to inspections during that time. Inspections were considered necessary, especially in the public sector (health care, education, social welfare etc.), which was dependent of funding by the taxpayer, and therefore public money. Inspections were seen as a means of ensuring accountability for citizens. This need for evaluation arose during the nineteenth century, briefly after education systems became publicly funded, in industrialised countries, but especially in Great Britain and the USA (cf. Kiely & Rea-Dickins 2005: 18). Evaluations in this understanding necessitated a certain approach – often external evaluations based on measurable, objective criteria.

The history of evaluations begins at a time frame when evaluations in the realm of education focussed on research design, data collection methods and clear evaluation outcomes that are measurable and can be used directly for decision-making (cf. Kiely & Rea-Dickins 2005: 17). But already at the early onset of evaluations and the study thereof researchers and practitioners both were faced with two seemingly “opposing concerns”: “subjective values” vs. “scientific, objective measures” (Kiely & Rea-Dickins 2005: 19). This ties in with an aspect mentioned at the very beginning of the chapter – the paradigm dialog (Lynch 1996:12): the question was (and still is) if the scientific way to go is by addressing and studying aforementioned “subjective values” and thereby paying attention to the actors within a programme and their needs or to pursue a different goal and study a programme in a way that will allow only quantifiable measurements and avoid any subjective statements.

3.2.1. The Tylerian model & first reactions

Ralph Tyler (1950: 48) developed a framework which was applied in the evaluation of curricula and continues to be assigned major importance today. As previously pointed out, evaluation frameworks such as the Tylerian model were not originally developed for application in the realm of education but for many other fields like management.

The key to understanding how successful a program was required detailed specification of purposes, processes and, especially outcome measures which informed on these. (Kiely & Rea-Dickins 2005: 20).

Tyler's framework therefore was based on measurable criteria or outcomes which had to be clearly set beforehand and the ensuing measurement thereof. This framework basically consists of 4 defining questions that have to be answered in the course of a Tylerian evaluation:

- 1) What educational purposes should the school seek to attain?
- 2) What educational experiences can be provided that are likely to attain these purposes?
- 3) How can these educational experiences be effectively organised?
- 4) How can we determine whether these purposes are being attained?

(Tyler 1950: 48).

Two main aspects were put into practice when conducting such an evaluation: defining intended learning outcomes and teacher behaviours as objectives and the measurement thereof (cf. Kiely & Rea-Dickins 2005: 20).

The focus lay on what Lynch (1996: 14) called "preordained" criteria. This process satisfied the scientific desire for empirical validation through evaluation procedures (cf. Taba 1962: 314). Hilda Taba (1962) further explains that evaluation is carried out in the following steps:

- acknowledging that theory alone is not enough to prove the merit of certain procedures and strategies
 - identifying behaviours which represent the curriculum strategy
 - measuring these behaviours
 - developing comparisons with a parallel context where the strategy is not used
 - making sure that the evaluation process takes a broad, open view, documenting curriculum processes so that unanticipated data can be interpreted and understood
- (cf. Taba in Kiely & Rea-Dickins 2005:21)

Taba already pointed out the two-sidedness of evaluation procedures, which resulted from different demands that were expressed towards evaluation at the time (1960s). These

remind us of the different aims and purposes of evaluations as discussed in section 3.1. She uses the terms “external” and “internal evaluation” to hint at issues such as stake-holding in evaluation and the dichotomy between external and internal views in the evaluation process. External evaluation, very simply put, encompasses the concept of an outsider to the evaluated programme carrying out the evaluation. Internal evaluation on the other hand is carried out by a person/ a group of people within the programme that is being studied and evaluated. This thesis makes use of the methods of internal evaluations in its empirical part.

Returning to Taba (cf. Kiely & Rea-Dickins 2005: 21f.) – she already took notice of the divide between different characteristics of evaluations depending on viewpoints and positions of stakeholders and Kiely & Rea-Dickins comment on this as follows:

Reconciling these divides became a major task for evaluation theorists and practitioners in evaluation during the following decades, in language education as in general education programs and innovations. (ibid: 22).

With this established, I now want to proceed to tracing the developments in the history of evaluation from Tylerian ideas onward, starting with the first reactions to the model proposed by Tyler.

The first major step happened at a time when the common paradigm, especially in the USA, was the quantitative experimental method. This evaluation design consists of

test[ing] the effectiveness of a particular strategy [...] by comparing two groups: an *experimental group* which experience the strategy [...]; and a *control group* which have the normal educational experience. (ibid: 23).

This form of evaluation design had gained considerable popularity for a number of reasons. Theorists like Beretta (in Kiely & Rea-Dickins 2005: 23) argue that, especially in the USA, this development was due to the fact that countries wanted to be more competitive, particularly on a technological level and “reduce social disadvantage” (ibid: 23). This, in turn, necessitated new approaches in education whose effectiveness had to be proven and if necessary, furthered.

Moreover, the fact that this approach still is, up to this day, the main technique in medical research allows the conclusion that it has proven its aforementioned desired effectiveness over time. Since it is heavily supported by quantitative data, it allows making generalisations as well as “clear, objective comparison” (ibid: 24). It has also been used in the field of language programmes in an educational context various times. Two examples were the

Pennsylvania Project (Smith 1970) and the Colorado Project (Scherer and Wertheimer 1964), which both employed the experimental approach to “scientifically show the merits of audio-lingual and cognitive code methods in foreign language classrooms” (Kiely & Rea-Dickins 2005: 24).

With reference to what has been said earlier about stakeholding in evaluation one must add that evaluations can be undertaken for various reasons and purposes. The experimental approach was not always best tailored to the needs of all the stakeholders in an evaluation and did not produce results that were meaning- and useful to everybody. For a teacher or student within a language programme for example, will have greatly differing demands to an evaluation and will assume a different viewpoint. Reasons such as this were the reason for critical reactions to the experimental approach and for the development of a different focus in evaluation.

Kiely and Rea-Dickins (2005: 26) refer to this new development in the theory of education, that actually consisted of a shift towards a different focus in the purpose of evaluation and happened during the 1970s, evaluation with “[...] a focus on worth and program development”. Theorists like Worthen and Sanders (in Kiely & Rea-Dickins 2005: 27) gave evaluation a new definition and purpose by stating:

Evaluation is the determination of the worth of a thing. It includes obtaining information for use in judging the worth of a programme, product, procedure, or object, or the potential utility or alternative approaches designed to attain specific objectives. (Worthen & Sanders 1973: 19 in Kiely & Rea-Dickins 2005: 27).

So instead of only designing means to measure the outcomes of a programme and comparing these to others and, consequently, basing a maybe financial or political decision thereupon, the shift in purpose allowed for much more orientation toward a single programme and the works and processes within it, thereby assigning increasing importance to different stakeholders, namely the more immediately affected ones.

Worthen and Sanders (1973, in Kiely & Rea-Dickins 2005: 27) make a distinction between evaluations of the *static* characteristics as opposed to evaluations of the *dynamic* characteristics of a programme. Static characteristics, in essence, comprise the resources a programme can make use of. These include human and material resources and allow, according to Worthen and Sanders (1973), a judgment on the quality of the programme.

Theorists such as Brown (1989, in Kiely & Rea-Dickins 2005:27) draw a connection between the evaluation of static characteristics and “procedures for institutional accreditation”.

In the light of the aforementioned new developments towards focussing on the worth of a programme, evaluations with a sole interest in the static characteristics seemed inadequate and inappropriate. The wish to study the dynamic characteristics, however, gave rise to process-oriented approaches such as Scriven’s (1967, in Kiely & Rea-Dickins 2005: 27) “*goal-free evaluation*”, which looked broadly at a program as an evolving social construct rather than as a set of Tylerian objectives”.

Seeing a (language) programme as an “evolving social construct” facilitated the move away from evaluations whose only function it is to base political decisions upon (e.g. regarding the continuation of funding etc.). This allowed for individual programmes to be studied more closely and comprehensively, in a way that an evaluation could include the effort to understand the internal processes of an educational programme and, even further, to include this understanding in any final judgment made about the programme as a whole rather than only using measurements of outcomes as a basis for decision-making.

3.2.2. Stake’s “countenance model” or “responsive model”

One of the most influential models for evaluation in the field of education that came up during the late 1960s was developed by Robert Stake (1967) and is called “the countenance model”. The basic motivation behind Stake’s model is what I already alluded to above, namely that evaluating by solely relying on the measurement of programme outcomes is not enough. He proposed an expansion of the evaluation focus to include what he refers to as “antecedents” (1967: 527) and processes within the programme alongside its outcomes. Thereby he wanted to increase the relevance of the information for stakeholders. In Stake’s view “an antecedent is any condition existing prior to teaching and learning which may relate to outcomes” (ibid: 527) and along with the outcomes these two categories make up the more static parts of the evaluation data. This draws a connection to researchers mentioned earlier (Worthen & Sanders) because for Stake the dynamic element in evaluation is the third category mentioned – the processes within the programme, or as Stake labels them – “transactions”. These, according to Stake, include “the countless encounters of students with teacher, teacher with teacher [...] – the succession of engagements which comprise the process of education” (ibid: 527).

Stake's model is partly based on the idea that at the time "[...] too little effort [was made] to spell out antecedent conditions and classroom transactions [...] and too little effort to couple them with the various outcomes [...]" (Stake 1967: 524). He propagates approaches that are tailored to each specific location (in his case: different schools) by claiming that "[t]he purposes and procedures of educational evaluation will vary from instance to instance" (ibid: 524). A further thought brought forward by Stake to support his approach is the importance of dealing with "the rationale of the program" (ibid: 524) before embarking on any evaluation project. According to Stake, the rationale is the basis of any sound evaluation since only by examining the "basic purposes of the program" (ibid: 524) it makes sense to make judgments about how the programme has been implemented.

When implementing an evaluation according to Stake's model, an evaluator will make use of a data matrix consisting of 13 cells. The first cell comprises aforementioned rationale. The other 12 are split into cells for descriptive and judgment data. The first set of cells – on descriptive data – include intents and observations. Evaluators gain information regarding intended outcomes of the programme from the stakeholders and complement these with data of the outcomes actually observed by the evaluators themselves. The judgment data matrix (6 cells) is divided into standards and judgments.

In order to complement the very brief presentation of Stake's programme, the following graphic will illustrate the concept of Stake's matrices. All four categories – intents, observations, standards and judgments are split into antecedents, transactions and outcomes.

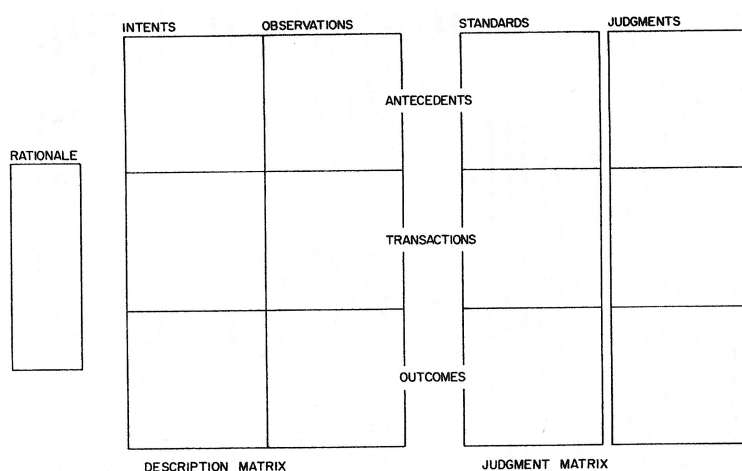


Figure 2 - Stake's data matrices

What Stake (ibid: 528) refers to as “intents” comprises all aspects of a programme that have previously been “planned-for” – these include environmental conditions, student behaviour, intended subject matter but also anticipated effects. In Stake’s model the term “intents” is synonymous to the frequently used terms “goals” and “objectives”.

Another aspect that distinguishes Stake’s concept from others is the clear difference between description and judgment. He explains that no useful evaluation can do without judgmental statements, yet he also notes that many evaluators seem to be very little inclined to assume the responsibility and authority to make judgements (cp. ibid: 526). Judging the specifics of a programme can be done in a twofold manner: (a) keeping absolute standards in mind as indicated by personal judgments (= absolute judgment) and (b) by relating standards to those of programmes similar to the one being evaluated (= relative judgment). Judging therefore happens by taking the observation data, then making comparisons to standards, checking in how far these have been met and then coming up with judgments (cp. ibid: 534ff.).

To round this description off – it was Stake’s opinion at the time that evaluations should not only focus on the objectives and outcomes of programmes (as done in the Tylerian model). With the goal in mind to take a more holistic stand in approaching evaluations, his countenance model that provides the evaluation matrix briefly discussed above, he wanted to depart from formal approaches to evaluation and include a richer variety of data into the scope of evaluation. Stake’s approach is based on

[...] an essentially naturalistic perspective and methodological approach toward evaluation. The evaluator investigates the program as it exists in the field and arrives at an understanding of its process and objectives through observation and interaction with program participants. (Lynch 1996: 80f.)

Another approach that was developed in the wake of this movement was Stufflebeam et al.’s “CIPP (Context, Input, Process and Product)” (1971) approach whose goal it was to study the implementation aspects of a programme while maintaining an external perspective.

Beretta’s “adversary approach” (1992) attempted to achieve the same study purpose but from an internal perspective which allowed evaluators and participants to support their

viewpoints about a programme and the course of its development by gathering their own data (cp. Kiely & Rea-Dickins 2005: 29).

One last evaluation model from this period I would like to mention, since it seems relevant to what the later part of this thesis will do, is Eisner's (1977) "educational connoisseurship approach". In order to explain it properly one has to consider that Eisner was also an art scholar who applied concepts such as connoisseurship in the field of education. This is how he describes connoisseurship:

Connoisseurship is the art of appreciation. It can be displayed in any realm in which the character, import, or value of objects, situations, and performances is distributed and variable, including educational practice. (Eisner 1998: 63)

Eisner's point of view motivated the quest for the value or worth of a programme by disregarding quantitative data but rather by making use of a "rich impressionistic narrative" and thereby moving even further away from Tylerian ideas (cf. Kiely & Rea-Dickins 2005: 29).

We also have to be able to draw upon, and make use of, a wide array of information. We also have to be able to place our experiences and understandings in a wider context, and connect them with our values and commitments. Connoisseurship is [...] not a technical exercise.⁵

Along with other evaluation models that I will introduce in the ensuing paragraphs these ideas come down to the following common denominator: during the 1970s theorists were trying to find feasible alternatives to the experimental approach and devised approaches which drew less from the "principles of industrial engineering and classical management, and more on their key social and interpersonal characteristics" (Kiely & Rea-Dickins 2005: 29).

3.2.3. Standards

The new developments in the field of evaluation theory during the 1960s and 1970s led to a number of different ideas and opinions as to how to approach evaluation, be that in the field of education or elsewhere. Evaluations had themselves become programmes whose methods were questioned in a way that searched for verification of their worthiness and value. The question was, with the number of different approaches and their different focal points in mind, which of them would provide reliable information. A demand for evaluation

⁵ Smith, Mark K. 2005. <http://www.infed.org/thinkers/eisner.htm> , accessed May 13th, 2012

standards evolved which led to the formation of a joint committee at the Center for Evaluation Study at the University of California in Los Angeles. The researchers developed “30 criteria in four categories, stated in order of priority as *utility, feasibility, propriety and accuracy*” (Joint Committee on Standards for Educational Evaluation 1981 in Kiely & Rea-Dickins 2005: 31). The use of these standards was then interpreted differently: while some researchers focussed more on utility and responsibility for use, others felt that the standards had been established with the decision-making process after an evaluation in mind. Kiely & Rea-Dickins (2005: 31) comment that “generally [standards were] intended for external evaluations [...]” and that these had significant influence in the field of language programme evaluations. However, they also note, even in spite of the development of the agreed-upon standards, the thought that evaluation is more concerned with “monitoring and accountability” (ibid: 32) in a way which very often “lost the benefits for practice” (ibid: 32).

3.2.4. Illuminative evaluation & Triangulation

In 1972 Parlett & Hamilton introduced a concept called “illuminative evaluation” which is yet another evaluation model that was developed to provide strong contrast to the evaluation models that were based on the experimental, psychometric approach. Parlett & Hamilton (1972: 59-60) provided extensive reasoning for their counter-approach to what they called the “agricultural-botany evaluation” paradigm. Their label for the experimental approach was due to the perceived “failure to address differences between the behaviour of humans and plants” (Bennett 2003: 25). Using methods stemming from social anthropology allowed them to study a programme in context and without the need for a parallel control group. The “primary concern is with description and interpretation rather than with measurement and prediction (Parlett & Hamilton 1976: 88). “Illuminative evaluation seeks to generate hypotheses and theories from within the data which have been gathered.” (Bennett 2003: 27). Parlett & Hamilton suggest educational evaluation to be carried out in three steps: (1) loose data collection which allows the researcher to determine issues at hand, (2) a phase where these previously identified issues are investigated into in more detail and (3) a final phase which consists of searching for patterns and explanation. In order to be able to conduct an evaluation in the fashion devised by Parlett & Hamilton it is suggested to make use of four sources of evidence:

- observation of events to identify common incidents, recurring trends and issues

- interviews with participants to probe their views
 - questionnaire and test data where appropriate
 - documentary and other background information to set the innovation in context
- (Bennett 2003: 28)

In essence, compared to more traditional evaluation models, illuminative evaluation draws from very different sources. What Parlett & Hamilton (1972) see as the result of illuminative evaluation is an in-depth study of a particular programme in use and that is one that is not necessarily comparable to other case studies. This is exactly the point where critics of the innovative approach see a clear fault – the lack of comparability, which in turn is considered evidence for lack of validity of a study. Several theorists have pointed out that case studies allow for too much room for the evaluator or the person conducting the study to exert some kind of influence on the study, data, and interpretation thereof thereby making it less reliable and valid (cf. Bennett 2003: 28).

A concept whose main purpose it is to reduce these concerns is *triangulation*. It comprises the notion that both credibility and reliability of an investigation or evaluation can be increased by drawing from multiple sources (cf. Lichtman 2006: 195). While theorists propose at least 5 different kinds of triangulation (data triangulation, investigator triangulation, theory triangulation, methodological triangulation, and environmental triangulation), data triangulation is the type that is most relevant here. It simply proposes incorporating data from different sources to make the study as a whole more objective and thereby more scientific.

Triangulation is based on the idea that something (e.g., a submarine, a cell phone) can be located by measuring the radial distance or direction from three different points. Some writers hold the view that validity can be established by triangulating data, investigators and methodologies. (Lichtman 2006: 85).

Another point of critique brought forward against illuminative evaluation is the question whether findings of a study can be generalized, a notion, which is considered imperative within the experimental approach. Stake (countenance model) notes (1995: 95) that “all evaluation studies are case studies” and that putting the results of evaluations to use can only happen in a meaningful way if one particularises instead of generalising. It is much more important, according to Stake, to understand the programme and how certain planned elements contribute to the process of the complete programme “and the lives of those

participating in it” (Kiely & Rea-Dickins 2005: 33) than being able to come to conclusions that are valid for more than the particular programme in question.

3.2.5. Lawrence Stenhouse

Lawrence Stenhouse (1975) took Parlett & Hamilton’s ideas and concept and developed them further in the field of curriculum development. He proposed the “teacher-as-researcher model” (Benett 2003: 31) and, consequently, a change of the role of evaluation in the field of education. One of Stenhouse’s major arguments to devise this model was motivated by the observation that many teachers who were involved in developing curricula could not put evaluation results that specialist or external evaluators had provided to use in their specific context. Too much generalisation was a problem that made it difficult for practitioners to apply evaluation results meaningfully. Stenhouse therefore suggested that evaluation be integrated into the process of what he called “curriculum betterment” (1975: 99), evaluation should therefore be merged with and become an instrumental element in the continuous development of a curriculum. Thereby he destroyed the conceptual division between development and evaluation (1975: 122).

In Stenhouse’s view, the primary focus of evaluation is to help lead the way in curriculum development of a school or a programme rather than to simply come up with results and recommendations that can be applied in other settings (cf. Kiely & Rea-Dickins 2005: 34). It should not be merely an assessment of worth (see sections above) but „a philosophical critique“ (Stenhouse 1975: 118) that gives insight into the intentions of the curriculum. It should also aim at identifying the potential a programme has, while also finding interesting problems. Uncovering problems is a necessary part of curriculum betterment since these trigger further investigation and understanding of the local conditions, which in turn then helps to generate possible solutions and therefore create improvement.

An important factor in Stenhouse’s model is that teachers become researchers themselves. „It is not enough that teachers’ work should be studied: they need to study it themselves.“ (ibid: 143). The data relevant for evaluation should come from classroom observation and interpretation of lessons. Stenhouse’s approach therefore complements Parlett and Hamilton’s concept of illuminative evaluation by adding the point that the teacher assumes a dual role – teacher and researcher/evaluator (cf. Bennett 2003: 31).

As with illuminative evaluation there are a number of potential points of critique, especially due to the fact that in Stenhouse's model the data is collected and interpreted by someone who is a participant as well as an observer and might therefore be more prone to bias than an external evaluator.

The contributions of Parlett and Hamilton and Stenhouse and his colleagues in many ways rewired evaluation theory. The ways in which evaluation processes interfaced with practice on the one hand, and educational research on the other, posited radically different roles for all participants in the curriculum. (Kiely & Rea-Dickins 2005: 35f.)

3.2.6. 4th generation evaluations

As a final step of this survey of the history of evaluation theory and the development of evaluation designs and methods I want to list a movement, most principally represented by the researchers Guba & Lincoln (1989). Taking all the previously suggested models another step forward, Guba & Lincoln assume a constructivist⁶ position and build their concept of evaluation called "Fourth Generation Evaluation" on the understanding that every single participant in a programme perceives the programme differently, interprets his/her experience differently and therefore brings a unique perspective into evaluation. What "fourth generation evaluation" intends to do is use these unique perspectives and experiences and make sense of them without the goal of arriving at "a single, universal, objective truth" (Kiely & Rea-Dickins 2005: 40). Guba & Lincoln see evaluation not as a technical form of investigation drawing heavily on statistical methods, in fact, they do not even define evaluation as a scientific endeavour since they believe that "[...]to approach evaluation scientifically is to miss completely its fundamentally social, political, and value-oriented character" (Guba & Lincoln 1989: 7).

Before explaining the defining elements of "fourth generation evaluation" I would like to provide an insight into the reasons for the development of this term. A "fourth generation evaluation" obviously raises the question about the first three generations of evaluation.

⁶ "Constructivism as a paradigm or worldview posits that learning is an active, constructive process. The learner is an information constructor. People actively construct or create their own subjective representations of objective reality. New information is linked to prior knowledge, thus mental representations are subjective." (taken from: <http://www.learning-theories.com/constructivism.html>, accessed May 20th, 2012)

Guba & Lincoln came to compare the development of evaluation procedures with the development from one generation to another. Every generation or phase in the development of evaluation (with special attention paid to its application in the field of education) has already been touched upon in the course of this overview but providing Guba & Lincoln's view serves as a good summary.

First generation evaluation is what Guba & Lincoln refer to as the *measurement* generation (1989: 26). The evaluator assumes a merely technical role and is equipped with the techniques to measure outcomes of a programme. In the context of education these were mostly test results.

Second generation evaluation added a new aspect to evaluation: *description*. The evaluator's role was expanded, he was not only a measurer anymore but also a describer. This is where aforementioned Tylerian objectives in evaluation appeared. At the time it was considered insufficient to merely carry out measurements and call that evaluation. Researchers stated it would only be fair to measure according to previously described desired (learning) outcomes, called objectives. Ralph Tyler was, in a way, the founding father of programme evaluation since it was mainly him and his research team who devised the term objectives. The evaluator then did not only measure outcomes, he compared strengths and weaknesses of a programme by comparison to specifically previously devised objectives.

The key characteristic of the first two generations is a reliance on preordained categories, while third and fourth generations work with actual program implementation and stakeholder experience. (Kiely & Rea-Dickins 2005: 41).

Third generation evaluation came in response to the demand for evaluation to include *judgement*. The evaluator now included the role of a judge into his function, along with being the measurer and describer. The main researcher to spark off this line of argument was Robert Stake (countenance model) who regarded judgement and description as the two basic "faces" or "countenances" in evaluation (Stake 1967: 523).

Guba & Lincoln identified three main problems of the first three generations in evaluation. These are (1) "a tendency toward managerialism", (2) "a failure to accommodate value-pluralism", and (3) "an overcommitment to the scientific paradigm of inquiry" (1989: 31f.).

The first point of criticism covers the relationship between the manager who, simply put, is the person or institution which demand that the evaluation be carried out (e.g. headmaster,

school board members etc.), and the evaluator who is appointed to actually lead the evaluation. Guba & Lincoln point out that this relationship is always necessarily lopsided since the manager always retains the power to decide on the methodology that will be employed in the evaluation, which aspects are to be covered and, finally, if and which parts of an evaluation will be published and released. The issue at hand is also that most managers will want to avoid any kind of criticism or loss of face as a result of possible evaluation outcomes and therefore exercise influence on the evaluator.

The second issue (failure to accommodate value-pluralism) ties in with the theory of constructivism. “The very term *evaluation* is linguistically rooted in the term *value*.” (ibid: 34). As mentioned earlier, evaluation came to be defined as an endeavour that would determine the worth or value of a programme. Guba & Lincoln acknowledge the fact that today’s society is made up of a vast range of different viewpoints and values which in turn raised the question “*whose values* would dominate in an evaluation” (ibid: 34). Personal values need to be given room in an evaluation since a value-free evaluation is virtually impossible to attain according to Guba & Lincoln. Many argued that evaluation findings could only be trusted (or considered valid and reliable) if they had been reached by a methodology that is scientific and entirely void of any values. What Guba & Lincoln point at is the fact that none of the first three generations take into account that there might be (and most certainly always are) value differences even among the different stakeholders of an evaluation (cp. 1989: 34f).

Overcommitment to the scientific paradigm of inquiry is the criticism that has been dealt with and presented in more detail in the course of this review. The description of the gradual movement away from the experimental (positivist) approach in evaluation (especially in the field of social sciences). Using scientific methods in evaluation or – in other words – adhering to positivist beliefs accepts the existence of a universal truth that can be found by means of scientific methods. In order to briefly explain the term of positivism in this context I want to make reference to theorists such as August Comte who maintained that

all genuine knowledge is based on sense experience and can only be advanced by means of observation and experiment. Following in the empiricist tradition, it limited inquiry and belief to what can be firmly established and in thus abandoning metaphysical and speculative attempts to gain knowledge by reason alone [...] (Cohen et al. 2000: 8)

Truth therefore is “non-negotiable” (ibid: 37). This is a belief that the theorists behind the first three phases or generations of evaluations cling to, in a too pedantic way, according to Lincoln & Guba.

Fourth generation evaluations focus on the individuals within the programme and their unique views and experiences within the programme (which stakeholder-group each person belongs to is considered irrelevant at this point). After measurement, description and judgement the new key term is *negotiation*. For Guba & Lincoln the results of an evaluation (what theorists before them considered “facts” that are “true” because found by scientific methods) are “*created* through an interactive process that *includes* the evaluator [...] as well as the many stakeholders that are put at some risk by the evaluation.” (ibid: 8). These findings that are created, or constructed, are then obviously influenced by the values of the constructing individual. As in other evaluation methods (Parlett & Hamilton, Stenhouse etc.) before Guba and Lincoln point out that all these “constructions” are only meaningful if put in perspective with the specific context. Guba & Lincoln on the necessity of context (for an evaluation): “In a very literal sense the context gives life to, and is given life by, the constructions that people come to form and hold.” (ibid: 8).

Fourth generation evaluation consists of two phases – discovery and assimilation. Discovery is the phase in the evaluation process where data is collected. This data can be taken from a variety of sources (former evaluations, programme descriptions, interviews with programme participants etc.). During the assimilation phase it is the evaluator’s task to draw the various data together, put it in perspective with the programme’s intentions and purposes and possibly come up with suggestions for alternatives or future steps in order to further the programme’s development.

Since the evaluation carried out in the course of the empirical part of this thesis will adopt and adhere to the methodology of a fourth generation evaluation, a more detailed account of what exactly fourth generation evaluation entails will be provided in chapter 5 (“Methodology”).

The purpose of this section was to provide an overview of how evaluation, especially in education, has been seen and used in the US and Britain in the past decades up to now. The concept and notion of effective evaluation has certainly come to be scrutinized by a vast amount of theorists as well as practitioners and has undergone a continuous development

process. The definition of evaluation has had to be altered, added to and modified in the course of this development, as well as the use evaluation has been put to in the last years compared to the current situation. The overview and review of relevant literature ends with fourth generation evaluation since it is the one that is frequently used and adhered to in education nowadays and it is the model that this thesis will make use of.

After the first two chapters have dealt with the necessary theory relevant for this thesis the following chapters make up the more practical part, beginning with providing a description of the school and its language programme that will be evaluated.

4. Setting the scene: the school, the programme, the researcher

This chapter will deal in detail with the school and the programme where the evaluation of this thesis has been carried out. In a first part the school will be introduced and described from a viewpoint that is intended to be as objective as possible using information provided on the school's internet site. The second part will explain the role I have as the person doing this evaluation. Being a teacher at this school myself I believe there are good grounds for me to provide an insight into where I stand within the school and the CLIL programme at hand and what it is that I believe qualifies me to undertake this evaluation.

For reasons of simplicity and readability the school (Sportrealgymnasium Maria Enzersdorf) will from now on be referred to as either "the school" or "the SRG".

4.1. The school & the CLIL-programme

With the Austrian school system in mind, the first important fact is that the school in question is a secondary school that leads the students up to school-leaving examinations which qualify them for further studies at universities. The school comprises lower (years 5-8) and upper secondary (years 9-12), therefore students between the ages of 10 to 18 attend. Students entering the school start in year 5 (age 10) and complete either 4 or 8 years of education at this school. After 4 years (so after year 8) the students can either choose to remain at the school and continue their secondary education for a further four years or decide to leave and continue their education at other institutions with more vocational foci, some of which ending in school leaving exams that are comparable to IB-level examinations (ones that provide university qualification).

Furthermore it has to be noted that the school places its main focus on sports. This is achieved by a comparatively high number of weekly sports lessons and a special subject called "Theory of Sports" during the upper secondary years.

Generally, the national curriculum applies to all subjects that are taught at this school, including "Theory of Sports" which is not entirely exclusive to this particular school but taught at every school with a focus on sports.

With the focus on sports in mind, the following has to be considered: Austrian educational law grants every school only one specific focus ("Schwerpunkt") – so while the SRG has its

focus on sports, other schools have theirs on music, bilingual education, natural sciences, etc. For the SRG this means that the CLIL programme that is being evaluated in the course of this thesis, is not an official focus and therefore under no formal supervision and there are no extra resources provided by the board of education. The CLIL programme therefore is entirely organized and funded independently of any local authority.

In addition to the sports focus that is compulsory for every student attending the school, the school offers two “projects” – namely the EAA-programme (“Englisch als Arbeitssprache” ≈ CLIL) and the IPP-programme (IPP = Internet, Projekt, Präsentation), both of which technically have no legal foundation (since they are not designated foci of the school and its curriculum), but are rather extra features that the school offers to its students. The IPP-project and the EAA-programme exist parallel to each other, but the IPP-project was founded two years after the introduction of the EAA-programme, which was introduced in the academic year 2004/05. The IPP-project aims at familiarising students with information technology and presentation techniques as well as project organization from year 5 onwards.

Again, it is important to realize that all the students receive the same formal education, independently of the programme (either “EAA” or “IPP”) they have chosen. They have the same amount of weekly lessons in every subject and the two programmes are incorporated into the existing lessons. This means that there are neither any specific lessons solely devoted to EAA nor to IPP. The EAA (CLIL) lessons are always part of regular Maths, Biology, etc. lessons and equally the IPP elements are worked into the regular lessons.

In the following sections the CLIL-programme (or: EAA-programme, EAA-project), its aims and internal organisation will be described in closer detail. All the information provided in the following can be readily obtained from the website the “EAA-Verein” (comments on which are to follow in section 4.1.2) has set up⁷. It is connected to the school’s main web presence and provides information about the EAA-project itself, as well as its organisation and, among other things, reports from students on trips that EAA-classes have taken abroad. Furthermore it provides prospective students and their parents with the slides of a Microsoft PowerPoint presentation that is used for parents’ evenings as well as the open day the

⁷ <http://eaa.sportgymnasium.net/wordpress/?pageid=309> , accessed 5 February 2012

school has every year, in order to provide information on the EAA-programme to prospective students and, more importantly their parents.⁸

4.1.1. Organisation and goals of the CLIL-programme

Seit dem Schuljahr 2004/05 bietet die Schule den Neueinsteigern "Englisch als Arbeitssprache" (EAA) an. Dabei wird Englisch in natürlicher Weise als Lern- und Arbeitsmittel in nichtsprachlichen Gegenständen eingesetzt und langfristig ein deutlich höheres Kompetenzniveau im Gebrauch der Fremdsprachen erreicht.⁹

The CLIL programme that is evaluated in the course of this research project has been in place at the "Sportrealgymnasium Maria Enzersdorf" since 2004/05. EAA – "Englisch als Arbeitssprache" at the SRG was devised and implemented mainly by an English teacher who has been teaching at this school for more than 10 years now and by the native speaker who has worked at this school the most time, both of whom will be among the interviewees in the empirical part that is to follow. In the course of the development of the EAA-programme a steadily increasing number of native speakers of English has been employed, the one mentioned above was the first in the school and has seen the development of the programme from its early stages onward.

In practice, there are four to five new year 5 classes (1.Klassen) every year, at least two of which are what is called an "EAA class", which means that two (sometimes 3) of the year 5 classes are run within the EAA-programme, whereas the others are part of the aforementioned second programme offered at the school, the "IPP-programme". Starting from the first year at the SRG (year 5) students get used to having a native speaker of English accompany some of their subject teachers and working along their side. In non-language subjects such as Geography, Biology and Mathematics the subject teacher and the native speaker prepare short sequences (max. 15 minutes) in English. These short sequences take into account the fact that the students have only just started receiving formal instruction in English at the same time. In Austria teaching English as a regular core subject is introduced at the beginning of lower secondary education and only very vaguely included in primary

⁸ http://eaa.sportgymnasium.net/wordpress/wp-content/uploads/2010/08/eltern_1klassen_2010_neu.pdf , accessed 5 February 2012

⁹ http://eaa.sportgymnasium.net/wordpress/wp-content/uploads/2010/08/eltern_1klassen_2010_neu.pdf , accessed 5 February 2012

educations (years 1-4) and even then in many cases very differently, depending on the primary school and teaching staff. This means that the students are, in nearly all cases, not fluent English speakers and therefore CLIL sessions (parts of a lesson) are kept brief for the duration of year 5, and the focus is more on familiarizing the students with very basic concepts and vocabulary of the content subject in English.

Gradually the duration of English sequences in the lessons is increased, also the number of subjects that can use English as working language increases with time. Two new subjects, History and Physics, which are introduced in year 6 (2.Klasse) can also be taught with the help of native speakers. Especially in the case of History this happens very frequently at this school. Year 8 adds Chemistry to the list of core subjects of the students and during year 9 (= first year of upper secondary) the subject "Theory of Sports" becomes a core subject. Another addition to the list of core subjects is a second foreign language (either Spanish or French) starting in year 9. Both "Theory of Sports" and the second foreign language are compulsory subjects for every student, no matter if part of the "EAA"- or "IPP"-programme has to take, since it is part of the obligatory sports focus at this school.

The amount of time that these content subjects are taught in English (with or without native speaker assistance) varies depending on the content subject teachers' readiness and confidence in teaching their subject in English but also on the availability of native speakers.

The school now (April 2013) has 4 native speakers working on different days of the week aiming at providing all the EAA classes with an equal share of EAA lessons a week, ideally working according a regular weekly schedule that both students and teachers can get used to. This means for example that, ideally, one of the three weekly mathematics lessons a class has (e.g. the Friday slot) is a regular EAA lesson.

The description of the target group as well as the explicit "official" goals can also be found in the presentation slides referred to above – I have translated them and put them into a comprehensive list leaving their order as indicated in the presentation unchanged.

Target group: students who have a good command of the German language (not necessarily native speakers of German) and possibly already some command of English; also bilingual students

Goals:

- cultivating students' language competence in English in many diverse areas (everyday conversation as well as more specific)
- assigning importance to knowledge of specific subject-related terminology (only at a later stage)
- complementing the regular foreign language lessons
- intensifying and deepening the competence of the English language through consistent application and use in class but also in everyday school life
- fostering a certain level of pleasure and positive attitude towards foreign languages and the study and acquisition thereof
- swifter and more natural access to the English language
- preparing students for the new qualifications that today's world of work demands
- providing education as required by the Austrian curriculum in both language all the way to the school leaving examinations (Matura)
- preparing students for the Cambridge Certificate examinations¹⁰

4.1.2. Funding for the programme

The native speakers within the EAA-programme are employed neither by the state nor by the board of education but by an organisation, which in its full title is called "Verein zur Förderung der Fremdsprachenausbildung – Englisch als Arbeitssprache am SRG Maria Enzersdorf" – in short and for the remainder of this thesis it will be referred to as the "EAA-Verein". This means that it is an organisation run solely by parents of students attending this school. Its main goal is to maintain the EAA-programme – first and foremost – on a financial level. The native speakers are paid on an hourly basis and this is the major expense the 'EAA-Verein' has every year aside from providing a budget for photocopies that teachers make for EAA lessons, buying English language subject books that serve as aids for the teachers in preparing the lessons and providing financial assistance for students going on school & language trips that EAA-classes embark on (destinations: London in year 8; England / USA / Canada during year 10). The "EAA-Verein" draws its income mostly from payments that every student participating in the EAA-programme has to make every term (so twice every school year). This payment is now € 70,00 per student, parents with more than one child in the EAA-programme only pay once. Additionally, the 'EAA-Verein' has managed to attract regular sponsors who reliably provide important financial aid on a yearly basis.

¹⁰ (cf. http://eaa.sportgymnasium.net/wordpress/wp-content/uploads/2010/08/eltern_1klassen_2010_neu.pdf)

The third main source of income for the “EAA-Verein”, aside from the payments made by the parents and funding from external sponsors is a summer course. This is a programme offered during the last two weeks of the summer holidays before the new academic year commences. During the first of the two weeks students registered to enter the school at the beginning of term (prospective students for year 5) can sign up and get to know the school, its facilities, the English native speakers in a relaxed atmosphere full of games, sports and English language before the entry into real school life starts a week later. The second week is open to students who have completed either their first or second year at the school and the idea is again to strengthen the relations the students have with the native speakers in a relaxed environment with sports and language. The summer course is a component that complements the whole “EAA-experience” and is intended to lower the prospective students’ affective filters before they come to the SRG and to make English as a means of communication more commonplace and natural for all the students (both prospective and current ones).

4.1.3. Admission to the EAA-programme

Since the school focuses on sports and both a profound practical and theoretical education in this field, all students enrolled must pass a physical aptitude test that is one of the main criteria for admission to the school. After applicants have been accepted, they need to declare their choice between the two aforementioned ‘projects’ the school offers: the EAA-programme or the IPP-programme. The further procedure is as follows:

The students enrolled for the EAA-programme are invited to an interview, which is conducted towards the end of the summer term before the new admissions start school. The sole purpose of these interviews is to determine the level of *German* prospective students have, since this is considered a vital requirement for admission into an EAA-class.

Die gute Beherrschung der deutschen Sprache ist eine Grundvoraussetzung für eine Aufnahme in die EAA-Klassen. (Dies gilt nicht für Kinder mit Muttersprache Englisch.)¹¹

¹¹ http://eaa.sportgymnasium.net/wordpress/?page_id=14 , accessed 5 February 2012

This is in slight contradiction to what the Eurydice report mentioned in chapter 2 (2005: 7) suggests, namely that there are “[...] no official admission requirements [in general]”. What the Eurydice reports for the Vienna Bilingual School (VBS) is that an “introductory talk” is part of the admission procedure. At this particular school (the SRG) these talks are used as a means to determine how many of the numerous applicants can be accepted into the limited number of EAA-classes every year group can offer. This means that there are normally between 4 and 5 year-5-classes open for new incoming students and that for as long as the programme has existed there have been more applicants for the EAA-programme than there were spaces available.

The interview comprises two parts: the first is a short reading test for which the time allowed is 4 minutes. The candidates are given statements in German and have to assess their truth-value by ticking either ‘true’ or ‘false’.

The second part is a conversation with a panel of three German teachers. In order to prepare for this, candidates receive a short text (1 page in length) or a sequence of pictures and are given about 15 minutes of preparation time before 3 to 4 students enter the talk with the teachers. The assessment criteria used at this point are abilities of expression, flexibility in lexis and grammar as well as an assessment of how well the students present themselves in this situation. If a candidate passes the first part at a level above average, he or she is exempted from the second part and accepted into the EAA-programme.

The following is an example for the aforementioned sentences of the first part of the German language assessment¹². Translations into English have been added:

Satz	Richtig	Falsch
Hamster können lesen. (<i>Hamsters can read</i>).		X
Eines der beliebtesten Getränke in der Welt heißt Koala. (<i>One of the most famous beverages in the world is called “Koala”.</i>)		X
Vögel können fliegen. (<i>Birds can fly</i>).	X	
Wenn eine Kerze brennt, schmilzt das Wachs. (<i>If a candle is burning, the wax melts.</i>)	X	

¹² http://eaa.sportgymnasium.net/wordpress/?page_id=14 , accessed 5 February 2012

The results of this language aptitude assessment are not made public but are used to create a ranking among the students. Since there is only a limited amount of spaces available in the two (sometimes three, depending on various factors, most importantly a decision made outside of the school, by the board of education) EAA-classes that are opened very year, there are cases of students who meet the requirements of the German language assessment but cannot be accepted into the EAA-classes.

As indicated before, the programme has now been running for one full cycle, which was completed at the end of the academic year 2011/12. This means that the students who were in year 5 when the EAA-programme was initiated graduated in the summer of 2012 and have therefore spent 8 years in the EAA-programme.

4.2. Approaching internal evaluation & validity – negotiating the two roles of the researcher

In this section I want to deal with an issue that needs to be discussed pre-emptively before commencing the next chapter, which will explain exactly how the empirical part of this thesis is structured. The issue at hand is one where ‘conservative’ evaluation theorists would strongly object, namely the one that the researcher of this project is at the same time part of the programme that is being evaluated.

I have taught at the school described in the preceding section for two and a half years now and I have been a part of the EAA-programme previously described right from the start. I started teaching at this school in September 2010 with a nearly full teaching load of Mathematics lessons. With English being a first language for me, I was invited to teach EAA-classes only, which I saw as a challenge right from the beginning. Since then I have taught Mathematics my core subject using English as the medium of instruction in a part of my lessons. Many of the aforementioned EAA-lessons I have taught in cooperation with the school’s native speakers, others I have taught by myself, thereby assuming the role of the core subject teacher as well as the one of the language teacher. It is therefore that I obviously assume an internal perspective in the course of this research project, one that I would like to comment on in a more detailed way.

As already said, I am the researcher evaluating a programme that I am a part of and that comes with certain aspects that have to be kept in mind as well as points of possible criticism.

Looking back to chapter 3, the theory and history of evaluation, one will remember the paradigm dialog (Lynch 1996: 12) and the great disparity between the two conflicting goals in evaluation. Are internal evaluations objective enough? Is this kind of evaluation where the researcher is part of the evaluated even scientific or does the researcher's possible bias cloud any form of objectivity and therefore leads to the research not being considered "acceptable qualitative research" (Lichtman 1996: 189).

Lichtman (2006: 191f.) asks a series of very important questions

Why is the self so important? Shouldn't the self remain outside the research? If the self is involved, can you trust what you read? I believe strongly that the role of the researcher is critical to the work. He or she should not try to remain outside the system. He need not try to achieve objectivity, since that is an assumption of quantitative research and not of qualitative research. (Lichtman 2006: 191f.).

Lichtman puts the question at hand in very straightforward words – the question is, which role "the self" plays in evaluation. Obviously, the research carried out here can only be qualified as qualitative research, mostly because there will be no quantification of any kind in the course of the empirical part of this paper and the interviews. Also, quantitative research is very much driven by the desire for validity. This is a concept that has only been touched in very brief instances so far in the course of this paper. This is also exactly the question that chapter 3 has raised and traced – do evaluations lose their validity if they do not rely on quantifiable results obtained from the measurement of certain preordained criteria but on including individuals' opinions rather than scientifically proven facts.

Traditionally, sound research has been and still is based on the following criteria: *internal validity* (Does the research project give a full, complete picture of the programme in question or does it omit certain aspects?), *external validity* (Can the results of the study be applied to other similar scenarios?), *reliability* (Would a different researcher in the same programme arrive at the same conclusions?) and *objectivity* (Are the results influenced or biased in any way?). (cp. Lichtman 1996: 193).

Lichtman (ibid: 192) "[...] argue[s] that an understanding of the other does not come about without an understanding of the self and how the self and the other connect." She also

claims that “[...] we need to recognize that the researcher shapes the research and, in fact, is shaped by the research [and that] [a]s a dynamic force, she [the researcher] constantly adapts and modifies her position with regard to the research topic [...]” (ibid: 206). She demands that qualitative researchers “[...] should not strive to be objective and look for ways to reduce bias [but] [r]ather, they need to face head on the subjective nature of their role.” (ibid: 202f.).

This summarizes what this section is all about – putting the researcher into perspective with the research and making clear that the position I have within the EAA-programme of the school will affect some of the conclusions I draw from my data simply because I have my own experiences with the programme and my own personal professional opinion on it.

External validity is an aspect that this study is concerned with to a very small degree, for the simple reason that the evaluation conducted in the course of this thesis is carried out for the sole purpose of the completion of this thesis. There are no stakeholders involved in the design of the evaluation other than myself, and none of the stakeholders of the programme have asked me to do this evaluation. The results of this study will, however, be shared with and brought to the attention of the parties involved, but whether or not they will have a decisive impact on the further course of the EAA-programme will remain to be seen – at least this can be said from my subjective perspective today. Also, external validity is an issue not addressed because the EAA-programme as it is carried out in the SRG is, as previously mentioned, not a regular bilingual programme that is the focus of a school. Therefore this demand on a piece of research can be disregarded to a great extent, although I do strongly believe that programmes such as this can benefit from the internal evaluation approach.

Internal validity is another means of measuring the quality of research, which addresses the comprehensiveness of the study and therefore seeks to examine whether or not the researcher has painted a complete picture of the matter. From an internal evaluation perspective I do believe that this demand has been met for the most part. Since qualitative interviews are the main source of information of this study an effort was made to include all the relevant stakeholders and this aim was met. Parents, teachers, the headmaster and students were involved to provide as many viewpoints as possible. The participants for the interviews were chosen either because of their function within the programme or because of an anticipated possibility for differing viewpoints. This was pursued in order to achieve

aforementioned complete picture. However, one party was specifically omitted – no student from the IPP-programme has been asked to be a part of the interviews. This was done for a certain reason that will become apparent in a later part of this section.

Reliability and *objectivity* as briefly illustrated above are the reasons for this entire section of this chapter. The question is whether the researcher's position, my position, combined with my personal and professional views within the programme affect the outcomes of the study and the conclusions drawn from the results. The answer is very straightforward: yes, most likely. But with Lichtman's arguments in mind (2006: 191ff.) I believe that this does not reduce the possible usefulness and validity of this study as long as the demand that Lichtman makes for self-reflection and self-awareness of the researcher within the study is met appropriately.

In the following I want to assume position within the EAA programme and in doing so I want to express my personal and professional opinion on the programme at the school I work at in order to accomplish two things. Firstly, I want to make sure that the position and opinion I have are clearly stated before I commence with the empirical part of this thesis. With the aim in mind of generating outcomes that are biased as little as possible I feel that this is a vital step before anything else happens in the actual study. Secondly, I am also using this section for purposes of reflexion and self-awareness. Lichtman (1996: 207f.) claims, thereby citing DeVault (1990), that

[a]s researchers, we are 'situated actors' and we need to understand the nature of our participation in what we know. We need to include ourselves in our research texts in visible ways in order for the reader to discern our interpretations.[...] Thinking about how [the researcher] serve[s] as the instrument of research is critical as [they] gather data, analyze it, and construct meaning.

I started out being very excited about the chance to teach a non-language subject in English since I saw the value and potential of using a foreign language outside of the regular foreign language lessons on a regular basis. The idea is for students to gain confidence speaking the foreign language, expand their language skills in a more relaxed setting (from a language perspective at least, since in a Mathematics lesson the quality of the language produced is not used for any kind of assessment), speak with native speaker assistants and grasp, revise and/or deepen concepts in the non-language subject at the same time. The benefits need no further explanation I would assume, since it goes without saying that proficiency in a

language is always a valuable asset. My attitude in this respect remains completely unchanged to this day. During my first year of teaching I saw how eager most of the students were to participate in class using English as medium of communication, how readily they worked in lessons with the native speaker assistants and how aptly they could handle even questions in written tests set in English. The linguistic outcomes are very obvious, the average English language proficiency among the EAA-students is noticeably higher compared to the average of non-EAA-students – a perception shared among all the teachers at the school. These results can be seen particularly well in the school leaving exams at the end of year 12 where all students who choose English as an exam subject have to pass a standardized and nation-wide centralized language test that is comparable to the FCE (First Certificate of English) in its design and level.

However, for me as a teacher at this school who sees more than just the students' increasing language abilities, I have, over time, identified a number of setbacks the EAA-programme creates. These points are not influenced in any way by possible results from the empirical part of this thesis, but I was aware of them before, which is why I am listing them here. Again, please be aware, this section has no claim for objectivity or comprehensiveness.

a) After about three months into my first year of teaching I had the first written examinations (in Mathematics students are required to sit 4 written examinations every year, these exams are spread over the academic year, 2 per term; this is standard procedure in every upper secondary school in Austria and not EAA- or school-specific). The results were astoundingly good and upon sharing them with colleagues I got comments like “Well, with your students that is no surprise.” or “I would never get results like that with my classes” and these reactions made me wonder. It did not take long for me to discover that these comments came from colleagues who were mostly or only teaching IPP-classes and not like myself, only EAA-classes in Mathematics. It became obvious that there was a number of teachers who saw (and still sees) a great disparity in aptitude and general level of grades, results, and motivation between EAA-classes and IPP-classes. In the 2.5 years that I have taught at this school, I have only taught one IPP class in English for half a year, which means that I have very limited basis for comparison of results. I have, however, seen noticeable differences in teachers' attitudes among the whole staff, those teaching in EAA groups only, those teaching both forms, those teaching only in IPP classes.

What has to be kept in mind here, is that not all the teachers agree to teach their non-language subject in English (with or without the help of a Native speaker). This generally has two reasons: either they do not feel comfortable enough with their own level of English to do so or they refuse to put in the extra effort it takes to turn a regular (e.g. Biology) lesson into a EAA Biology lesson for free. Let me remind you at this point that the subject teachers receive no additional payment for the EAA related preparation work, financially it therefore makes no difference at all to the subject teacher. These two reasons then lead to a number of teachers being unwilling to teach any lessons as EAA lessons and a clear consequence of this is that in the planning of the upcoming academic year it is frequently tried to put teachers into the EAA-classes who will be willing and ready to do EAA lessons in their subject. Thus it tends to be the same teachers who get EAA-classes or IPP-classes and the fact that EAA-classes tend to score better across the board seems to be widely accepted among the staff and for some this is very frustrating. Teachers teaching in IPP-classes (might) fear that they are seen as worse teachers than their colleagues teaching in EAA-classes, simply because the results their students achieve are tendentially lower. At first this was a problem I saw only among teachers and not at all among students, but again, one has to keep in mind here that for the first two years of my teaching at this school I only taught EAA-classes.

b) Before I explain the second setback that I see within the EAA-programme, I briefly need to explain some idiosyncrasies of the Austrian school system one more time. After four years of lower secondary school students can decide to remain at our school type for a further four years, which conclude in school-leaving exams similar to “A-levels”, and provide qualification to study at universities in Austria. Another option would be to leave school and pursue different educational paths such as more vocationally focussed upper secondary schools which also lead up to aforementioned school leaving exams but place a lot more focus on specific vocational training rather than providing a profound general education with a more academic focus.

This decision that students make during their fourth year (year 8) of their time at our school then leads to the fact that for the following year the grouping into classes might change. Depending on how students decide to continue their education, in some year 8 classes most students might remain at school, in others many students might decide to leave. In any case

the classes for year 9 will have to be set up from students staying and, in very rare cases, students applying from outside who want to complete their upper secondary education at our school after having passed through lower secondary education at a different school. In the past generally there have been 4 to 5 year 8 classes merging into 2 or 3 year 9 classes. When the EAA-programme had been running for four years (so from year 5 up to year 8), it was decided to keep the programme running up to year 12 and the same decision was made for the IPP-programme 2 years later (since the IPP-programme had only been introduced two years after the introduction of the EAA-programme).

There are two last aspects that need to be mentioned before describing the actual problem: Firstly, Austrian educational law requires there to be 60 or more registered students to establish 3 different classes in the same year 9 group. If there are 59 or fewer only 2 different classes may be created. Secondly, the recommended maximum number of students who can be put into one class is 26 but this number can be exceeded to up to 36 students in special scenarios.

The situation I actually want to describe at this point needed some initial explanation and occurred as follows. In the planning of the next academic year it was found that there would be enough students remaining at our school to plan for three classes in the year 9 yeargroup of the next academic year – two EAA-classes and one IPP-class, since both those programmes should be continued in upper secondary as well. The next academic year, however, brought some unexpected change in the number of students for various reasons whose explanations are not relevant at this point. It suffices to say that 10 days into the new school year the number of students in year 9 had dropped below 60 and the board of education decided to close one of the three year 9 classes and continue the year with only two. In order to keep the EAA- and IPP-foci upright the two EAA-classes were merged, leaving a class of 35 students. This caused massive protest among parents, students and teachers, which in turn led to the decision to move 5 students from the large EAA-class over to the smaller IPP-class in order to create an equal number of students in both classes. It was determined that the 5 students to be moved from one class to the other would be those with the lowest grade averages in the previous year. This decision lead to further resistance so that after some more intervention from parents with the board of education and higher

institutions it was decided to leave the EAA-class as large as it was (35) and maintain the EAA-focus for all of them.

The result of this was a lot of talk throughout but also outside of the school and IPP-students who had heard of the process by which basically the 5 weakest EAA-students would have to move to the IPP-class were raising the question that – to my knowledge – had only seriously been raised by teachers so far and that was: “Are we less intelligent than EAA-students?” “Is that why the 5 weakest would have had to be moved over to the IPP-class?” Needless to say that this whole situation caused quite some turmoil, not only among the two affected classes and for the 5 students who were supposed to switch classes from one day to the next and leave their classmates of the last 4 years behind. It has generated, I believe, more perceived disparity between the two programmes among all parties – students, parents and teachers. This, to my mind and that of some colleagues, might lead to the detrimental effect that IPP-students are less motivated because they might feel inferior to EAA-students.

The two points (a and b) described so far are reasons why I personally have become more critical of the EAA-programme as it is now, but not because of how the programme itself is run or organized (I still firmly believe in the aims and goals the programme and its organisers have) but because of its juxtaposition with the other programme which has a lot of merits too.

For the current academic year a new solution has been found and that comprised creating a year 9 class that is run as a combination of both programmes – in essence offering parts of both programmes (these students are also charged a lower EAA-fee than the full EAA-classes). In this way the problem of a too large EAA-class could be avoided.

c) The third point I want to bring up in the course of explaining my personal view of the EAA-programme is the fact that I feel that the quality of the programme is suffering slightly due to a slowly falling number of native speaker assistants who come to work at our school. With an increasing number of EAA-classes throughout all the year groups the demand for native speakers in EAA-lessons has risen significantly and – from my perspective – the organisers of the programme are faced with some difficulties maintaining an equal amount of EAA-lessons with native speaker assistants in all the EAA-classes, especially in the upper classes. I do, however, believe that this is an issue that could only be current for a while; the organisers of the programme are constantly looking for new native speaker assistants, so this difficulty

might be a thing of the past even before this thesis is completed. What has to be pointed out in this context is that native speaker assistants are available comparatively easily for this school due to its proximity to Vienna, the country's capital. Schools in more rural areas are therefore less likely to find equal numbers of native speakers.

At this point I will refrain from making any suggestions for the future, I will add some possible thoughts in the conclusion, thereby also taking the results of the empirical part into account.

In closing, chapter 4 has intended to set the scene which has been done on a number of levels – the reader has been introduced to the school and to relevant aspects of the Austrian educational system and law. The programme that will be studied in the course of the empirical part of this paper was introduced and its goals, aims and internal organization were thoroughly explained in order to give the reader a comprehensive picture of the situation. The second part of this chapter had the main purpose of inscribing the researcher into the scene and explaining his position within the research but also within the programme. It was important for me as the researcher to highlight my own personal and professional opinion at this point so that the reader can later detect possible bias in the interpretation of the results or is more easily able to distinguish between the researcher's opinion and the ones of the individuals interviewed. I have also briefly touched the issue of validity in qualitative research, studied the researcher's role in qualitative research and this will now be complemented and added on to in chapter 5. This will introduce the reader to the methodology used in the empirical part of this research project in order to clarify exactly how the results have been found and thereby justify possible conclusions drawn from the study.

5. Methodology

This section will discuss the part that can be labelled the “how” of this research project. As such, it will give a detailed insight into how exactly and based on which principles the empirical part of this study has been carried out. It will therefore draw from the theoretical issues of evaluation, especially language programme evaluation, presented in chapter 3 but will be complemented with ideas from a relatively new field of research, that is especially popular in education, namely ‘Action Research’.

Action Research is defined as “[...] a reflective process that allows for inquiry and discussion as components of the “research” (Ferrance 2000: 1) and “[...] in which participants examine their own educational practice systematically and carefully, using the techniques of research.” (ibid: 1). In other words, it is a technique that can be compared to the concept of internal evaluations simply because the researchers in this kind of research are the practitioners themselves. Without wanting to go into too much detail, it can be said that generally there are four types of educational action research, which can be distinguished according to the level and scope (starting from research affecting one individual teacher reaching to district-wide action research). The one most relevant for this study is “school-wide action research” since the CLIL-programme evaluated here affects nearly all of the individuals in the school. Some of the leading Austrian researchers in this field, Altrichter and Posch have published a guidebook (2004) on how to carry out action research studies and the section on qualitative interviews is one that has provided the basis for the development of the questions for the interview conducted in the course of this study.

With this in the background, I will define what the empirical part of this thesis aims to do, what I intend to find out and which means I will use to achieve these specified goals.

5.1. Turning theory into practice – defining the evaluation

In this section I would like to briefly reflect on the aspects of chapter 3 (Theory of Evaluation) which have shaped the development of the evaluation of this project. I want to do that by referring back to selected parts of that chapter and joining them into the idea of a concept of evaluation that will be pursued in the course of this thesis.

In his “countenance model” Stake (1967) proposed that an evaluation has to be tailored to the location where it is carried out and therefore rejected the notions of comparability and

transferability – these are both ideas that are not sought to be followed in the course of this paper. Furthermore he suggested that – what he referred to as – antecedents (1967: 527) be included in any study of a programme and this is an aspect that will be included here too. All the participating interviewees will be asked to recall the very beginnings of the EAA-programme at the school, how and why it was designed and introduced and what brought about the idea in general.

Parlett & Hamilton (1972) introduced “illuminative evaluation” into the field of research which “[...] seeks to generate hypothesis [...] from within the data [...]” (Bennet 2003:27) which is a further aspect to this study, namely that possible interpretations and conclusions will only be drawn from the data provided by the interviews and not according to any preordained criteria. Again, as Stake did before them, Parlett & Hamilton point out that the results and studies themselves need not be comparable to studies at other places. This issue was also addressed in the previous chapter when the notion of validity was discussed. Parlett & Hamilton also introduced the concept of triangulation. As explained in chapter 3 this comprises the idea of drawing from multiple sources when gathering the data relevant for the evaluation. When compiling the data for this study, i.e. when conducting the interviews, attention was paid to incorporating all the relevant stakeholders and to ensure a variety of viewpoints within the stakeholder groups.

Stenhouse (1975) believed that evaluations carried out internally are what should be mostly used in what he called “curriculum betterment” (1975: 99). This is an aim that is not at the forefront of this research study but rather left a possible option, although it has to be said that any “betterment” originating from this study would probably affect the organisation of the programme and its standing in the school and with the stakeholders but not directly the curriculum itself.

Guba and Lincoln (1989) introduced the concept of “**Fourth Generation Evaluation**” which basically consists of the idea that every individual with their views and perceptions of something assumes a unique perspective that forms part of the “truth” which is constructed by all the participants.

[...] precisely because they are the product of human thought, consensual constructions are subject to human error; none can be considered “true” in an absolute sense, nor even an approximation of “truth.” (Guba & Lincoln 1989: 9).

By propagating the notion that there is no such thing as a “single, universal, objective truth” (Kiely & Rea-Dickins 2005: 40) Guba & Lincoln paved the way for evaluations that are driven by the individuals involved in the programme to be studied.

However, they also argue that this construction is only meaningful if put in perspective with the specific context. In the course of chapter 4 a large part of the perspective demanded by Guba & Lincoln has been laid out. The school and the programme to be evaluated were thoroughly described along with the special role of the researcher within the programme and the evaluation. Guba & Lincoln (1989: 71) claim that merely by examining the word “evaluation” linguistically they see the inherent word “value” and therefore raise the question whose values are being “e-value-ated” in such a process. For the study here the values of the stakeholders within the school are the only relevant parties, apart from the parents who are technically not in the school but very closely linked to the EAA-programme and its maintenance and up-keeping.

Guba and Lincoln suggest a path consisting of 9 steps to be followed in the course of conducting a fourth generation evaluation. They, however, do not suggest following it in a linear manner but rather in a cyclical one. In the following I want to list those 9 steps and comment on each of them with reference to the study to be carried out here and indicate how these demands have been addressed in the course of my evaluation and give reasons for possible deviation. The steps are taken from Guba & Lincoln (1989: 72ff.) and are consecutively numbered.

(1) Identifying the full array of stakeholders who are at risk in the projected evaluation

This has already happened in chapter 4 – the stakeholder groups are:

- a) *students*: those participating in the EAA-programme
- b) *teachers*: both those teaching EAA-classes and those who do not
- c) *parents*: here one has to differentiate between parents who are running the “EAA-Verein” mentioned in chapter 4 and those who are merely parents of students in EAA-classes with no affiliation or active participation within the “EAA-Verein”
- d) *founders and heads of the programme*: these are also teachers but play a significantly more important role in running and maintaining the programme
- e) *native speakers*: here a person has been interviewed who is both a native speaker and one of the directors of the programme; he has been part of the programme from the very beginning
- f) *headmaster of the school*: he played an important role in the process of founding the programme and is now involved in all personnel related matters (i.e. which teachers teach which classes)

Representatives of all of these groups have been selected for an interview. Guba and Lincoln recommend for the evaluator (1989: 72) to “[...] always be open to the inclusion of new stakeholders whenever and however they come to the evaluator’s attention.” This has been maintained right from the beginning of the interview cycle. Initially, the plan was to interview 5 persons and in the course of the first interviews the need to include more representatives from certain stakeholder groups became evident.

(2) Eliciting from each stakeholder group their constructions about the evaluand [(=what ever is being evaluated),FS] and the range of claims, concerns, and issues (CC&I) they wish to raise in relation to it

This is what I consider one of the most vital steps of the evaluation. Here Guba and Lincoln (1989: 72) demand that the “[...] elicitation must take place in an open-ended way, to guarantee that it is an emic – an insider’s – view that emerges rather than an etic – outsider’s – view.” Therefore the questions in the interviews are designed in a way that will allow the interviewees to express candidly their position within the programme and their respective claims, concerns and issues (henceforth CC&I). More on question design is to follow at a later stage.

(3) Providing a context and a methodology through which different constructions and different claims, concerns, and issues, can be understood, critiqued, and taken into account

The methodology referred to here by Guba & Lincoln is what they refer to as the “hermeneutic dialectic process” (1989: 149).

It is *hermeneutic* because it is interpretative in character, and *dialectic* because it represents a comparison and contrast of divergent views with a view to achieving a higher-level synthesis of them all, in a Hegelian sense. [...] The aim of this process is to reach a consensus when that is possible; when it is not possible, the process at the very least exposes and clarifies the several different views and allows the building of an agenda for negotiation. [...]

In a nutshell, Guba & Lincoln suggest the formation of a hermeneutic dialectic circle that allows the construction of meaning within a stakeholder group before then proceeding to comparing the group constructions and their claims, concerns, and issues. What the hermeneutic dialectic circle includes is that the evaluator builds a construction after every interview before moving on to the next interview and then creating yet another construction of meaning. In the evaluation here this has been carried out on a considerably smaller scale.

During the first couple of interviews notes were taken regarding the effectiveness of certain questions, how well interviewees responded, whether or not certain aspects to the issue were overlooked when preparing the question guideline, etc. but there was no building of constructions between the interviews, at least not in written form.

(4) Generating consensus with respect to as many constructions, and their related claims, concerns, and issues, as possible

First this will be done within the stakeholder groups (unless there is only one representative) before then cross-referencing them with other stakeholder groups. What Guba & Lincoln suggest here is dismissing any item where full consensus can be achieved. They propose this with the “agenda for negotiation in mind” (see step 5), with negotiation being what adds the aspect of 4th generation to evaluation. Since the process is a circle, which is implied by step number 9, namely to “recycle an evaluation” only those items where no full consensus can be reached are placed on the agenda for negotiation, so that a re-run of the circle can begin.

(5) Preparing an agenda for negotiation on items about which there is no, or incomplete, consensus

In this evaluation I will propose something that is similar to an agenda for negotiation and this will include a summary of the items where consensus between the stakeholders can be established, as well as enlarging on and pointing out areas of differences. The latter will then be complemented by suggestions for future developments.

(6) Collecting and providing the information called for in the agenda for negotiation

This would in most cases require a new set of interviews with stakeholders on the issues where no, or incomplete consensus, has been achieved.

(7) Establishing and mediating a forum of stakeholder representatives in which negotiation can take place

(8) Developing a report, probably several reports, that communicate to each stakeholder group any consensus on constructions and any resolutions regarding the claims, concerns, and issues that they have raised

(9) Recycling the evaluation once again to take up still unresolved constructions and their attendant claims, concerns, and issues

Steps 7 through 9 do not really apply to the evaluation in this research project. Neither stakeholders nor the evaluator demand for a consensus to be reached that is as complete as possible. On the contrary, I believe that listing a number of inconsistencies in opinions and

perceptions is a means to catch the attention of possible readers of the results of this study. Furthermore, there will be no report for any stakeholder specifically, but there will be one complete report accessible to any stakeholder upon request. This procedure was agreed upon with every interviewee individually upon completion of the interview.

This section was meant to provide the necessary theoretical background from the theory of evaluations as well as the more in-depth study of Guba & Lincoln's (1989) concept of fourth generation evaluation, which is the most guiding principles of this study. The next section takes a closer look at the exact questions of the interviews and gives a brief reflective insight into the choice of the individual questions.

5.2. The interview questions

In the following the process of determining the interview questions will be explained and traced using guidance from the aforementioned field of action research. Altrichter & Posch (2004) address the topic of interviews within action research in their book that serves as a guidebook for action researchers as to how to approach an internal evaluation.

One of the most important reasons for choosing interviews as the main source of information for this study is very effectively explained by Altrichter & Posch (2004: 164):

Die spezielle Stärke von Interviews liegt ‚in der zweiten Frage‘. Auf die erste Frage bekommen wir oft ‚Oberflächen‘-Antworten, die diejenigen in einem Fragebogen entsprechen und bei denen man sich bei der Auswertung oft fragt: *Und was meint der Befragte eigentlich damit?* Das Interview bietet die Möglichkeit diese zweite Frage zu stellen und so die Gedanken, Einstellungen, Wünsche, Begründungen und Haltungen zu erschließen, die „hinter“ dem aktuellen Verhalten stehen.

This sums up the main advantage of conducting interviews – the possibility to ask a follow-up question, a second question that allows clarification of the answer to a first question, which might not always be as conclusive or allow correct and immediate interpretation and elicitation of the interviewee's claims, concerns and issues. It is through interviews that I felt it best possible to arrive at and interpret honest, easy to understand data that then lends itself well to concepts such as triangulation and hermeneutic dialectic circles.

The interview questions and the individual setting of the interviews were intended to allow for enough freedom on the part of the interviewee.

In the design of the questions it is recommended to start an interview with open, non-invasive questions – these are questions that allow the interviewee to ease into the situation and give “easy” answers. Furthermore Altrichter & Posch (2004: 166) point out that the sequence of questions is vital and that important, more detailed questions that are at the heart of the evaluation must not be kept for the end of the interview. Another important issue raised by Altrichter & Posch (2004: 168) is that the interview question guideline developed for the interviews is not binding in its order – it should always be on the part of the interviewee to develop the interview with only little intervention from the interviewer (who should not be talking for more than 10% of the interview). Another recommendation provided is to never interrupt the person interviewed or bring about a change of topic without knowing that the interviewee has finished saying what he/she wanted to say.

I now want to move on to presenting the questions used in the interviews of this evaluation. They will all briefly be commented on and reflected upon. All but one interview were conducted in German, all the interviews started with a few preliminaries such as asking the interviewee for how long they have been connected to this school (be that as a teacher, student or parent), explaining to them the purpose of the interview, etc. In the following the questions used for interviewing the teachers and the parents involved are listed, for the students there is a modified version of the questions. Brief comments and rationale on those are to follow.

1) When and how did the idea for “Englisch als Arbeitssprache” come up?

This was chosen as the opening question firstly for reasons of chronology and secondly because it is a very open question that just asks the interviewee to remember as much as possible about the initial events of the EAA-programme. It is also meant to elicit possible reasons and motivations of the different parties (stakeholders) to do so. Furthermore it allows the interviewees to “get a feel for” the interview situation and does not require them to assume a critical position. They are just asked to recount their recollections of these events if possible (some might not have been affiliated with the school or the programme at the time).

2) How was this idea first implemented?

This question is in fact a follow-up to the first question and is only relevant if the first question can be answered. It aims at the transition between the plan for the programme

and the actual first year of the programme and slowly invites the interviewees to not only remember but to start making evaluative comments, maybe making remarks on differences between the two steps of the development, or highlighting certain events or unexpected problems that had to be dealt with during the initial phase.

- 3)** Looking back to the time the programme was first devised and implemented 8 to 9 years ago, what were the goals/expectations/aims of the programme? What were your personal expectations at the time?

This is, in fact a question that is twofold. First it asks the participants to give a reasonably objective answer and consider goals, aims and expectations of the founders of the programme and only in a second step remember their personal attitude towards the introduction of the programme. Moreover, the results of this question will complement the data that was presented in chapter 4, where the official website of the “EAA-Verein” was used as a primary source for the programme’s goals and aims.

- 4)** Have these expectations/goals been met? If yes, in how far?

This is a question that is meant to be answered entirely subjectively – when asking the question that is emphasized, that is made clear that what I am looking for is the individual interviewee’s viewpoint and not an opinion influenced by an intention on the part of the interviewee to assume a more objective role.

- 5)** What has worked out? Which aspects of the programme would you maintain?
6) Were there any problems that arose and if so, which were they? How were they approached/solved?

Again these two are questions, as the previous one, that solely target the interviewee’s perspective. Making sure that the interviewee understands that it is only a subjective answer that is relevant here ensures a greater diversity of thoughts and possible critical viewpoints.

- 7)** (a) Which challenges do you see the EAA-programme has posed for you personally?
(b) Which challenges do you see the EAA-programme has posed for the school as a whole?
8) How is the quality of the programme maintained? Are there any quality standards for this particular programme that are being adhered to?
9) How is the EAA-programme situated and placed in the context of the whole school and what it has to offer? In other words, what role does the EAA-programme play officially and what role does it actually play in regular day-to-day school life?

10) What can a prospective student of an EAA-class expect? What can parents of such a student expect?

For these questions it will be interesting to see how the answers differ. Those interviewees who are mainly involved in the organization of the programme and those who teach hardly any or no EAA-classes are expected to have diverging answers to these questions. These questions will very likely also show how effectively the programme and its inner processes are communicated throughout the school.

11) Is the future development of the EAA-programme currently being worked on? How do you see the future of the EAA-programme?

The reasoning for the first of the two parts of this question is similar to the one of question 8. Its purpose is to determine how well these notions are being communicated within the school and how much somebody who is part of the school but not necessarily regularly involved in the programme knows about these internal procedures of the programme.

The second part is again one that seeks to elicit diverging viewpoints, as previously explained.

12) Which piece of advice would you give to schools that are planning to introduce a similar programme? What recommendations would you give a fellow teacher, who wants to start teaching within the EAA-programme?

In order to bring the interview to a logical and chronological end this last question was included because it allows and also in a way asks the participants to provide a summary of what they have said, thereby possibly repeating aspects that are particularly important to them.

13) Is there any issue/topic that I have not asked you anything about but you consider important to mention?

The purpose of this question is self-explanatory.

The questions used for the two students interviewed are significantly different, especially because the viewpoint is considerably different from a student's perspective. As explained in previous chapters both students are students from EAA-classes and no students from IPP-classes was asked to participate. The questions aim, even more so, at the subjective reality of the students, their perceptions and their feeling of how well the programme is working for them, i.e. how much they are noticing the real effects on their language skills. Also, the students were specifically asked about their views of the difference between the programme

they know and the other one (IPP). The students were also asked for their view of a possible future of the EAA-programme and for any possible suggestions for innovation.

5.3. The interviewees

In this last section of chapter 5 the individuals asked for interviews will be introduced briefly. Also their positions in relation to the EAA-programme and within the school will be mentioned.

As indicated in section 5.1. when Guba & Lincoln's steps for a fourth generation evaluation were explained there are generally three stakeholder groups (teachers, students, parents). I have subdivided these groups into subgroups with more attention to detail and position within the programme. The interviewees will not be referred to by their names but only by their initials. These will occasionally be used throughout the next chapter, which is why I want to highlight and explain these in more detail.

5.3.1. Teachers

Firstly, there is the headmaster of the school, the initials used for him are **WS**. He has been the head of this school for about 10 years now, which means that he has seen the EAA-programme evolve from its inception to its present-day status.

Then there is the official head or director of the EAA-programme, also referred to as the EAA-coordinator. His initials are **AZ** and he has been an English and History teacher at this school for over 10 years now and has been the driving force behind the programme from the very beginning.

Then there is **TW** who is the native speaker who has been at this school the longest. He was the first native speaker participating within the programme and was taken on board immediately after the decision was made to create such a programme. He has been at the school for 9 years now, is originally from the UK and has taken up university studies in Austria to get a local teaching degree. After a couple of years working only as a native speaker, he was asked to teach English classes of his own due to the shortage of teachers in that area of Austria. He has therefore been in the relatively unique position of being a teacher as well as a native speaker at the school. His importance in the administrative part of the programme, its coordination, has steadily grown and he is now the organizer of the

summer courses and in charge of devising the timetables for all the native speakers and creating the liaison between the subject teachers and the native speakers.

The next teacher interviewed is **PS**. She has taught History and Geography at this school for over 15 years now and has been heavily involved in the teaching of EAA-classes from early on. She was invited to be an interviewee for a very particular reason: she holds an interesting position at the school. For one part she teaches many EAA-classes in her content subjects but at the same time she is the programme director of the IPP-programme. She therefore has a very unique position and her insights and viewpoints are therefore expected to be highly interesting.

The last teacher interviewed is **AL**. She has been a part of the staff for nearly two decades now and has not taught any EAA classes in her content subject (Mathematics) yet. She was selected to provide critical arguments from a perspective outside the actual EAA-programme.

5.3.2. Students

The first student interviewed is **VL**. She was in year 5 (1st class) in the first year of the EAA-programme's existence. She graduated from this school in June 2012, the interview was conducted 5 months before her graduation, which means that at the time of the interview she was in final year of the first complete cycle of the EAA-programme.

The second student **MB** is in year 10 (6th class) at the time of writing, during the time of the interview he was in year 9 and he attends the class that was the result of merging two year 9 classes (see section 4.2.). This is the reason why he was chosen for these interviews.

Both of these students were a part of the EAA-programme at the time of the interview and have at least 5 years of experience with the programme.

5.3.3. Parents

The first parent interviewed assumes a double role as well. **IB** is the head of the EAA-Verein and the mother of a student who is due to complete his secondary education in June 2013. She therefore has some insight as a mother but also into the organization of the EAA-programme.

The second parent is “just” a mother of a student who had been part of the EAA-programme for 4 years at the time of the interview. She was picked to provide the “parent only” perspective, one that could be constructed while the programme was already fully established.

5.4. Data analysis

This section will briefly describe how the data analysis has been carried out and how the findings will be presented in the next chapter.

The interviews have been transcribed and sorted according to stakeholder groups (e.g. students, parents, teachers). It was my intention to provide a summary of the findings in a twofold manner – one by stakeholder group and another by categories. The first summary (by stakeholder group) essentially comprises the answers to the interview questions sorted by category. Certain category labels were assigned (through underlining in the text) according to the content and this was done to put certain statements in order. This became necessary because some aspects were discussed on more occasions, i.e. when answering different questions. The main purpose of this section is to provide a “narrative” that summarizes the stakeholders’ positions.

The second summary was intended to present the findings according to categories only and not by stakeholders. It is a means of connecting all the claims, concerns and issues (CC&I – see section 5.1.). When stakeholders are mentioned it is only to point out and explain differences in opinion and viewpoints.

The final step is to present what is previously (section 5.1.) called the “agenda of negotiation”. According to Guba and Lincoln (1989) this comprises all the issues and categories where diverging opinions and perceptions have arisen and need further discussion and attention on the part of the programme directors in cooperation with all the stakeholders. The issues on this “agenda” have been summarized by the researcher in a concise way in order to be presented to programme directors and the headmaster of the school. It will include the most relevant issues & categories, i.e. those that seem most important to all the stakeholders.

The purpose of chapter 5 was to familiarize the reader with the design of the evaluation to be carried out here and to provide information as to which strands of theory I have chosen to adhere to and why.

It should now be obvious that pursuing Guba & Lincoln's approach is the right way to go, because it allows room for all the individuals involved in the study. The goal is not to provide mono-perspectival results that allow judgements on how effectively the programme is running and achieving its goals. The goal is more to achieve an impression of the current status quo. This status quo is an amalgam of viewpoints, that is the result of triangulation and comparison and contrast. This status quo should provide information of how people affected by the programme currently see the programme and then move on to providing certain suggestions and views of the future.

6. Findings

In the following the results from the interviews will be presented. The path chosen is to present the results by stakeholder group. Within every stakeholder group categories will come up and the findings will be summarized and worded as objectively as possible. At this point again I would like to point out that this part of the analysis will solely consist of recounting the things that were said. Conclusions that will be drawn afterwards will logically be influenced by the researcher's viewpoint and experiences. The second part, as already indicated, will take a closer look at the individual categories, study them how controversial the stakeholders' opinions are and take a closer look at the ones where there is less agreement. In the third and final part I will then try to paint a picture of the entire theoretical as well as empirical part of the study by integrating the results with what has been said about the theory of CLIL, the recommendations given for the implementation of a CLIL programme, as well as the previous presentation of the school and the programme and the viewpoints of the researcher.

6.1. Findings by stakeholder groups

6.1.1. Students

Two students were interviewed, one of them in her final year, only months before graduation (VL) and one of them (MB) having spent a bit more than half of his time at this school, having completed lower secondary and just started upper secondary.

Both students make statements regarding the entrance exam: they had to pass a reading screening in German and determine whether or not the sentences were true or not and the more sentences they completed in a certain amount of time the better. Neither of them recall a speaking part to the admission exam. Both remember a clear difference between the sports exam, which they both consider as the real test that had to be passed to get a space in this school and the other exam at a much later time (in June) that was a prerequisite for EAA-placement.

They were then both asked to look back on their time with EAA and comment on the intensity (how frequently EAA-lessons took place), which subjects and other relevant experiences made in their years within the EAA-programme. Both state that in lower secondary the programme was very intense and regular, most content subjects on the

timetable had at least some EAA teaching in it. Most of the EAA-teaching mentioned by the interviewees occurred in the subjects Geography, History, Biology and Mathematics. Both of them remember the strategy that content that was at first taught in German, through normal standard teaching was then revised and complemented by English, by introducing certain content-relevant technical terms (e.g. for fractions in Mathematics) and by discussing the topic in English as a summary. One of them also recalls having English lessons where they were split into groups and worked with a native speaker assistant outside the classroom. Both report that the intensity of EAA-teaching decreased over time but the change especially noticeable by them was between lower and upper secondary, where especially in Mathematics the lessons were now mainly in German. They both comment on the fact that EAA-teaching really depended on the content teacher they had and how ready and able they were to teach their subjects in English and maybe even without the help of a native speaker. They both also attribute this decline to the rising number of EAA-classes every year and a seemingly stagnant number of native speakers. One student notes that he perceived the EAA-intervals as a welcome break from standard teaching. They both remember that they were asked to keep a vocabulary log with all the different subject-related words and that the content of this log was regularly checked by teachers.

Assessment in EAA is mentioned only very briefly – the student reports that any content assessment carried out in English was rewarded with extra credit and that students could always chose to take an oral exam in English or in German. In written exams English questions always resulted in extra points added to the total score of the test.

Another aspect brought up right at the beginning when looking back into the first years of EAA, was that right from the start there was a “real” EAA-class and a “test” EAA class (with less EAA exposure). The student (VL) started out in year 5 in the “test” class and moved to the “real” or “full” EAA class (with more regular exposure) because she was a good student and felt she could learn more there and also because she got along better with the students from the “real” EAA class.

An aspect both enlarged on quite a bit is the perceived benefits of the programme. Both students feel that they are much more confident when using the foreign language than friends who are equally old. They feel, especially when abroad (London trip in year 8, Canada trip in year 10, or private travelling), that speaking English and interacting with native

speakers of English is very natural and not at all challenging for them. They attribute this and their general confidence in everyday conversation in English to the EAA-programme and both express how happy they are for having chosen this path in their education.

The more experienced student also comments on the level of English of some of the content teachers. She mentions that some of the content teachers lacked the English and also had bad pronunciation, which, especially during the high time of the students' adolescence lead to some disciplinary but also motivational problems when some students felt their level of English was higher than the one of the teacher. She points out that teachers need to be self-confident enough to not let language deficiencies stand in their way. She compliments some teachers for their willingness to work with native speaker assistants even though it did seemingly not come easy for them to run the lesson in English and states that what it takes is self-confidence in order to run a successful EAA classroom even when their English language skills are not perfect.

Upon being asked about how the EAA-programme could be improved a number of issues are raised. However, both point out immediately at the beginning that they have a good general impression and are happy about the results of the programme (their language skills).

An issue that is raised by the more experienced student (VL) is that teachers should be prepared better for the challenges of teaching EAA classes so they have the necessary confidence and knowledge. She also points out that some teachers need to remember that EAA is not about endless lists of vocab or the teaching of grammatical structures in the EAA-classroom (both these should take place in the regular English class) but about the ability to speak confidently. She adds that this increased preparation would lower inhibitions on the part of the teachers and therefore add to student motivation.

The number of native speakers at the school is addressed. The number and frequency of EAA lessons declines on average the higher up the students go – in upper secondary they are much less frequent than in lower secondary. This is attributed to a lack of native speakers which affects some content subject teachers who then are not willing to teach EAA lessons without the help and support of native speakers. Both students report having had consecutive months without a single EAA lesson with a native speaker.

The questions that addresses the two different programmes run at the school brought quite different answers: the older, more experienced student (VL) points out that the difference in

the level of English between her class and the other class of the same year group (a “normal class” – VL) is phenomenal and that the general achievement level is quite different. She attributes that in part to the fact that content and language were integrated and that having to process the content material in the foreign language has always made it easier for her to memorize and retain certain content information. She also says that she feels that the attitude that students of the EAA-programme are brighter and smarter than the other students seems to have become an accepted fact. She does point out that students of her class do lack certain presentation skills, which is a shortcoming that she attributes to the fact that these skills were never trained as much due to the concentration on the EAA-classes.

The younger student mentions no such feeling between EAA- and IPP-students. He says that he perceives no difference between the two programmes and also not in the attitude towards each other. He does, however, point out that a logical difference for him is that while the students of the EAA-programme have an advantage in English, the students of the IPP-programme excel in presentation skills as well as project planning and implementation.

Both suggest a mixture of both programmes for the entire school. They express a desire to learn the skills taught within the IPP-programme as well as the EAA-programme and see an amalgam of the two programmes as a good way to go.

6.1.2. Parents

Two parents were interviewed, one of them a mother of a student who has been part of the programme for 7 years (IB) and the other is the mother of a student in the programme for 3 years (AH). IB is also the head of the previously mentioned EAA-Verein.

Both claim that the reason for enrolling their child in this programme was the expressed wish to ensure a quality foreign language education for their child. They consider fluency in English a requirement in nearly every job today and therefore they both discuss the importance to foster that skill. Sports was the driving force to apply for a space in this school, but EAA was an additional feature which, for one of the two parents, would have been a decisive factor for withdrawing an application had the child not been accepted into the EAA-programme.

Expectations when applying for the space in the EAA-programme are discussed by both parents. It was expected at the time that the students would be educated in two languages

and English would be learned from the immediate beginning. What both expected was that almost half of the content subjects would be taught in English on a regular basis (e.g. 1 of 2 weekly Geography lessons in English). One parent expected project work within the EAA-teaching (drama performances, etc.). The expected focus was on fluency and not on an abundance of topic-specific vocabulary. Also it was considered as a vital expansion of and addition to the regular foreign language classes.

For both parents these expectations were met to a large extent: both see that their child is confident in using English in everyday situations without any inhibitions, that the “barrier” had been overcome. One parent (AH) comments that in the first couple of years the provision of native speaker lessons was less than expected and that the situation had been improved after some inquiries had been made. The subjects EAA was mainly taught in were History, Geography, Biology and partly Mathematics. One parent comments that she expected EAA across the board and suggests EAA in Physical Education as well as Music classes.

IB points out that the financial aspect to having a child participate in the EAA-programme is not to be underestimated. 70€ per child per term, plus costs for language trips must be factored into the decision when enrolling but she considers the effect the programme has had on her child good value for the money.

Both parents report that they never perceived EAA as an additional challenge for their children. They both maintain that their children never had any difficulty because of the EAA programme but thoroughly enjoyed the challenge. One parent (IB), however, does point out that she believes that EAA requires increased cognitive effort. She also claims that the cognitive demand regarding the use of English in class always depends on the teachers of the content subject. Some follow a more playful approach while others design more demanding courses.

Regarding the entrance test, both mention the reading test, one mentions a test of logics and a short talk in English between an English teacher and the prospective student. One comments that it is in a way an aptitude test that determines whether or not the child’s level of German is high enough for the it to be able to handle the increased cognitive demand of the EAA-programme.

IB, being the head of the EAA-Verein is asked for a brief explanation of what the EAA-Verein does. It was originally founded because the EAA-programme needed a legal foundation, especially for the financial aspect to it. The main task of the EAA-Verein is to collect the 70€ fee that has to be paid for every student per term, administrate this money, buy teaching resources with it, as well as paying the native speakers (on an hourly basis) and providing financial assistance to those students who could otherwise not afford participating in costly language trips abroad. The members of the board of the EAA-Verein meet on a regular basis (2 to 3 times a year) and the coordinating teacher gives his report. The EAA-Verein serves a mainly financial function and does generally not interfere with decisions that affect the organization of EAA-teaching in class, these are left for the coordinator to handle. As the representative of the EAA-Verein she reports that there is hardly any feedback from the part of the parents to the EAA-Verein and that attendance at board meetings, to which all EAA-parents are invited, is generally low. She interprets that as parental approval for the programme and that people would show up and give feedback if there was something negative to be discussed. For her the challenge in the EAA-Verein is making sure that the students get what they expect – native speaker assistants in as many classes as possible. The EAA-Verein also has a presence in the world wide web, but she states that this is not a vital tool in making the EAA-programme known. On the contrary, she says that the EAA-Verein do not pay much attention to it because the number of applicants is very high every year as it is, that this is hard to cope with and that the best propaganda for the programme is by word of mouth.

As far as the situation regarding the coexistence of the EAA- and the IPP- programme is concerned one parent notes that she knows nothing about the IPP-programme apart from its focus, while the other questions why the IPP-programme even had to be introduced. She feels that a situation of competition could have been avoided here. She has heard of rumours from parents who claim that “IPP students have bad teachers” which she considers ludicrous. She does also mention, however, that there definitely is competition between the two programmes.

Regarding the future of the EAA-programme both parents point out that the programme should be continued the way it is but to ensure more EAA-lessons in upper secondary by investing more into hiring new native speakers. It is pointed out that the number of EAA-

lessons decreases in upper secondary classes due to a lack of native speakers. Another aspect to this is also ensuring that the content teachers stay with the programme and are constantly motivated and invited to be a part of the programme in order to stop the programme from coming to a halt. On a side note both parents mention that a combination of the two programmes would merit some further thought but they both want to leave that with the professionals.

6.1.3. Teachers

In accordance with the questionnaire prepared for the interviews the first category I would like to address is the very beginning, the inception and implementation of the programme. Again, as much as possible, I will try to report in a condensed version that presents all the opinions/thoughts expressed on the same issue in one comprehensive section.

The idea had been in the programme coordinator's head for years (AZ) since he thought and still does that English (and any foreign language) is best learned in a natural, non-stressful environment, that allows the learner to lower the affective filter and learn without any fear of exam stress. In a school, outside of the regular foreign language classes this is possible through an EAA-programme, the regular use of English in other subjects in order to make speaking the language a natural activity. In the year 2004, through his and the headmaster's input, motivated also by the British foreign language assistant at the time (TW) the idea to implement such a programme in the school came up and was quickly carried out, after a programme had been devised in cooperation with other schools and discussed with and approved by the board of education. In a next step teachers had to be found who could imagine teaching EAA-lessons (one teacher – PS – points out that no teacher was really personally asked if they could imagine teaching in an EAA-setting but that the only question was whether the staff could imagine EAA at the school). Other reasons for installing the programme at the school were issues such as making the school a more attractive option for a wider audience by adding "an intellectual aspect" to the sports focus of the school. Upon starting the programme the number of applications was surprisingly high and instead of creating one EAA-class as originally planned, two classes were opened for EAA-students. One of the classes was seen as the "100% EAA"-class and the other was seen as an "EAA light" or "test class". The idea was at the time that maintaining two classes throughout lower secondary was a sensible step in order to ensure that the programme could be upheld in

upper secondary even after students leave (to pursue different academic paths, as discussed in chapter 4) to create one continuing EAA-class for upper secondary. The second reason was to allow for changes after one introduction year – when it was found that the additional demand that EAA causes was too much for students to handle and provide the opportunity to leave the programme after a year to avoid creating further problems.

In the beginning there were only a very limited number of content teachers who agreed to teaching in an EAA-setting with a native speaker assistant (only TW at the time) but the number slowly increased when teachers saw, after trying to run an EAA-lesson with the help of the language assistant, that their expectations were different to reality. Many then slowly started to let go of inhibitions which were caused by a fear of losing face in class through not speaking English perfectly, or by opening the classroom for a native speaker. This again is one more reason mentioned by the programme coordinator why the EAA-setting is ideal – the students speak English in a class where they might notice that the teacher himself is not safely grounded in the language and thereby also feel less under pressure because even the teacher is not perfect. The number of content subjects slowly increased over time, ranging from Geography, Biology, Mathematics, Physical Education to Religion.

The entrance test / admission exam that is required before being allocated a place in an EAA-class is also described and commented upon by nearly all of the 5 teachers. It is perceived by everybody and intended to be a test of German language skills that seeks to ensure a certain level of German. This is considered a precondition in order to avoid creating any deficiencies in German through the increased instruction in a foreign language and moving away from the first language. The more critical teachers see the exam similar to a filter that finds those with the better language skills and places them in EAA classes and leaves those with lower scores in the “normal”/“general” classes.

The challenges the programme has been faced with over the years were manifold. Finding teachers who were willing to put in the extra effort needed for EAA-teaching without extra payment, who understood that including an EAA-aspect into their teaching would mean reallocating time resources away from some content breadth to EAA and teachers who were confident that teaching in a foreign language which some students might know better than them would not be a problem or source of intimidation.

All teachers agree that the programme has successfully met the originally planned intentions. This is indicated by a perceived notably high level of English language skills, a high readiness and high confidence to use English in everyday conversation with teachers and native speakers as well as in the classrooms, but also on language trips to London and destinations in the USA. All these demonstrate how the original goals (increasing language skills) have definitely been met and, as some say, even exceeded.

Employing native speakers in the programme is seen entirely positive by all teachers involved for a number of reasons. Some say that opening their classroom to native speakers has broadened their horizon tremendously, others comment that communicating with a person who is not the teacher who is responsible for grading makes it a lot easier for the students to practice the language without feeling self-conscious. Others feel that the content is learned better and more effectively and permanently after being dealt with in two languages. This very simple, and “practical” and “performance-oriented” (AZ) goal is considered met by all parties.

One of the first challenges was calibrating the teachers’ expectations of the aims and goals for EAA lessons with those of the parents. One teacher explains that she then used a lot of English and also demanded English in revisions and other assessment. This met with some resistance from parents who expressed concerns that the demands were too high and that they had had a more playful and less academic approach in mind. The issue was resolved through some exchange of feedback and discussions. Consequently, the teachers whose EAA teaching approach had been the reason for these complaints made changes to their methods and altered their approach in their EAA classes.

During the first two years after the programme was initiated a first major challenge arose which nearly threatened the programme was that somehow, what a number of interviewees called a “2-class-system”, came into place. Teachers were noticing a significant increase in level in the EAA-classes and a noticeable drop in level of academic achievement in the then non-EAA classes. One teacher attributes this to the fact that there was a lack of heterogeneity in the class rooms – the brighter, more talented students seemed to have ended up in the EAA classes leaving those achieving lower results in the non-EAA classes. Teachers who never got to teach in EAA classes (because they did not want to, etc.) saw more and more comparatively weak results, felt insecure and feared being seen as “bad

teachers” due to the results they saw in their classes. “You are considered a good teacher if you teach EAA-classes” (PS) and similar statements illustrate this and also explain a certain amount of animosity caused among teachers who also wanted to teach in EAA-classes and those who did. One teacher refers to a division in the teaching staff into two subgroups. The EAA and non-EAA classes were somehow seen as separate streams and students as well as teachers started to call the EAA classes “the smart kids” and the non-EAA class “the weak kids”.

This situation was fed back up to the board of education through various canals and a solution to this situation was demanded if the programme was to be granted approval to be continued. The idea of creating a second programme that would run alongside the EAA-programme and provide an alternative that also offered inviting aspects to academically bright students came up and was quickly implemented. While the IPP programme was only slowly accepted and developed, a number of teachers report that this feeling of there being two different camps or streams has decreased markedly, yet nobody claims that this problem is fully dealt with. The more critical of teachers (AL) mentions that this gap has markedly narrowed and she feels that even though she is a teacher who happily agrees to teach IPP classes only because she does not feel comfortable enough using English in her class, she knows that if she wanted to try to teach an EAA class, she would be allowed to do so. The headmaster is perceived to make a point now of ensuring that every teacher teaches both in EAA- as well as IPP-classes.

The school has seen a great increase in enrolment figures since the inception of the EAA-programme, but with the IPP-programme gaining ground and increasing popularity these figures have shifted to a roughly equal amount of applications for both programmes. Teachers report that students (first years) outspokenly claim to have chosen the IPP option for the content that it offers or because they felt that the EAA option would have demanded too much of them.

As far as ensuring and maintaining quality is concerned it is said that it is the coordinator’s task to work towards an equally spread amount of EAA lessons throughout all the EAA classes. The shortage of native speakers manifests itself in situations when classes do not have lessons with native speaker assistants for weeks on end. While there is no required amount of EAA-lessons a week or per subject, there are, however, expectations on the part

of the parents and students that are to be met. Another aspect that has to be addressed by the coordinator is that a sufficient amount of teaching material in English is available for all subjects. In this context all teachers feel that this demand is fully met by the coordinator.

Some of the quality control work is passed on to the teacher who teaches the regular English language classes and who regularly assesses the level of English. At this school this is done by using standardized test materials from lower secondary onwards (PET, CAT, FCE, etc.) and also the A-level results show – in a standardised manner – how results differ as far as the level of English is concerned. Assuring quality by means of tests in English in the contents is legally not possible because that would require approval from the board of education and the guarantee that more than 50% of the annual teaching time was conducted in English. The native speaker (TW) calls the current quality situation a “good level” that could only be furthered by moving on to a bilingual programme, which as explained is not possible. All the teachers interviewed see little need for improvement on the part of the quality of the programme itself.

All the teachers who mention the topic say that the marketing/promotion that is done for the programme is truthful in that it does not promise something that it cannot keep. The expectations that are created are met – it is not a bilingual education that is to be expected and that is communicated from the very beginning.

The expectations created through marketing of the programme include a number of issues that are openly communicated to prospective applicants for the EAA-programme. Parents and children are informed to expect a comparatively high level of English language skills at a much younger age (1,5-2 years ahead of students of equal age in non-EAA schools). Having committed teachers for EAA-students and providing increased attention for the students in the foreign languages are expectations that are instilled with parents. They are told to expect a playful approach to using English in a content subject classroom but at the same time to be aware of the increased cognitive demands that students have to face and also that extra time might be needed during the first years. Also the EAA-programme is portrayed as a means of preparing students for the demands of later working life as well as possible trips and student terms abroad, etc. All teachers feel that what is to be expected is clearly communicated before parents enrol their children in the programme.

Concerning the future of the EAA programme only few teachers see necessary changes within the programme itself. The coordinator as well as the native speaker expresses the need to constantly keep searching for native speakers to keep the programme running. The native speakers should ideally also be provided with contracts by the board of education and thereby some security in the workplace (insurance, etc.). This would also make the native speaker job more attractive and might lead to less fluctuation within the native speaker staff. They also envision real EAA-teams of teachers within the school since the varying willingness of teachers to include EAA in their teaching can cause the programme to come to a halt, at least in some classes. It is also the coordinator's opinion to consider returning to the option of reviewing a student's performance after a year at the school and allowing the students to make changes and move from an EAA-class to an IPP-class and vice versa.

A number of teachers experience that the motivation for EAA among the students drops during adolescence (years 8 through 10) and the coordinator sees a remedy in increasing the number of EAA lessons during the first three years in order to make English such a natural tool in a content subject class that it will not be considered an additional burden.

Maintaining and upholding the EAA-programme during the years of upper secondary is considered a challenge by all teachers and is an issue that some want addressed on a number of levels. Especially science subjects pose a challenge here – teaching mathematics, physics, etc. on the level of A-level exams comes increasingly difficult to subject teachers as well as native speaker assistants. This is one front of the issue. The other is the situation described in chapter 4 which is how classes are constituted in year 9 in order to uphold both programmes but avoid unfair conditions (unequal number of students in a class, etc.). Here it will take some new ideas and flexibility within both programmes and the school to handle this situation in annually changing scenarios.

Most comments regarding the future of the programme are made on the co-existence of the two programmes and on how to make that easier and beneficial for all parties involved (students, parents, teachers). The representative of teachers who teaches mainly IPP classes speaks about working hard to foster a feeling of achievement in IPP students. Everybody mentions that it will take continuous hard work to improve the situation between the standing of the two programmes within the school. Suggestions for this process are: introducing an entrance test/requirement for the IPP programme so that this loses the

image of being a “plan b that requires no special abilities from prospective students” or being something that “everybody can do”. Teachers “on both sides” have to work harder to abolish the discrepancy in attitudes on the two programmes. Teachers who mainly teach EAA or IPP should be made teach both programmes in an equal amount in order to raise awareness and allow them to see the quality of both programmes.

A last question asked in all interviews aimed at tips/comments/recommendations the interviewees would give other teachers/schools before implementing a similar programme. Making sure that one had a sufficient amount of native speakers available, taking enough time considering pre-existing conditions at the school in order to create an agreement among the whole staff as well as providing an alternative to the programme within the school as a viable other option that is considered equal in worth and educational value in order to avoid the situation this school experienced at the beginning are mentioned. Ensuring that the whole staff are on the same page and providing teachers with training in the principles of EAA (e.g. in-service training) and reducing any possible inhibitions they might have about not speaking English perfectly. Another criterion mentioned is that a programme requires a programme director who is willing to oversee the programme and represent it within the school (staff, etc.), with parents and also on higher levels (board of education, etc.). Creating a positive attitude towards the programme among the teachers is mentioned on numerous occasions since that is a prerequisite of a good professional community that will allow the exchange of ideas and materials for example.

6.2. Results by category

In the following section I want to take a different perspective in the analysis of the interview data. Rather than analysing the content of the interviews by stakeholder as was done in the previous section, this section now looks at some categories (see underlined key words in previous section) and presents a comprehensive, all-data-encompassing viewpoint and summary.

When asked about the inception and implementation of the programme it is very much clearly understood that the initial impetus for the programme came from its current programme director who then in cooperation with the headmaster and the board of education devised a plan for the programme. What is commented on by one teacher is the fact that the implementation might have been carried out too quickly without actually

asking a sufficient amount of individual teachers if they wanted to participate rather than asking a broad question (“Can you imagine EAA at this school?”) in the course of a survey among teachers preceding the introduction of the programme.

The entrance exam that is a requirement before receiving a place within the programme is not commented upon in all the interviews and if so, mostly without any attitude, rather than just a report of what it consisted of. Only one participant questions if the effect of the entrance exam is taking all the academically bright students out of the general mass and grouping them in one class and thereby causing a loss of heterogeneity in the classroom. Also she considers the fact that there is no entrance requirement for the IPP programme highly questionable since it creates the impression that becoming a part of the IPP programme is something everybody can do while becoming a part of the EAA programme is only for the bright and gifted.

The benefits of the programme mentioned in the course of the interviews are very much similar – an increased competence in the foreign language. This means that it is natural for the students to use the foreign language in familiar contexts and their inhibition level is very low, even if they are abroad, surrounded by native speakers.

The goals of the programme are considered to have been met by all parties. This view is gathered from the results from standardized testing as well as from how natural it is for students to switch from one language to the other and communicate on a very effective level in the foreign language.

Regarding assessment the different parties involved report a variety of approaches taken by various teachers but there seems to be no disagreement or problematic view. Students are given the choice to take oral revisions in English or German, they seem to ask for the materials possible test questions will be derived from in German but have no problem dealing with the whole topic in English otherwise. Some teachers make the English aspect part of their assessment (e.g. asking one of the test questions in English), others do not make English a part of their assessment at all. Approaches taken are different but there seems to be general agreement that these are fine as long as they do not become too demanding (long factual texts in English with numerous technical terms, etc.).

Very soon after the introduction of the EAA-programme, problems arose regarding the level of the results among the non-EAA-students. A certain stigma came to be placed on non-EAA

students as to be less talented and bright, which in turn led to the impression that non-EAA classes must be taught by less motivated, less hard-working teachers. These facts led to a perceived two-class-system within the school, an issue that is addressed in a number of the interviews and that has been tried to be remedied by introducing the IPP-programme. Some improvement has already been noticed but the way to perceiving the programmes as equal and having the teachers and students involved in the two programmes see, understand and propagate that is still not completed. Some interviewees suggest a merging of the two programmes in order to avoid the competition between the two programmes and to teach all students the skills offered in both programmes.

The level of English content subject teachers have seems to be an issue of low relevance to students and parents provided there is a native speaker accompanying the content subject teacher in the lesson. Students consider confidence and showing no signs of self-consciousness a much more important quality for a content subject teacher in the EAA-setting. On the part of the programme directors it is a particular goal to win an increasing number of teachers to teach in an EAA class by demonstrating that it is not as difficult as it may seem and thereby lowering possible inhibitions.

The number of native speakers is an issue brought up in a number of interviews. It seems that there are phases where the programme cannot cater for the expected number of EAA-lessons in all the classes, which cuts back on the probability of reaching the programme's potential and also causes negative feedback from parents.

The future of the programme is an aspect discussed by every interviewee. Everybody made recommendations or wishes for the future and these ranged from maintaining and continuing the programme in its current state, to making sure to recruit more native speakers in order to be able to provide as many EAA-lessons as possible for all year groups, as well as investing more time to involve more teachers in the programme and raising awareness that EAA does not require perfect pronunciation on the part of the content subject teacher. A distinct wish was to carry out work in order to smoothen the coexistence of the two programmes at the school. It is also understood that for the time being the size of the project should be maintained and well administered, instead of trying hard to reach even higher goals at the risk of potential quality loss.

This section was intended to analyse the categories that arose in the course of the interviews. These labels were assigned to certain thoughts and issues that were brought up regardless of the stakeholder interviewed. These categories are also the driving force of the next and final section of the data analysis, which is what Guba & Lincoln (1989) refer to as the “agenda for negotiation”. This includes what I perceive from the interview analysis to be the most pressing aspects for the future development of the programme.

6.3. Agenda for Negotiation & Reflection

The findings generally indicate four main areas of possible future development of the programme. While two of these are entirely concerned with the EAA-programme itself, the others address the position of the programme within the school, parallel to another programme on offer at the school. These are the four topics I feel all the stakeholders consider the most relevant at the time and the ones which should most urgently be attended to.

The first is the question of native speakers. Native speaker co-teachers are considered a vital aspect of the programme by all stakeholders and it is therefore everybody’s wish to make the programme and the school attractive to native speakers in order to maintain the high quality of the programme. While the desired goals (fluency, lowering the inhibition to speak the foreign language in general, expanding the students’ vocabulary in the foreign language, etc.) are being met there is a perceived localized problem with the distribution of native speakers across all the subjects and year groups.

The second idea for the future that I suggest in order to reduce different perceptions on the part of parents and teachers is to devise a more unified approach to the teaching of EAA – in a way to seek to apply standards to EAA-teaching at the school. These standards would comprise the amount of time taught per subject per year group in an EAA-setting, the mode of assessing EAA-course content (if at all), developing materials that all teachers have access to, etc.

Within the school the need to work hard to resolve the issue of the perceived gap between the two programmes is expressed by nearly all parties most firmly, however, by teachers. Raising awareness of the differences between and advantages of both programmes and ensuring the instilment of a feeling of worth and value in students of both programmes

seems a common goal that everyone can agree to. Some (students, parents) even go as far as suggesting a merging of the two programmes in order to lower competition between the two programmes and increasing heterogeneity within all the classrooms. This is an issue that would affect the entire school and its current concept. Changing it would take extensive work on a whole-school level, especially making sure to include a large number of staff in the development.

The last aspect that is brought up by all three stakeholder groups is that in order for the EAA-programme to be upheld and successfully continued it takes an increased effort to sensitise teachers towards the goals of the EAA programme. This could be achieved by providing in service training that helps lower inhibitions and reservations some teachers might have regarding joining the EAA-programme as teachers. The teachers in a school have to share the same common goal, the coexistence of the two programmes and represent them both to a certain extent.

7. Conclusion

Content and Language Integrated Learning – a concept that has taken many shapes and forms throughout Europe and a concept that has yet to find its clear-cut, well-defined incarnation with common features and guiding principles is one that has created a lot of impetus and movement in classrooms over the past two decades.

The purpose of the thesis was to examine more closely the CLIL-programme of a secondary school in lower Austria. The evaluation carried out, given that it is an internal evaluation did not seek to generate numeric facts and quantitatively analysed statements, had as its core goal to determine a status quo. By that I mean an evaluation that is mostly intended to give an overview of how the stakeholders directly involved in and affected by the programme see the programme today after having been part of it for varying amounts of time. The goal was not to prove how effectively the programme is achieving its goals by carrying out language tests that show an improvement that can be attributed to the EAA-programme. It was much more the aim to compare and contrast stakeholders' views and perceptions of the current situation of the programme.

Before embarking on the evaluation itself it was necessary to provide a sound review of relevant literature in both the fields of CLIL and evaluation. While studying the linguistic and didactic dimension to CLIL I have also focussed on the situation of CLIL in Austria and briefly discussed some recommendations made for the implementation of a CLIL programme. The latter were also incorporated in the final segment of the data analysis section.

The study of the theory of evaluation traced relevant history of the concept of evaluations in order to understand how the term evaluation had come to include the evaluation type used in the course of this thesis. It was a priority for me to provide solid ground for adopting the kind of evaluation used for the purpose of this thesis. Evaluating in a qualitative way without measuring previously fixed criteria but by giving a voice to representatives of all stakeholder groups and letting them illustrate their view of reality. It took a great paradigm shift before evaluations such as “4th generation evaluations” proposed by Guba & Lincoln (1989) gained acceptance as a scientific method. Constructivism paved the way for this development and is also its foundation.

In a next step this project moved from the more theory based part onward to the practical part, which commenced with a chapter entitled “setting the scene”. This chapter presented the school, the relevant details regarding the CLIL programme that was the core of the evaluation and in a second part discussed the roles of the researcher, who as a teacher within this programme assumes a double role. It was a vital step for me to clearly point out my position and explain my viewpoint on the programme before embarking on the close analysis of the interview data. The greatest point of criticism within the theory of evaluation is that an internal evaluation such as this is always viewed as more subjective than objective and while this is what defines it – not only accepting subjectivity, but making it the central part of the evaluation, I wanted to make sure in this chapter that my own position was well-reflected upon before moving on to my interviewees’ opinions in order to avoid influencing my understanding of the data provided in the interviews to an unacceptable degree. While, of course, every piece of research and every interpretation of data is influenced by the researcher, I do believe that the way I have chosen is a path that makes it possible for the reader to distinguish between the two.

Inspired by my reading on the theory of evaluation, most prominently by the method of a 4th generation evaluation, the design of my evaluation was devised. Again, while conducting the interviews and especially between the interviews I took time to distinguish between my position and the one of the interviewee. What was done, however, and I see that very much in accordance with the principles of 4th generation evaluation is that the interview questions were refined in more detail after every interview and following interviewees were sometimes also asked to comment on statements that had previously been made.

The findings, though not entirely intended, did not come surprising to me, being a part of the programme and knowing some of what people perceive as pressing matters within the school and programme. Originally it was the intention to evaluate the programme and its internal workings, but the issue that moved to the forefront within all the interviews, regardless of the approach taken, was the coexistence of the two programmes offered at the school and the implications this had.

The agenda for negotiation proposed for the school then obviously includes this aspect among other programme-internal ones. These will be passed on to the headmaster, as well as the programme director(s).

Looking back on previous chapters of this thesis it can be said that the methods of 4th generation evaluation were loosely adhered to, since the purpose was rather to arrive at similar goals and that is a picture of a reality that is co-constructed. It was not the purpose to arrive at a truth that is scientifically quantifiable, it was the purpose of this study to bring stakeholder opinions together and compare them and make them one complete take of “reality” as it is for the individuals involved in the programme. Looking back I can say that the evaluation went in a way that I had not planned but increasingly anticipated as the interviews progressed. The issue of the problematic coexistence between the two programmes offered at the school was in my mind when I came up with the idea for the evaluation but originally my focus was more on the EAA-programme itself and its “interior design” and “inner workings” but the interviews clearly showed what the topics were that are most relevant to the stakeholders.

In accordance with what Guba & Lincoln (1989: 149) called the “hermeneutic-dialectic circle” I have spent time between the interviews to prepare certain questions regarding issues that the original questionnaire might not have catered for but necessary to prompt a further interviewee to make a statement on an issue brought up by a previous interviewee. Thereby I believe to have co-constructed the evaluator’s take on the matter and merged the stakeholders’ “claims, concerns and issues” (1989: 149) into the desired “agenda for negotiation”, which I have laid out above.

Referring back to the principles of CLIL discussed in chapter 2 it can be said that while I am personally unsure that the core principle of CLIL (applying the dual focus – teaching content and teaching language items) is fully executed and achieved at this school, it can be said that its goals (those of CLIL and the programme alike) are met within the programme at this school. In other words, some of the EAA-teaching may be more similar to teaching “in” a foreign language rather than teaching “through” a foreign language, which would include special attention to a language skill as well as a content skill in the EAA-lesson. However, students graduating from this programme show a generally astounding level of foreign language competence that seems very natural to them. Native speakers consider them equal partners in discussions and deem the students fit for life abroad and this is clearly the goal of the programme.

8. References

8.1. Graphs, Pictures

Figure 1: Stake's cells

taken from Stake, Robert E. 1967. "The countenance of educational evaluation". *Teachers' College Record* 68(7), p.529.

Figure 2: Integrating the 4Cs of CLIL

taken from Coyle, Do; Hood, Philip Marsh, David. 2010. *CLIL – Content and Language Integrated Learning*. Cambridge: Cambridge University Press, p.41.

8.2. Books & Articles & Reference works

Abuja, Gunther (ed.). 1996. *Fremdsprachen als Arbeitssprachen: Die Rolle des 'native speakers'*. (ZSE-Report: 15) Graz, Wien: Zentrum für Schulentwicklung. Bereich III: Fremdsprachen.

Abuja, Gunther (ed.). 1998. *Englisch als Arbeitssprache: Modelle, Erfahrungen und Lehrerbildung*. Bericht Reihe III, Nummer 4. Graz: ZSE III.

Admiraal, W.; Westhoff, G.; de Bot, K. 2006. "Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English." *Educational Research and Evaluation*, 12, 75-93.

Airey, John. 2009. "Estimating undergraduate bilingual scientific literacy in Sweden." *International CLIL Research Journal*, 1, 26-35.

Airey, John; Linder, Cedric. 2006. "Language and the experience of learning university physics in Sweden." *European Journal of Physics*, 27, 553-560.

Alderson, Charles J. (ed.). 1985. *Evaluation*. Oxford: Pergamon Press Ltd.

Altrichter, H.; Messner, E.; Posch, P. 2004. *Schulen evaluieren sich selbst – Ein Leitfaden*. Seelze: Kallmeyersche Verlagsbuchhandlung GmbH.

Badertscher, Hans; Bieri, Thomas. (Hrsg.). 2009. *Wissenserwerb im Content and Language Integrated Learning*. Bern-Stuttgart-Wien, Schweiz: Haupt Verlag.

Bennett, Judith. 2003. *Evaluation methods in research*. London: Continuum.

Beretta, Alan. 1992. "Evaluation of language education: an overview". In Alderson, Charles J; Beretta, Alan. (eds.). *Evaluating Second Language Education*. Cambridge: Cambridge University Press, 5-24.

Brown, James D. 1989. "Language programme evaluation: a synthesis of existing possibilities." In Johnson, Robert K. (ed.). *The Second Language Curriculum*. Cambridge: Cambridge University Press, 242-261.

Bruton, Antony. 2010. "Are the differences between CLIL and non-CLIL groups in Andalusia due to CLIL? A reply to Lorenzo, Casal and Moore (2010)." *Applied Linguistics*, 32(2), 236-241.

<http://applied.oxfordjournals.org/content/32/2/236.full.pdf+html?sid=1ec3eadb-3bfc-43e0-bb2a-31d8b71c42a5> (6 April 2013).

- Bruton, Antony. 2011. "Is CLIL so beneficial, or just selective? Re-evaluating some of the research." *System*, 39(2011), 523-532.
- Butler, Melanie. 2012. "A questions of maths, science ...and social class?" *ELgazette*, 2 (2012), 13. <http://mag.digitalpc.co.uk/Olive/ODE/ELGAZETTE/> (6 April 2013).
- Caspari, D.; Hallet, W.; Wegner, A.; Zytadiß, W. (Hrsg.). 2007. *Bilingualer Unterricht macht Schule*. Frankfurt am Main: Europäischer Verlag der Wissenschaften.
- CLIL-Compendium. <www.clilcompendium.com>. 30 November 2011.
- Cohen, Louis et al. 2000. *Research Methods in Education*. (5th edition). London: Routledge.
- Cohen, Louis et al. 2007. *Research Methods in Education*. (6th edition). London: Routledge.
- Cook, Guy. 2003. *Applied Linguistics*. Oxford: Oxford University Press.
- Coyle, Do; Hood, Philip Marsh, David. 2010. *CLIL – Content and Language Integrated Learning*. Cambridge: Cambridge University Press.
- Dalton-Puffer, Christiane. 2002. "Content and language integrated learning in Austrian classrooms: Applied Linguistics takes a look." *VIEWS*, 11(1&2), 4-26. http://www.univie.ac.at/Anglistik/views/02_1&2/CDP.PDF (20 October 2011).
- Dalton-Puffer, C.; Smit, U. (eds.) 2007. *Empirical Perspectives in CLIL classroom discourse*. Frankfurt am Main: Europäischer Verlag der Wissenschaften.
- Dalton-Puffer, Christiane. 2008. "Outcomes and processes in Content and Language Integrated Learning (CLIL): current research from Europe." In Delanoy, Werner; Volkmann, Laurenz (eds.). *Future Perspectives for English Language Teaching*. Heidelberg: Carl Winter. 139-157.
- Dalton-Puffer, Christiane. 2011. "Content-and-Language Integrated Learning: From Practice to Principles?" *Annual Review of Applied Linguistics* 31, 182-204. http://anglistik.univie.ac.at/fileadmin/user_upload/dep_anglist/ARAL_2011_CLIL_article.pdf
- DeVault, Marjorie. 1990. "Talking and listening from women's standpoint: Feminist strategies for interviewing and analysis." *Social Problems*, 37(1), 96-116.
- Eisner, Elliot W. 1977. "On the uses of educational connoisseurship and criticism for evaluating classroom life." *Teacher's College Record* 78(3), 345-58.
- Eisner, Elliot W. 1998. *The enlightened eye: qualitative inquiry and the enhancement of educational practice*. Upper Saddle River, N.J.: Merrill.
- Ellis, Rod. 1997. *Second Language Acquisition*. Oxford: Oxford University Press.
- Eurydice (eds.). 2004/05. *Content and Language Integrated Learning (CLIL) at School in Europe: Country Reports. National description: Austria*. http://eacea.ec.europa.eu/resources/eurydice/pdf/070DN/070_AT_EN.pdf (16 August 2011).
- Ferrance, Eileen. (Northeast and Islands Regional Educational Laboratory At Brown University). 2000. *Themes in Education: Action Research*. Providence: Brown University. http://www.lab.brown.edu/pubs/themes_ed/act_research.pdf

- Gierlinger, Erwin. 2005. "Fremdsprache als Arbeitssprache, EAC, EMI, CLIL – Österreich auf dem Weg von Insellösungen zu Strukturen." In Krechel, Hans/Ludwig (ed.). *Mehrsprachiger Fachunterricht in Ländern Europas*. Tübingen: Günther Narr Verlag, 89-102.
- Gierlinger, E.; Hainschink, V.; Spann, H. 2007. *Content and Language Integrated Learning (CLIL) auf der Sekundarstufe I*. Linz: Trauner Verlag.
- Guba, Egon; Lincoln, Yvonna. 1989. *Fourth Generation Evaluation*. Newbury Park, CA, USA: Sage Publications, Inc.
- Jäppinen, Aini-Kristiina. 2005. "Thinking and content learning of mathematics and science as cognitional development in content and language integrated learning (CLIL): Teaching through a foreign language in Finland." *Language and Education*, 19, 148-169.
- Joint Committee on Standards for Educational Evaluation. 1971. *Standards for Evaluation of Educational Programs, Projects and Materials*. New York: MacGraw Hill.
- Kiely, Richard; Rea-Dickins Pauline. 2005. *Program Evaluation in Language Education*. New York: Palgrave Macmillan.
- Krück, B.; Loeser, K. (Hrsg.). 2002. *Innovationen im Fremdsprachenunterricht: 2 – Fremdsprachen als Arbeitssprachen*. Frankfurt am Main: Europäischer Verlag der Wissenschaften.
- Lasagabaster, David; Ruiz de Zarobe, Yolanda. (eds.). 2010. *CLIL in Spain: Implementation, Results and Teacher Training*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Lichtman, Marilyn. 2006. *Qualitative Research in Education. A User's Guide*. Thousand Oaks, CA, USA: Sage Publications, Inc.
- Llinares, Ana; Morton, Tim; Whittaker, Rachel. 2012. *The Roles of Language in CLIL*. Cambridge: Cambridge University Press.
- Lynch, Brian K. 1996. *Language Program Evaluation - Theory and Practice*. Cambridge: Cambridge University Press.
- Lorenzo, Franzisco; Moore, Pat; Casal, Sonia. 2010. "The Effects of Content and Language Integrated Learning in European Education: Key Findings from the Andalusian Bilingual Sections Evaluation Project." *Applied Linguistics* 31(3), 418-442.
- Maljers, A.; Marsh, D.; Wolff, D. (eds.). 2007. *Windows on CLIL*. The Hague: European Platform for Dutch Education.
- Marsh, David; Wolff, Dieter. (eds.). 2007. *Diverse Contexts – Converging Goals. CLIL in Europe*. Frankfurt am Main: Peter Lang GmbH Internationaler Verlag der Wissenschaften.
- Mehisto, Peeter. 2007. "What a school needs to consider before launching a CLIL programme: the Estonian experience." In Marsh, David; Wolff, Dieter. (eds.). *Diverse Contexts – Converging Goals. CLIL in Europe*. Frankfurt am Main: Peter Lang GmbH Internationaler Verlag der Wissenschaften, 61-77.
- Mehisto, Peeter. 2008. "CLIL Counterweights: Recognising and Decreasing Disjuncture in CLIL." *International CLIL Research Journal* 1(1), 93-119.

- Mehisto, Peeter; Marsh, David; Jesús Frigols, María. 2008. *Uncovering CLIL – Content and Language Integrated Learning in Bilingual and Multilingual Education*. Oxford: Macmillan Education.
- Mewald, Claudia. 2004. "Paradise lost and found: A case study of content based foreign language education in lower Austria". Unpublished PhD Thesis, University of East Anglia, Norwich.
- Nezbeda, Margarete. 2005. *Überblicksdaten und Wissenswertes zu Fremdsprache als Arbeitssprache*. EAA Serviceheft 6. Graz: Sprachen-Kompetenz-Zentrum.
- Nevo, David. 1983. "The conceptualization of educational evaluation: An analytical review of the literature." *Review of Educational Research* 53, 117-128.
- Nikula, Tarja. "On effects of CLIL on a teacher's language use." In Dalton-Puffer, Christiane; Nikula, Tarja; Smit, Ute (eds.) 2010. *Language Use and language learning in CLIL classrooms*. Amsterdam, Niederlande: John Benjamins, 105-124.
- Oestreich, Katja; Grogger, Günther. 1997. "Der Einsatz einer Fremdsprache als Arbeitssprache in nichtsprachlichen Gegenständen." In ZSE Zentrum für Schulentwicklung. *ZSE Report (31)*. Graz: Zentrum für Schulentwicklung.
- Parlett, Malcolm; Hamilton, David. 1976. "Evaluation as Illumination: a new approach to the study of innovatory programmes." In G.V. Glass (ed). *Evaluation studies review annual (1)*. Beverly Hills, California: Sage.
- Rea-Dickins, Pauline; Germaine, Kevin. 1992. *Evaluation*. Oxford: Oxford University Press.
- Rossett, Allison; Sheldon, Kendra. 2001. *Beyond the Podium: Delivering Training and Performance to a Digital World*. San Francisco: Jossey-Bass/Pfeiffer.
- Ruiz de Zarobe, Yolanda; Jiménez Catalán, Rosa M. (eds). 2009. *Content and Language Integrated Learning – Evidence from Research in Europe*. Bristol: Multilingual Matters.
- Saville-Troike, Muriel. 2006. *Introducing Second Language Acquisition*. Cambridge: Cambridge University Press.
- Scriven, Michael. 1967. "The methodology of evaluation". In Tyler, R.W.; Gagné, R.M; Scriven, M. (Eds.). *Perspectives of curriculum evaluation*. Chicago, IL: Rand McNally, 39-83.
- Smith, Mark K. 2005. "Elliot W. Eisner, connoisseurship, criticism and the art of education". *the encyclopaedia of informal education*, 2005, <http://www.infed.org/thinkers/eisner.htm>, (13 May 2012).
- Stake, Robert E. 1967. "The countenance of educational evaluation". *Teachers' College Record* 68(7), 523-40.
- Stake, Robert. E. 1995. *The Art of Case Study Research*. Thousand Oaks, CA, USA: Sage.
- Stenhouse, Lawrence. 1975. *An Introduction to Curriculum Research and Development*. London: Heinemann.
- Stufflebeam, D.L.; Foley, W.J.; Gephart, W.J.; Hammond, L.R.; Merriman, H.O.; Provus, M.M. 1971. *Educational Evaluation and Decision-Making in Education*. Ithaca, NY: Peacock Press.

- Taba, Hilda. 1962. *Curriculum Development: theory and practice*. New York: Harcourt, Brace and World, Inc.
- Tyler, Ralph W. 1950. *Basic Principles of Curriculum Development*. Chicago: University of Chicago Press.
- Vollmer, Helmut Johannes; Heine, Lena; Troschke, Randi; Coetzee, Debbie; Küttel, Verena. 2006. *Subject-specific competence and language use of CLIL learners: The case of geography in grade 10 of secondary schools in Germany*. Paper presented at the ESSE8 Conference, London, UK.
- Wilde, Judith; Sockey, Suzanne. 1995. "Evaluation Handbook". Albuquerque: New Mexico Highlands University.
http://www.ncela.gwu.edu/files/rcd/BE020502/Evaluation_Handbook.pdf (31 October 2011)
- Wildhage, Manfred; Otten, Edgar. (eds.). 2009. *Praxis des bilingualen Unterrichts*. Berlin: Cornelsen.
- Wilhelmer, Nadja. 2008. *Teaching Mathematics in English*. Saarbrücken: VDM Verlag Dr. Müller.
- Wolff, Dieter. 2007. "CLIL: bridging the gap between school and working life". In Marsh, David; Wolff, Dieter. (eds.). *Diverse Contexts – Converging Goals. CLIL in Europe*. Frankfurt am Main: Peter Lang GmbH Internationaler Verlag der Wissenschaften, 15-26.
- Worthen, Blaine R.; Sanders, James R. 1973. *Educational Evaluation: Theory and Practice*. Worthington, OH: C.A.Jones Publishing Company.

8.3. Online resources, webpages mentioned

- <http://www.nwlink.com/~donclark/hrd/sat6.html>, accessed 30 December 2012
- http://www.nwlink.com/~donclark/hrd/isd/types_of_evaluations.html, accessed 30 December 2012
- <http://www.learning-theories.com/constructivism.html>, accessed 20 May 2012
- <http://eaa.sportgymnasium.net/wordpress/?pageid=309>, accessed 5 February 2012
- http://eaa.sportgymnasium.net/wordpress/wp-content/uploads/2010/08/eltern_1klassen_2010_neu.pdf, accessed 5 February 2012
- http://eaa.sportgymnasium.net/wordpress/?page_id=14, accessed 5 February 2012
- http://eacea.ec.europa.eu/education/eurydice/thematic_studies_archives_en.php, accessed 16 August 2011
- <http://eacea.ec.europa.eu/education/eurydice/>, accessed 22 March 2013
- <http://mag.digitalpc.co.uk/Olive/ODE/ELGAZETTE/>, accessed 6 April 2013

Appendix

„Content and Language Integrated Learning“ ist ein Unterrichtsprinzip, das europaweit seit über zwei Jahrzehnten in vielen Formaten erfolgreich eingesetzt wird. Ziel ist es, durch den Einsatz von lebenden Fremdsprachen (zumeist Englisch) im Unterricht von sogenannten Sachfachgegenständen (also nicht sprachlichen Gegenständen) zu einer deutlich erhöhten und flexibler und deutlich selbstverständlicher einsetzbaren Fremdsprachenkompetenz zu führen. In Österreich wird diese Methode oft unter dem Namen EAA (Englisch als Arbeitssprache) in Schulen eingesetzt.

Im Rahmen der vorliegenden Arbeit wird ein derartiges Programm an einer österreichischen Schule evaluiert. Nach einer umfassenden Diskussion und Präsentation der relevanten Theorie, nämlich der sprachwissenschaftlichen & fachdidaktischen Kontextuierung dieser Unterrichtsmethode sowie einer genauen Behandlung der Theorie, die hinter Evaluationen an sich steckt, wird dazu übergegangen, die Situation an der konkreten Schule und die Doppelrolle des Evaluierenden, der gleichzeitig auch Lehrer an derselben Schule ist, zu beleuchten. Die genaue Abgrenzung seitens der Person, die die Evaluation durchführt zum Gegenstand der Evaluation ist ein wichtiger Schritt.

Die Methode der Evaluation ist eine qualitative. Abgeleitet aus den Prinzipien einer „4th generation evaluation“ nach Guba & Lincoln (1989) entspringen die Daten offenen Interviews mit einer Gruppe an Personen, die nahezu alle im Programm involvierten Parteien, abbilden. Ziel der Evaluierung ist es einen Gesamteindruck aller direkt betroffenen Personen herzustellen und diese oft auseinander klaffenden Meinungen in Kontext zu stellen und zusammenzufassen, um eine möglichst einheitliche Darstellung der derzeitigen Situation zu schaffen.

Im Rahmen der Untersuchung stellt sich heraus, dass das Programm an sich und seine Funktionsweise hohes Ansehen und Akzeptanz genießt, aber seine Koexistenz zu einem weiteren Angebot in der Schule als sehr kritisch gesehen wird und die meisten Wünsche nach Veränderung für die Zukunft generiert. Ziel der Arbeit ist es hier auch nicht unbedingt Lösungsvorschläge zu liefern, sondern mehr diese verschiedenen Positionen, sofern vorhanden, einander gegenüber zu stellen und eine Art „Momentaufnahme“ der Meinungen aller beteiligten Parteien zu liefern.

Curriculum Vitae

Fabian Szymanek

31 March 1988

Education

2007 – 2013

University of Vienna, Austria

teaching qualification programme for English and Spanish (*secondary school teaching*)

2007 – ongoing

University of Technology in Vienna, Austria

teaching qualification programme for Mathematics (*secondary school teaching*)

1998 – 2006

Bundesgymnasium & Bundesrealgymnasium Gmunden, Upper Austria

lower & upper secondary education; graduated with honours; focus on foreign languages (English, French, Spanish, Italian, Russian)

1994 – 1998

Volksschule Regau – upper Austria; primary education

1991 – 1994

American British Academy in Muscat, Oman

Key Stages K1, K2

Relevant Work Experience

2010 – ongoing

Upper & Lower Secondary Teacher of English, Mathematics & Spanish

secondary school teacher; A-level examination experience; form tutor – pastoral & academic guidance; organization of language trips (USA, Berlin, etc.)

2010 – ongoing

SFA Sprachreisen

manager and coordinator of student language camps during the summer months; administrative skills; team-building as well as leadership skills; responsible supervision for students aged 10-15; supervision, coordination and support in language teaching (& teachers) at the camp

2009 – 2011

IFL Institut für Lernen – Nachhilfeinstitut Dr. Rampitsch

regular tuition in English, Mathematics, Spanish for students aged 10-18

2007 – 2012

Lernhilfe Gmunden

regular tuition in English, Mathematics, Spanish for students aged 10-18; management and leadership skills;

2007 – 2008	Jugendzentrum Attnang-Puchheim youth counsellor for teenagers aged 14-18; organization of free time activities
2006 – 2007	Civil Service – Jugendzentrum Attnang-Puchheim youth counsellor for teenager aged 14-18; organization of free time activities
2006 – ongoing	regular private tuition for English, Mathematics & Spanish;

Relevant Additional Skills and Interests

- ✓ MS Office, Apple Mac
- ✓ Online learning platforms (Moodle, Fronter, etc.)
- ✓ Seminars on teenage violence, teenage alcoholism, migration