Dissertation

zur Erlangung des Grades eines Doktors der Philosophie, angenommen vom Fachbereich Sprach- und Literaturwissenschaft der Universität Osnabrück

Assessment of subject-specific task performance of bilingual geography learners: Analysing aspects of subject-specific written discourse

vorgelegt von

Debbie Coetzee-Lachmann

2 Mai 2007

Hauptberichterstatter:	**Prof. Dr. J. H. Vollmer**
Mitberichterstatter:	**Prof. Dr. D. Siepmann**

Table of contents

TABLE OF CONTENTS ..... **I**

LIST OF FIGURES AND TABLES ..... **IV**

1 INTRODUCTION **1**

1.1 AIMS AND INTENDED CONTRIBUTION OF THE STUDY 2
1.2 OVERVIEW OF METHODOLOGY 4
1.3 STRUCTURE OF DISSERTATION 6

2 THE IMPORTANCE OF WRITING IN SUBJECT-SPECIFIC CLASSROOMS **9**

2.1 SUBJECT-SPECIFIC DISCOURSE: A SOCIAL VIEW 9
2.1.1 SFL: Basic theoretical concepts *9*
2.1.2 Subject-specific discourse *10*
2.2 BILINGUAL SUBJECT-SPECIFIC EDUCATION IN GERMANY: DEVELOPING SCHOOL-BASED LITERACY IN A FOREIGN LANGUAGE 12
2.2.1 Views on the nature and aims of bilingual subject-specific education in Germany *13*
2.2.2 School-based literacy as a prerequisite for participation in institutionalised discourses *14*
2.2.3 School-based literacy as a prerequisite for autonomous participation in administrative, political and commercial activities *18*
2.2.4 Reflecting about the meaning-making process as a prerequisite for the transferability of school-based literacy *20*
2.3 THE NEED TO ASSESS BILINGUAL LEARNERS' USE OF THE SUBJECT-SPECIFIC REGISTER 23
2.4 FURTHER STUDIES ON BILINGUAL LEARNERS' PROCESSING OF GEOGRAPHY TASKS 26
2.5 CONCLUSION 28

3 A SOCIAL SEMIOTIC FRAMEWORK FOR THE RATING OF SUBJECT-SPECIFIC WRITING **30**

3.1 RATING TASK-BASED SUBJECT-SPECIFIC WRITING: THE NEED FOR AN INTEGRATIVE AND SOCIALLY-BASED THEORY 30
3.2 BASIC IMPLICATIONS OF A SOCIAL SEMIOTIC APPROACH TOWARDS THE ASSESSMENT OF LEARNERS' SUBJECT-SPECIFIC TASK-BASED WRITING 32
3.2.1 Developing frameworks of expectations: The importance of context *32*
3.2.2 Constructing meaning as a member of a community: the role of social semiotic systems and formations *36*
3.3 THE CONTEXT OF CULTURE: IDENTIFYING PATTERNS IN THE MEANINGS AND MEANING-MAKING PROCESSES OF THE SUBJECT COMMUNITY 40
3.3.1 Thematic formations *41*
3.3.2 Activity structures *42*
3.3.3 Genres *45*
3.3.4 Summary of the theoretical constructs involved in developing a framework of expectations based on an analysis of the context of culture *46*
3.4 CONTEXT OF SITUATION AND REGISTER THEORY: IDENTIFYING PATTERNS OF MEANING AND LANGUAGE USE IN THE CONTEXT OF TEST-BASED TASK PROCESSING 49
3.4.1 Situational variables *51*
3.4.1.1 Field of discourse 51
3.4.1.2 Tenor of discourse 52
3.4.1.3 Mode of discourse 56
3.4.2 Metafunctions *57*
3.4.2.1 The importance of regarding all three types of meaning in assessing learners' task-based subject-specific writing 60
3.4.2.2 Ideational meaning 62
3.4.2.3 Interpersonal meaning 69
3.4.2.4 Textual meaning 74
3.4.3 Framework of expectations for learners' language use in task-based subject-specific writing *76*
3.4.4 Summary of the theoretical constructs involved in developing a framework of expectations based on an analysis of the context of situation and register theory *77*

3.5 MERGING THE CONTEXT OF CULTURE AND CONTEXT OF SITUATION: THE CONCEPT OF A REQUIRED DISCOURSE FUNCTION 80
3.5.1 Defining discourse function *80*
3.5.1.1 Discourse function as defined by Zydatiß 80
3.5.1.2 Defining a discourse function: a social semiotic and systemic function linguistic reinterpretation 82
3.5.2 Basic outline for a framework of expectations related to the required discourse function *86*
3.5.4 Defining subject-specific task-based written discourse competence *88*
3.5.4.1 Vollmer's competence model for geography 89
3.5.4.2 The role of volition 93
3.5.4.3 Summary of the aspects involved in subject-specific task-based discourse competence 96
3.6 A FRAMEWORK FOR THE DEVELOPMENT OF RATING SCALES USED TO ASSESS LEARNERS' TASK-BASED SUBJECT-SPECIFIC WRITING 98

4 RESEARCH DESIGN AND PROCEDURES 102

4.1 OVERVIEW OF EMPIRICAL AIMS 102
4.2 BASIC PRINCIPLES OF QUALITY ASSURANCE IN TESTING 103
4.2.1 Reliability *104*
4.2.1.1 Definition 104
4.2.1.2 Strategies involved in promoting reliability 104
4.2.2 Validity *107*
4.2.2.1 Traditional views: different types of validity 107
4.2.2.2 Current views: validity as a unitary concept 110
4.2.3 Practicality *111*
4.3 OUTLINE OF PROCEDURES FOLLOWED IN SUPPORT OF VALIDITY AND RELIABILITY 112
4.3.1 Aspects related to validity: using authentic tasks designed in co-operation with experts *112*
4.3.2 Aspects related to reliability: number of performances, applying rating scales and double rating *113*
4.3.3 Generalisability of the results *116*
4.4 THE SAMPLE 118
4.5 THE ELICITATION TASKS 122
4.5.1 Development and refinement of a first draft for piloting purposes *122*
4.5.2 Pilot testing *124*
4.5.3 Data elicitation sessions *126*
4.5.4 The nature of the tasks *126*
4.5.4.1 Importance of variation between the tasks in terms of the required discourse function 126
4.5.4.2 Variation between the six tasks in terms of thematic focus 128
4.5.4.3 Variation between the six tasks in terms of input material 129
4.5.4.4 Variation between the six tasks in terms of the required discourse function 131
4.6 RATING SCALES 136
4.6.1 General overview *136*
4.6.1.1 Requirements the rating scales had to meet 136
4.6.1.2 Outline of the three types of rating scales used 137
4.6.1.3 Challenges involved in developing the scales 139
4.6.1.4 General overview of development processes 140
4.6.2 Analytic scales: Appropriateness of the textual realisation of meaning *142*
4.6.2.1 Description of scales 142
4.6.2.2 Process involved in developing the scales 144
4.6.3 Dichotomous and three-point scales: Completeness and correctness of meaning constructed *149*
4.6.3.1 Description of scales 149
4.6.3.2 Process involved in developing the scales 153
4.6.4 Holistic scales: rating the appropriateness of the discourse function / the degree of task fulfilment *157*
4.6.4.1 Description of scales 157
4.6.4.2 Process involved in developing the scales 161
4.7 RATING 163
4.7.1 Rating of the completeness and correctness of the meaning constructed *163*
4.7.2 Analytic rating of the textual realisation of meaning *165*
4.7.3 Holistic rating of the appropriateness of the discourse function *166*
4.8 CONCLUSION 166

5 SCORING RESULTS: ANALYSIS AND DISCUSSION 167

5.1 APPROACH FOLLOWED IN DATA PROCESSING AND INTERPRETATION 167
5.2 … HOLISTIC SCORING OF THE APPROPRIATENESS OF THE DISCOURSE FUNCTION: PRESENTATION AND ANALYSIS OF THE RESULTS 168
5.2.1 Task 1c 171
5.2.2 Task 2a 176
5.2.3 Task 3 181
4.2.4 Task 6 187
5.2.5 Task 7 191
5.2.6 Task 8 195
5.3 COMPLETENESS AND CORRECTNESS OF THE MEANING CONSTRUCTED: PRESENTATION AND ANALYSIS OF THE RESULTS 200
5.3.1 Incompleteness of the meaning constructed 202
5.3.2 Incorrectness of the meaning constructed 206
5.4 PRESENTATION AND ANALYSIS OF RESULTS OF ANALYTIC SCORING: APPROPRIATENESS OF THE TEXTUAL REALISATION OF MEANING 210
5.4.1 Problems in terms of structuring and linking 212
5.4.1 Problems in terms of language use 216
5.5 CONCLUSION 221

6 CONCLUSION AND PERSPECTIVES 223

6.1 DEVELOPING LEARNERS' TASK-BASED SUBJECT-SPECIFIC WRITTEN DISCOURSE COMPETENCE: SUGGESTIONS FOR THE CLASSROOM 223
6.1.1 Demystify the discourse 224
6.1.2 Give learners ample opportunity to construct meaning 226
6.1.3 Give learners effective feedback 228
6.1.4 Suggestions for the bilingual geography classroom 231
6.2 PERSPECTIVES FOR FURTHER RESEARCH 233
6.3 CLOSING REMARKS 235

APPENDIX A 238

Elicitation tasks 238

APPENDIX B 245

Summary of task categories and input material 245

APPENDIX C 246

Rating scales 246

Holistic scales: Degree of task fulfilment – appropriateness of the discourse function 247

Completeness and correctness of meaning constructed: 268

Appropriateness of the textual realisation of meaning 275

APPENDIX D 278

Guidelines for text production and analyses taken from geography textbooks 278

BIBLIOGRAPHY 283

List of figures and tables

Figure 3.1: The semiotic contexts relevant in rating the learners' written responses 34
Figure 3.2: The meaning-making processes involved in the writing and rating of a written response to a subject-specific task. 39
Figure 3.3: The role of activity sequences in a learner's production of a written response to a subject-specific task 44
Figure 3.4: Concepts from the context of culture involved in developing a framework of expectations 47
Figure3.5: The relation between language, register and genre 59
Figure 3.6: Concepts from the context of situation involved in developing a framework of expectations 78
Figure 3.7: Required discourse function: basic outline of aspects involved in identifying expectations 88
Figure 3.8: Schematic representation of subject-specific task-based written discourse competence 96
Figure 3.9: Developing rating scales: incorporating social semiotic and systemic functional linguistic constructs 101
Table 4.1: Number of semesters of bilingual subject-specific education the learners had followed in geography and history 119
Table 4.2: Percentage of learners per group identified in terms of number of semesters of bilingual education followed in geography and history respectively 120
Table 4.3: Comparison between means of groups' scores on holistic rating 121
Table 4.4: Summary of input material and aspects constituting the required discourse function 135
Table 4.5 Summary of process involved in the development of the analytic scales used to rate the textual presentation of the meaning constructed 148
Table 4.6 Summary of process involved in the development of the scales used to rate the ideational meaning constructed 156
Table 4.7 Summary of process involved in the development of the holistic scales used to rate the appropriateness of the discourse function 162
Table 5.1: Task 1c: Summary of data – appropriateness of the discourse function 171
Figure 5.1: Task 1c: Distribution of scores - appropriateness of the discourse function 172
Table 5.2: Task 2a: Summary of data – appropriateness of the discourse function 176
Figure 5.2: Task 2a: Distribution of scores - appropriateness of the discourse function 176
Table 5.3: Task 3: Summary of data – appropriateness of the discourse function 181
Figure 5.3: Task 3: Distribution of scores - appropriateness of the discourse function .. 181
Table 5.4: Task 6: Summary of data – appropriateness of the discourse function 187
Figure 5.4: Task 6: Distribution of scores - appropriateness of the discourse function .. 187
Table 5.5: Task 7: Summary of data – appropriateness of the discourse function 191
Figure 5.5: Task 7: Distribution of scores - appropriateness of the discourse function .. 191
Table 5.6: Task 8: Summary of data – appropriateness of the discourse function 195
Figure 5.6: Task 8: Distribution of scores - appropriateness of the discourse function .. 195
Table 5.7: Completeness and correctness of the meaning constructed: summary of the results 201
Table 5.8: Appropriateness of the textual realisation of meaning: summary of the results 210

1
Introduction

Bilingual subject-specific education[1] is widely practiced in Germany, especially in grammar schools. Teachers and researchers see it as an effective way of developing learners' foreign language proficiency. In the light of increasing global mobility and international information networking, it is perceived to harbour the potential for developing a key qualification: the ability to participate in various discourses, embedded in recreational, educational and professional activities of the European and world community. Although there is general agreement that the bilingual subject-specific classroom should develop both learners' general language proficiency and their ability to use the subject-specific register, research-based assessment initiatives undertaken in the past were mainly concerned with the first aim. Consequently, little is known about the extent to which these learners can use the foreign language appropriately for subject-specific purposes.

In this regard, Vollmer (2002c: 101-102) for instance writes:

> "Die 'Sprachlichkeit' ... jeglichen schulischen Lernens und Leistens erfordert es, dieser untrennbar mit dem Aufbau von fachlichen Konzepten und fachlicher Kompetenz einhergehenden Dimension größte Bedeutung beizumessen und ihr mehr Aufmerksamkeit zu widmen, als es bislang der Fall ist. Dies gilt allerdings sowohl für den muttersprachlichen wie für den fremdsprachlichen Sachfachunterricht. Darüber hinaus gibt es Spezifika des bilingualen Lernens, die es erforderlich machen, die Erweiterung der *fachlich-sprachlichen Ausdruckfähigkeit* bis hin zu einer sich entfaltenden *Text- und Diskursfähigkeit* in der L2 systematischer als bisher geschehen aufzubauen, zu üben und zu *überprüfen*; es handelt sich also um die Hinführung zu einem fachlich zusammenhängenden Denken und Argumentieren im Sinne von zunehmender Kohäsion, Kohärenz und 'Sachfachliteralität' ... die sich in den Schülerprodukten dokumentieren müssten." (Vollmer 2002c: 101-102. Italics: DC.).

In its aims and intended contributions, this study attempts to respond to calls such as this one for an increased focus on subject-specific language use and effective ways of assessing it in bilingual subject-specific education in Germany.

[1] Although the term "bilingual subject-specific education" is seen to be misleading and the term "subject-specific training in a foreign language" is preferred, it is applied here, because it is so widely used in the scientific discourse on this matter in Germany.

1.1 Aims and intended contribution of the study

In order to address this need for insight into bilingual learners' use of the subject-specific register, this study set out to assess the subject-specific task-based written performances of a large group (n=84) of 10th grade learners following geography in English at four different grammar schools in Lower Saxony. The focus is on Geography, since it is one of the most widely taught "bilingual subjects" in Germany. Furthermore, the study is concerned with the written task-based performances of 10th grade learners, because the 10th grade is the final year of lower secondary education in Germany. It thus represents the end of a particular educational phase, so that the findings give an insight into the strengths and weaknesses of these learners' writing produced at the end of the first phase of geography instruction in a foreign language and allow feedback about their future training needs.

The research interest is both theoretical in nature and practice oriented. In terms of its theoretical contribution, the study presents a definition of subject-specific task-based written discourse competence that takes the social nature of task-based writing in a subject-specific context into consideration. In developing this definition, Zydatiß's (2005) notion of a discourse function, which he puts forward as the basis for a curriculum construct for bilingual subject-specific education, as well as Vollmer's (2007) model of subject-specific geographical competence are both reinterpreted from a social semiotic point of view. The resulting definition of task-based subject-specific discourse competence explains how knowledge of subject matter, subject-specific methodological competence, subject-specific communicative competence and volitional and motivational aspects interrelate in the production of a written response to a subject-specific task. Moreover, theoretical constructs from social semiotics and systemic functional linguistics are applied in analysing and explaining the context of task-based subject-specific writing in order to contribute to an understanding of the expectations learners' task-based subject-specific writing is to meet.

The practice-oriented contribution of this study includes instruments that can be used to rate learners' task-based subject-specific writing and to give them clear and structured feedback. Furthermore, using the results of the rating processes carried out and the theoretical insights gained in the study, suggestions about possible ways of addressing

learners' difficulties in using the subject-specific register appropriately and effectively are made for the subject-specific classroom.

In short, the aims of the study can be summarised as follows:
The study aims to:

1. present a definition of subject-specific task-based written discourse competence that takes the role of contextual factors into account.
2. present analytic scales that can be used across different tasks and subjects to rate learners' language use in their subject-specific writing and give them diagnostic feedback.
3. present frameworks that can be used to develop task-specific scales for scoring written responses to subject-specific tasks. These frameworks firstly make provision for the development of scales with which appropriateness of the discourse function, or degree of task fulfilment can be rated and secondly for the development of scales that allow a comprehensive and detailed rating of the content of learners' written responses.
4. outline suggestions for the development of bilingual learners' ability to use the subject-specific register appropriately.

Rooted in the principles of social semiotics and systemic functional linguistics, the scales and instruments presented allow more than a mere identification of problems in the form of a classical error analysis. Since these instruments treat language use as functionally motivated and socially oriented, they provide a basis for feedback that firstly explicates the ways in which a learner's text deviates from the conventions of the subject community and that secondly also explains the functions these conventions serve in the discourse of this community. In this respect, the instruments can thus contribute to an approach towards the development of learners' subject-specific discourse competence that goes beyond the mechanical application of rules to include an understanding of the norms and values that underlie the discourse of the subject community.

The set of seven analytic rating scales developed in the context of this study can have immediate practical value beyond the context of this study. Since these scales are not task-specific, but focus on general conventions of subject-specific language use on the

level of the lexico-grammar and discourse-semantics, they can be applied across different tasks and even across different subject-specific contexts. Because each scale focuses on one of seven different aspects of subject-specific writing, the application of these scales can enable teachers to give learners detailed feedback on their task-based subject-specific writing that draws focused attention to specific aspects of language use and text structuring.

The frameworks used to develop the task-specific holistic and content scales and the criteria applied can again be of value for teachers or researchers also wishing to develop task-specific scales with which learners' written responses can be rated.

1.2 Overview of methodology

The six tasks used to elicit the written responses were developed in close co-operation with experts on geography education within the context of the project: Subject-specific learning and (foreign) language use: Task-based cognition, communication and co-operation[1], sponsored by the German Research Foundation[2]. The tasks, which were part of a geography competence test administered during a 90-minute test session, all cover aspects related to the tropical rain forest. All of the tasks also involve the integrated application of higher order subject-specific competences and the production of an extended constructed response.

The tropical rain forest was chosen as a central theme for the test, due to its importance in the subject's discourses, which again stems from its value as a unique ecosystem with an immense biodiversity and the serious effects its destruction has on the global climate. Furthermore, the tropical rain forest serves as an example of a specific environment that can be studied from different perspectives, for example in terms of its climate, its nature as ecosystem, its value as natural resource, but also as a living space and the conflicts this leads to etc. Its multi-dimensional nature thus allows a number of concepts central to geography to be illustrated.

The learners' responses to the six tasks were rated in three scoring processes involving routine and random sample double marking and the application of three different sets of

[1] *Fachlernen und (Fremd)sprachlichkeit: Aufgabenbasierte Kognition, Kommunikation und Kooperation*
[2] *Deutsche Forschungsgemeinschaft – DFG*

scales especially developed for the purposes of this study. The first set of scales consists of six holistic scales. These scales were specified per task and used to determine the appropriateness of the discourse function or degree of task fulfilment of a learner's response to a specific task. Each of these holistic scales makes provision for identifying seven levels of performance. The second set of scales includes seven analytic scales with five levels each. These scales were applied to the learners' answers to all six tasks and served to identify general strengths and weaknesses in the learners' subject-specific written language use that occur across all of the tasks. The third set of scales consists of collections of dichotomous and three-point scales, again specialised per task, and used to assess the correctness and completeness of the content of the learners' answers to each of the tasks.

The development of these three sets of scales firstly involved creating frameworks of expectations on the basis of a social semiotic and systemic functional linguistic interpretation of the context of subject-specific task processing. These frameworks included expectations in terms of the meanings to be constructed in learners' texts, as well as to their language use on the level of the lexico-grammar and discourse-semantics. On the basis of these frameworks, preliminary scales were developed and presented to the experts on geography education involved in designing the test, as well as to geography teachers. The educational experts and teachers commented on the clarity of the scales and on the extent to which they cover aspects regarded as relevant by the subject community in the rating of learners' written responses to subject-specific tasks. The feedback received was used to refine the scales. Furthermore, in an extensive process involving geography teachers and English graduates, the preliminary scales were tested on random selections of learners' texts and further improved in terms of their clarity and practicality and the extent to which they allow the level of a performance to be determined unambiguously.

The holistic scoring of the texts involved random sample double marking by two raters who were both involved in developing the scales and in identifying benchmark texts that illustrated the different levels of performance on each of the tasks. In terms of the analytic scoring of the language use, a random sample of eight learners' responses to five of the tasks was double marked applying the analytic scales. In rating the completeness and correctness of the content of the learners' answers, all responses were

double marked by raters who received training in applying the content scales. The scores of the raters were averaged in order to reach a final score per task.

1.3 Structure of dissertation

In the next chapter the need for assessment initiatives that focus on learners' *subject-specific language use* and discourse competence in the context of bilingual subject-specific education is explained in more detail. The chapter starts with an outline of the basic theoretical principles of systemic functional linguistics. A number of leading authors' views on the aims of bilingual subject-specific education in Germany are then related to views of school-based literacy and its role in preparing learners for participation in institutionalised discourses in public contexts. The chapter concludes with a brief overview of the other three studies conducted within the context of the project: Subject-specific learning and (foreign) language use: Task-based cognition, communication and co-operation.

Chapter 3 illustrates how theoretical concepts from systemic functional linguistics and social semiotics can be used in developing frameworks of expectations on which different types of rating scales can be based. This development process involves analysing the context of culture and context of situation of subject-specific task-based text production in order to identify the conventions the subject community follows in constructing meaning. It is firstly shown how these conventions pertain to the standardised models of experience with which subject communities operate and that repeatedly occur in their discourses as so-called "thematic patterns" (e.g. Lemke 1990). Secondly, attention is paid to conventions related to the standardised procedures subject communities follow in constructing meaning. These include forms of technical action, such as experiments, but also ways of interacting with texts. The textual realisation of these activities in the form of genres is then explained. The notion of register, as defined by Halliday (e.g. 1989) is then analysed and used to identify stylistic norms and conventions of subject-specific language use.

In the second part of chapter 3 the definition of a discourse function as presented by Zydatiß (e.g. 2004b, 2005) in the discussion on bilingual subject-specific education in Germany is reinterpreted from a social perspective. The resulting definition of the

concept of a discourse function is then used to define subject-specific task-based written discourse competence. In doing so, a structural model of geographical competence, developed by Vollmer (e.g. 2007, in preparation), is adapted on the basis of social semiotic views of the nature of the meaning-making processes. The chapter concludes with a framework outlining the processes and theoretical concepts involved in creating frameworks of expectations used to design scales that can be applied to rate learners' task-based subject-specific writing.

Chapter 4 includes a description of the sample and gives an overview of the processes involved in developing the elicitation tasks and rating scales used in this study, as well as of the procedures followed in administering the test and applying the scales. The chapter starts with a summary of guidelines for quality assurance in testing presented in the literature on language and performance assessment. Special attention is paid to the nature and potential of performance assessment, but also to the challenges that this form of assessment poses in terms of its reliability and the generalizability of results. The specific measures taken in the context of this study in order to promote the reliability of the assessment and the validity of the interpretations of the results are discussed in detail. This firstly involves a description of the process involved in developing authentic and interactive tasks in close co-operation with experts on geography education. The tasks used are also described in terms of the discourse function the responses are required to fulfil. Secondly, the extensive processes followed in developing the three types of rating scales used in the study are explained. The scales are also described and the criteria that underlie them are outlined. Moreover, the measures taken in support of reliability during the administration of the test and during the rating sessions are explained.

In chapter 5 the results of the three rating processes are presented and discussed. Frequency distributions are used to give an overview of the group of learners' performances on the different tasks and to identify deficient aspects of their written performances. The problems identified in each of the three rating processes are discussed and examples of learners' responses to the tasks are used to illustrate the problems found.

In the concluding chapter, the results of the study are used to make suggestions for the development of learners' subject-specific discourse competence in the subject-specific classroom. The suggestions are rooted in a social semiotic view of subject communities' language use. Finally, an outline is given of further research needed in the area of bilingual subject-specific training in Germany.

2
The importance of writing in subject-specific classrooms

In this chapter, I relate concepts from systemic functional linguistic theory to views on the aims of bilingual subject-specific education in Germany and social interpretations of school-based literacy. I argue that the role school-based literacy plays in preparing learners for participation in other discourses necessitates research on learners' ability to use the subject-specific register.

2.1 Subject-specific discourse: A social view

A number of basic theoretical concepts operated with in systemic functional linguistics (SFL) are important for the discussion on the aims of bilingual subject-specific education in Germany and the ways in which these relate to social interpretations of literacy. A brief overview of the principles of systemic functional linguistic views are therefore first given (a more in-depth discussion follows in chapter 3). This is followed by a definition of the term subject-specific discourse and a short outline of social views on the nature of learning in subject-specific contexts.

2.1.1 SFL: Basic theoretical concepts

Systemic functional linguists follow a social semiotic approach towards the study of language (e.g. Halliday 1978, 1985, 1989, 1998, Martin 1984, 1985, 1992, Halliday / Martin 1993, Lemke 1990, 1998, Eggins 1994, Hasan 1985, 1989, Martin / Veel 1998, Christie 1984, 1985, 1998 and Ventola 1987, 1991, 1997). In systemic functional linguistic views, the focus is on the *functional nature* of language and the ways in which it is used by different *discourse communities*. Language is seen as a *social semiotic tool*, or a *meaning-making system* that individuals apply to construct meaning as they *participate in the activities* of the communities of which they are members. The construction of meanings shared by different members of a discourse community is seen to be facilitated by an *adherence to meaning-making patterns*, or *conventions* followed by the community in its application of social semiotic systems. These patterns are related to the *meanings* that are made, as well as to the *structures used to realise these meanings.* Put differently, these patterns are firstly related to *that which is repeatedly said and written* - or in traditional terms to the content - and to *the ways in which* this

content is repeatedly expressed through the use of language and other social semiotic systems.

Members of communities are seen to use language and all other social semiotic systems, such as visual and mathematical representations, actions and gestures in order to construct three types of meanings: the *ideational*, in which language is seen to be used to represent a physical and biological reality; the *interpersonal*, where language is viewed in its function of constructing an orientation towards other interactants and towards the meaning one is making; and the *textual*, focusing on language's potential for organising meanings into texts. These three types of meanings are regarded as being *made simultaneously*, so that a particular instance of language use can be analysed from all three perspectives. Together, they realise a particular *register* that is associated with a specific situational context, such as the subject-specific classroom.

2.1.2 Subject-specific discourse

Applying the theoretical concepts outlined above, subject-specific discourse is defined for the purposes of this study as the conventionalised use of social semiotic systems by a specific subject community in constructing particular meanings as part of its activities. Subject communities are seen to be *school based*. In this respect, this study applies Berkenkotter's and Huckin's (1995: 13), as well as Christie's (1995: 155) view of subject-specific discourse as discourse that has been taken from other contexts, for instance from the predominantly university-based disciplinary contexts, and that has been situated in the activities of subject-specific communities operating in primary and secondary educational contexts. Bernstein (e.g. 1990) uses the term *pedagogical discourse* to refer to this *recontextualised discourse* that has been taken from other contexts, primarily from disciplinary contexts, and that has been *adapted for the purposes of primary and secondary education*. Christie (1998: 174) explains this phenomenon as follows:

> "The school is an important institution in which pedagogic discourses operate, taking forms of knowledge and their discourses from other institutional sites, and relocating these for the purposes of their selective transmission to learners. Teachers are the agents of the process of relocation, and these processes bring about transformations in the nature of the discourses because of the recontextualisation involved. Where the discourse of physics, for example, may be found in the institutional site of a university, it will be recontextualised and

> hence changed for the purposes of its transmission to school students." Christie (1998: 174).

One of the characteristic features of subject-specific discourse, taken over from disciplinary discourse, is the *pursuit of universality* and the consequent *standardisation of the conventions applied*, e.g. the use of the subject-specific terminology. In systemic functional linguistic views, subject-communities are seen to firstly operate with *standardised models of experience*, or models of subject-specific content, which are always realised in some or other social semiotic form, such as language. Lemke (1990: 202 – 203) refers to these networks of basic semantic relations that are repeatedly realised between particular concepts or thematic items as a subject-community's *thematic patterns* or *thematic formations*. Subject-specific terminology, for example, represents a large collection of thematic items that are repeatedly combined in particular semantic relations in the community's discourses, e.g. in textbooks, in teacher talk and in learners' answers.

Members of subject-communities are also seen to engage in particular *standardised activities*, or procedures in constructing meaning that also involve *standardised ways of using social semiotic systems*. These activities are firstly technical in nature, such as experiments, but also include those activities in which meaning is constructed on the basis of and through a variety of text types that can vary from being highly subject-specific, for example climate graphs for geography, to texts that are used across subject-specific contexts, such as tables. *Genres* are the "textualisations" of particular activities. Examples of genres that can be found in subject-specific contexts are descriptions, explanations and evaluations. These genres are realised by using the lexico-grammar and discourse semantics in ways that are conventional for the subject community.

Relating these theoretical constructs to a learner's production of a written text in response to subject-specific task, the meaning constructed in a learner's response is conceptualised as the result of a process during which a learner engaged in cognitive-semiotic activities. The learner constructs meaning on the basis of his or her knowledge of the subject's thematic patterns and of the social semiotic systems used in the task's input material, e.g. language, visual depiction, mathematical representations etc. The process further involves that the meaning constructed is presented in a written response.

Should the learner follow the conventions, the written response will realise a specific genre that is appropriate in terms of the cognitive-semiotic activities that were involved in the process of constructing the meaning.

Views held by systemic functional linguistic theorists about the *nature of learning* in subject-specific contexts are based on the theories of Bernstein (e.g. 1990) and Vygotsky (e.g. 1978). The aim of subject-specific education is seen to be concerned with initiating learners into the subject community's representations of experience and its practices involved in constructing these. Learners are seen as apprentices (e.g. Christie 1998) involved in a process entailing the building of identity, through their induction into various institutionally sanctioned values and practices. The teacher's role is a mediating one, guiding the learners into new ways of thinking, doing and talking through pedagogic discourse, which is either instructional or regulative (e.g. Bernstein 1990). Whereas instructional discourse is concerned with the transmission of specialised skills or competences and the ways in which they are related to each other, regulative discourse is used to create specialised order, relation and identity.

Against the background of systemic functional linguistic views of the dynamics of subject-specific discourse and the mechanisms involved in developing learners' ability to use it, two of the central aims defined for bilingual subject-specific education in Germany are now related to the notion of school-based literacy.

2.2 Bilingual subject-specific education in Germany: developing school-based literacy in a foreign language

Although authors writing on the subject of bilingual subject-specific education in Germany differ in terms of their views on the nature and aims of this form of education, there is agreement in the literature about the fact that *learners should be taught to use the register of the subject* and that the *competences developed in the subject-specific classroom should be transferable to other contexts*. In the following discussion, the differences and similarities between the views on bilingual subject-specific education are summarised. The concept of a transferability of competences is then related to views on the nature of school-based literacy. Finally, attention is paid to aspects that facilitate this transferability of competences to contexts beyond the classroom.

2.2.1 Views on the nature and aims of bilingual subject-specific education in Germany

Three different conceptualisations of the nature and aims of bilingual subject-specific education in Germany can be identified in the literature. These have been discussed at length by a number of other authors (e.g. Zydatiß 2002, Vollmer 2002b, 2002c, Abendroth-Timmer 2002, Hoffmann 2004) and are therefore only briefly summarised here. In its original conceptualisation this form of education is viewed as *general foreign language proficiency development on the basis of subject-specific content* (Wode 1995: 12). In this view the interactive use of the foreign language in a subject-specific context leads to a focus on content or meaning, rather than form. It is therefore seen as an effective way of developing learners' ability to use the language for real communicative purposes in contexts beyond the classroom.

The second view of bilingual subject-specific education turns the focus around and operates with the notion of *subject-specific education in a foreign language* (Hallet 1998: 117). In this view the development of subject-specific competences, including the ability to use language in order to participate in the discourses of the subject, is regarded as central.

The third view of bilingual subject-specific education also acknowledges the importance of developing learners' ability to participate in the discourses of the subject. The supporters of this view, however, call for a more conscious focus on the interaction between language and content in the context of the subject-specific classroom and consequently use the term *Content and Language Integrated Learning* (*CLIL*) to reflect their conceptualisation of the central aim of bilingual subject-specific education.

In spite of the differences between these conceptualisations of the nature and aims of bilingual subject-specific education in Germany, there is general agreement in terms of two aspects: firstly, that bilingual subject-specific education should develop learners' ability to use the foreign language in order to *participate effectively and actively* in the *activities of the subject*, in the sense of Hallet's (e.g. 2004) "Arbeitssprache", Vollmer's

(e.g. 2002b) "wissenschaftspropädeutischer Denk- und Arbeitsfähigkeit" and Zydatiß's (e.g. 2000) subject-specific literacy ("Sachfachliteralität"). Secondly, there is agreement that bilingual subject-specific education should develop a more generalised ability to use the foreign language in *participating in discourses beyond the subject-specific classroom*. This consensus is reflected in the following statement by Vollmer (2002c: 107):

> "Dass bilingualer Unterricht zu einer allgemein – *und* fachsprachlich basierte fachlichen Handlungskompetenz führen muss, ist allgemein akzeptiert." (Vollmer 2002c: 107. Italics in original).

If one accepts that bilingual subject-specific education should enable learners to participate in the activities of discourse communities beyond the school context, the following two questions arise: firstly, which types of contexts can subject-specific discourse competence transfer to most effectively? Secondly: which factors can support or facilitate this transfer of competences from the classroom to other contexts? The first question will now be addressed by relating socioliterate conceptualisations of school-based literacy to views on the nature of subject-specific discourse presented in the literature on bilingual subject-specific education. The second question is addressed by drawing from Zydatiß's (2002b, 2004, 2005) view on the importance of text competence in bilingual subject-specific education, Vollmer's (2002c) call for a focus on language awareness and Hallet (1998, 2002, 2004) and Breidbach's (2001, 2003, 2004) concern with cultural awareness.

2.2.2 School-based literacy as a prerequisite for participation in institutionalised discourses

In social views of *literacy*, reading and writing are seen to be embedded in the activities of particular communities. Consequently, the aspects involved in being able to read and write effectively is regarded as being context dependent. Communities' use of language in written form is seen to adhere to particular functionally-based conventions, facilitating the achievement of individual goals in a social context. Johns (1997: 17) for example, using the term literacy in its plural form to indicate its dependence on context, defines it as follows:

> "Literacies involve *'socially-approved acts'*, taking place within environments in which an *informal and unstated contract* exists between readers and writers. Successful text processing and production involves understanding the terms of this contract, terms that include text content, form, register, quality of paper, context and many other factors. We understand and use texts because we are part of a community that accepts this contract as an efficient means for comprehending and producing particular written discourses. We read and write successfully because we work within the terms of the community contract." (1997: 17. Italics: DC.).

Literacy theorists, such as Cope and Kalantzis (1993) and Gee (2002) distinguish between *primary* and *secondary discourses*. They relate the first category of discourses to processes involved in socialising an individual into a specific linguistic and cultural community. The second category of discourses is associated with a larger cultural context and is seen to facilitate interaction in the regulated activities of institutions, e.g. educational, professional and governmental or administrative. The school is regarded as instrumental in developing learners' ability to participate in these secondary discourses, because of the similarities between the conventions regulating discourse in these different contexts.

Put simply, the main similarity between language use in the subject-specific context and other institutionalised public contexts is a reduced reliance on shared knowledge between interactants, resulting in language use that is verbally explicit. As already pointed out in outlining the theoretical concepts of SFL, systemic functional linguists refer to subject communities' pursuit of universality and their consequent standardisation of the discourse conventions applied. In his in-depth analysis of the discourse of science classes, Lemke (1990), for instance, lists stylistic norms that teachers and learners operate with in subject-specific contexts. He (Lemke 1990) shows that not only teachers require these norms to be applied, but that learners also react when teachers do not adhere to the conventions. One of the norms he (Lemke 1990: 133) identifies is:

> "Be as *verbally explicit* and *universal* as possible. This means that verbal, rather than gestural or other nonverbal signs are required, and that implicit forms of grammar (e.g., use of pronouns of any kind, or ellipses) are not fully acceptable. Teachers do not accept statements like 'It reflects it' or 'One left in each' as complete answers. Statements also need to be expressed in the form of propositions that seem to have universal validity (e.g. using the indefinite present tense) and make no reference to the here and now, to accidental features

> of the immediate classroom situation, or to specific people, time, and places." (Lemke 1990: 133).

As Pellegrini and Galda (1998: 1, 9) point out in their discussion on school-based literacy, this need to be as verbally explicit and universal as possible can be attributed to the fact that the classroom register requires learners to operate with constructs of themselves and others as generalized others, with a minimum of shared experiences. This is reflected in the use of the lexico-grammar and discourse-semantics in ways that realise the intended meanings as explicitly as possible. Furthermore, in spoken interaction, using the classroom register additionally involves reduced reliance on contextual cues, so that meaning is predominantly constructed through linguistic means.

This feature of subject-specific and other forms of secondary discourses clearly distinguishes these discourses from primary discourses. Interaction in the family and amongst friends is based on a higher degree of shared knowledge and shared experiences allowing language use to be verbally less explicit than in discourses taking place in public spheres. In the case of institutionalised discourses found in educational, professional and administrative contexts, the conventions regulating the discourse are also standardised. These standardisations range from the lexical level, for example in the form of standardised terminology, to the textual level, for example in the form of the specified text types of these discourse communities.

This distinction between the language of the classroom and that of the family is also found in the literature on bilingual subject-specific education. Especially Cummins's (1979, 1981) distinction between *CALP* (Cognitive / Academic Language Proficiency) and *BICS* (Basic Interpersonal Communicative Skills) clearly points to the differences in the demands placed on learners' communicative proficiency in primary and secondary contexts. In addition to the notion of a reduced reliance on contextual support, CALP is also related to the ability to use language for cognitively demanding tasks.

In following Cummins (1978, 1979), Zydatiß (2004b: 95) summarises the nature of the ability involved in using language in a school context as follows:

> "Im Gegensatz zur situations- und handlungsgebundenen Alltagsprache erweist sich die schulrelevante Sprachkompetenz üblicherweise als schriftsprachlich fundiert, linguistisch explizit (da paralinguistische Indikatoren und

> Situationskontext fehlen), sprachlich elaboriert (da kognitive Operationen zu verbalisieren sind) und inhaltlich komplex (weil auf vielschichtige Sachzusammenhänge verwiesen wird)." (Zydatiß 2004b: 95).

In socioliterate views, the notion of tasks that are cognitively more demanding, is related to the fact that a learner is less familiar with the thematic patterns (content) the task deals with and consequently with the language in which these patterns are realised verbally. Christie (1998: 152), for example, points out that much of school knowledge is "uncommonsense", because it is not habitual and not constructed in familiar language.

One of the unfamiliar ways in which language is used by subject communities is the use of subject-specific terms that originate in the discourses of the disciplines and which are then adapted for the teaching and learning activities of the classroom. In communities concerned with the scientific observation and explanation of experiential meaning, the phenomena that are observed are classified into taxonomies and are labelled involving the use of technical terms. For the members of these communities, the taxonomies, or networks of semantic relations and the technical terms serve to organise an interpretation of experience based on scientific observation and functions as a way of distilling meaning. In other words, the purpose of scientific terms for these communities is to compact meaning, but they simultaneously change the nature of more commonsense meanings by which they are defined (e.g. Halliday / Martin 1993: 29-30, Wignell *et al*. 1993: 164-165).

For the initiated members of the scientific or subject community, the use of these terms is functional. It serves to allow interaction without constantly having to elaborate in order to explain explicitly and precisely what is meant. However, for learners who are still in the process of being initiated into the subject community, these terms may obscure meaning and the thematic patterns constructed through the language use may represent an interpretation of experience that is far removed from what they know from personal experience.

This view of the *alienating effect* of subject-specific discourses pointed out by systemic functional linguists is also addressed by Breidbach (e.g. 2001, 2003, 2004) in his

contribution to bilingual subject-specific education in Germany. In addressing the differences between the models of experience with which learners operate and those introduced in the classroom, Breidbach (2004: 156) writes:

> "Für jeden wissenschaftsorientierten Fachunterricht stellt sich das Problem, dass wissenschaftliche Abstraktionen und Denkweisen kognitiv außerhalb des alltagspraktischen Handlungs- und Erfahrungshorizontes der Lernenden liegen und demzufolge im Lernprozess nicht aus diesem hergeleitet werden können." (Breidbach 2004: 156).

Thus, based on these views on the nature of the school-based register and the difficulties involved in mastering it, developing subject-specific literacy can be said to involve the development of learners' ability to use *a specific register* that is different from the registers embedded in the discourses of family life and interaction with friends. It means developing learners' ability to use language to construct meaning *in explicit ways* for a generalised audience and to engage with *unfamiliar interpretations of experience* that are different from their own.

Subject-specific literacy is not only paramount to successful participation in the activities of school, but as pointed out earlier, also forms the basis for active participation in all other forms of institutionalised discourses. These firstly include the discourses of other educational institutions and professional communities. Furthermore, as will be argued in the next section, the ability to use the school-based register has become important for participation in the activities of discourse communities situated in a wider social context.

2.2.3 School-based literacy as a prerequisite for autonomous participation in administrative, political and commercial activities

The discourse of science has gained an immense social power in most industrial societies. As Hallet (2004: 144) for instance points out in quoting Böhme *et al.* (2000), in so-called modern societies, science has also come to shape commonsense experiences through its alliance with technology and because it serves as a general source of models that explain experience. Recognising the powerful position that scientific discourses have in society today, Halliday and Martin (1993: 11) write:

> "Every text, from the discourses of technocracy and bureaucracy to the television magazine and the blurb on the back of the cereal packet, is in some way affected by the modes of meaning that evolved as the scaffolding for scientific knowledge.
> In other words, the language of sciences has become the language of literacy. Having come into being as a particular kind of written language, it has taken over as model and as norm. Whether we are acting out the role of the scientist or not, whenever we read and write we are likely to find ourselves conjured into a world picture that was painted, originally, as a backdrop to the scientific stage. (Halliday & Martin 1993: 11).

In other words, the public discourses in industrial societies have come to be based on what is referred to as rational thought and logic. These ways of thinking represent a way of using language in which only particular semantic relations between particular concepts and not others are seen to be making sense. These modes of constructing experience have their genesis in the sciences and have become naturalised – they have become the predominant ways of knowing and of making meaning in these societies. Scientific modes of constructing meaning are regularly used by politicians and companies in their communication with voters and consumers, for example when a politician cites statistics to support his suggestions for reforms of the social system, when marketing companies use images of experts in laboratories explaining the features of a specific product, or when oil companies explain rising petrol prices by drawing a causal relation with events in the Middle East.

The power and status that the discourse of science enjoys in society today, thus makes school-based literacy, developed in the different subjects, a key qualification. It is in the context of the subject-specific classroom that the learner is guided in a planned and structured way into a scientific view of the world. In this view, each phenomenon can be analysed through the application of social semiotic systems of technical action, such as the experiment. Furthermore, phenomena can be explained through the use of a social semiotic system such as language that allows the expression of a network of semantic relations, regarded to be acceptable or true. It is the responsibility of the subject-specific teacher to guide learners towards a disposition where they can make informed political decisions, based on the ability to understand how the politician uses language to construct a particular interpretation of experience. They must be able to inform themselves about genetically modified food, alternative methods for cancer treatment and the advantages and disadvantages of the newest communication technology by

consulting the documents available on the Internet and published in the mass media. Learners must be empowered to engage in critical dialogue with local authorities about rising energy costs, using patterns of reasoning that are regarded as valid by the authorities. Developing these abilities means developing transferable competences that lay the foundation for empowered participation in discourses in public spheres.

In terms of bilingual subject-specific education, the aim should be to develop learners' ability to participate in institutionalised, technological and political discourses of a globally connected world, promoting their personal and professional mobility and their independent and critical use of international networks of communication, such as the Internet. This aim is described by Zydatiß (2002: 37) as follows:

> "Das curricular geplante, didaktisch strukturierte und relativ langfristig angelegte Konzept des bilingualen Sachfachunterrichts scheint ein erfolgversprechender Weg zu sein, Heranwachsende zu einem fortschreitend eigenverantwortlichen inhaltsbezogenen Gerbrauch einer Fremdsprache in der Kommunikationspraxis des Schulalltags anzuleiten – in der Hoffnung, dass der gesellschafts- und bildungspolitisch motivierten kommunikativen Interaktion im Klassenzimmer ein 'autonomes' kommunikativ-kulturelles Handeln in der Lebenspraxis der Jugendlichen und später (bzw. werde kann)." (Zydatiß 2002: 37).

In order for this transfer of abilities from the classroom to other educational, professional, political, administrative and commercial contexts to occur, learners need to be enabled to reflect about their meaning-making processes and the social contexts in which they are embedded. The following section combines three views on the aspects learners should be brought to reflect about and which are presented in the literature on bilingual subject-specific education in Germany. It is argued that a construct making provision for reflection on the level of the lexico-grammar, the text and the larger cultural context, is needed to empower learners for reflective and autonomous participation in various contexts.

2.2.4 Reflecting about the meaning-making process as a prerequisite for the transferability of school-based literacy

In their contributions to the literature on bilingual subject-specific education in Germany both Vollmer (2002a, 2002b) and Zydatiß (2002b, 2004, 2005) explicitly pay attention to aspects, which according to them, promote the transfer of subject-specific

competences to wider situations of use. Vollmer (2002a, 2002b) calls for the development of learners' language awareness and ability to reflect about language use and Zydatiß (e.g. 2002b, 2004, 2005) suggests a focus on text competence. *Language awareness* is associated with the ability to reflect about the relations between content and form, or structures and their meanings (e.g. Gnutzman 1995, 1997). These structures can be lexical, grammatical or textual. Zydatiß's (e.g. 2004b, 2005) notion of text competence specifically draws attention to reflection on the level of the text.

Zydatiß (2004b: 94-95) views *text competence* as the ability to construct meaning on the basis of and through specific genres (text types) that deal with specified meanings (content). The focus is on the interaction with and the production of longer stretches of language that are packaged as coherent and cohesive thematic and functional units. He (e.g. 2002, 2004b, 2005) equates discourse functions with categories such as description, conclusion, explanation and evaluation. In his (Zydatiß 2005) view discourse functions establish areas of interaction between language, cognitive processes and cognitive structures, and form a basis for transfer across contexts, due to their generic nature[1]. The sufficient development of the subject-specific competences involved in constructing meaning on the basis of and through texts embedded in class-based subject-specific discourses, is seen to be a prerequisite for this transfer (Zydatiß 2004b: 97).

Zydatiß's (e.g. 2004b, 2005) and Vollmer's (2002c) views on the aspects that facilitate a transfer of competences from the bilingual subject-specific context to participation in discourses beyond the classroom should be extended to include concepts from cultural theory. Hallet (e.g. 1998, 2002, 2004) and Breidbach's (e.g. 2001, 2003, 2004) contributions to the literature on bilingual subject-specific education in Germany draw from this theory. Both call for the development of learners' ability to critically reflect upon the ideological nature of discourses and the ways in which social semiotic systems like language are used by particular communities to construct specific interpretations of experience. They attribute the potential of bilingual subject-specific education to foster

[1] In chapter 3, Zydatiß' view of discourse functions is re-interpreted from a systemic functional linguistic and social semiotic point of view. In the definition of discourse function operated with in the context of this study, categories such as descriptions, explanations and evaluations are seen as genres, or as text categories realising specific cognitive-semiotic activities. Discourse functions are viewed to be more situation specific, as they combine particular meanings with a specific genre.

critical cultural awareness and reflection to the differences, even conflict, between learners' personal interpretations of experience and those of the subject community. According to them, these differences are further emphasised by the use of a foreign language. Hallet (e.g. 1998, 2002, 2004) and Breidbach's (e.g. 2001, 2003, 2004) views in this regard relate to the alienation effect learners experience in subject-specific contexts, referred to by Halliday and Martin (1993) and discussed earlier in this chapter.

Following Bhabha's (Bhabha / Rutherford 1990) concept of a third space, Hallet (e.g. 2004) views the bilingual subject-specific classroom as a *hybrid space* in which there is an interplay between the disciplinary discourse, which is re-interpreted for the classroom as educational discourse, the common sense ways of using language and the discourses of the different national cultures represented by the languages involved, e.g. German and English. According to Hallet (2004: 150) this hybrid space creates the basis for a transcultural discourse in which meanings not made before can be constructed. Hallet (2004: 150) consequently argues for an approach towards bilingual subject-specific education in which focused attention is paid to the development of an awareness of the constructed nature of culture and a reflection about personal cultural identity.

Breidbach (e.g. 2004) is primarily concerned with a reflection about the ways in which scientific discourses construct particular representations of experience. As cultures with practices that are determined by specific values and norms, scientific communities are seen to construct models of reality that can be critically analysed. Breidbach (2004: 157) calls for subject-specific education to provide a basis for critical analysis of the representations of experience with which a specific disciplinary community operates and of the ways in which these have been constructed over time.

The contribution that Hallet (e.g. 1998, 2002, 2004) and Breidbach's (e.g. 2001, 2003, 2004) views make to a conceptualisation of the dynamics involved in a transfer of competences from the subject-specific classroom to contexts beyond school, is the introduction of an ideological concern, of *a critical awareness of norms and values*, those with which one personally operates, as well as those that underlie the discourses in which one wishes to participate (or not). Such an awareness is needed in order to ensure an independent and self-regulatory participation in discourses. It is not enough

for an individual to be able to reflect about the functionality of a specific text structure in a specific context, or the relation between content and form when using language in participating in the activities of a specific discourse community. Individuals need to be enabled to negotiate between their personal values and goals and the values and functionally-based conventions of a particular discourse community. This implies an awareness of the ideological nature of all discourses and an ability to reflect upon the values regulating discourse in a particular context.

Integrating these views about the role of reflection in empowering learners to participate in discourses beyond the classroom, leads to a construct that makes provision for reflection on three levels of increasing abstraction: Firstly on the level of the specific social semiotic code, or the language system (cognitive language awareness), secondly on the level of the text that represents a specific text structure (genre) that is a conventionalised textual realisation of particular cognitive processes (awareness of the generic nature of texts) and finally on the level of the larger meaning-making practices of the discourse community and the ideological constructs that determine these (reflection about the constructed nature of culture and discourse).

The following section addresses the lack of research into learners' use of the subject-specific register in bilingual subject-specific education in Germany and explains the need for the development of instruments with which learners' use of the subject-specific register can be assessed.

2.3 The need to assess bilingual learners' use of the subject-specific register

The fact that school-based literacy functions as a basis for empowered participation in a number of other discourses in public contexts, necessitates the development and assessment of learners' *subject-specific* language use. Looking at the literature on bilingual subject-specific education in Germany, it is surprising that very little research has been done in this regard so far. Studies conducted in terms of learners' language ability have focused on general language proficiency ignoring aspects of their use of the subject-specific register (e.g. Vollmer, 2002b, Zydatiß, 2004a: 89).

One study that does indeed address subject-specific language use in bilingual subject-specific education in Germany was conducted be Zydatiß (e.g. 2004b, forthcoming). In this study, he Zydatiß (e.g. 2004b, forthcoming) compared the ability of bilingual and monolingual learners to construct meaning on the basis of subject-related texts and through the production of coherent written responses. The study involved learners from geography, history and biology. Consequently, Zydatiß (e.g. 2004, forthcoming) did not assess the learners' written discourse competence related to a specific subject, but rather operated with a construct of cross-curricular text or discourse competence. The results of the assessments are to be published soon (Zydatiß, forthcoming).

Furthermore, the study of Dielman (in preparation) aims to address the current lack of research on German bilingual learners' use of the subject-specific register. In her study, Dielman (in preparation) analyses the written responses of 10th and 12th grade learners following history in English to subject-specific tasks requiring the production of a text that generates a specific discourse function. Based on an assessment of learners' performances on these tasks, Dielman develops a competence model that enables her to design tasks addressing specific deficiencies in the learners' subject-specific discourse competence. Results in this regard are, however, not yet available.

Except for Zydatiß's (e.g. 2004b, forthcoming) and Dielman's (in preparation) studies, only two brief discussions on the assessment of task-based subject-specific oral and written production could be found in the literature on bilingual subject-specific education in Germany. Both these discussions show a limited view of the dimensions involved in effective participation in the discourses of the subject community and focus exclusively on the penalisation of errors, further ignoring issues related to qualitative differentiation between performances.

In the first of these two discussions, Ernst (1995) presents a classification of language errors, developed on the basis of a categorisation of errors that Giese (1994) identified for foreign language teaching. The second contribution by Lenz (2004) focuses on the assessment of written performances in geography and suggests particular adaptations to Ernst's (1995) classification. Lenz (2004) further argues for the increased use of assessment tasks that require the integrated application of higher level subject-matter,

procedural and communicative competences. However, he (Lenz 2004) does not present any framework for the rating of performances on such tasks.

The study presented here consequently aims to address the need for more insight into bilingual learners' use of the subject-specific register, as well as the need for instruments with which their task-based written products can be assessed. In order to develop bilingual learners' ability to use the subject-specific register, it is necessary to assess their language use in response to subject-specific tasks and to identify aspects that deviate from the subject-specific conventions. The study therefore sets out to present instruments with which learners' use of the subject-specific register can be assessed in a comprehensive way. It also aims to identify aspects of the task-based written performances of a group of bilingual learners that deviate from the subject's conventions.

As will be argued in the last chapter, bilingual subject-specific teachers should address problematic aspects of learners' subject-specific language use in a focused way as part of the process of initiating learners into the discourse practices of the subject community. This process entails instructing learners in the subject's standardised meaning-making practices, including the standardised models of subject-specific content and standardised ways of using language and other social semiotic systems to present this content. This process also involves bringing learners to understand the functional nature of these conventions, so that they do not simply mimic them. Learners need to be able to reflect about the meanings they construct and their use of social semiotic systems in constructing them. They also need to be able to reflect upon the values that underlie the discursive practices of the subject community and the dynamics involved in achieving individual goals in the subject-specific context. Only by paying attention to all of these factors can learners be prepared for effective and autonomous participation in other educational, professional, administrative, political and commercial discourses.

As already mentioned in the previous chapter, this study has been carried out in the context of the project “Fachlernen und (Fremd)sprachlichkeit: Aufgabenbasierte Kognition, Kommunikation und Kooperation” (Subject-specific learning and (foreign) language use: Task-based cognition, communication and co-operation). The next

section of this chapter outlines the three remaining studies conducted as part of this project. All three of these studies also focus on bilingual learners' use of language in a subject-specific context and consequently promise to address the current lack of knowledge and understanding in this respect.

2.4 Further studies on bilingual learners' processing of geography tasks

Vollmer's (e.g. 2006a, 2006b, 2006c, 2007, in preparation) study is specifically concerned with a *comparison between aspects of the subject-specific competence of bilingual and monolingual learners*. Final results are not available yet, but a first analysis of the data suggests that there are no significant differences between these two groups of learners' subject-specific competences. In order to enable a comparison between these groups of learners, Vollmer (in preparation) created comparative groups on the basis of data elicited about a large number of variables such as general scholastic achievement, socio-economic background and general proficiency in German and English.

Furthermore, Vollmer (e.g. 2007) has designed a structural competence model that identifies three basic areas of competence: knowledge of subject matter, methodological competence and subject-specific communicative competence. In his model, Vollmer (2007: 288-289) defines knowledge of subject matter as declarative knowledge of basic subject-related concepts and structures, as well as of the ways in which these can be extended and developed. This knowledge is firstly seen to be applied in naming, describing and explaining geographically relevant phenomena, events and relations. It is secondly regarded as the basis for the ability to combine knowledge of different concepts into conceptual networks that can be expanded. Methodological competence, in Vollmer's (2007: 289) view, is the ability to define a problem and to decide on procedures for addressing it. This competence is seen to include the activation of existing knowledge, but also the construction of new knowledge, e.g. by linking information extracted from input material. Subject-specific communicative competence is seen to involve the discursive use of the L1 or the foreign language in subject-specific contexts in oral and written reception, production and interaction. Vollmer (2007: 290)

equates this competence with the ability to use a specific variety (referred to as register in the discussion above) of school language that is subject based.

Following Zydatiß (e.g. 2004b, 2005), Vollmer (e.g. 2007: 291) postulates a second competence level that represents the integrative application of these competences. The integration, in his view, is stimulated by a task with a specific discourse function, which he, as Zydatiß, equates with categories such as description, explanation and evaluations.

Whereas this study and the one conducted by Vollmer e.g. (2006a, 2006b, 2006c, 2007) are concerned with measurement issues and focus on the learners' written products, the studies carried out by Heine (2007) and Troschke (in preparation) focus on the production processes. Heine's study (2007) deals with the production processes of bilingual and monolingual learners working on subject-specific tasks on their own. Troschke (in preparation), on the other hand, investigates the dynamics of bilingual learners' co-operative task-processing.

The study conducted by Lena Heine (2007) is specifically concerned with *the effect that the use of a foreign language can have on cognitive processes* occurring during task processing. Applying a model of cognitive-linguistic task processing, developed on the basis of an in-depth study of the literature on cognitive theory, Heine (2007) analyses think-aloud protocols recording learners' processing of a number of subject-specific tasks.

Based on the results, she formulates seven hypotheses. The first four are related to using language in spoken production and in writing and can be summarised as follows: firstly, the expression of conceptual relations in linguistic form leads to a deeper semantic processing of the content. Secondly, learners who are using a foreign language in their processing experience more problems in expressing their thoughts than learners using the L1. Thirdly, problems in terms of expressing meaning lead to a reconstruction of the semantic field involved and can lead to a deeper understanding of the conceptual relations. Fourthly, the use of the foreign language leads to more reflection about the language use and consequently to a deeper semantic understanding.

The last three hypotheses are related to using language in reading. These hypotheses are: Bilingual learners have more comprehension problems than learners using the L1. Secondly, bilingual learners compensate for a lack of understanding by making use of other contextual cues and they continue attempts to understand the input material, even when they are experiencing considerable problems in understanding the texts. Thirdly, learners using the L1 also show a significant number of problems in their language use.

Finally, Troschke's (2006, 2007, in preparation) study is concerned with *the negotiation of meaning between learners working on tasks in pairs*. Analysing video recordings of the learners' task processing, she (Troschke 2006, 2007, in preparation) applies a model that enables her to identify three areas in which negotiation of meaning can occur: the content area, the linguistic area and area of discursive organisation. Final results are not yet available.

The overall aim of the project is to make an important contribution to an understanding of the nature of bilingual learners' use of the foreign language for subject-specific purposes. In doing so, the project sets out to address the lack of research into bilingually trained learners' use of language for subject-specific purposes and to lay a foundation for further systematic investigations into aspects related to subject-specific literacy and discourse competence.

2.5 Conclusion

In order to effectively develop bilingually trained learners' ability to use the subject-specific register and to prepare them for institutionalised discourses beyond the classroom, more insight into their use of language in subject-specific contexts is needed. This requires structured assessment initiatives. This specific study aims to contribute to the understanding of bilingually trained learners' use of the subject-specific register through an extensive assessment initiative, involving the application of three types of rating scales to a large number of learners' written responses to subject-specific tasks.

In rating the written responses of a group of bilingually trained geography learners to a selection of subject-specific tasks, the study sets out to identify strengths and weaknesses in the learners' use of the subject-specific register. Furthermore, through the

definitions of discourse function and subject-specific discourse competence presented in chapter 3, theoretical constructs are offered that can be used in other assessment initiatives involving the development of rating scales. The practical implementation of these constructs in an assessment initiative is exemplified by the scales that have been designed for the purposes of this study. Finally, some of the challenges involved in promoting reliability and validity in performance assessment are illustrated in the discussion on the development of the tasks and the scales presented in chapter 4.

In the next chapter, the theoretical constructs used in developing the rating scales applied in this study are discussed in detail. In applying theoretical constructs from SFL theory to task-based subject-specific writing, the focus is not specifically on *bilingual* subject-specific education. Rather, the discussion is concerned with developing an understanding of a *school-based* and *subject-specific* register, regardless of the language involved.

3
A social semiotic framework for the rating of subject-specific writing

In this chapter, I apply principles and constructs from social semiotic and systemic functional linguistic theory in order to develop frameworks of expectations in terms of learners' written responses to subject-specific tasks. Further, I explain how these frameworks are used to develop holistic and analytic scales that can be applied in rating learners' texts. It is firstly illustrated how these expectations are identified through analyses of the context of the subject community's culture, the context of test-based task processing and the context of the task. Secondly, I present the definition of the construct of a discourse function applied for the purposes of this study and explain how it merges expectations from the context of the subject community's culture and the context of test-based task processing. Furthermore, I present a definition of subject-specific task-based discourse competence, which is operationalised in the holistic scales applied in the context of this study. Finally, the chapter concludes with a summary of all social semiotic and systemic functional linguistic constructs involved in developing frameworks of expectations.

3.1 Rating task-based subject-specific writing: the need for an integrative and socially-based theory

Reliability and validity are two important quality criteria for assessment initiatives. The application of rating scales promotes *reliability* in the assessment of learners' task-based subject-specific writing, since all learners' texts are rated against the same descriptions of levels of performance. In order to make *valid* decisions about learners' performances, these scales should firstly reflect the conventions that regulate the subject's discourse and secondly make it possible to determine the extent to which a learner's text adheres to or deviates from these conventions.

Although the application of rating scales promotes reliability, it does not make the assessment process an objective and value-free endeavour. Designing rating scales involves subjective decisions made by individuals who determine the categories, criteria and values that underlie the assessment process. In support of a valid use of assessment

results, it is therefore necessary to *explicate the process* involved in identifying categories and criteria. This promotes *transparency* in terms of the basis for the decisions made in the rating process. Furthermore, as a *contribution to the discourse on procedures* involved in assessment, descriptions of the development process create platforms for theoretical discussions and debates that can lead to methodological advancements in the testing field.

Systemic functional linguistics (SFL) provides a socially-based theoretical framework for the identification of rating categories and criteria. Furthermore, the procedures followed by these theorists and the concepts applied by them in the analyses of texts, provide guidelines for the processes involved in designing rating scales.

Over the last thirty years, systemic functional linguists, such as Halliday (1978, 1985, 1989, 1998), Martin (1984, 1985, 1992, Halliday / Martin 1993), Lemke (1990, 1998), Eggins (1994), Hasan (1985, 1989), Veel (Martin / Veel 1998), Christie (1984, 1985, 1998) and Ventola (1987, 1991, 1997) have been developing a theory and models that attempt to explain the discursive practices of disciplinary and subject-specific communities. In doing so, they have been applying the principles of social semiotics in analysing the ways in which members of these communities use language in order to construct meaning.

The collection of concepts and models developed by this group provides a basis for an *integrative theoretical account* of task-based subject-specific writing. It is integrative in the sense that it explains the interaction between the meaning constructed in task-based subject-specific writing, the use of various systems of representation and the processes involved in the construction of meaning. In more cognitively-oriented terminology, this approach allows an understanding of a written product as the result of cognitive processes involving content and language knowledge. Thus, not only language's role as a meaning-making system, but also that of subject-matter models and subject-specific methodologies and procedures are taken into account. Applying the core theoretical principles of systemic functional linguistics enables the development of rating scales that make provision for the multi-faceted nature of the meaning-making processes involved in task-based subject-specific writing. Integrative rating scales are needed in

order to gain a comprehensive insight into the nature of learners' written responses to subject-specific tasks.

Being a social theory of the contextualised use of language, SFL acknowledges that language use conventions are functionally motivated. It therefore also creates *a platform for feedback* about task-based subject-specific writing that goes beyond the mere listing of instances of inappropriate language use, to include explanations of the social motivations behind standardised patterns of form-function relations. This type of understanding is needed if learners are to be empowered to use language creatively, but also appropriately in constructing individual meaning in social contexts.

In the following discussion, the concepts developed by social semiotic theorists and systemic functional linguists that are relevant for the purposes of this study, are defined and explained in relation to task-based subject-specific writing in geography. These concepts are used in developing socially-oriented scales that can be used in rating task-based subject-specific writing.

3.2 Basic implications of a social semiotic approach towards the assessment of learners' subject-specific task-based writing

3.2.1 Developing frameworks of expectations: The importance of context

As already briefly explained in chapter 2, the basic principles of social semiotics can be summarised as follows (e.g. Lemke 1990, Eggins 1994): Social semiotic theorists study processes by which individuals *construct meaning*, using *social semiotic systems*, e.g. language, in participating in the activities of the communities of which they are members. In other words, specific instances of meaning-making processes and products are studied *in relation to the social contexts* in which they occur and are compared to the *patterns* according to which semiotic systems are *conventionally used* in these contexts in order to make the construction of shared meanings possible. These patterns are referred to as *semiotic formations*. The particular ways in which meaning-making systems are used by individuals and communities are seen to be *functionally motivated*

and *three basic functions* are distinguished: the ideational, also called the presentational, or thematic, the interpersonal, also referred to as the orientational and the textual or the organisational function. In its ideational function, language is used to represent a physical and biological reality. The interpersonal function serves to construct a social reality and in its textual function, language is used to organise ideational and interpersonal meaning.

The implications these social semiotic principles have for the development of a theoretical framework for the scoring of learners' task-based subject-specific writing in geography can be outlined as follows: A learners' *written response* to a geography task is viewed as a *semiotic product* – an instance of language use and possibly also of the use of visual and mathematical forms of representation in order to construct meaning. The meaning-making processes do not only occur as the learner writes his or her answer. These processes naturally also occur during the reading process, as the learner interacts with a task's instruction and input material, which in the case of geography tasks may include a wide variety of text types such as definitions, maps, graphs and tables with statistics.

The assessment of a learner's response to a subject-specific task entails *relating it to its context* in order to identify ways in which this specific instance of language use is similar to and differs from the patterns of meaning and of using language as a meaning-making system found in the subject community. Just as members of communities approach texts with particular *expectations*, based on assumptions about ways in which meaning is conventionally constructed in the type of context the text is embedded in, raters also approach learners' responses to subject-specific tasks with particular expectations. These expectations arise from relating the demands of the task to the subject community's semiotic formations. The appropriateness of the learner's response is then determined on the basis of the extent to which it meets these expectations.

These expectations can be formalised and explicated in the form of *rating scales*. Both holistic and analytic scales can be developed on the basis of these expectations. Holistic scales allow a rater to score a performance based on an overall impression of the text as a whole and can be applied to assess the degree of task fulfilment. Analytic scales focus raters' attention on distinguishable aspects of a text. The language use in a text can for

instance be divided into aspects related to coherence and cohesion, as well as to the appropriateness of the language use. These two categories can again be broken down into more detailed categories. In terms of coherence and cohesion, it is possible to distinguish between sentence order, organisation of meaning into different sentences and the linking of these sentences. As far as the appropriateness of the language use is concerned, a distinction can be made between the use of subject-specific terms, formal language, clear and succinct expressions, general academic vocabulary, structures realising semantic relations and a standard grammar.

In developing the rating scales used for the purposes of this study, *three contexts* showing a decreasing level of abstraction and generality have been analysed: Firstly the *context of the subject culture*, in which the second context, *the assessment activity* or *context of situation*, is embedded. The assessment activity is again a larger context in which the third context, namely *the task* is embedded. Figure 3.1 visually represents the way in which these three contexts are interrelated.

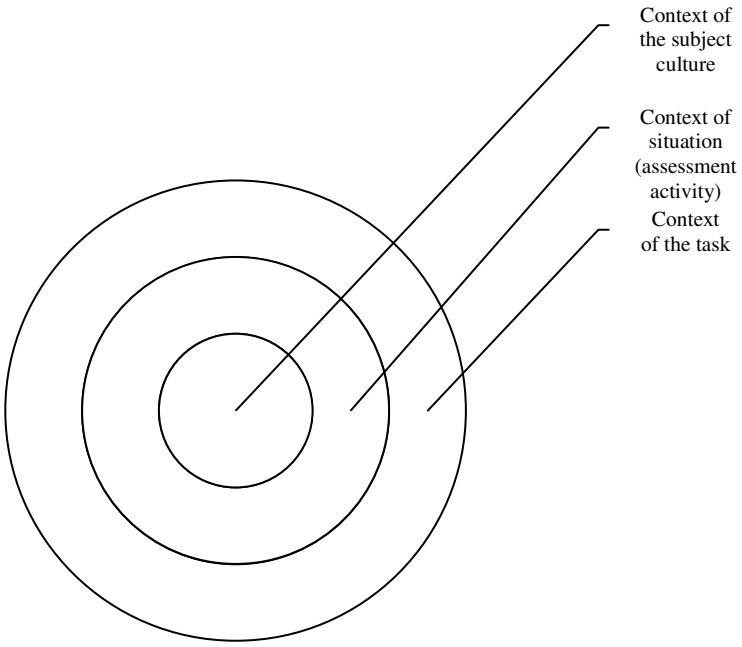


Figure 3.1: The semiotic contexts relevant in rating the learners' written responses

The expectations identified through an analysis of these contexts have firstly been used to develop holistic scales with which the degree of task fulfilment can be rated. Secondly, the expectations formed the basis of a set of analytic scales used to rate learners' language use and they have thirdly been applied in designing a set of analytic scales focusing on the content of learners' texts. The expectations have been applied in formulating the *highest possible levels of performance* on the different scales. Departing from these levels of "idealised performance", criteria have been identified and used in defining lower levels of performance that increasingly deviate from the expectations.

It is possible to identify a fourth level of context that is one level of abstraction higher than the context of the subject community's culture. A community's culture and the nature of its meaning-making practices are driven by its *ideology*. This ideology regulates the meaning potential and meaning-making systems of the community by determining what can be said, by whom and in which ways. Whereas an analysis of the subject community's culture, the assessment activity and the context of the task leads to the identification of expectations in terms of learners' written responses, an analysis of the context of ideology gives an insight into the values and biases prevailing in a discourse community and regulating its meaning-making processes. In other words, whereas an analysis of the context of culture, context of situation and the task provides descriptions of "this is how things are done and said in the community", a study of the context of ideology explains "why things are the way they are".

A study of the discourses of the larger culture of formal education for instance reveals the particular power structures at work in this community. To illustrate: the fact that learners can be expected to demonstrate subject-specific knowledge and ability in a test and that a teacher is given the power to make judgements about the levels of learners' knowledge and ability, is the result of an ideology that constitutes a community in which learners are apprentices and teachers are experts. Relating learners' texts to the context of ideology thus provides insights into issues of authority and power in educational contexts.

Although the context of ideology does not directly contribute to identifying expectations in terms of learners' written responses, the role of ideology in meaning-making processes involved in test-based task processing will receive attention in the discussion

that follows. I will especially highlight the complexities involved in the relationship constructed between a learner as writer and teacher as reader and rater in a test situation. In defining subject-specific discourse competence in section 3.5.4, I will also refer to the aspect of volition and the willingness to take on a specific identity in a subject-specific context.

I will now define the concepts *social semiotic system* and *social semiotic formations* in more detail. This will be followed by an analysis of the context of the subject's culture and of the context of situation and the implications these contexts have for the frameworks of expectations that form the basis of the rating scales. Furthermore, in both these analyses, I outline the role of task requirements in identifying the expectations.

3.2.2 Constructing meaning as a member of a community: the role of social semiotic systems and formations

Social semiotic resource systems are the cultural tools that individuals use in order to construct meaning as they engage in the activities of the communities of which they are members (e.g. Lemke 1990: 194, Eggins 1994: 13-16). These resource systems are used in making sense *of* experience and sense is made *with* them. Applied to language as a social semiotic system, constructing meaning thus means using language in the modes of listening and reading in order to decode meaning and in the modes of speaking and writing to encode meaning. In addition to language, communities use an array of other social semiotic resource systems. Within the context of the subject geography, for instance, purely verbal texts, such as definitions, descriptions and evaluations are used alongside texts that are purely visual, e.g. cartoons and photographs. A great number of texts used in the geography classroom also consist of a mixture of verbal, visual and mathematical representations, e.g. climate charts, diagrams, graphs, tables with statistics etc.

A social semiotic system consists of a finite collection of discrete *signs* (e.g. Lemke 1990: 202-204, Eggins 1994: 13-16). Each sign combines a meaning or a function with a representation or a specific form. Signs acquire their meaning from the ways in which they are different from other signs in the system. In a description of climate, for instance, the meaning of the word "humid" is dependent on its opposition to the word

"arid". A semiotic system thus presents its users with choices and individuals construct meaning by selecting particular signs. In analysing a learner's written response to a geography task then, the choices the learner has made, e.g. in terms of the lexical items and grammar structures, the stylistic devices and rhetorical structures, are seen in relation to the range of choices available and in terms of the subject's semiotic formations.

The concept of semiotic formations explains why certain choices made by a learner in terms of the application of social semiotic resources are more appropriate than others. *Semiotic formations* are patterns of meaning and meaning-making processes that are repeatedly made and followed in a particular community (e.g. Lemke 194-198). Semiotic formations represent the conventionalised ways in which social semiotic systems are used by communities in the construction of meaning. Lemke (1990: 194) explains the difference between semiotic resource systems and semiotic formations as follows:

> "Semiotic resource systems tell us what you *can* meaningfully do or say in a community; semiotic formations describe what repeatedly *does* get done and said." (Lemke, 1990: 194. Italics in original).

In order to make sense to and of one another, members of communities must share meanings and meaning-making practices. In other words, in order for an individual's use of a social semiotic system to be socially meaningful, the choices he or she has in terms of the meanings to be constructed and the ways in which the social semiotic system can be used to do so, are limited by a community's semiotic formations.

Communities differ regarding the degree of freedom their semiotic formations allow in terms of the meanings its members can make and the ways in which they can make them. The *semiotic formations of subject communities are standardised*, allowing less freedom than those of communities whose activities are less regulated (e.g. Lemke 1990, Halliday / Martin 1993, Martin / Veel 1998). An example of the standardised nature of subject communities' semiotic formations can be found in their use of subject-specific terms. Within the subject community, these terms carry specified meanings that are introduced to learners in the form of definitions presented in textbooks. Particularly in the sciences, these terms are often also presented in the form of taxonomies that reflect the standardised nature of the semantic relations drawn between them. These

taxonomies and definitions found in textbooks function to stipulate the use of these terms in the construction of meaning within the context of the subject community. Standardised semiotic formations and formalised regulations in terms of the use of social semiotic systems are features of all communities involved in specialised forms of activity. Disciplinary communities (from which subject communities adopt their semiotic formations) and communities of legal, medical and technical professionals also standardise their meaning-making processes.

In responding to a subject-specific task requiring a written answer, a learner thus constructs meaning on the basis of his or her knowledge of social semiotic systems and of the ways in which these are used by the subject community. This meaning is then encoded in a written text in response to the specific task. In assessing the written text, a rater also constructs meaning on the basis of his or her knowledge of social semiotic systems and by decoding the learner's text against the background of the task and the subject community's semiotic practices. Both learner as writer and rater as reader are thus involved in a *socially-embedded mutual meaning-making process* involving the contextualised use of social semiotic resources. The larger context is that of the subject-community's culture. The task itself represents a second context embedded in the subject-community's meaning-making practices. This process is visually depicted in Figure 3.2. The two squares represent the contexts in which the meaning-making activities of both the learner and the rater are embedded. The activities themselves firstly involve the use of social semiotic systems against the background of the subject's standardised ways of using social semiotic systems and the specific task's requirements. The processes secondly involve the encoding and decoding of meaning in the form of a written response to the task.

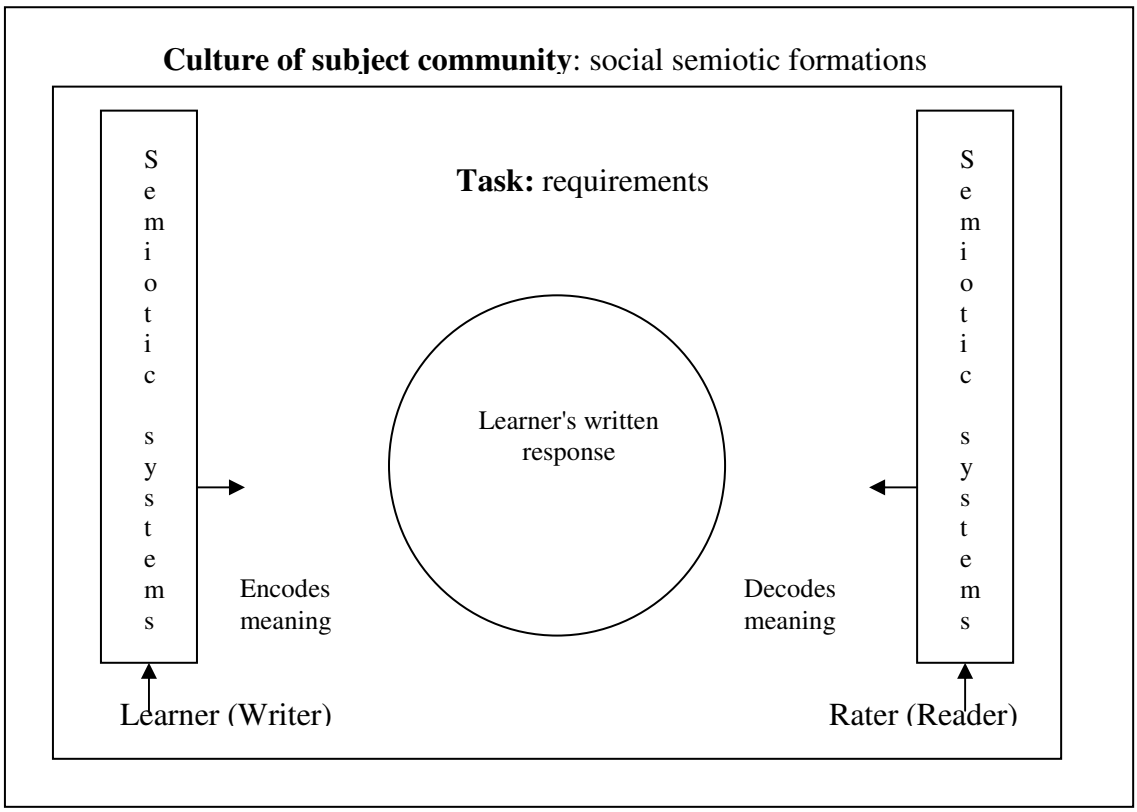


Figure 3.2: The meaning-making processes involved in the writing and rating of a written response to a subject-specific task.

The responsibility of the learner and rater in this situation is further explained by Bazerman's (1998: 27) description of the dynamics involved in the construction of meaning on the basis of a written text:

> "The joint construction of meaning [between reader an writer] is more than simply a calculus of symbols, it is a rhetorical sharing of thought within a generically identifiable space using the accepted resources appropriate to be deployed in that space in ways that keep enlisting the meanings and understandings and procedures the readers will bring to that space." (Bazerman 1998: 27).

The "meanings and understandings" that the rater brings to his or her reading of a learner's text form the basis of the rating decisions. The learner's choices in terms of the use of social semiotic systems are inappropriate when they deviate from the subject community's semiotic formations and consequently do not allow the construction of a meaning shared by the learner as writer and rater as reader. This is for instance the case

when a learner unintentionally chooses a sign that is used with a different meaning by the subject community than the meaning intended by the learner, for example when he or she confuses the meanings with which "arid" and "humid" are used in the community and uses the first in describing the climate of the tropical rain forest. Another example, this time on the level of the text, would be when the task requires the production of an evaluation and the learner's response follows the rhetorical structure of a description, thus performing a function not associated with the task.

In the following discussion, I discuss the three types of semiotic formations found in a community's discourses and I give an outline of their roles in task-based subject specific writing and its assessment. The first type of formation is the subject's thematic formations, the second its activity structures and the third its genres. All three formations are related to the larger context of the subject community's culture.

3.3 The context of culture: identifying patterns in the meanings and meaning-making processes of the subject community

An analysis of the context of culture leads to an identification of expectations about the ideational meaning to be realised in a written response, or in traditional terms, to an identification of the required content. Furthermore, it contributes to identifying the required genre to be realised. Both the construction of ideational meaning and the production of a text belonging to a specific genre result from following particular procedures in interacting with input material and constructing the written response. Identifying these procedures consequently also contribute to identifying expectations.

In the following four sections, I explain the relation between the activities involved in processing a subject-specific task and the resulting content and schematic structure of an individual written response. Furthermore, I outline the ways in which standardised procedures and conventions applied by subject-communities contribute to identifying expectations.

3.3.1 Thematic formations

When subject-specific teachers rate learners' written responses to tasks, the content of a learner's response usually enjoys significantly more attention than his or her language use. In rating the learners' responses, the teacher normally compares the meaning he or she constructs on the basis of a learner's response to a recognised model of subject matter as for instance found in the textbook. In social semiotic terms, this *standardised model of subject matter* is called a thematic formation.

A community's *thematic formations* are the meanings that are repeatedly constructed by members of the community and these are standardised in subject communities (e.g. Lemke 1990: 87, 202-204). Thematic formations consist of patterns of semantic relations that are recurrently established between key concepts (*thematic items*) in the meaning-making processes of a specific community. These semantic relations between items can be expressed in different ways. The tropical rain forest is for instance an important thematic formation in the subject geography. It includes thematic items, such as tropical climate, equatorial zone and slash and burn technique and established networks of semantic relations between these items, e.g. the causal relation that is repeatedly drawn between local farmers using the slash and burn technique and the destruction of the tropical rain forest. Ways in which this relation can be expressed in language include: "In order to gain farming land, farmers use the slash and burn technique, which contributes to the destruction of the tropical rain forest." or "The destruction of the tropical rain forest is partly caused by the fact that farmers use the slash and burn technique to clear land for agricultural purposes."

Subjects' thematic formations are thus standardised models of experience of the physical and biological world. Each subject is more concerned with some aspects of a physical and biological reality than with others. Geography is for instance partly concerned with the historical development of different types of landscapes; history with the historical development of different types of civilizations. Whereas biology focuses on the interaction between living creatures, for example, chemistry is concerned with the interaction between chemical elements. Because *subject communities focus on different*

aspects of experience, differences between these communities' discursive practices manifest themselves particularly clearly in their *respective thematics*.

However, although the *items or topics* contained in the thematic patterns of each subject *differ* from one subject to the next, subject communities make use of the *same types of basic semantic relations* between different items or topics in their thematic patterns.

When language is used as a social semiotic system, the construction of meaning about experience of the physical and biological world is referred to as ideational meaning (see section 4.3 for a detailed discussion on ideational meaning). In assessing a learner's response to a subject-specific task, the ideational meaning constructed on the basis of the learner's text is compared to subject-specific models of organising meaning about the specific phenomenon in question. In other words, the content of the learner's text is assessed in relation to the way in which the subject's thematic formation demarcates the relevant concepts involved and organises them into semantic networks.

A subject community's thematic formations are constructed through its activity structures and genres. Each of these two concepts will be explained now.

3.3.2 Activity structures

In generating content - in social semiotic terms - in constructing meaning in response to a geography task, a learner has to go through a sequence of actions. The learner must for instance read the instruction, analyse the input material, plan his or her response, write the response down and maybe edit it. Social semiotic theorists refer to such action sequences as *activity types*. When activity types are often repeated in a particular community as part of its meaning-making processes, they enact *activity structures* (e.g. Lemke 1990: 198-200).

There are various activity structures in subject-specific contexts and different activity structures are *embedded in one another*. Analysing a climate chart is just one of a number of actions that can be embedded in the processing of a particular task, which may again be embedded in the activity of writing a test. The analysis of a climate chart itself involves distinct actions such as identifying the total amount of annual

precipitation and the average temperature, identifying differences in monthly precipitation and temperature and comparing the precipitation line to the temperature line in order to reach a conclusion about humidity.

The different activities that constitute an activity structure stand in *functional relationships* to one another, so that the same activity structure can be realized in different ways. In processing a subject-specific task for instance, learners do not all go through a routine-like linear process of reading the instruction, analysing the input material, planning a response, writing the response down and editing the response. One learner's approach towards reading the instruction might involve carefully analysing it and identifying key words, whereas another might first try to develop a general understanding of what the instruction and the input material entail. Furthermore, in analysing a climate chart, for instance, one learner may first focus on the information about temperature, before he or she pays attention to information about precipitation, whereas another learner may first pay attention to precipitation and then to temperature.

In subject communities, certain activity structures are *standardised*. A subject's *technical activities*, such as scientific measuring techniques involving the use of particular instruments, e.g. a compass or a thermometer, are examples of standardised activity structures. The actions to be performed in constructing meaning on the basis of scientific observation, and sometimes also the sequence in which they are to be performed are prescribed. Since subjects use different techniques in making observations about the physical and biological world, subjects also show clear differences in terms of the nature of their technical activities or more regulated activity structures.

Subject communities often also standardise procedures involved in *constructing meaning on the basis of texts*. This is particularly evident in the context of geography teaching, where textbooks include sections on subject-specific methodology in which procedures for the interaction with different types of texts are stipulated. See Appendix D for examples of such guidelines that have been taken from two textbooks that are widely used in geography teaching in Lower Saxony.

In the same way in which the thematic formations of a community are constructed through different activity structures, the learner constructs meaning in a text in response to a task as result of going through *action sequences*. These action sequences involve the use of language, depiction and mathematics in the construction of meaning and enact activity structures. This process is illustrated in Figure 3.3.

In rating written products, it is not possible to directly assess learners' enactment of the subject's activity structures. In order to get direct access to the sequences of actions learners go through in completing a task, their processing needs to be observed and recorded, as is done by means of think-aloud protocols used in process research. The enactment of activity structures in responding to a task is however reflected in the expectations held by the rater regarding the meaning the task should allow readers to construct. This can be done by formulating an "idealised ideational response" that firstly organises meaning on the basis of the particular thematic formation with which the task deals, and that secondly contains the ideational meanings resulting from following subject-based procedures in processing the task.

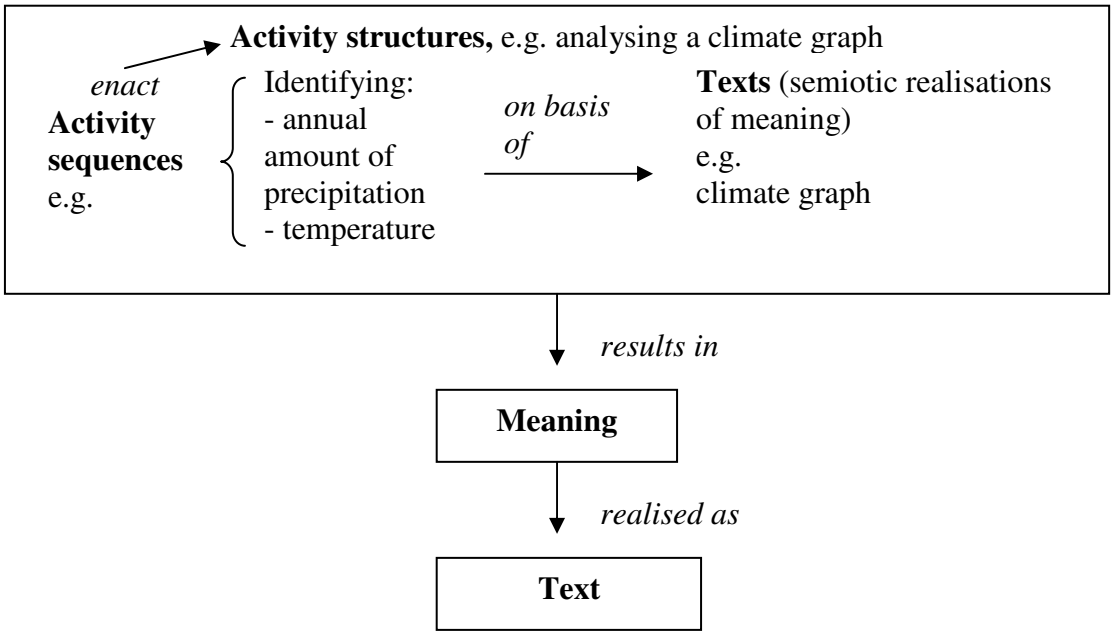


Figure 3.3: The role of activity sequences in a learner's production of a written response to a subject-specific task

3.3.3 Genres

A genre is the *linguistic embodiment of an activity structure* (e.g. Lemke 1990: 200-201, Eggins 1994: 25-48, Bazerman 1998: 23-25), or as Martin (1985: 250) formulates it: "Genres are how things get done, when language is used to accomplish them". Like thematic formations and activity structures, genres represent meaning-making patterns followed by a community in pursuit of their goals. Each genre has an identifiable schematic structure that is constituted of stages, each of which contributes to the genre's overall function. Furthermore, specific lexico-grammatical structures are used in the different stages of a genre, in order to realise meanings that are typical for each particular stage. Examples of genres found across subject-specific contexts, are the textbook, tests and triadic dialogue[1] (Lemke 1990), where the teacher asks a question to which a learner gives an answer that is then accepted or rejected by the teacher. The excursion report is an example of a genre that is typical for the subject geography.

It is possible to distinguish between *macro-genres* or major genres (e.g. Lemke 1990: 201, Eggins 1994: 47) and what Lemke (1990: 201) calls *mini-genres*, or *rhetorical structures*. These mini-genres constitute the stages, or the functional parts of communities' more specialised macro-genres. They are shorter, simpler genres, consisting of a limited number of functional elements or stages and are widely used over different macro-genres and meaning-making processes. Examples of these mini-genres are functional forms of organisation such as description, exposition, regulation and argumentation. Descriptions are for instance used in the major genres of different subjects and form part of different genres used by one subject community. Descriptions can for example be found in geography's excursion reports and in chemistry's laboratory reports. Textbook introductions to a new unit, or a learner's response embedded in triadic dialogue may contain a description as a functional unit.

The concept of mini-genres is useful in terms of developing a theoretical framework for the assessment of task-based subject-specific writing, since tasks mostly require the organisation of meaning according to one or a combination of these rhetorical structures. Learners are rarely expected to produce more complex and specialised genres and when this is the case, it is usually restricted to the senior grades (11^{th} grade onwards). In rating a learner's text, the rhetorical structures realised by the learner's use

[1] Lindfors (1981), for instance, referred to IRF (Initiation, Response, Follow-up).

of the lexico-grammar is compared to the activities required by the task, e.g. to describe, to explain, to evaluate, to compare, and the genre conventions associated with these types of activities. In addition to guidelines for interaction with input texts, geography textbooks also sometimes offer guidelines for the construction of a number of these mini-genres. In the 2006 -edition of the textbook "Erdkunde für Gymnasien in Niedersachsen 11", published by Westermann, for example, guidelines for the production of texts in response to specific instructions are given (see Appendix D). Such guidelines indicate the extent to which a subject community can also regulate the linguistically realised texts it uses in pursuit of its goals.

Some caution is needed in teaching learners these guidelines, however, since these mini-genres are rarely realised as "pure rhetorical" forms in actual spoken and written interaction. Even the simpler texts that learners are expected to produce in response to subject-specific tasks are normally hybrids. In the guidelines given in the textbook, descriptions, for instance, are identified as a possible constituent of explanations. However, it is not further explained that descriptions and explanations are often merged in actual texts. In other words, guidelines like those presented in the textbook in question, although well intended, may also be harmful, because they simplify the nature of the meaning-making processes that usually operate in subject-specific and all other meaning-making contexts.

3.3.4 Summary of the theoretical constructs involved in developing a framework of expectations based on an analysis of the context of culture

The theoretical constructs involved in developing a framework of expectations related to the context of culture and the interrelationships between these constructs are schematically outlined in Figure 3.4. The development process can be summarised as follows: an "idealised response", representing the *ideational meaning* to be constructed and *genre* to be realised, is designed.

The required ideational meaning is firstly identified on the basis of the particular part of a *subject's thematic formation* with which the task in question deals, e.g. geography's standardised subject-matter models on the climate of the tropical rainforest. Identifying

the required ideational meaning secondly entails determining the meaning to be constructed when *subject-specific methodologies and procedures*, e.g. in terms of the analysis of input material, are followed in processing the task. The particular activities required by the task enact the subject's standardised activity structures.

These activities also form the basis of the required functional structure of the response, or the *genre* involved. The genre is consequently identified on the basis of the task's activity requirements related to the use of language, for example to describe, to comment, to evaluate etc.

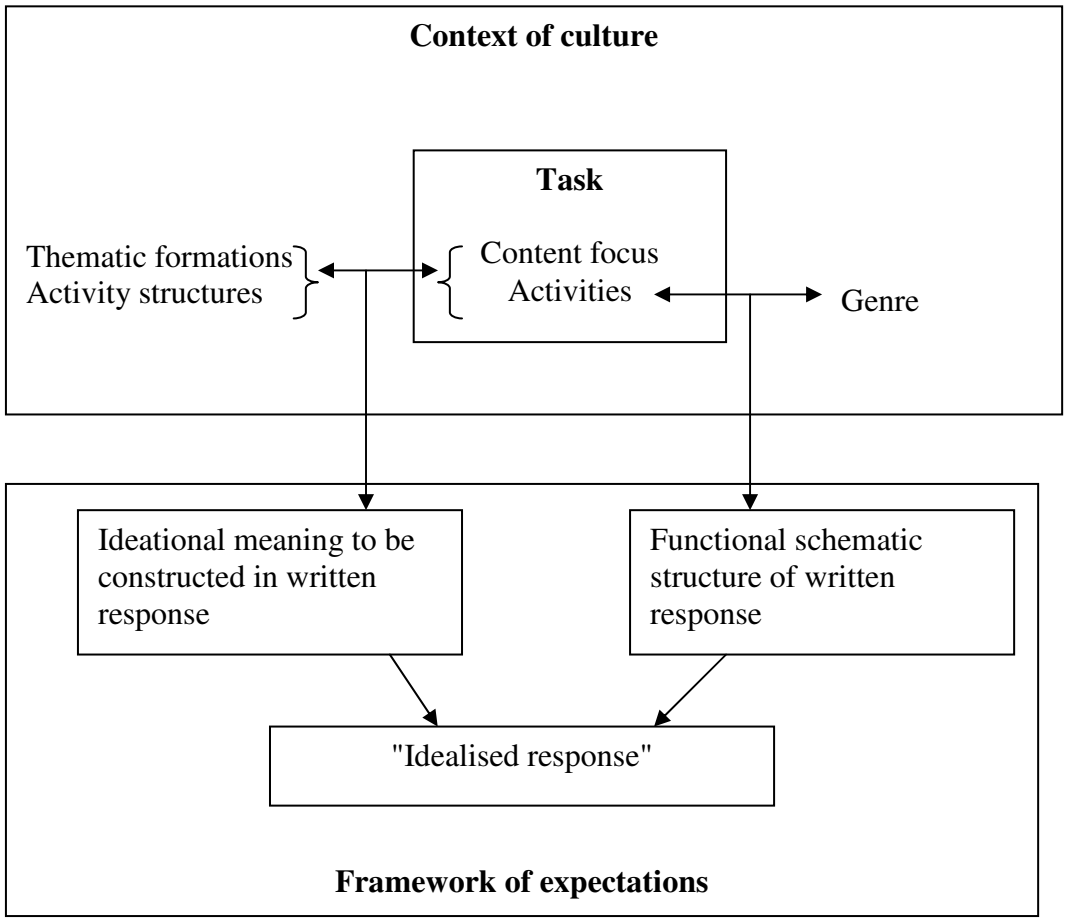


Figure 3.4: Concepts from the context of culture involved in developing a framework of expectations

Task instructions usually explicitly indicate demands in terms of the ideational meaning to be constructed in the response and the activities involved in constructing this meaning. The instruction usually however does not explicate the procedures to be

followed in enacting activity structures, since the aim of the task is partly to assess learners' ability to independently solve problems following these procedures. The genre to be produced is however normally pointed out by the instructional verb used, e.g. describe, explain, evaluate etc.

The following task instruction from the geography competence test applied for the purposes of this study illustrates these features of subject-specific task instructions as they are normally presented in the context of geography.

Task 2a

Characterize the living conditions in the three layers of the tropical rain forest with the help of Figure 3. Start with the giant trees.

This task explicitly refers to the ideational meaning the learner is required to construct in response to the task. Put differently, it clearly points out the subject matter the written response is required to deal with, namely the living conditions in the tropical rain forest. The task also explicate the activities learners should engage in, in order to construct this meaning, namely to characterize on the basis of an analysis of a diagram. The whole task with input material is presented in Appendix A.

Expectations in terms of the ideational meaning to be constructed can be used to develop analytic scales that permit a detailed identification of the *thematic items* and *semantic relations* to be included in a learner's response to a task. In developing these scales, the thematic pattern involved is broken down into its parts and raters are given two to three options, making it possible to indicate whether a particular aspect of the thematic pattern has been realised or not and in some cases whether a particular aspect has partly or implicitly been touched upon. The scales that have been used to rate the ideational meaning constructed in the learners' responses to the six test tasks appear in Appendix C.

As will be illustrated in section 3.5, expectations stemming from an analysis of the context of culture, together with expectations stemming from an analysis of the context of situation, can be used as the basis for the *holistic rating* of a written response

focusing on the degree of task fulfilment. The expectations are based on the idealised response and on norms that regulate language use in subject-specific contexts.

The *register concept* is used to identify these norms that regulate language use and to develop an understanding of their function in subject-specific contexts. In addition to contributing to the development of holistic scales, the register concept also serves as the basis for the development of *analytic rating scales* with which learners' language use can be rated in a detailed way.

Whereas thematic formations, activity structures and genre function on the level of the context of culture, register operates on the level of the *context of situation*. In the case of this study, the task to which the learners had to respond was embedded in the context of a test, so that the context of situation of their written responses can be described as *test-based subject-specific task processing*. Subject-specific tasks are of course not only used in test situations. The other predominant use of tasks by subject-specific communities is to facilitate learning. However, as will be demonstrated below, the register of task processing in assessment contexts as they are realised in school, is different from that of task processing situated in the teaching and learning activities of the subject.

The context of test-based task processing and its implications for the development of a framework of expectations functioning as the basis for analytic scales will now be examined in detail. This is followed by a definition of the term discourse function in which it is shown how this concept links the expectations arising from context of culture to those arising from the context of situation to allow the development of holistic scales.

3.4 Context of situation and register theory: identifying patterns of meaning and language use in the context of test-based task processing

Applying Halliday's (1989) notion of register allows the identification of norms that regulate *language use* in the context of test-based task processing. This contributes to

identifying expectations in terms of structuring and linking and the application of various stylistic devices in written responses to test tasks.

Halliday's register theory is concerned with explaining systematic variation of language that corresponds to systematic variation of situation (e.g. Halliday 1989: 38-43, Eggins 1994: 49-80). *Register* is thus a generic concept that functions on a more abstract level than that of a single text, such as a specific task, or a learner's written response to this task. A text is seen as an instance of a particular functional variety or register and related to this register when analysed (Halliday 1989: 42). Halliday (1989: 39) defines register as:

> "a configuration of meanings that are typically associated with a particular situational configuration of field, mode and tenor [and] include the expressions, the lexico-grammatical and phonological [and graphological] features that typically realise these meanings." (Halliday 1989: 39).

In Halliday's (1989) register theory, *three types of situational variables* are seen to affect the ideational meaning (representing experiences), interpersonal meaning (creating interpersonal relationships) and textual meaning (creating texts)[1] constructed in a specific context. These situational variables are the field of discourse (the topic and activities involved in a meaning-making process), the tenor of discourse (the interpersonal relationships between the participants) and the mode of discourse (the role language is playing in the situation). The *three types of meaning* are seen to be realised by using the *lexico-grammar* and *discourse-semantics* in particular ways.

Patterns in the manifestation of the three situational variables of field, mode and tenor are socially determined and due to the ideology prevalent in formal education, the ways in which they are realised in school contexts are *controlled by the educational institution*. In other words, in subject-specific contexts conventions in terms of "who can say what in which way and under which circumstances" are predominantly controlled by the teacher who represents the educational institution. Since these discourse patterns are linguistically realised through conventionalised ways of using the lexico-grammar, the educational institution also determines the stylistic norms. These

[1] Other social semiotic systems are seen to construct the same basic meanings. The terms used to refer to these functions when discussing other social semiotic systems may, however, be different. The ideational meaning is sometimes referred to as the representational or thematic meaning, the interpersonal as the orientational and the textual as the organisational meaning.

norms are introduced to learners through subject-specific genres, such as the textbook. However, the norms are rarely made explicit. When learners produce written texts in subject-specific contexts, these norms are also an implicit aspect of the framework of expectations against which learners' texts are rated.

In the following discussion, I define and explain the three situational variables of field of discourse, tenor of discourse and mode of discourse in more detail. This is followed by a discussion on the three types of meaning, namely the ideational, the interpersonal and the textual and an explanation of the ways in which these realise the situational variables through the lexico-grammar and discourse-semantics. I also outline the role considerations about register play in rating learners' texts.

3.4.1 Situational variables

As already mentioned earlier, the context of situation against which learners' written responses are rated for the purposes of this study, is that of *test-based task processing in a subject-specific context*. This context can be identified on a higher level of abstraction than the context of the task, because of the fact that the manifestation of the tenor and the mode of discourse remain similar over class-based test situations in a specific subject. The field of discourse, however, varies from one task to the next.

3.4.1.1 Field of discourse

The *field of discourse* refers to the topic of the communication and the nature of the social action that is taking place (e.g. Halliday 1989: 12, Eggins 1994: 9, Lemke 1998: 93). Social action can be categorised as either technical and specialised, involving expert knowledge, or as unspecialised involving participants who only share common knowledge about the particular topic, e.g. cars, playing football, childcare, gardening etc. (e.g. Martin 1992: 536, Eggins 1994: 67-76). Subject communities have specialised activities and the aim of subject-specific education is to initiate learners into these specialised activities. Learners are taught to solve problems using subject-specific interpretations of experience of the physical and biological world and subject-specific methods of observation. The field of discourse in school-based test situations involves the construction of meaning on the basis of subject-matter knowledge (knowledge of

models that organise and explain experience of the physical and biological world from a subject-specific perspective) and subject-specific procedures and techniques.

In tests, learners are expected to demonstrate the extent to which they have acquired knowledge of a specific area of subject-matter and have mastered a particular selection of procedures and techniques. In terms of the six tasks that were used for the purposes of this study, the overall field of discourse involved a subject-specific modelling of the natural phenomenon of the tropical rain forest and subject-specific techniques of text analysis. Each task realised its own more focused field of discourse by concentrating on a particular feature of the tropical rain forest, for example on climatic aspects, and by including specific input material, e.g. a climate chart.

As I pointed out in the discussion on thematic formations and activity structures (3.3.1 and 3.3.2), subject communities focus on different aspects of experience of the physical and biological world and use different procedures and techniques in constructing meaning about these experiences. Consequently, there are also clear differences between the subject communities regarding the field of discourse involved in test-based task processing. On the other hand, the nature of the interpersonal relationships between interactants (tenor of discourse) and the role of language in the interaction (mode of discourse) remain similar over test situations in all content subjects.

3.4.1.2 Tenor of discourse

The *tenor of discourse* is related to the nature of the interpersonal relationships between interactants (e.g. Halliday 1989: 12, Eggins 1994: 9, Lemke 1998: 93). This situational variable has three dimensions. The first is related to a speaker or writer's *social orientation towards other participants*. This depends on a combination of the aspects of power (e.g. equal, superior, subordinate), affective involvement (e.g. intimate, emotionally uninvolved) and contact (e.g. being well acquainted, being strangers) (e.g. Poynton 1985, Eggins 1994: 63-67). A particular constellation of these three aspects allows a situation to be placed on a continuum ranging from formal (unequal power, low affective involvement and infrequent contact) to informal (equal power, high affective involvement and frequent contact).

The second dimension of the tenor of discourse is the specific *speech role* the speaker or writer takes on in the communication event. Two basic speech roles are identified, namely giving and commanding. Depending on whether the commodity of the exchange is information or goods and services, four basic initiating speech functions can be distinguished. These are statement, question, offer and command. Based on whether the responding speech function is supporting or confronting, a further eight speech function classes can be identified, namely acknowledgement and contradiction (initiating speech function: statement), answer and disclaimer (initiating speech function: question), acceptance and rejection (initiating speech function: offer) and compliance and refusal: (initiating speech function: command) (e.g. Halliday 1985: 68-71, Eggins 1994: 149-154).

The third dimension of the tenor of discourse is the writer or speaker's *orientation towards the meaning he or she is constructing and towards actions and events*. This aspect of the context of situation is realised by the area of grammar known as modality (e.g. Halliday 1985a, Eggins 1994) and is concerned with degrees of probability (how likely something is to happen), usuality (how often something happens), obligation (the extent to which something is obliged to happen) and inclination (the extent to which something is preferred to happen).

Whereas exactly the same the field of discourse can be realised in task processing embedded in learning and in assessment activities, the tenor of discourse involved in *task-processing for assessment purposes* is different from the tenor involved in other situations in which tasks are used in subject-specific contexts. Across all subjects, school-based test situations show particular complexities in terms of the tenor of discourse realised. This is firstly because of the social consequences that tests can potentially have for learners and secondly because the roles learners and teachers are expected to take on are in conflict with their usual roles within the larger context of classroom activities.

Test situations across all subject-specific contexts are formal in nature. Although the interactants, learner and teacher, know each other, there is a clearly unequal distribution of power in favour of the teacher and it is expected that any form of affective involvement that might exist be disregarded for the sake of fairness. The ideology of

formal education makes it possible for a teacher to expect a learner to present a solution to a particular problem and then to make decisions about the learner's abilities, based on a judgement of the acceptability of the solution. The *decisions that a teacher makes* on the basis of test performances have *material consequences for a learner*. They are used by the teacher in his or her assessment of the learner's achievement in the subject. Teachers' assessments are again used in making decisions about aspects such as whether the learner is allowed to proceed to the next level in the educational system and in a country like Germany, it also affects what type of school form the learner is advised to attend (which can have social consequences for a learner). The grades that learners receive at the end of their school training also play a role in the type of career opportunities that are open to them.

Responding to a test task, the learner is expected to give information to a reader, who is to be regarded as an expert member of the community, but who nevertheless needs to be advised about a possible solution to a specific problem. However, in effect the reader is the teacher, who does not already only have a conceptualisation of a solution, but who has a notion of a *preferred solution*. Since the aim of subject-specific education is to develop knowledge and understanding of subject matter and mastery of subject-specific procedures, this conceptualisation involves an idealised response to the task as it would be presented by a knowledgeable and competent pupil – one that has been initiated into the practices of the subject community.

The tenor of discourse involved thus firstly presents a conflict between the *projected speech role* of the learner as writer in the test situation and the *actual nature of a learner's speech role* in this situation. The learner's projected speech role is to inform and the teacher's is to be informed. However, both writer and reader are aware of the fact that the purpose of the written text is not to present a reader with information he or she does not yet have, but to demonstrate knowledge and ability which is then assessed by the reader. Thus, in actual fact, the commodity exchanged is not information, but a service realised by a written text.

The tenor of discourse secondly presents a conflict in terms of the social roles of writer and reader in a test situation and the nature of the *interpersonal relationship* that exists between them in most other situations in the classroom context. The learner, who is

usually in the role of an apprentice in the subject-specific context, is now expected to take on a role as a member of the community who is not in need of guidance, but who "gives guidance" in the form of a solution to a specific problem. This conflict between an occasional and a usual interpersonal relationship is typical for assessment situations and can cause learners problems, also in terms of the *orientation they are to adapt in terms of the meaning they are making*.

As Meyer (1993: 46-47) suggests in discussing problems student writers experience in producing essays, assessment situations require apprentices *to take on a voice of authority*, both in terms of the field involved and in terms of the reader, because they are expected to show mastery. Assessment situations thus require the learner to realise a particular orientation towards the meaning he or she is making. Learners, however, might not feel confident enough about their understanding of the subject matter and mastery of the skills in order to be able to take on this authoritative voice. Learners are also well aware of the fact that regarding their relationship to the reader (the teacher), they are not regarded as an authority of equal or almost equal expertise. This too, makes it difficult to take on an authoritative voice. Consequently, an assessment situation can be viewed as requiring learners to play a game, which can be daunting, if one does not know the rules.

This is not an argument against the use of tests in subject-specific contexts. It is also not a criticism of the fact that tests require learners to demonstrate not only *what* they know and *what* they can do, but also to demonstrate *that* they know and *that* they can do, thus to show competence. The belief is held that tests have their place in educational contexts as a way of determining the level of a learner's progress in the process through which he or she is instructed in the subject's ways of thinking and doing. What is criticised here are *the uses to which test results are usually put in school contexts*. It is believed that tests should be used less often in their gateway function and more often as a way of giving constructive feedback in terms of levels of progress reached and areas in need of more focused attention. If tests are to support learning, learners and teachers have to reach agreement about which tests written in the subject-specific classroom context will be put to which use.

Furthermore, the *complexities* involved in the interpersonal relationships to be constructed through their responses to test tasks *must be explained to learners*, as well as the reason for this particular construal of social roles. This does not mean that learners should play a game in which they pose as an authority, even if they do not feel confident in terms of the meaning they are making and their ability to inform an expert reader. Rather, learners should be brought to understand that in the context of a test situation, they are expected to indicate which areas of the subject matter and which subject-specific procedures they *do* feel that they have mastered. Simultaneously, a test culture should be established in which learners have the freedom to indicate when they do not feel confident about the meanings they are constructing and they should be taught to handle the modal devices involved in realising this and other orientations towards the meanings they are constructing.

3.4.1.3 Mode of discourse

Mode of discourse refers to the *role language is playing* in the situation and to the *possibility of immediate feedback*. The role of language in a particular situation could be performative, to comment, to recounting or to construct experience (e.g. Halliday 1989: 12, Eggins 1994: 9, Lemke 1998: 94). The possibility for immediate feedback depends on whether there is visual and aural contact between the participants, e.g. face-to-face interaction, where immediate feedback is possible due to visual and aural contact, as opposed to written texts where there is no aural and visual contact and where no immediate feedback is possible[1]. A learner's written response to a test task serves as a particular construction of experience, recorded in writing and consequently allows no immediate feedback.

In school-based test contexts, the use of written responses constructed independently by each learner is motivated by a number of factors. The first is that a written record of a solution *can be more carefully assessed* than if the solution had been given orally, because the rater can repeatedly return to the text. Naturally, the same is possible if answers that are produced orally are recorded, as is often the case in the oral sections of language proficiency tests. Due to the number of learners, this option is however not practical in the school-based classroom context. A second reason for making use of

[1] Communication via written texts involving the use of technological communication modes, e.g. e-mails and faxes, of course constitutes a unique category where there is no visual or aural contact, but where rapid feedback is possible.

written responses to test tasks is that it gives learners the *opportunity to plan their responses*, which is not possible to the same extent in face-to-face interaction. A third reason for using written responses in test contexts is that it involves the application of competences that are important in all school-based learning contexts and that also form the basis for the competences needed to participate in the discourses of other institutionalised public contexts. As I pointed out in chapter 2, cognitive academic language proficiency, or CALP, is central for participation in the activities of the subject community. This form of language proficiency is also needed in other educational contexts, as well as in professional contexts. Subject-specific tasks requiring the production of a written response provide an ideal opportunity for learners to develop the ability to communicate in an explicit and clear way about complex matters. The use of these tasks in tests can also serve to determine the extent to which learners have mastered *context-reduced communication on cognitively demanding tasks*.

A disadvantage of using written responses in test situations is that a written text is a rather *reduced synopsis of a complex process*. It does not allow teachers to observe learners' control of aspects that are important for effective participation in the activities of the subject community. Written responses, for instance, do not allow teachers access to the procedures learners follow in constructing an interpretation of the instruction, in analysing input texts and in planning a written response. This points to the importance of process research for a more thorough understanding of the difficulties learners may experience in learning to do and talk as a member of the subject community.

I will now examine the three types of meanings that realise the three situational variables in a specific way in a particular situation more closely.

3.4.2 Metafunctions

The *ideational* function of language serves to relate experience, the *social* function to create interpersonal relationships and the *textual* to organise information. These three types of meaning can be made through language, or any other social semiotic system (e.g. Lemke 1998).

From a systemic functional linguistic point of view, the three metafunctions are not seen to be performed by different parts of a text, but represent three distinctive ways of analysing the different *meanings simultaneously made* with language or another social semiotic system. In analysing a written text, it is thus not possible to identify an instance of language use that is purely ideational, or exclusively interpersonal, or textual and ideational, but not social. Texts are seen as multi-functional units of meaning that can be studied from different angles. Each perspective then contributes towards a total interpretation of the text (Halliday 1989: 23).

As explained in the previous section, the three metafunctions of language realise one of the three *situational variables*: field, tenor and mode of discourse and a particular combination of meanings conventionally associated with a specific configuration of situational variables constitute a distinct *register* (e.g. Halliday 1989: 15-23, Halliday / Martin 1993: 27-36, Eggins 1994: 11-13,76-79, Lemke 1998: 91). A register again represents the situational specifics of a particular *genre* (e.g. Eggins 1994: 33-36, Halliday / Martin 1993: 36). To illustrate: a specific genre such as an argument has the same schematic structure over different contexts of use. Its obligatory or defining elements are a statement of the writer's position and support for this position. The field, mode and tenor realised in arguments embedded in different situations can however differ from one situation to the next. The field of discourse realised in an argument written in the context of geography, for instance, differs from that of an argument written in biology or history. When a learner presents an argument orally to a fellow pupil in response to a task completed in pair work, the tenor and mode of the situation differ from the tenor and mode of a test situation where a learner has to present an argument in written form to a teacher. The differences in the situational variables influence the meanings constructed in the various situations. These meanings are realised by the *discourse-semantics,* which again are realised by applying the *lexico-grammar* in specific ways. The relation between these factors are illustrated in Figure 3.5.

<table>
<tr><td colspan="4">Genre</td></tr>
<tr><td rowspan="2">↑
realises
Register</td><td>Field of discourse</td><td>Tenor of discourse</td><td>Mode of discourse</td></tr>
<tr><td>↑
realises
Ideational meaning</td><td>↑
realises
Interpersonal meaning</td><td>↑
realises
Textual meaning</td></tr>
<tr><td rowspan="2">↑
realises
Language</td><td colspan="3">↑
realises
Discourse-semantics</td></tr>
<tr><td colspan="3">↑
realises
Lexico-grammar</td></tr>
</table>

Figure3.5: The relation between language, register and genre

The *lexico-grammar* refers to a language's lexical and grammatical resources. These resources allow language users to construct an infinite number of meanings, using a finite number of phonemes and graphemes. Sounds and letters, which constitute the level of expression, are combined to form words, which can again be arranged in different grammatical structures. It is this intermediate stratum of language (the lexico-grammar) that distinguishes it from other semiotic systems where there is a one-to-one relation between the meaning constructed and the form of expression.

Discourse-semantics refers to the resources language users have in order to create texts. It involves *coherence*, which is connected to the way a group of sentences or clauses relate to the context, and *cohesion*, which refers to the internal organisation of a text (e.g. Halliday / Martin 1993: 29-32, Eggins 1994: 82-87, 114-117).

At this point in the discussion, it is important to draw attention to the almost exclusive focus on the ideational meaning in subject-specific contexts and to present an argument for rating practices that pay attention to the construction of all three meanings in learners' texts.

3.4.2.1 The importance of regarding all three types of meaning in assessing learners' task-based subject-specific writing

Ideational meaning is most closely related to subject-specific content. Consequently, the ideational function of language tends to be foregrounded in subject-specific contexts. On the other hand, the use of language in establishing an orientation towards others and the meaning one is making (interpersonal function), as well as to organise meanings into texts (textual function) tend to be neglected. The question that now arises is why a framework for the assessment of task-based subject-specific writing should make provision for rating the linguistic realisation of all three types of meanings, instead of focusing on what is regarded as "important" in subject-specific contexts.

The first reason is that registers consist of particular *combinations* of situation-specific ideational, interpersonal and textual meanings so that the inappropriate realisation of one of these metafunctions consequently adversely affects a text's *situational coherence*. In order to develop a comprehensive framework that can be used as the basis for developing learners' ability to use a specific register, inappropriate use of the lexico-grammar and discourse-semantics related to the realisation of all three meanings must thus be accounted for.

By excluding expectations related to the interpersonal and textual functions, some aspects of learners' responses that may be identified as inappropriate for the specific situation cannot be fully understood, so that effective feedback can also not be given. Examples of problems that may be caused by an inappropriate realisation of the interpersonal and textual function in response to test tasks are language use that expresses an attitude (interpersonal meaning) in a strong way, e.g. *terrible, fantastic, wonderful, great,* or language use that is context dependent (textual meaning). Further examples from learners' responses to the tasks used in this study are given in the following discussion.

The second reason for including expectations arising from the tenor and mode of discourse is that these two situational variables tend to manifest themselves in similar ways in comparable contexts of situation occurring in the cultures of all subject communities, e.g. tests. The "mini-genres" (e.g. describe, explain, comment etc.) and

the realisation of the interpersonal and textual function, are in effect what language use across different subjects have in common. Realising these meanings in appropriate ways is thus important over all subject-specific contexts.

In this respect, the ability to realise the tenor and mode of discourse in an appropriate way in school contexts is related to *school-based literacy* referred to in chapter 2. The register of school language realises a tenor of discourse that assumes little shared knowledge between interactants, so that there is a demand for explicitness. This has implications for the application of devices on the level of the lexico-grammar and discourse-semantics specifically related to interpersonal and textual meaning. For example, the register of school language entails creating a generalised self and others in texts, through the use of third person pronouns, creating lexical cohesion and realising the information structure of the text on the level of the sentence by encoding given information in the subject and new information in the predicate. In order to give learners support in acquiring and using this register, learners thus need to receive feedback on the inappropriate realisation of these two functions, in addition to feedback on the realisation of the ideational function.

As argued in chapter 2, the ability to use the register of school is also central to successful participation in discourses of institutions of further education and professional communities. It is also central to participation in discourses with authorities and enables responsible and independent political participation in public life. Regardless of the context, any meaning-making process always entails the construction of ideational, interpersonal and textual meaning. Particular realisations of interpersonal and textual meaning in school contexts are also relevant in contexts of further education, in professional contexts and in contexts of public interaction, e.g. using formal, succinct and clear expressions in written texts, stating an opinion in "emotionally neutral language" and organising information into a coherent and cohesive text. Completely disregarding the construction of interpersonal and textual meaning would consequently imply that learners are not fully prepared for the demands of communication situations beyond the classroom and that school partly fails to develop literacy.

In the case of foreign-language subject-specific education, the language of the school and that of the broader society is of course not the one used in the classroom. However, attention to all three types of meanings is also of paramount importance in these classrooms, especially if this form of education aims to prepare learners for future participation in academic and professional contexts where the foreign language is used. Especially the inappropriate realisation of the tenor of discourse could have adverse effects on interaction with L1 speakers. Learners therefore need to be made aware of the ways in which the foreign language is structured for use in this respect and also need to be able to use the structures involved.

Thus, by including the expectations related to the construction of interpersonal and textual meaning in the framework, a *comprehensive platform for feedback* is created. Furthermore, an assessment basis for *literacy-oriented* subject-specific teaching is established.

In the following three sections, it will be shown how the lexico-grammar is used to construct ideational, interpersonal and textual meaning in subject-specific contexts. Special attention is paid to conventions followed by subject communities in constructing these three types of meaning and to the functions these particular conventions have. Each section also includes an outline of the effects these conventions have for the rating of learners' task-based subject-specific texts.

3.4.2.2 Ideational meaning

As already explained in the discussion on a subject's thematic formations (see section 3.3.1), language is used in its ideational function to *construct meaning about experience of the physical and biological world* (e.g. Halliday 1989: 18-21, Halliday / Martin 1993: 27-28, Eggins 1994: 11-12, 78, Lemke 1998: 91). This metafunction is subdivided into two closely related categories: the *experiential* and the *logical*. In its experiential function, language is used to represent processes, agents and circumstances observed, or "experienced" in the real world (e.g. Halliday 1989: 19). The logical function is used to express relations between the different aspects of our experiences, thus between the processes, participants and circumstances observed (e.g. Halliday 1989: 21).

The *lexico-grammar realises ideational meaning* through patterns of processes, represented by verbs, participants, involving nouns and circumstances entailing prepositional phrases of time, manner, place etc. (Eggins 1994: 77). A comparison between the sentences below illustrates the way in which ideational meaning is constructed through the use of the lexico-grammar. One of the tasks used for the purposes of this study requires learners to produce a description of the climate of Kisangani, a city in the Democratic Republic of the Congo, on the basis of a climate graph. The task (task 1c) is presented in Appendix A. Below are examples of sentences that one could expect in responses to this task:

a) Kisangani has a tropical climate[1].

b) Kisangani's climate is humid.

c) Kisangani is in Africa.

All three sentences construct meaning about a particular place, linguistically represented by the name *Kisangani*. However, whereas the first two sentences construct meaning related to Kisangani's climate, the third sentence constructs meaning about its location. In sentence (a), this is done through the use of the verb *has*, which constructs a possessive relation between *Kisangani* and a particular attribute represented by *tropical climate*. In sentence (b), a possessive relationship is established between *Kisangani* and *climate*, through the use of the possessive *'s*. *Kisangani's climate* is then related to an attribute *humid*, through the verb *is*. In sentence (c), the verb *is* relates *Kisangani* to a particular location, *in Africa*.

Thus, through lexical choices, different representations of reality are constructed. All three sentences involve a linguistic representation of a particular location (*Kisangani*) that really exists and that can be identified on a map (thus in another semiotic mode), or physically discerned, should one travel to the Tshopo Province in the Democratic Republic of the Congo and go to where the Luluba River becomes the Congo River, north of the Boyoma Falls. In all three sentences, this location is linguistically related to other phenomena which can also be observed in the physical and biological world. Sentence (a) relates it to a type of climate (*tropical climate*), which is identified through technical activities, involving observations of mean weather conditions over a particular period of time, usually thirty years. Sentence (b) first relates this location to the general

[1] Letters of the alphabet are used for sentences that have been specially formulated as examples of aspects explained in the discussion. Numbers are used for extracts from learners' texts used to illustrate aspects that are explained.

concept of climate, thus not to a specific category of mean weather conditions, but to a phenomenon that can be experienced and recorded in the form of a climate chart. This combined (*Kisangani's climate*) is then related to an atmospheric condition (*humid*), which is also observed through technical activities, and which is related to the amount of water vapour in the air. In sentence (c), this location is related to a larger region, *Africa*, which like Kisangani, is semiotically represented on maps and can be physically identified as one of the seven continents.

In addition to constructing a representation of a particular aspect of the physical and biological world, the lexical choices also *indicate the social activity involved in constructing these representations of reality*. Many activities of subject communities are of a *technical nature*. Furthermore, in observing the physical and biological world, the focus in subject communities is exclusively on the specific phenomenon observed, leading to an *exclusion of human agents*. These features of the activities of subject communities manifest themselves in the communities' use of the lexico-grammar. Instances of language use in learners' written responses to subject-specific tasks that do not reflect this technicality and "dehumanised" view of the physical and biological world are consequently inappropriate in this context.

The reflection of technicality in the language use is illustrated by the differences in lexical choice between sentences (a) and (b). *Tropical climate* is the technical term used by the geography subject community to refer to the predominant type of climate found along the equator. *Humid* is also a technical term used by the community, but not in categorising types of climate. In its "commonsense" use, however, the word is associated with weather conditions that in everyday language use is often described as "hot and wet". Since a tropical climate is indeed characterised by a high total annual rainfall and average temperature, *humid* has, outside the circles of the subject and disciplinary communities, gained the meaning of a type of climate similar to the one found in the tropical rain forest. In discourses set in the activities of the subject community, the use of the word *humid* to describe the climate of the tropical rain forest is however not fully appropriate.

Differences in the level of technicality expressed are also illustrated by the following three extracts taken from learners' responses to task 1c:

Task 1c, 2008[1]

1) *The climate is humid. The average annual temperature is 25°C. The total annual precipitation is 1804mm.* [2]

Task1c, 4005

2) *Because of the location where Kisangani lies (see 1b)) it has to be a mediterranean climate. This means you have high temperatures throughout the year and the rainfall is a very high one, too.*

Task 1c, 2030

3) *It's very hot and wett. I can imagine that the air is humid.*

In spite of the incorrect use of the subject-specific term *humid,* text 1 expresses a clearly higher level of technicality than texts 2 and 3. This technicality is realised by the use of the subject-specific terms *average annual temperature* and *total annual precipitation*, and choices in terms of the construction of meaning using numbers and units of measurement: *25°C* and *1804mm*. By presenting information in a different semiotic form than language, namely that of numerical measurements, the expressions become more exact in the meanings they are making about quantity. It consequently more effectively reflects the nature of the activities involved in making these observations. It also reflects the fact that precise and objective observation is valued by the subject community. In short, text 1 shows a deeper understanding of subject-specific activities and procedures. In text 2, the term *mediterranean climate* is used incorrectly. Furthermore, this text contains the "commonsense" term, *rainfall*, as opposed to text 1, where reference is made to *precipitation*. The use of *high* and *very high* also constructs a meaning about the average temperature and the amount of rainfall that is less precise than the one constructed in the first text. It is, however, more precise still than text 3,

[1] For the purposes of the data elicitation, each learner was given a four-digit code. The first two digits represent the class the learner belongs to and the last two digits constitute the particular identification number the learner was given in his or her class.

[2] Learners' answers are presented as they were written, except in terms of the words that were crossed out, which have not been included in the representation of the texts here, because editing phenomena occurring in learners' texts are not central to the current discussion. No corrections of learners' writing have been undertaken though. Inaccuracies in terms of grammar, spelling and punctuation, or socially inappropriate language use are presented as they occur in the texts. Those instances of incorrect and inappropriate language use not directly relevant for the particular issue discussed, are also not pointed out or commented on in the current discussion.

where no references are made to either temperature or precipitation. What exactly the text describes as being *hot* and *wet*, is only clear in reference to the task.

The type of reality constructed by sentences (a) and (b) and text 1 presented above exclude human agents. Text 2, on the contrary, refers to a generalised individual, referred to by the personal pronoun *you* and text 3 refers to the writer by means of the inclusion of the personal pronoun *I*. Thus, the focus in sentences (a) and (b), as well as in text 1 is on natural phenomena. This "dehumanised" presentation of the world is typical for the science subjects: chemistry, physics, biology and geography. Naturally, there are particular areas of study in biology and geography where humans are the object of study and where they occur as agents. However, in describing climate in the context of subject-specific activities, it is unconventional to choose lexical items that realise a world in which humans figure. This is illustrated by the following two extracts, taken from learners' responses to task 1c, in which personal pronouns represent human agents:

Task 1c, 5016

4) *In Kisangani they have the whole year a constant climate of 25°C. They have the highest precipitation rate in October. But they have a high precipitation rate the whole year…*

Task 1c, 5020

5) *They have always the same temperature, but not always the same rainfall. In the summer months there is less rainfall because of the time of the year there is in every country less rainfall and in autumn they have the most rainfall.*

In addition to a presentation of the world in which human agents are excluded, so that the focus is on the natural phenomena, the *use of general academic vocabulary and grammar structures that relate different semantic relations to each other*, are characteristic for language use across all subject-specific contexts. As mentioned in the discussion on thematic formations, the basic semantic relations between items in the thematic formations of the different subject communities are the same, although there are clear differences between the items included in the formations. Examples of such basic semantic relations are cause and effect, addition, condition, similarity and opposition and taxonomic relations (class and members). These relations can be realised in different ways through the use of basic academic vocabulary and expressions. In

sentences (d) and (e) below, a relation of cause and effect is expressed, whereas sentences (f) and (g) illustrate relations of class (ecological effects of the destruction) and members (concrete examples of ecological effects).

d) The increase in the greenhouse effect is partly caused by the destruction of the tropical rain forest.

e) The destruction of the tropical rain forest contributes to the greenhouse effect.

f) The extinction of species and a decrease in soil fertility are examples of the ecological effects of the destruction of the tropical rain forest.

g) Ecological effects of the destruction of the tropical rain forest include the extinction of species and a decrease in soil fertility.

These basic semantic relations can also be realised by means of grammatical structures such as conditionals and comparisons involving the use of conjunctions. The effective use of these structures can realise ideational meaning in which the phenomena referred to are not simply presented as unrelated entities, but are brought together into a semantic network that expresses the ways in which they are connected. This realises meaning that indicates an approach towards the interpretation of experience valued by subject communities, namely the quest to understand the world *as a system* in which phenomena are related to each other in predictable ways. Compare in this regard the following two extracts from learners' responses to a task on the international destruction of the tropical rain forest (the task is presented in Appendix A). This task required learners to comment on a statement claiming that the destruction of the tropical rain forest is worst in Asia. This had to be done on the basis of a comparison between a depiction of the international destruction of the tropical rain forest, as presented in a type of bar chart (where the bars are presented in the form of trees of different heights) and a table with information about the size of the tropical rain forest, given in km^2:

Task 3, 2014

6) The Asians use the wood for their industry. They don't care very much about the conditions of the rainforest. They only think about money. They ignore the danger of natural catastrophes.

Task 3, 2020

7) If the destruction continues the biggest decrease will be in America because the annual change is much bigger in Brazil and the other Soutern American

> *states than in Asia. The rain forests in Asia are smaller so the percentage is higher.*

The interpretation of experience realised in the first text presents the aspects involved as unrelated entities. No attempt is for instance made to indicate the possible causal relations between the use of wood for industrial purposes and financial considerations, or between a lack of ecological awareness and possible ecological consequences. The use of the personal pronoun *they* also places the focus on human agents, rather than on the events. This use of the lexico-grammar has the effect of creating a construction of experience in which the emphasis is on identifying those who are allegedly responsible for the destruction, "the Asians", rather than on describing the nature of the destruction itself.

The construction of an interpretation of experience realised in the second text, however, presents the issues involved in a network of relations in which the entities are connected. Possible tendencies in the future decline of the tropical rain forest are related to current practices, through the use of the conjunctions *if* and *because*. The conjunction *than* realises a comparison between the current annual changes in the size of the tropical rain forests in South America and those in Asia. The conjunction *so* realises a relation between the size of the rain forests in Asia and America and the extent of the destruction presented in percentage terms.

What are the *consequences* that the use of these lexico-grammatical devices by subject communities have *for the rating of learners' written responses* to subject-specific tasks? These conventionalised ways of using language occurring in the discourses of the subject community give rise to a number of expectations in terms of learners' texts. In scoring a learner's response to a task, it would be expected that the text *represents a particular aspect of experience* and not another, e.g. to construct meaning about Kisangani's climate and not about its location, should the task focus on climatic and not topographical aspects. In traditional writing assessment terms, one would expect the response to be on topic. There would also be the expectation that the response *reflects*, at least to some extent, *the technical nature of the subject community's activities*. Naturally, as apprenticing members, learners cannot be expected to have an in-depth understanding of all of the subject community's activities and extensive knowledge of its taxonomies. Consequently, the technical or subject-specific terms learners are

expected to be able to use, should be made clear, both during teaching and learning activities and in terms of the framework of expectations used in rating learners' texts. Finally, it could be expected that a learner's text reflect the interconnectedness of the phenomena reported about through the inclusion of words belonging to *general academic vocabulary* and *structures that realise logic relations*.

Whereas language is viewed as a *mode of thinking* in its ideational function, it is viewed as a *mode of doing* in its interpersonal function (language as reflection as opposed to language as action) (e.g. Halliday 1989: 20).

3.4.2.3 Interpersonal meaning

In its interpersonal function, language is used to *construct a particular social reality and a point of view*. (e.g. Halliday 1989: 20, Halliday / Martin 1993: 27, Eggins 1994: 12-13, 78, Lemke 1998: 93-94).

As already pointed out in the discussion under 3.4.1, the nature of the tenor of discourse in the context of test-based task processing presents learners with a number of difficulties. The *features of the tenor of discourse in test-based task processing* can be summarised as follows: A learner's written response to a test task should reflect the *formal nature of the situation*. It should further *be written to inform*, although it is in actual fact offering a service, by demonstrating knowledge and ability in the form of a written text that is assessed by the teacher in order to make a decision about a learner's progress. In other words, it is expected to express a stance that can be described as "this is how things are", whereas its real function in the community from the writer's point of view can be summarised as "Would you accept this as an appropriate account of how things are?" Finally, it should realise a particular orientation towards the meaning constructed. In test situations, this orientation can be said to be authoritative in nature, reflecting certainty that is supposed to result from an understanding of the subject matter and mastery of subject-specific procedures. Learners may struggle to take on this *authoritative voice*, since they may not yet fully understand the subject matter or be familiar enough with subject-specific procedures. How do these features of the tenor of discourse in test-based task processing manifest in language use and which difficulties do learners have in the linguistic realisation of these features?

Formal language is, among other aspects, characterised by neutral and formal lexis, the use of full forms rather than abbreviated forms and the avoidance of all colloquial forms (Eggins 1994: 65-66). Compare the following two sentences in terms of the degree to which formal versus informal language is used:

a) *Kisangani has a both a high total annual precipitation and a high average temperature.*
b) *It really rains very, very much in Kisangani and it's extremely hot there.*

In referring to the total amount of rain and the average temperature, sentence (a) contains the neutral lexical item *high*. It is neutral in the sense that it does not express a personal judgement and particular attitude in terms of the amount of precipitation or in terms of how high the temperature is. In comparison, sentence (b) expresses a personal judgement and attitude towards these aspects through the modal intensifiers *really*, *extremely* and *very* (which is also repeated). The use of the short form *it's* also contribute to the fact that sentence (b) realises an informal context.

The next two extracts from learners' responses to task 1c contain modal intensifiers that express a personal judgement and attitude towards the stability of the temperature, the extent of the humidity and the amount of precipitation.

Task 1c, 2029

1) *Kisangani has a very stable temperature of about 25°C every month. It has a very humid climate, because its value of precipitation (in mm / m^2 / month) is more than twice higher than the average temperature (in °C).*

Task 1c, 3011

2) *Kisangani has a relief about 460m and the amount of percipitation is in winter and summer very low, (about 100m) and in spring and autumn especially very high (about 220m at the highest point).*

The following two extracts were taken from learners' responses to task 3 (see Appendix A). In both texts, the learners express strong attitudes towards the destruction of the tropical rain forest through their lexical choices.

Task 3, 2029

3) *Yes, Asia had most of the rain forest in it's country. It declines a lot. This is very sad and tragic because the living rooms from many animals are destroys and the destruction goes on and on.*

Task 3, 2001

4) As to my mind the destruction of the rainforest is absolutely terrible.

In both instances, modal intensifiers (*very*, *absolutely*) are used in combination with lexical items that express strong negative evaluation: *sad*, *tragic* and *terrible*. Lexical items that also express negative evaluation, but which have come to be associated with a more "objective", or "rational" and less emotional stance are for example *unfortunate* and *regrettable*.

As far as its *orientation towards the meaning constructed* is concerned, the sentence below is a categorical assertion, because it is an unmodulised expression. This is established through the use of generic reference, realised by the verb *has*. Through generic reference, propositions are expressed as universal truths, without a specific time reference (e.g. Simpson 1990: 77-76).

c) Kisangani has a tropical climate.

In contrast, sentence (d) suggests a degree of tentativeness in terms of the truth value of the statement, due to the use of the modal verb *may*.

d) Kisangani may have a tropical climate.

The following sentences, taken from learners' responses to the task on Kisangani's climate, illustrate other devices involved in expressing a lower degree of probability and thus less certainty in terms of the truth value of the meaning constructed. The meaning constructed in sentence 5 is modulised by the use of the modal verb *seems*:

Task 1c, 5012

5) The climate in Kisangani seems to be very constant.

In sentences 6, 7 and 8 Mood Adjuncts (e.g. Eggins 1994: 167), *I think*, *rather*, *Maybe* and *relatively*, modulise the meanings made:

Task 1c, 5015

6) I think the type is much precipitation.

Task 1c, 2029

7) Kisangani is located in the rain forest, so it has a rather humid climate.

Task 1c, 2029:

8) Maybe Kisangani has got (Tropen –oder Subtropenklima?), because it is relatively hot all year through and there is a high amount of precipitation.

In terms of the last sentence, the writer's level of uncertainty regarding the meaning constructed is further signalled by the lexical uncertainty regarding the correct term for the type of climate. This is firstly represented by the use of the German terms and secondly by presenting two possible terms followed by a question mark.

In addition to indicating one's orientation towards the truth value of the meaning one constructs, it is also possible to express *judgements about the frequency with which something is regarded to happen* (e.g. Eggins 1994: 179). The following extract has been taken from a learner's response to a task in which learners had to evaluate a suggestion to turn the tropical rain forest into a national park. Learners had to base their evaluation of the suggestion on the principles of sustainability. The learner's repetitive use of the word *always* gives the meaning the status of a sweeping generalisation. This type of language use is problematic in subject-specific contexts, since subject communities strive for precision of expression.

1c, 3006

9) *It's like the national parks in Africa to save animals. There is <u>always</u> a group of people who get into the park and kill an animal. With the parks in the rain forest it will be nearly the same. There is <u>always</u> a group who get into the park and gain some wood.*

What are the *implications* of the complexities posed by the tenor of discourse in test situations *for the rating of learners' texts*? It is suggested here that a framework of expectations used to rate task-based subject-specific writing should include adherence to stylistic norms related to the *use of formal language*, thus the use of a formal and attitudinally neutral lexis and avoidance of colloquialisms. These are not only the norms of the subject community's written genres, but also of writing in all public contexts, thus in contexts where there might be an unequal distribution of power between the writer and reader, where they have only occasional contact and low affective involvement. By expecting learners to adhere to the stylistic norms of this type of writing, the subject-specific classroom contributes towards preparing learners for the demands that future institutional contexts (e.g. educational, professional, public administrative institutions) will place on their language use.

Expectations in terms of the *realisation of a particular speech function* (e.g. to give information) and a particular *orientation towards the meaning constructed*, should be based on the particular task. Tasks differ in terms of the amount of support they give learners concerning the meaning to be constructed through the formulation of the instruction and the nature of the input material. This, in addition to learners' expected competence levels, influences the degree of certainty about the meaning constructed the reader can expect a learner to express in his or her answer. Furthermore, tasks differ in terms of the genre they acquire learners to produce and genres show variation regarding the speech functions conventionally associated with them, as well as the patterns of modality involved in their realisation. Subject-specific descriptions of natural phenomena for instance, are conventionally realised through the declarative voice and generic reference (unless they describe completed processes, such as the geological effects of a past ice age). Rational arguments, however, may very well contain other speech functions, such as offers, e.g. "I would like to argue that …", or commands, e.g. "In assessing the situation, bear in mind that …". They can also entail the use of quite a number of modal devices, in order to distinguish between what is regarded as empirically supported and what is presented as a possible hypothesis.

Feedback should definitely be given on instances of language use that are inappropriate in terms of the speech function and the writer's orientation towards the meaning constructed. However, these instances should not lead to the subtraction of marks. As in terms of the use of formal language, the ability to effectively and appropriately use lexico-grammatical structures in realising particular speech functions and orientations concerning probability, usuality, obligation and necessity is required in discourse contexts beyond the subject-specific classroom. Learners should thus be given opportunities to acquire and practice the strategies involved. However, test results have material consequences for learners and the tenor of discourse in the context of test-based task-processing is particularly complex, requiring learners to write in an authoritative voice on subject matter. Learners should be given opportunities to learn, also in tests. Consequently, inappropriate language use related to the learners' speech role and modal orientation should not affect grades in subject-specific contexts.

Whereas language is used as a means of *constructing a physical and biological reality* in its ideational function and to construct a *social reality* in its interpersonal function, the textual function constructs a *semiotic reality* (e.g. Halliday / Martin 1993: 29).

3.4.2.4 Textual meaning

In its textual function, language is seen to *organise and link the ideational and interpersonal meanings* constructed into a coherent functional whole and to *relate these to the context* (e.g. Halliday 1989: 20, Halliday / Martin 1993: 28-29, Eggins 1994: 12-13, 78, Lemke 1998: 94). The lexico-grammar is thus used to establish cohesion (text-internal structure) and coherence (text-context organisation).

The following example serves to briefly illustrate a number of ways in which textual meaning is realised:

a) *Kisangani has a tropical climate. The average temperature is 25°C and its total annual precipitation is 1804mm. It is humid, since the precipitation is higher than the evaporation.*

This short text can be said to have both *generic and situational coherence* (e.g. Eggins 1994: 87-94). It is possible to identify it as a description and not as a narrative or an argument (generic coherence). The text also has register values that are firstly stable over the whole text and that secondly relate the text to the situation in which it is embedded, including the task to which it is a response (situational coherence). The whole text is about aspects related to climate and the lexical choices and numerical references to measured quantity and temperature suggest it is embedded in specialised activities (field of discourse). Furthermore, it consistently adopts a formal tenor and a declarative voice (tenor of discourse) and constructs experience using language associated with written rather than spoken language (mode of discourse).

Generic and situational coherence are both in part realised by the cohesive ties established in the text. The same participants are referred to in the different sentences and semantic links are established between them. In the second sentence, the pronoun *its* is an anaphoric reference to *Kisangani* and the pronoun *it* in the third sentence anaphorically refers to the concept of Kisanagni's climate, also introduced in the first sentence. Lexical cohesion is established by the lexical items semantically related to climate: *tropical climate*, *average temperature*, *total annual precipitation*, *humid*,

evaporation. Finally, logical relations are established through the conjunctions *and* and *since*.

The following three examples of learners' responses to task 1c show *problems in terms of the appropriate realisation of aspects of textual meaning*. The first text below (already discussed in terms of the ideational meaning it realises) and second text both show features of a spoken answer given in direct response to a question posed in face-to-face interaction.

1c, 2030

1) It's very hot and wett. I can imagine that the air is humid.

1c, 5015

2) I think the type is much precipitation.

Without being familiar with the task, it is not possible to discern with certainty what these two texts are about. The referents for *It*(*'s*) in the first text and *the type* in the second text are lacking. Although conventional in spoken interaction, this lack of lexical explicitness is problematic in formal context-reduced communication where language is used in written form. Furthermore, it contributes to a lack of cohesion found in the first text, because the semantic ties between these two sentences only become clear when the reader fills in the missing concept "climate".

The next example is comparable to the first two, in the sense that the text is not presented as a semantically independent unit.

1c, 2016

3) As I've already said Kisangani is an humid area because of the high rainfalls and his also high temperatures. The whole year it is over 24°C but it also rains the whole year.

In this answer, the reader is referred back to a previous response to another task (*As I've already said*). The lexical choice *said*, rather than for instance "explained", contributes to the construction of textual meaning associated more with spoken language found in a conversation than with formal written communication lacking the possibility of immediate feedback. Unlike text 1, cohesion is however established through the semantically related lexical items connected to climate: *humid*, *high rainfalls*, *high temperatures* (which is semantically also related to the numeric information about temperature) and the verb *rains*. The reference to the time period *The whole year* is also

repeated. Furthermore, semantic ties are established by the conjunctions *because*, *and* and *but*. Finally, although used incorrectly in terms of a standardised grammar, the pronoun *his* anaphorically refers to *Kisangani*.

The *implications* of the textual meaning for developing a framework of expectations as the basis for *rating scales* can be summarised as follows: a learner's written text is a response to a subject-specific task and as such records a solution to a specific problem. The written response thus textualises a construction of experience in a situation where immediate feedback is not possible. In order to function as a text, the learner's written response should be cohesive and show generic and situational coherence. In addition to a clear organisation of the different functional units in the text (thus following the schematic structure of a specific genre), and the establishment of semantic ties between different elements in the text, it should *consistently* adhere to the norms associated with formal writing, namely thematic focus, formal lexis, clarity and succinctness of expression and the use of a standardised grammar. By requiring learners to produce written texts that function as cohesive and coherent units of meaning, learners can be prepared for the demands placed on their writing competence in further educational and professional contexts.

In the three sections above, I have illustrated how conventional uses of the discourse semantics and stylistic norms for the context of test-based task processing are identified on the basis of the register concept and an understanding of the ways in which the situational variables field, tenor and mode manifest themselves linguistically in this specific context of situation. I have also explained that these norms function as expectations with which a rater approaches a learner's written response to a test task. In the following two sections, the expectations identified on the basis of the register concept are outlined and the process involved in identifying these expectations is summarised.

3.4.3 Framework of expectations for learners' language use in task-based subject-specific writing

The following devices contribute to constructing a subject-specific register and learners can be expected to use them in their written responses to test-based tasks:

- The use of cohesive ties in order to link the different parts of a text.
- The realisation of generic coherence through the establishment of a clearly identifiable rhetorical structure that is appropriate in terms of the requirements of the task to which the text is a response (this expectation corresponds to expectations in terms of genre discussed under 3.3.3).
- The realisation of situational coherence by adhering to the following stylistic norms:
    - The inclusion and correct use of subject-specific terms and precise forms of expression as a reflection of the technical nature of the subject community's activities.
    - The consistent use of formal language, including the use of a neutral and formal lexis and the avoidance of colloquialisms.
    - The appropriate use of speech functions and modal devices, seen in relation to the task and required genre involved.
    - The consistent use of clear and succinct expressions.
    - The consistent use of words from a general academic vocabulary and structures realising semantic relations.
    - The consistent use of a standard grammar.

The theoretical constructs involved in identifying these expectations are summarised below.

3.4.4 Summary of the theoretical constructs involved in developing a framework of expectations based on an analysis of the context of situation and register theory

In identifying expectations in terms of the register a learner's text should realise, the text is placed in its *context of situation*. An analysis of the particular manifestation of the *three situational variables*: the field, tenor and mode of discourse in this context is used to make decisions about the ways in which the *discourse-semantics* and *lexico-grammar* should be used in order to realise the relevant ideational, interpersonal and textual meanings involved. These decisions are applied in formulating a framework of expectations that is used to develop *analytic scales* with which learners' language use

can be rated. The factors involved in developing the framework of expectations and the interrelations between them are visually depicted in Figure 3.6.

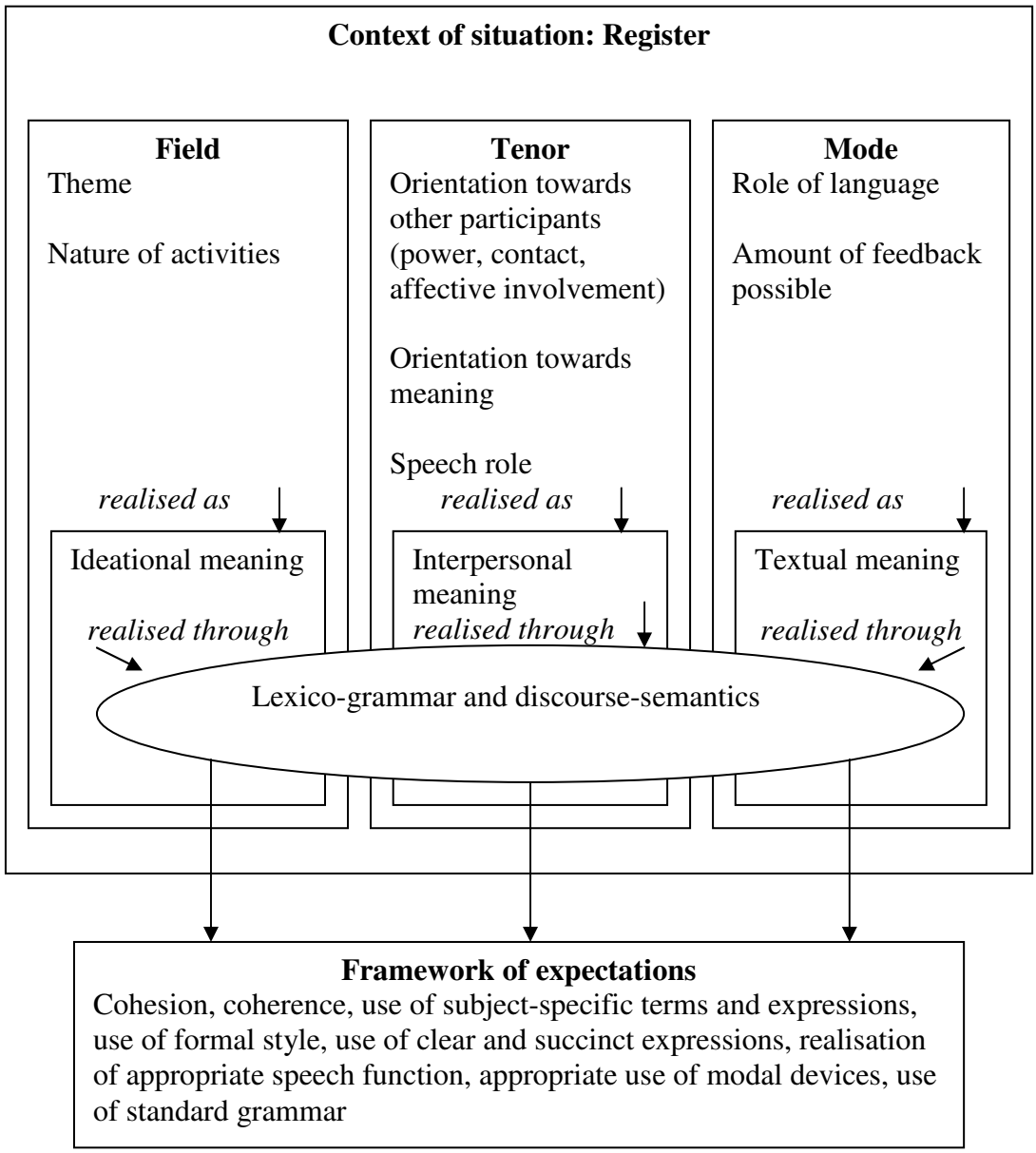


Figure 3.6: Concepts from the context of situation involved in developing a framework of expectations

Although *task instructions* normally give clear indications in terms of the ideational meaning to be constructed in a learner's response, demands in terms of the types of interpersonal and textual meanings to be constructed are usually not spelled out. It is taken for granted that learners understand the nature of interpersonal relations between themselves as writers and the teacher as reader, as well as the role language is playing in

the context of task processing, namely to record a solution to a problem in writing so that it can be repeatedly consulted. In the face of the difficulties that especially the tenor of discourse in the context of test-based task processing can cause learners, subject-specific teachers should draw learners' attention to these unstated demands. It is also possible to use tasks that give clear instructions in terms of the interpersonal and textual meanings to be constructed, in order to raise learners' awareness of the interrelation between tenor and mode and language use in subject specific discourse. These are tasks where learners are expected to construct ideational meaning in a specific text type and for an audience other than the teacher. An example of such a task could for instance involve designing a poster (mode of discourse) that informs members of the local community (tenor of discourse) about possible ways in which they can contribute to the protection of the tropical rain forest (field of discourse).

The *analytic scales* that have been developed in the context of this study in order to rate learners' language use are presented in Appendix C. In developing the analytic scales, the expectations identified by means of an analysis of the context of situation (listed under 3.4.3) were divided into two categories: expectations related to the effectiveness of the textual organisation of meaning and expectations in terms of the appropriateness of the language use. The expectations were then used to design seven scales: three under the first category and four under the second.

In section 3.3, I have illustrated how concepts from the context of the subject community's culture give rise to expectations in terms of learners' responses to test-based subject-specific tasks. Section 3.4 focused on the role of the context of situation in identifying expectations and I have pointed out that the context of situation and the register concept contribute to the development of analytic scales needed to rate learners' language use. In section 3.5, I will explain how the concept of a *required discourse function* merges the context of culture and context of situation and serves as the basis for *holistic scales*.

3.5 Merging the context of culture and context of situation: the concept of a required discourse function

3.5.1 Defining discourse function

In order to develop scales with which the degree of task fulfilment can be rated, it is necessary to operate with a *construct that combines the context of culture and context of situation* in which a learner's written response is embedded. The concept of a required discourse function integrates the expectations that arise in these two contexts in terms of a learner's written response against the background of the demands of a task. The definition of discourse function applied for the purposes of this study draws from Zydatiß's (2005: 162-165) definition of this term, presented in his discussion on the integration of content and language learning in bilingual subject-specific education. Zydatiß's (2005: 162-165) definition has however been reinterpreted on the basis of a social semiotic view of the relation between content, language and cognitive processes and to include interpersonal and textual meaning. In the following two sections, Zydatiß's (2005: 162-165) definition is first presented, followed by the reinterpretation of his definition.

3.5.1.1 Discourse function as defined by Zydatiß

Zydatiß (2002, 2005) himself applies sociocultural theory and systemic functional linguistics in addition to the genre approach in his discussions on a curricular model for bilingual subject-specific education. In essence, the view of the meaning-making process involved in task processing reflected in Zydatiß's (2005: 162-165) discussion is that ideational meaning, resulting from cognitive operations involving knowledge structures, is realised in the form of a text that belongs to a specific genre.

He (Zydatiß 2005: 163) defines discourse function as follows:

> "In unterrichtlichen Kontexten greift die jeweilige Aufgabenstellung (=*task*) eines bestimmten Lern- oder Ausbildungsbereichs bzw. eines spezifischen Schul- oder Studienfaches – fokussiert auf eine spezifische Text- oder Materialsorte (=*genre*) – auf die drei Parameter des inhaltlichen, kognitiven und sprachlichen Lernens zu ... Die gemeinsame, zentrale Schnittmenge von Sprache, Inhalt und Denken bilden die so genannten Diskursfunktionen ... Die stärkste Affinität haben Diskursfunktionen sicher zur kognitiven und zur sprachlichen Ebene, weil sie kognitive Operationen in einer spezifischen

> lexikogrammatischen Realisierung widerspiegeln (die inhaltliche Dimension kommt über die fachlichen Gegenstände mit den entsprechenden Konzept- und Begriffsbildung zum Tragen)." (Zydatiß 2005: 163. Italics in original.)

The concept of a discourse function, as defined by Zydatiß, is firstly useful for the development of holistic scales, because it attempts to *integrate the concepts* genre ("spezifische Text- oder Materialsorte"), language use ("Sprache"), subject-specific content ("Inhalt") and cognitive operations ("Denke") into an integrated construct against the background of the demands of subject-specific tasks. Although not reflected in the quote presented above, Zydatiß (2005: 162-163) does relate content to ideational meaning and following Mohan (e.g. 1986, 1998), views ideational meaning "transported" (Zydatiß 2005: 163) in a particular genre to have a cognitive equivalent in the form of particular knowledge structures.

A further useful aspect of Zydatiß's (2005: 163) definition of a discourse function is that it *applies a wider understanding of texts*, making provision for the use of different social semiotic systems in the textual realisation of meaning ("verbale und/oder visuelle Unterstützungssysteme"). In explaining the value of the concept of a discourse function for bilingual subject-specific education for instance, Zydatiß (2005: 163) writes:

> "Diskursfunktionen sind im bilingualen Sachfachunterricht besonders zu reflektieren, da sie eine zentrale Gelenkstelle zum integrierten Sach-Sprachlernen darstellen; und zwar um einen über ihre linguistischen Exponenten und zum anderen über graphische Repräsentationen, die in vielen Fällen möglich sind. Beide Aspekten können in Unterrichtssituationen (im Sinne des *scaffolding* der soziokulturellen Theorie) als verbale und/oder visuelle Unterstützungssysteme für die kognitive wie sprachliche Verarbeitung der Inputmaterialien bzw. für die sprachlich-inhaltliche Realisierung des Outputs der Lernenden dienen." (Zydatiß 2005: 163. Italics in original.)

Including non-verbal social semiotic systems in defining a discourse function is particularly useful in the context of the subject geography, where a great number of texts include a combination of linguistic, visual and mathematical forms of representation.

The relations drawn between a task, language, content and thought in Zydatiß's (2005: 163-165) discussion on discourse functions can be summarised as follows:

(1) A task that focuses on a specific genre involves content, language and cognitive learning.
(2) The area where language, content (cognitive structures) and thought (cognitive processes) intersect, constitutes a discourse function.
(3) Discourse functions have the closest relationship to language and thought as they are lexico-grammatical realisations of cognitive operations.
(4) Content is related to language through the ideational function and realised through the terminology used.
(5) The textual "transportation" of ideational meaning has a cognitive equivalent in the form of knowledge structures on which cognitive operations are based.

3.5.1.2 Defining a discourse function: a social semiotic and systemic function linguistic reinterpretation

For a definition of a discourse function that fully reflects social semiotic and systemic functional linguistic views on meaning-making processes, Zydatiß's (2005: 163-165) definition needs to be adapted so that it points out that:

(1) The cognitive operations involving knowledge structures are embedded in activity sequences that enact a subject community's standardised activity structures.

(2) An individual text that can be related to a specific genre is the product of activity sequences carried out according to the conventions of the subject's activity structures.

(3) These activity sequences involve cognitive operations mediated by social semiotic systems.

(4) Not only ideational meaning, but also interpersonal and textual meaning is involved in realising a genre.

(5) In the context of task processing, the ideational meaning to be constructed focuses on a particular aspect of the subject community's thematic patterns.

As explained in sections 3.3.2 and 3.3.3, systemic functional linguists see meaning as constructed through a serious of *action sequences* that realise *activity structures*, which are standardised in subject-specific contexts. A genre is seen as a socially identifiable

and functionally defined schematic structure that results from the enactment of particular activity structures (Lemke 1990: 199). To illustrate: should a task require the description of a particular region's climate on the basis of an analysis of a climate graph, the learner has to go through a series of actions in constructing the description. These actions include various steps in analysing the climate graph and then verbalising the meaning constructed on the basis of this analysis in written form. The subject community has developed activity structures or conventionalised procedures both in terms of the analysis of climate graphs and the production of descriptions. The activity structure for the analysis of climate graphs for instance involves four sets of actions: identifying aspects related to temperature, identifying aspects related to precipitation, comparing changes in temperature seen over a period of twelve months to changes in precipitation and describing the weather station's location. In analysing aspects related to climate, the maximum temperature, the minimum temperature, the temperature range, the annual average temperature, and changes in temperature over a year should be identified. Provided that a learner applies these procedures in analysing the climate graph and procedures for the production of a description, his or her text will realise the ideational meaning and genre required by the task.

From a social semiotic and systemic functional linguistic point of view, it is important to point out that the activities the learner engages in are *simultaneously cognitive and semiotic*. These activities always entail cognitive operations, e.g. to remember, to understand, to apply, to analyse, to evaluate and to create (Anderson / Krathwohl 2001) and cognitive structures, e.g. knowledge of the subject matter. In order for these cognitive operations and structures to be socially meaningful though, they have to involve the use of social semiotic or meaning-making systems, such as language and behavioural patterns. Put differently, the only way possible to get access to a learner's thoughts and thinking processes or to the mental concepts he or she operates with and to the logical networks activated between them is through the "mediation" of social semiotic systems.

Christie's (1985: 25-26) opinion on the importance of language in the context of school subjects illustrates the social semiotic view of the dependence of socially recognisable cognitive activity on semiotic activity:

> "Most mother tongue teachers – whether specialists in the teaching of the school subject English or specialists in the teaching of other school subjects – focus not upon language, but upon what they think of as the 'issues', 'ideas', or 'content' to be dealt with, or the mental skills to be developed in their students. Yet, issues, content or ideas are realized in language; they do not have identity apart from language patterns, any more than the skills of concern have an identity apart from the behavioural patterns in which they find expression." (Christie 1985: 25-26).

Lemke (1990: 194) also points to the interrelatedness between cognition and social semiotic systems:

> "What we call thinking is simply material processes which enact the meaning-making practices of a community: the use of language and other semiotic resources." (Lemke 1990: 194).

This interrelatedness of mental activity and semiotic activity is for instance illustrated by the meanings associated with the instructional verbs usually used in the formulation of tasks. On the one hand, describe, explain, evaluate etc. are associated with mental activity, as can be seen in the use of these terms in taxonomies of learning objectives that organise types of knowledge and cognitive processes, for instance in Bloom's widely used taxonomy (Anderson / Krathwohl 2001). On the other hand, these terms have also come to be associated with particular text types, or genres, thus with social semiotic structures, as can be seen in the use of the terms in their nominalised forms. Descriptions, explanations and evaluations, are all socially recognisable ways of organising meaning in a text.

A social definition of a discourse function should therefore not only reflect the role of a community's activity structures in the construction of meaning, but also the interrelatedness between cognitive and semiotic processes in the enactment of these structures. Furthermore, this definition should make provision for the construction of *all three types of meanings*: the ideational, the interpersonal and the textual and not only refer to the first, as is the case with Zydatiß's (2005: 163-165) definition.

As explained before, in systemic functional linguistic theory, a genre, embedded in a particular context of culture, is associated with a specific *register* in which ideational, interpersonal and textual meanings manifest themselves in distinct ways, realised by the

discourse-semantics and the lexico-grammar. All three types of meanings are always made simultaneously and the inappropriate linguistic realisation of any of the three situational variables: field, tenor, or mode adversely affects situational coherence.

The exclusion of the register concept from a definition of a discourse function contributes to the tendency to neglect the mechanisms involved in the construction of interpersonal and textual meaning already prevailing in subject-specific teaching context. A definition of a discourse function used in developing scales with which learners' task-based subject-specific writing can be rated, should therefore include the construction of all three types of meanings.

The last aspect that needs to be reflected in a definition of discourse function that is based on social semiotic and systemic functional linguistic views is that the ideational meaning constructed in response to a task deals with a particular aspect of the *subject community's thematic formation*. As pointed out in section 3.3.1, a community's thematic formations consist of networks of semantic relations repeatedly established between key concepts or thematic items. Subject communities standardise these formations as they standardise their activity structures and the textualisation of these structures in the form of genres. Thus, what a subject community refers to as its content is a collection of standardised thematic patterns. A task requires a learner to construct meaning on a particular aspect of a thematic pattern or formation. Should a learner's response deal with an aspect of the thematic patterns not required by the task, the response would be what is often referred to as "off topic" in rating terms used in the context of language assessment.

Based on this reinterpretation of Zydatiß's (2005: 163-165) definition, a discourse function is defined as follows for the purposes of this study: an individual text realises a particular discourse function by relating ideational meaning with a specific thematic focus in a socially-identifiable and functionally-defined schematic structure and in a register indicating its intended context of use. The ideational meaning and genre realised result from cognitive-semiotic activities involving the enactment of the community's activity structures and knowledge of its thematic formations.

This definition of a discourse function combines into one construct thematic patterns, activity structures and genre, which all function on the level of the context of culture, with the required register and the specific task, which function on the level of the context of situation. In other words, it reflects the interaction between the context of culture and context of situation. The definition of a discourse function presented above does this by connecting the specific cognitive-semiotic activities involved in processing the task to the subject community's activity structures. Furthermore, it relates the text produced to a particular genre and the ideational meaning constructed in a specific response to the subject's thematic patterns. In addition, it relates a specific combination of ideational, interpersonal and textual meaning realised in a text to a particular register.

On the basis of this definition of a discourse function, it is possible to analyse a task and identify the discourse function the task requires a learner's written response to realise. This *required discourse function*, or "idealised response" (see 3.3.4) can then be used to identify expectations that form the basis of holistic scales which can be applied in rating the *degree of task fulfilment*. The aspects involved in identifying these expectations are summarised in the next section.

3.5.2 Basic outline for a framework of expectations related to the required discourse function

Tasks differ in terms of the ideational meaning and genre they require a learner's response to realise. Whereas one task may require the production of a description (genre) of the climate of the tropical rain forest (ideational meaning), another may require a comparison (genre) between the ecosystems of tropical rain forests and deciduous forests. The interpersonal and textual meanings to be realised in especially test-based subject-specific tasks, however, tend to remain constant over all tasks, as explained in section 3.4.1. Due to the differences in terms of the ideational meaning and the genre to be realised, a specified framework of expectations should be developed for each single task and no general framework can be presented that can be applied over all tasks. It is however possible to present a basic outline that can be used to develop these specified frameworks of expectations.

Such a *framework, specialised per task*, must make provision for expectations in terms of the ideational meaning to be constructed and the genre to be realised. As far as the *required ideational meaning* is concerned, the thematic items to be included, as well as the semantic patterns to be realised between them, must be identified. This is done on the basis of the subject's standardised thematic patterns and by identifying the ideational meaning constructed when standardised procedures are followed in interacting with the input material. The thematic items and semantic relations identified can further be divided into main points, sub-points and supporting details, depending on the nature of the taxonomy involved. Furthermore, where tasks involve activities requiring detailed information to be extracted from the input material, e.g. information about exact measurements, such as the temperature and the amount of precipitation, the details to be included must also be specified.

As far as the realisation of the *required genre* is concerned, the schematic structure, including its possible stages, must be identified. To illustrate, if a task requires the production of an argument, the stating of an opinion is one stage and the arguments supporting this opinion constitute another obligatory stage.

Expectations in terms of the *interpersonal and textual meaning* to be constructed remain stable over assessment tasks and entail the production of a coherent and cohesive text, using a formal, clear and succinct style.

The factors involved in identifying expectations in terms of the required discourse function is outlined schematically in Figure 3.7 on the next page.

The ability to solve a subject-specific task in an appropriate way involves subject-specific discourse competence. In the next section, the knowledge and abilities involved in this competence will be explained. The definition of subject-specific discourse competence applied for the purposes of this study is based on a social semiotic and systemic functional linguistic interpretation of Vollmer's (e.g. 2007, in preparation) competence model for geography. The definition also integrates the notion of volition.

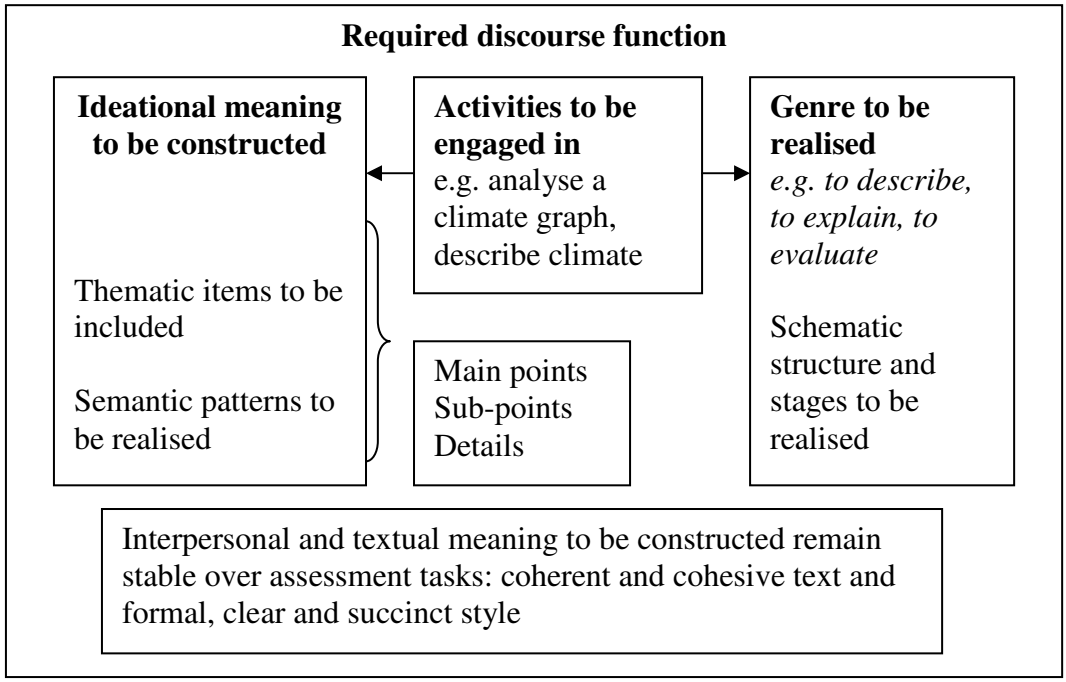


Figure 3.7: Required discourse function: basic outline of aspects involved in identifying expectations

3.5.4 Defining subject-specific task-based written discourse competence

Subject-specific task-based written discourse competence is defined for the purposes of this study as the willingness and ability to realise task-oriented and subject-based ideational meaning in written form in a register and in a schematic structure that realise the genre required by the task as the result of following subject-specific procedures in constructing meaning on the basis of the task instruction and input material.

The following *competence areas* are seen to be involved in task-based subject-specific discourse competence: subject-matter knowledge, or knowledge of a subject community's thematic formations, methodological competence or the ability to enact a subject community's activity structures, as well as subject-specific communicative competence. The ability to produce a response to a task that realises the required discourse function is also seen to involve the willingness and motivation to follow the conventions of the subject community.

In section 3.5.4.1 Vollmer's (e.g. 2007, in preparation) competence model for geography is related to the constructs from social semiotic and systemic functional linguistic theory presented in the previous sections of this chapter. By relating Vollmer's (e.g. 2007, in preparation) structural competence model to these constructs, *the involvement of knowledge of semiotic systems, such as language in all areas of subject-specific competence is emphasised*. A social semiotic and systemic functional linguistic interpretation of the competence model leads to an understanding of the competences involved in subject-specific task processing that acknowledges the interrelatedness between cognitive and semiotic activities. Not only the relation between knowledge of subject matter and knowledge of social semiotic systems is highlighted, but also the relation between subject-specific procedures and written responses as the semiotic products of these procedures. Furthermore, the interpretation of Vollmer's (e.g. 2007, in preparation) model leads to a distinction between subject-specific communicative competence and subject-specific discourse competence that relates instances of meaning-making processes to the specific contexts in which they are embedded.

In section 3.5.4.2, the role volition plays in task processing in subject-specific contexts is explained and it is shown how aspects from the ideological realm, e.g. values and identity are involved in subject-specific task processing. In section 3.5.4.3, all aspects involved in subject-specific task-based discourse competence and the relations between them are summarised.

3.5.4.1 Vollmer's competence model for geography

Vollmer (e.g. 2007, in preparation) presents a structural model of subject-specific competence for geography that identifies *three competence areas*: subject-matter knowledge, methodological competence and subject-specific communicative competence. These three competence areas are seen to *integrate in the processing of a subject-specific task* requiring the production of a written response that realises a discourse function.

Vollmer (2007, 288-289) defines *knowledge of subject matter* as declarative knowledge of basic subject-related concepts and structures, as well as of the ways in which these can be extended and developed. This knowledge is firstly seen to be applied in naming, describing and explaining geographically relevant phenomena, events and relations. It is

secondly regarded as the basis for the ability to combine knowledge of different concepts into conceptual networks that can be expanded.

Relating Vollmer's (2007: 288-289) definition of knowledge of subject matter to *constructs from social semiotic and systemic functional linguistic views*, knowledge of subject matter is equivalent to knowledge of the subject's thematic formations, consisting of thematic items and the semantic relations between them. Whereas Vollmer (2007, 288-289) refers to knowledge of basic subject-related concepts and structures, this study applies the notion of knowledge of the subject's thematic items and the standardised semantic relations drawn between them. Where this knowledge is applied in naming, describing and explaining geographically relevant phenomena, events and relations, it is possible to speak of the construction of meaning through the application of social semiotic systems on the basis of knowledge of the subject's thematic patterns. Vollmer's (2007, 288-289) notion of the ability to combine knowledge of different concepts into conceptual networks that can be expanded is equal to the ability to combine thematic items into new patterns, based on knowledge of the semantic relations the subject community operates with.

In essence, Vollmer (2007: 289) defines *methodological competence* as the ability to define a problem on the basis of a task instruction and to decide on procedures for addressing the problem. In explaining the aspects involved in defining a problem and selecting procedures, Vollmer's (2007: 289) writes:

> "Dabei wird ein Wissensraum aufgespannt, innerhalb dessen bestimmte Elemente durch Abruf rekonstruiert werden können, bestimmte andere Elemente jedoch aktiv neu konstruiert werden müssen. Während eine Reihe von *Wissensanteilen* durch die Art der Aufgabenformulierung, durch bereit gestellte Fachmaterialien sowie durch eigenes Vorwissen in das Arbeitsgedächtnis gestellt werden kann, muss anderes *Wissen durch Verfolgung geeigneter Prozeduren* erst neu geschaffen und gezielt aufgebaut werden. Dabei ist entscheidend, ob der Lerner Informationen aus den verschiedenartigen Quellen, die er zur Lösung einer fachlichen Aufgaben benötigt, so miteinander verknüpfen kann, dass damit ein selbständiger Bearbeitungs- und Problemlöseprozess in Gang gesetzt und erfolgreich zu Ende geführt wird." (Vollmer 2007: 289.Italics: DC.).

Applying a *social semiotic and systemic functional linguistic view* to this interpretation of methodological competence, the "knowledge elements" ("Wissensanteilen") that are reconstructed are seen to be knowledge of the thematic pattern involved, which are

made available to the working memory. Knowledge ("Wissen") that is constructed, is seen to be "new" meanings made in the context of processing the task. This involves the learner placing thematic items in semantic relations to create a thematic pattern he or she has not constructed before, or one that is not available to the working memory.

This activation of existing knowledge (reconstruction) and construction of new knowledge are both seen to entail social semiotic systems and are thus regarded as examples of *cognitive-semiotic activities* ("geeigneter Prozeduren"). Social semiotic systems, such as language, visual and mathematical systems of depiction are used in task instruction and the input material. Specific thematic patterns are realised in the instruction and input material through the application of these social semiotic systems. In the reconstruction of knowledge on the basis of these social semiotic resources, existing knowledge of thematic patterns is activated ("Rekonstruktion durch die Art der Aufgabenformulierung, bereit gestellte Fachmaterialien sowie eigenes Vorwissen"). Furthermore, the construction of new knowledge also involves constructing meaning on the basis of linguistic, visual, or mathematical representations, because the input material serves as the basis for this process ("Verknüpfen von Informationen aus den verschiedenartigen Quellen"). Both processes, reconstruction and construction of knowledge thus involve the application of knowledge of the ways in which particular meanings are conventionally realised in linguistic, visual and mathematical forms in the subject-specific context. This knowledge of conventions is seen to have been developed through previous experiences with similar text types.

The clearest differences between the view of the competences involved subject-specific task processing applied in this study and the one held by Vollmer (2007: 288 - 290) occur in the respective definitions of *subject-specific communicative competence* and *subject-specific discourse competence*. Vollmer (2007: 289 - 290) uses the terms interchangeably and defines subject-specific communicative competence or subject-based discourse competence as the discursive use of the L1 or the foreign language in subject-specific contexts in oral and written reception, production and interaction. Vollmer (2007: 290) equates this competence with the ability to use a specific variety.

For the purposes of this study, a distinction is made between subject-specific communicative competence and subject-specific discourse competence. *Subject-specific*

communicative competence is namely seen to be related to knowledge of the subject-specific register and genres realised in reception, production and interaction in the classroom. Furthermore, subject-specific communicative competence is not regarded as being limited to knowledge of language-related categories such as the lexico-grammar and discourse-semantics, but seen to include knowledge of all other social semiotic systems used by the community, e.g. visual and mathematical forms of representation. A wider interpretation of communication is thus applied than one that is restricted to the use of language in the construction of meaning.

The discursive use of the knowledge in a specific situation, on the other hand, is viewed as an aspect of *subject-specific discourse competence*. Subject-specific discourse competence, as applied to the situation of task processing, is seen as the ability and willingness to select and apply knowledge of the subject's thematic patterns, subject-specific procedures and knowledge of the subject community's use of social semiotic systems in order to construct meaning in response to a task's instruction and on the basis of input texts. To illustrate, in responding to a task, a learner has to produce a text on a specific topic. This involves the selection and application of knowledge of the thematic pattern the task deals with. The ideational meaning constructed in the response is also the result of the selection and application of subject-specific procedures in terms of interaction with the task's instruction and the input material. Where a task for instance requires the description of climate on the basis of a climate graph, knowledge of the subject's thematic pattern concerned with climatic aspects and the ability to analyse climate graphs in a subject-based way is necessary to construct the ideational meaning in an appropriate way. The ideational meaning constructed should then be presented in a text that reflects the genre required by the task, e.g. to describe, to explain or to evaluate and in a subject-specific register. This is done on the basis of subject-specific communicative competence. The process of producing a written response to a subject-specific task thus involves an analysis of the task instruction and the input material in order to identify the demands of the task. The demands are then related to the subject's formations: its thematic patterns, activity structures and genres, as well as to the register of the specific situation.

For a comprehensive understanding of the aspects involved in appropriately addressing a task in a subject-specific context, it is necessary to explain the role of volition. In

chapter 2, I have referred to the fact that social views of learning in a subject-specific context see it as a process that entails the initiation of the learner into the subject's ways of thinking and doing. I have also pointed out that this process is associated with the building of a specific identity. The role that identity and the willingness to take on this identity play in subject-specific discourse competence is explained in the next section.

3.5.4.2 The role of volition

In the expertise on the development of national educational standards in Germany (Klieme *et al.* 2004: 65) it is stated that:

> "Descriptions of competence and especially attempts to operationalise it, tend to focus on cognitive attributes (subject-specific memory, extensive knowledge base, automised skills). However, the concept of competence expressly also covers motivational and action-related attributes." (Klieme *et al.* 2004: 65).

The inclusion of the aspect of volition or motivation in the definition of subject-specific discourse competence applied in the context of this study serves to raise awareness of the ideological aspects involved in successful task processing. By acknowledging the role the willingness to accept certain values and to take on a particular identity plays in task processing, subject-specific teaching and rating activities can be critically evaluated in terms of the relations of power they construct. In other words, these activities can be placed on the same footing as all other cultural practices set within a specific context and operating with particular values and therefore become open to scrutiny. Furthermore, acknowledging that participation in subject-specific discourses involves the acceptance of specific values contributes to demystifying the mechanisms of the discourse for learners. This process of demystification is needed in order to empower learners to participate effectively in the subject-specific discourse.

Christie (1998: 174), in referring to the role of pedagogic discourse, states the following in terms of the process through which learners are initiated into a subject's ways of doing and thinking:

> "A successful operation of the pedagogic discourse in school will produce particular pedagogic subject positions, such that persons are apprenticed into ways of reasoning and valuing deemed of importance in a culture." (Christie 1998: 174).

The notion of the development of a specific identity, related to particular pedagogic subject positions, can also be related to Hallet's (2004: 149-150) notion of a third space

or a transcultural context. In describing the potential of bilingual subject-specific education as a space in which transcultural discourses can take place, Hallet for instance writes (2004: 150):

> "Natürlich wirkt der bilinguale Sachfachunterricht aufgrund der hier beschriebenen Eigenarten an der 'Auflösung' traditioneller, auf homogene Kollektive bezogener kulturelle Identitäten mit." (Hallet 2004: 150).

Both these views serve as a reminder of the fact that effective participation in classroom discourses entails more than the application of knowledge, it also entails the *willingness to take on a specific identity*. The subject-specific classroom, whether it entails the use of the L1 or a foreign language, presents an environment in which learners are expected to develop a particular subject position. The subject position aimed for in subject-specific communities is in its nature a transcultural one, as can be seen in the fact that its register requires the construction of meaning as a generalised self for generalised others. It is an identity that is different from the one developed in the discourses of family life and the immediate cultural context. Some learners may struggle to come to terms with this identity and may sometimes choose to reject the subject-specific ways of thinking and doing as valid interpretations of experience and action.

Consequently, the degree of task fulfilment reflected in a learner's response is also influenced by aspects related to volition and motivation, also referred to with the apt term "*social readiness*" in the expertise on the development of national educational standards in Germany (Klieme *et al.* 2004: 65). In order to appropriately respond to a subject-specific task, a learner must be willing to think and talk in a specific way. Knowledge of the thematic patterns and social semiotic systems and the ability to appropriately define the problem to be addressed and then to engage in socio-semiotic activities can only lead to an appropriate response, if the learner is willing to apply the knowledge and the ability in the context of processing a task. Learners can and sometimes do choose not to follow the conventions although they know them.

The *consequences* that ideological aspects and the willingness to accept this ideology have *for the subject-specific classroom* can be summarised as follows: Teachers and learners should be aware of the fact that the classroom is an environment in which particular ways of thinking and talking also involve particular ways of being. Subject communities should in other words not be presented as free of norms and values. The

nature of these norms and values and the influence the ways in which they regulate the subject community's discourses should be made clear to learners. They should be brought to an understanding that subject communities aim at constructing verifiable interpretations of experience that are based on systematic and standardised procedures of observation and analyses and that is realised in textual forms allowing discussion and debate with other members of the subject-specific community over time and space. Learners should also be given opportunities to challenge the norms and values of the subject community and to reflect about the differences between the discourses used in the context of the family and the circle of friends and those used in the classroom.

By understanding that the subject community thinks and acts and talks the way it does, because it is seen to support its central aims, demands for precision and explicitness in expression, the use of subject-specific vocabulary and production of cohesive and coherent texts become more than mere rules that need to be followed. They become specific ways of using language in order to realise a particular type of interpretation of experience that may be different from the ways in which other communities with which learners are familiar use language. Presented in this way, the subject-specific discourse becomes less difficult, less threatening, and less alien, because its underlying dynamics are demystified.

The aspects involved in subject-specific task-based discourse competence and the relations between them referred to in the discussion above are schematically presented in Figure 3.8 and summarised in the next section.

3.5.4.3 Summary of the aspects involved in subject-specific task-based discourse competence

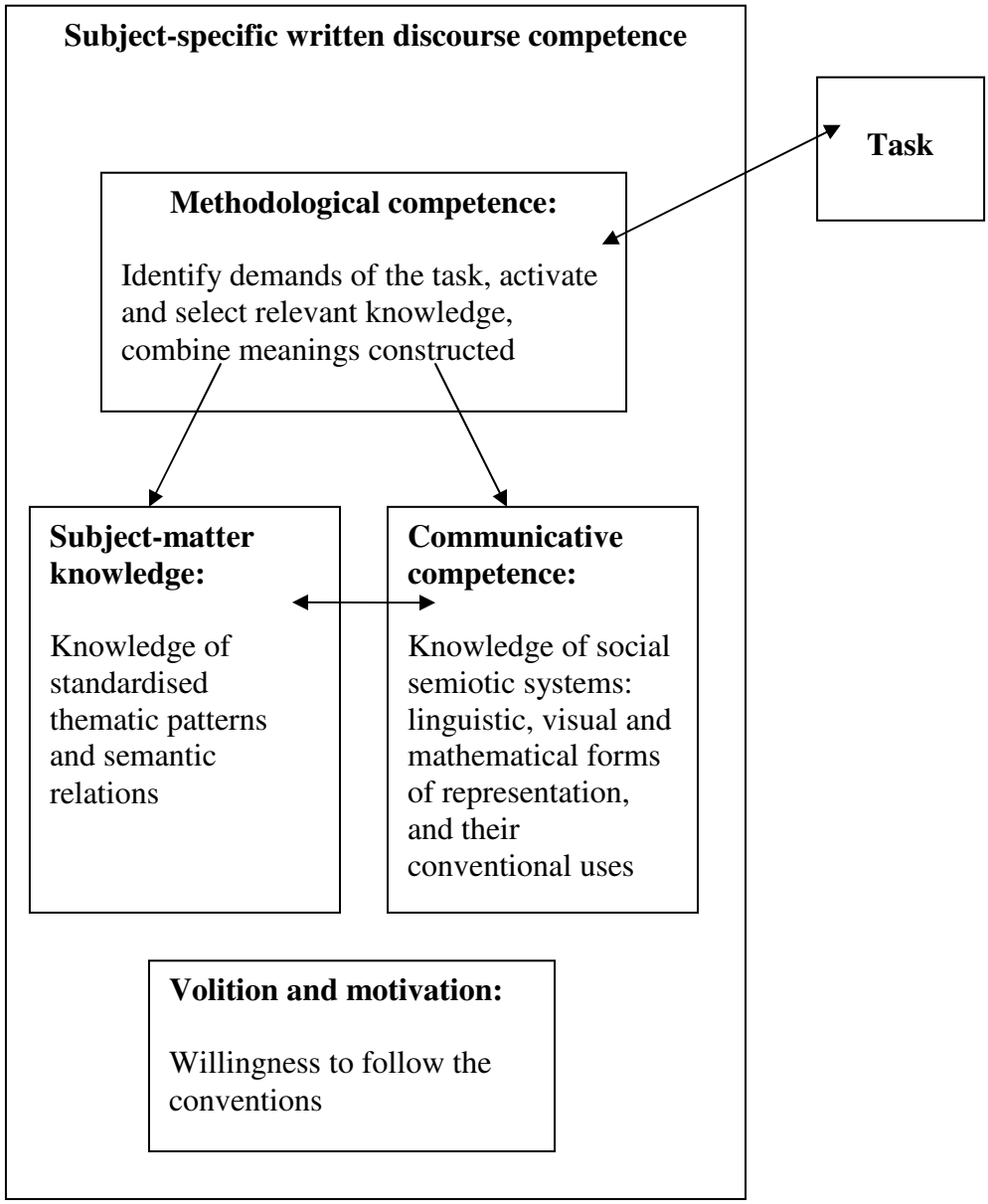


Figure 3.8: Schematic representation of subject-specific task-based written discourse competence

Methodological competence enables a learner to identify the demands of the task and to activate knowledge of the relevant thematic pattern the ideational meaning constructed should realise. It also allows the activation of knowledge of the social semiotic systems that realise meaning in the input material. This knowledge naturally includes knowledge

of language needed to construct meaning on the basis of the instruction and linguistically encoded input material, as well as knowledge needed to construct meaning in the form of a written text. Furthermore, different aspects of the thematic patterns involved can be combined in order to construct new meanings. This is firstly done on the basis of knowledge of the basic semantic relations used by subject communities in extending their thematic patterns, which is an aspect of subject-matter knowledge. It is secondly done on the basis of combining meanings constructed on the basis of the instruction and input material, which is a specific aspect of methodological competence.

There is a close interrelationship between subject-matter knowledge and knowledge of the social semiotic systems, since knowledge of the subject's thematic patterns has been acquired through social semiotic systems and can only be demonstrated through the realisation of meaning in social semiotic form.

Finally, a subject-based interpretation of the demands of the task and the activation and combination of the relevant knowledge of subject-specific thematic patterns and knowledge of social semiotic systems, depend on the willingness to engage in a task in a way that is appropriate in a subject-specific context. This aspect is part of volition and motivation. This willingness entails that the norms and values of the subject community are accepted which leads to the construction of a specific identity.

Holistic ratings of learners' texts, focusing on the extent to which the required discourse function has been realised appropriately in the responses, serve to provide an insight into learners' task-based subject-specific written discourse competence. The way in which the holistic rating of learners' responses to the six tasks has been done, is explained in the next chapter and the interpretation of the results follows in chapter 5. To conclude this chapter, the whole range of contextual factors involved in identifying expectations in terms of learners' language use and the discourse function their responses to a specific task should realise, are now summarised and schematically presented in Figure 3.9. The way in which these expectations are operationalised in rating scales is also explained.

3.6 A framework for the development of rating scales used to assess learners' task-based subject-specific writing

The process of identifying expectations in terms of learners' responses to subject-specific tasks is based on the social principle that members of communities approach texts with particular *assumptions about ways in which meaning is conventionally constructed* in the type of context the text is embedded. The process involved in rating learners' responses to subject-specific tasks is seen to function in the same way.

In identifying the expectations that form the basis of holistic and analytic rating scales, a learner's written response is related to the *context of the subject community's culture* and to the *context of subject-specific test-based task processing*. Analyses of these contexts in relation to the demands of the task, lead to the identification of an idealised response that will allow members of the subject community to construct particular ideational, interpersonal and textual meanings. The analysis of the context of the subject-community's culture entails identifying *semiotic formations* operating in the broader context of culture, including the subject community's *thematic patterns*, *activity structures* and its *genres.* Analysing the context of subject-specific test-based task processing involves identifying the *register* that should be realised in this context through conventionalised uses of the lexico-grammar and discourse-semantics.

The concept of a *required discourse function*, as indicated by the task's requirements, merges the expectations stemming from the context of a subject community's culture and the context of situation. The expectations identified on this basis can be used in developing *holistic scales* with which the *degree of task fulfilment* can be rated. On the basis of the concept of a required discourse function, it is possible to identify the *thematic focus* of the *ideational meaning* and the *specific genre* to be realised in response to a specific task, as a result of going through particular *action sequences* enacting activity structures.

The expectations related to the *register* and the *ideational meaning* to be constructed in response to a task can be used to develop *analytic scales*. Expectations based on an analysis of register form the basis of analytic scales used to rate language use and give rise to the categories *effectiveness of textual organisation of meaning* and *appropriateness of language use*. These two categories can again be broken down into

finer distinctions. The textual organisation of meaning is related to the use of devices that contribute to establishing coherence and cohesion. Learners' use of these devices can be rated with the help of the following categories: effective ordering of structural units, e.g. sentences in the text, effective organisation of meaning constructed into different structural units and effective linking of structural units. The appropriateness of the language use can be rated on the basis of the following categories: sufficient and correct use of subject-specific terms and expressions, sufficient use of formal language and succinct and clear formulations, sufficient use of general academic vocabulary and structures realising semantic relations and correctness of grammar, vocabulary and punctuation.

The expectations in terms of the *ideational meaning* can be used to develop analytic scales with which the correctness and completeness of the ideational meaning constructed in a learner's response can be rated. In developing these scales, the specific aspect of the subject's thematic formation on which the task focuses is broken down into thematic items and the semantic relations drawn between them. For example, in the case of a task focusing on the climate of the tropical rain forest, it is possible to break down the thematic pattern involved into thematic items and semantic relations related to temperature, those related to precipitation and those connected to the relation between temperature and precipitation. By breaking down the required ideational meaning in this way, a detailed rating of the content of learners' responses is made possible, because the rater's attention is drawn to specific aspects of the thematic pattern involved in a step by step way.

The analytic and holistic scales operationalise the frameworks of expectations as the highest possible level of performance. Departing from these levels of "idealised performance", *criteria* are identified and used in defining further levels of performance that meet the expectations to a decreasing extent. The nature of the criteria applied depends on the nature of the expectations. In terms of the ideational meaning to be constructed, the central criteria are the completeness of the representation of the thematic pattern, as well as the correctness of the semantic relations drawn between different items of this pattern. The relevance of the meaning constructed, seen in relation to the thematic items and semantic relations included in the subject community's thematic formation, can also be used to distinguish between different

levels of performance. In the case where a task requires meaning to be presented in the form of an argument, the consistency of the meaning in relation to the central position taken is also an important distinguishing criterion.

Regarding learners' language use, criteria such as the extent to which different structural units are effectively organised and linked to form a text are useful in defining different levels of performance. In terms of the use of subject-specific terms and expressions, correctness and sufficiency are central criteria. Sufficiency is also a criterion used to determine different levels of performance related to the use of formal language, clear and succinct formulations, general academic vocabulary and structures realising semantic relations. Accuracy is naturally a criterion in terms of the use of the grammar.

As will be shown in the next chapter, frameworks of expectations and a selection of relevant criteria provide a sound basis for developing rating scales that allow the identification of strengths and weaknesses in learners' task-based subject-specific writing. Whereas it is possible to develop analytic scales that can be used to rate aspects of learners' language use *across different tasks*, the scales used to rate the appropriateness of the discourse function and the completeness and correctness of the ideational meaning naturally need to be *specified for each task*. This is because the interpersonal and textual meanings realised in response to subject-specific tasks remain stable from one task to the next, unless the task experiments with aspects such as the intended audience and the purpose of the written text. The ideational meanings and the activities and genres involved, however, differ from task to task. Furthermore, *preliminary scales* developed for rating these two aspects of learners' writing also need to be *applied to a random selection of learners' texts* in order to ensure that they make provision for the possible range of performances and idiosyncrasies of learners' responses to tasks which cannot always be predicted.

In the next chapter, the procedures followed in developing these scales are discussed in more detail, in addition to a comprehensive description of the tasks and the processes involved in their design.

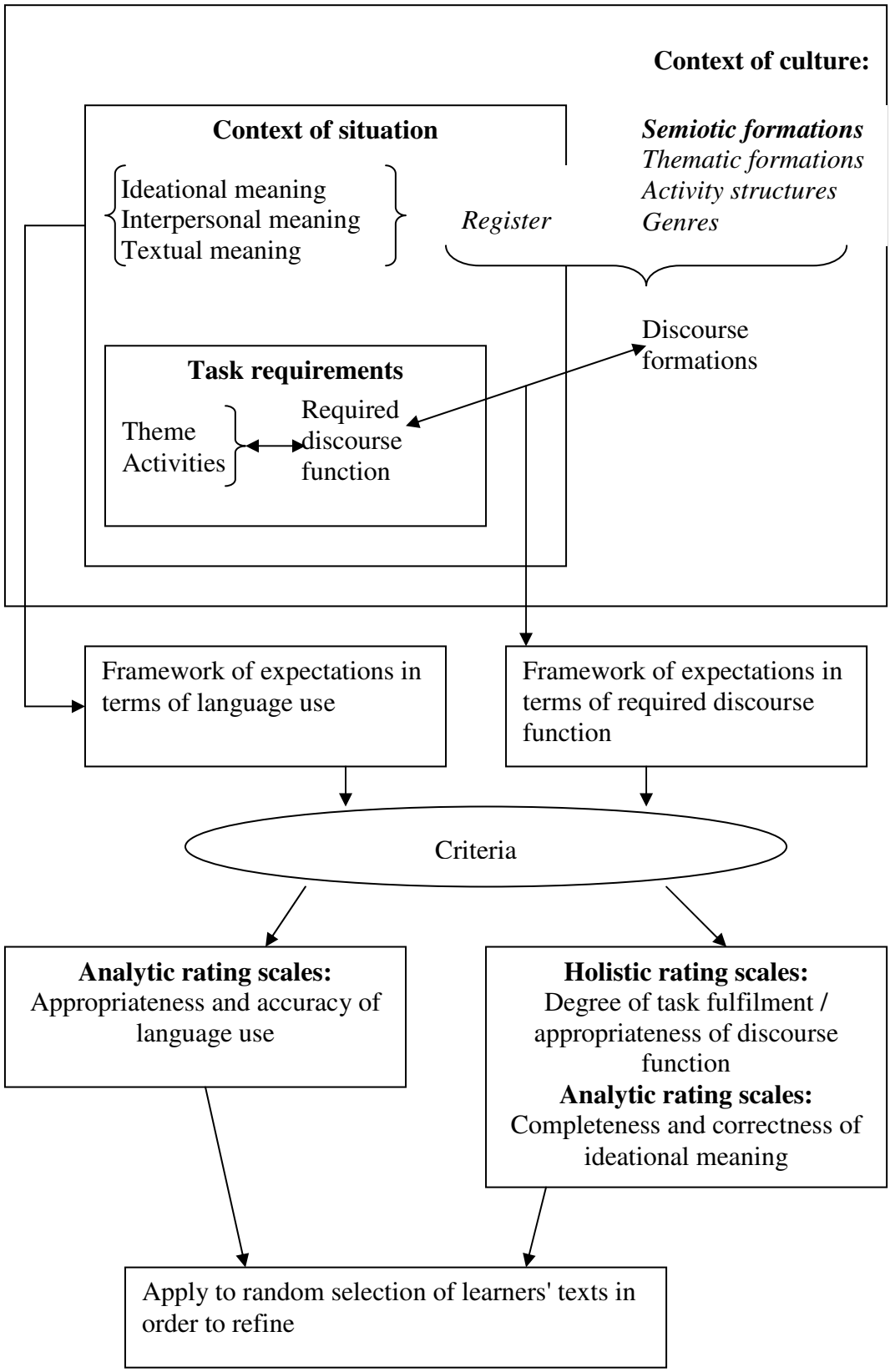


Figure 3.9: Developing rating scales: incorporating social semiotic and systemic functional linguistic constructs

4
Research design and procedures

This chapter focuses on the measures taken to enhance the reliability and validity of the assessment undertaken in this study. After giving a brief overview of the empirical aims of the study, I discuss the measures for quality assurance in testing, especially those relevant for performance assessment, criterion-referenced tests and the use of constructed response items. This is followed by a description of the sample, the data elicitation sessions and the tasks used. Finally, I present the three types of scales that were used to rate the learners' responses and explain the processes involved in their development and application.

4.1 Overview of empirical aims

As already explained in Chapter 1, the central empirical aim of this study has been to *gain insight into the task-based subject-specific discourse competence of German learners following geography in English and to identify possible deficiencies in their subject-specific task-based writing*. This involved the application of three different sets of rating scales to the written responses that 84 10th grade grammar school learners produced in response to six geography tasks.

The aim of the holistic rating of learners' responses has been to determine the *general level of task fulfilment* demonstrated by the group. The analysis of the results involved identifying the range of performances on a specific task and the number of times a specific level of performance on a particular task occurs. The analytic scoring of learners' responses aims to identify *areas of strengths and weaknesses in terms of language use* that occur across learners' performances on all six tasks. Strengths and weaknesses have been operationalised as the appropriate and accurate as opposed to the inappropriate and inaccurate use of language to construct meaning in response to a subject-specific task. The third rating is concerned with determining the extent to which the group of learners *realise aspects of the subject-specific thematic pattern related to the tropical rain forest* correctly and comprehensively.

Throughout the whole process involved in assessing the learners' subject-specific task-based writing, care has been taken to apply principles that contribute to quality assurance in testing. In the following section, a general overview of quality measures for testing is given with a specific focus on aspects that are central to performance assessment. This is followed by a discussion on the methodology used for the purposes of this study and an explanation of how principles of quality assurance in testing have been applied.

4.2 Basic principles of quality assurance in testing

The literature on quality assurance makes it clear that there is *no generic model for good testing*. The aspects involved in promoting high-quality assessment depend on the context in which the test is used. Bachmann and Palmer (1996: 17) for example state in their discussion on the development of language tests that the quality of a particular test needs to be seen in relation to its usefulness in the specific context in which it is applied. Gipps (1994: 103, 173), in writing on educational assessment, refers to this aspect as a test's "fitness for purpose".

Bachman and Palmer (1996: 18) also point out that different aspects that mark high-quality testing should not be evaluated independently, but that an appropriate balance must be found between them as to *maximise the overall usefulness* of a particular test for a specific testing situation. In their (Bachman / Palmer 1996: 17-42) model of test usefulness, serving as a basis for quality control during the test development process, they identify six test qualities. The six qualities include reliability, construct validity, three further aspects of validity, namely authenticity, interactiveness and impact, and finally practicality. Traditional approaches to testing tended to treat these qualities as more or less independent characteristics of a test that all needed to be maximised. Such an approach, however, causes considerable problems in test development. This results from the fact that two of the central qualities, namely reliability and validity, tend to be in conflict, as will be illustrated in the following discussion (e.g. Hughes 1989: 38, 42, Henning 1987: 91, Gipps 1994: 67, Alderson *et al*. 1995: 187, Wiersma / Jurs 1990: 197).

The nature of each of these qualities will now be examined in some detail and general procedures involved in supporting these qualities will be described.

4.2.1 Reliability

4.2.1.1 Definition

Reliability refers to *consistency of measurement*, both in terms of the performance of the test taker and the rating of this performance (e.g. Bachman 2004: 153, Bachman / Palmer 1996: 19-21, Brown /Hudson 2002: 149-151, Davies *et al*. 1999: 168-169, Gipps 1994: 67-69, 103, Henning 1987: 73-74, Hughes 1989: 29-31). In essence, a test score is seen to be reliable if a test taker receives the same score on a test on different occasions of taking the test, or if two test takers, who are seen to be equally competent on the aspect measured by the test, receive the same score. To achieve this, the tasks used should elicit performance of equal quality on different occasions of administering a test and the scoring of the performances should lead to the same score.

Reliability is also an issue in the use of different tests aiming to measure the same ability. In the case of norm-referenced tests, where a test taker's score is interpreted in relation to the performance of other test takers, the scores of these two tests are said to be reliable if they rank order the test takers in the same way. In terms of criterion-referenced tests, where a test taker's performance is measured against an external criterion, represented by specific target behaviours, the scores will be perceived as reliable if they consistently classify test takers as masters and non-masters (e.g. Bachman 2004: 30-32, Brown /Hudson 2002: 2-6, Davies *et al*. 1999: 138-39, 30, Gipps 1994: 79-83).

4.2.1.2 Strategies involved in promoting reliability

There are numerous strategies that can be followed in order to promote the reliability of test results. *Different strategies are however followed for different types of tests*. There are, for instance, distinctions between the strategies involved in norm-referenced tests as opposed to those followed in the case of criterion-referenced tests (e.g. Gipps 1994, Brown /Hudson 2002). Furthermore, the aspects involved in increasing the reliability of tests using selected response items, also referred to as fixed or forced response items are different from those using constructed response items (e.g. Gipps 1994, Wiersma /

Jurs 1990). Whereas selected response items (e.g. Davies *et al*. 1999: 64) require candidates to choose between response options, as in multiple-choice tests, constructed or extended response items, also called open-ended questions (e.g. Brown / Hudson 2002: 56-100, Davies *et al*. 1999: 32, Wiersma / Jurs 1990: 41-90) require test takers to formulate their own answers.

Due to the ways in which learners' scores are interpreted for the purposes of this study and because of the nature of the tasks used, the following discussion focuses on ways in which reliability is supported in criterion-referenced tests and in tests making use of constructed response items.

Criterion-referenced tests (CRTs) measure performance in terms of a certain criterion. This criterion can firstly be taken to refer to a specified set of tasks from a particular domain that the test was designed to sample and measure. Some authors, however, use the term criterion to refer to a particular level of performance required to pass a test, while a third way in which the term is applied, merges these two uses. Furthermore, CRT is also used as an umbrella term that covers domain-reference tests (DRTs) and objectives-referenced tests (ORTs). The first refers to a test that is based on a well-defined domain and that consists of items selected from that domain. ORTs assess the specific objectives of a training programme. Where these objectives also happen to represent a specific domain of behaviour, the distinction between these two types of tests is blurred. What all types of tests related to criterion-referenced testing have in common, is that *the interpretation of the test scores involves absolute decisions*, because each test taker's score is meaningful without reference to the scores of other test takers (e.g. Brown / Hudson 2002: 3-6).

The procedures followed in interpreting learners' scores for the purposes of this study can be related to criterion-referenced testing. Although the learners' performance levels are not interpreted in terms of a pass or a fail, or a mastery / non-mastery distinction, as is the case with true criterion-referenced tests, the frequency distributions of the learners' results are related to scale descriptors in order to describe the level of performance most often demonstrated in the group. The study thus does not make use of norm referencing, interpreting a learner's score in relation to the scores of other learners. Rather, learners' performances are placed on a continuum reflected by the scale and, in

reporting the results, the learners' scores are related to the level of performance they reflect.

The strategies involved in promoting the reliability of criterion-referenced tests and tests making use of extended response items can be divided into three categories. The first category is related to the nature of the test and the tasks, the second to the test administration and the third to the rating of test takers' performances.

Reliability is firstly supported by using the *highest possible number of test tasks*. Chance error is reduced by increasing the number of writing samples elicited from each test taker. In essence, the more samples of a learner's performance available, the smaller the chance that his or her performance on one specific task wrongly affects the score based on his performance over all tasks. Furthermore, since performance on constructed response items have been proven to be highly task dependent, reliability of test scores is supported by *not giving test takers choices in terms of the test tasks* (e.g. Gipps 1994: 104-108, Skehan 2001: 167-169, Weigle 2002: 60-70).

As far as the conditions of test administration is concerned, it is naturally important to minimise inconsistencies in the ways in which a test is administered to different groups, on different days and in different locations. *Standardising the test administration procedures* contributes to consistency in this regard and entails aspects such as the test instructions and the way in which they are given to test takers, the time that different test takers are given to complete the test, whether and the way in which remaining time is reported and the extent and nature of the interaction allowed between test administrators and test takers. Furthermore, creating a positive and supportive test environment that equally contributes to all test takers' physical and emotional comfort during the administration of the test is important. This involves reducing environmental disturbances to a minimum. All forms of disruption that can negatively influence test takers' concentration should for instance be avoided during all test administration sessions and the sessions should take place in locations with equal and optimal reading light and audio quality.

Constructed-response items require subjective scoring by raters, which poses challenges in terms of reliability. Not only can the rating behaviour of one single rater

vary during a rating session, but there also may be significant differences between different raters' rating behaviour. In the case of the first, it leads to intra-rater error variance, whereas errors arising from differences between raters is called inter-rater variance (e.g. Henning 1987: 76-77).

Variation in raters' rating behaviour is mainly addressed by making use of *rating scales and benchmark texts* that illustrate levels of performance on the different levels of the scale. Furthermore, where available resources cover the personnel costs involved, hiring competent raters who are then trained in using the scales also contributes to curbing rater error variance. Where more than one rater is available, performances can be scored by more than one rater in sessions during which raters are not allowed to see each others' scores. The different raters' scores can then be averaged to reach a final result. Finally, monitoring the rating process and getting further ratings where the scores given by two raters lie too far apart also contribute to optimising the reliability of the rating process. (e.g Alderson et al. 1995: 133-136, Gipps 1994: 104, Henning 1987: 32-34, 75-87, Hughes 1989: 36, 42, 97, McNamara 1996: 117-121, Weigle 2002: 127-130).

4.2.2 Validity

4.2.2.1 Traditional views: different types of validity

Traditionally, *validity* was conceptualised as the extent to which a test measures what it has been designed to measure and different types of validity were identified, the main ones being predictive, concurrent, construct and content or face validity (e.g. Brown /Hudson 2002: 212-248, Davies *et al*. 1999: 221-222, Gipps 1994: 58-59, Henning 1987: 89-100, Hughes 1989: 22-28).

In short, *predictive validity* relates to whether a test successfully predicts future performance. For instance, tests used in deciding whether students should be allowed entry to a specific institution or field of study rely heavily on predictive validity.

Concurrent validity is connected to whether the test gives the same results as some other form of measurement used to assess the same ability. The results obtained using a newly developed test can for instance be compared to the results obtained by

administering a well-established test, known to give reliable results and perceived to be sufficiently valid for a particular purpose. If the two tests test the same ability and are applied to the same group of test takers, the new test is said to have concurrent validity when the scores on the two tests correlate.

Predictive and concurrent validity are often combined to give *criterion validity*, since they are both concerned with predicting a test taker's performance on some or other criterion, either at the same time, or in the future.

Construct validity is concerned with whether a test adequately measures the underlying ability (the construct) involved and with whether the test tasks sufficiently represent tasks from the domain of performance to which the test results are to be generalised. This type of validity is firstly threatened when other abilities than those intended to be measured affect performance on the test tasks. When this occurs, there is *construct-irrelevant variance* (e.g. Davies *et al.* 1999: 33, Gipps 1994: 101-102). An example of construct-irrelevant variance would be when a test on language ability includes tasks that make use of input material involving other social semiotic systems in addition to language, such as cartoons, graphs and tables. In such a case, the test scores would not allow a valid interpretation of the test takers' language ability, because they do not only involve knowledge of language structures and the ability to apply this knowledge in constructing meaning. Should the construct definition underlying the test and guiding the interpretation of its results, however, include the ability to decode other forms of semiotic presentation, a valid interpretation is possible. In such a case, however, the construct would probably then not be language ability, but be related to a larger definition of literacy in a specific domain.

Another threat to construct validity is construct *underrepresentation*. This occurs when a test fails to sample parts of the domain it is meant to represent, or to cover aspects of the construct to be tested (e.g. Davies *et al.* 1999: 33, Gipps 1994: 101-102, 105-108). In other words, construct underrepresentation results when the tasks in the test do not systematically represent all possible categories of tasks from the domain the test is concerned with, or all possible aspects of the ability to be tested. An example in this regard would be a test that claims to assess medical doctors' English language proficiency for professional purposes, but that only focuses on communicative tasks

related to face-to-face interaction with patients. Such a test would exclude other important tasks involving communication with patients, such as consultation by phone, or tasks involving interaction with other medical professionals and with authorities. Consequently, the score a medical doctor obtains on such a test would be an invalid indication of his or her ability to use English for professional purposes.

The quality of *interactiveness*, one of the six test qualities identified by Bachman and Palmer (1996: 25-26), is concerned with construct validity. It refers to the extent to which a task involves *all* of a test taker's individual characteristics constituting the construct and *only* these characteristics.

Content validity is likely to be a result of construct validity and is related to whether the test is regarded to cover the necessary content, or the abilities involved in good performance in the domain. This aspect of construct validity is related to what Bachman and Palmer (1996: 23-24) call a test's authenticity. *Authenticity* pertains to the correspondence between test tasks and tasks from the domain of behaviours the test intends to assess. Decisions about content validity are usually based on judgements by professionals or experts from the domain, who analyse the test tasks in terms of their relevance, or the extent to which they are regarded to be a valid way of measuring performance in the domain (e.g. Henning 1987: 10).

A test's *face validity* involves judgements by test users, who have not been involved in the development, administration and rating of a test. Test takers, parents of learners who have taken tests, or institutions making decisions on the basis of test results, such as a university administration that uses a particular test for screening purposes, are examples of such test users. In terms of test takers, the perceived relevance of the test tasks is important, since it can have an effect on their affective response to the test, which can either facilitate or impede an optimum performance.

Construct and content validity require careful planning in the development of tests. Ensuring that the test measures the construct it is intended to measure and only this construct demands a detailed definition of the construct involved and of the domain to which the test scores have to generalise. It further requires writing task specifications

that define the construct operationally. These specifications stipulate the procedures and conditions under which performance will be elicited and are related to aspects such as the type and number of tasks that will be used, the amount of time test takers will get to complete the test and the ways in which their performances will be scored (e.g. Alderson *et al.* 1995: 9-39, Bachman 2004: 14-16, Hughes 1989: 22-23, Weigle 2002: 77-107). Tasks then have to be developed on the basis of the test specifications and need to be piloted or pre-tested and analysed by experts from the domain in terms of the extent to which they cover the domain involved. Where necessary, tasks need to be adapted on the basis of the results of the pre-testing and the expert opinion about their authenticity and suitability for measuring the construct they are intended to measure (e.g. Alderson *et al.* 1995: 40-104).

A final type of validity to be discussed here is what Bachman and Palmer (1996: 29-35) refer to as *impact*, which is related to notions of *consequential validity* and *washback* (e.g. Davies et al. 1999: 225, Gipps 1994: 102). These aspects all refer to a test's effect on individuals and institutions, which can be either positive or negative. The effect of tests on teaching and learning can, for instance, be seen in the tendency to teach towards a test. Where the effects a test has on educational practices are seen to lead to innovative teaching approaches that support learning and the development of higher order competences involved in solving tasks occurring outside the classroom, the test is said to have positive washback. Individuals are for instance directly affected by the feedback they get on tests and also by the decisions made on the basis of test scores, for instance whether a learner is allowed to carry on to the next grade.

4.2.2.2 Current views: validity as a unitary concept

Samuel Messick's (1989) conceptualisation of validity has had a major impact on the way in which validity is currently dealt with in educational and psychological testing. It has started a move away from focusing on different types of validity to a view of validity as a *unitary concept* with *construct* as the unifying theme and a focus on the *uses that are made of assessment results*. In his Messick's (1989: 13) view:

> "Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment." (Messick 1989: 13. Italics in original)

In terms of the purposes for which test scores are used, Messick emphasises the social values involved in considerations of validity and the consequences test use could have for individuals. This is particularly central in *high-stakes tests* where test scores are used to make major life-affecting decisions about large numbers of individuals, where these decisions are difficult to correct and where the costs caused by making wrong decisions are high. An example of a high-stakes test would be a university entry tests of which the results are used as the main criterion in deciding about admission to a university. An example of a *low-stakes test* would be a class test given by a teacher of which the results are used to decide if the class is ready to move on to the next learning unit (e.g. Bachman 2004: 11-13).

Construct validation is an ongoing theoretical and empirical process. It involves building an interpretative argument collecting evidence in support of this argument (e.g. Bachman / Palmer 1996: 22, Bachman: 257-258). The validation process for instance includes presenting a theoretical model of the construct that identifies all aspects (knowledge, abilities, skills, strategies etc.) that are part of it and that explains the ways in which they interact when a task is performed. It further includes careful task design on the basis of this model and collecting data through quantitative approaches that support the theoretical claims being made, for example by correlating test scores with scores achieved using other forms of measurement. Validity can, however, never be proven. All that test developers can do is to gain as much evidence as possible within the limits of the available resources in order to *demonstrate that the interpretations made on the basis of test scores are plausible*.

4.2.3 Practicality

The last of Bachman and Palmer's (1996) six test qualities is practicality. This quality is related to the *resources* available for developing and administering the test and influences the measures that can be taken to promote reliability and validity. Resources needed for developing and administering a test can be divided into three categories: personnel (test writers, test administrators, raters and administrative support), material resources (space, equipment, materials) and time. In developing a test, a balance needs to be found between the other five test qualities: reliability, construct validity, authenticity, interactiveness and impact, within the limits of the available resources,

while taking the intended purpose of the test in its specific situation of use into consideration. An overview is now given of the measures taken within the context of this study in order to optimise the usefulness of the test applied.

4.3 Outline of procedures followed in support of validity and reliability

In this outline, procedures followed in the context of this study are related to quality guidelines that apply specifically for performance assessment. Attention is paid to the measures taken in order to support the geography competence test's "fitness for purpose" (Gipps 1994: 103, 173), or its usefulness in the specific context in which it is applied (Bachman / Palmer 1996: 17).

4.3.1 Aspects related to validity: using authentic tasks designed in co-operation with experts

As mentioned earlier, the six tasks used for the purposes of this study all require the production of extended written responses. The approach followed in this study can therefore be described as performance assessment. The term *performance assessment*, sometimes also *constructed response assessment*, is applied to the use of an instrument that elicits a performance rather than a simple indication of choice, as is the case with multiple-choice items (e.g. Baker *et al*. 1993: 1210-1212, Davies *et al.* 1999: 144, Gipps 1994: 98-99, McNamara 1996: 1-10, 120 – 121, Weigle 2002: 46-48). The use of the term in discussions on assessment should, however, not be confused with Chomsky's notion of a competence-performance dichotomy. The word *performance* as it is applied in referring to this type of assessment, is meant to indicate that the tasks used require test takers to demonstrate their capabilities directly on tasks that are regarded as being relevant for a specific domain and that elicit actual performances, for example creating some product or engaging in some activity. Another distinguishing feature of performance assessment is that it requires relatively complex scoring procedures involving subjective decisions about the quality of a performance, usually on the basis of rating scales.

Performance assessments are generally regarded to have *high validity*, since they involve test items that resemble tasks from a specific domain (e.g. Gipps 1994: 100). They are thus seen to model "real-world" activities, rather than to fragment them, as standardised tests tend to do. This allows higher order competences to be tested. Performance tests are consequently also seen to support the validity of interpretations about behaviour in actual problem-solving situations made on the basis of test scores.

As explained above, collecting *feedback from people who are experts on the domain* to be tested is central to establishing content validity. The test used in the context of the project was therefore designed in close co-operation with educational experts on the subject geography. These experts did not only give feedback on the tasks, but were actively involved in the test development process, co-operating in developing the competence model on which the test is based, conceptualising outlines of tasks that incorporate all three competence areas identified and providing material that could be used as input texts in the tasks. Furthermore, these experts constantly gave feedback during the finalisation of the test and their input was also used in designing the frameworks of expectations used as the basis for scale development.

4.3.2 Aspects related to reliability: number of performances, applying rating scales and double rating

Although performance assessments are usually seen to be highly valid, they pose problems in terms of reliability and generalisability that need to be addressed in order to promote good quality testing. There are, however, a number of procedures that can be followed in order to enhance the reliability and generalisability of performance assessments.

One aspect of performance assessment that can negatively affect reliability, is the fact that fewer tasks can be used than in standardised forms of assessment, since performance assessment tasks take time to process and to rate. This is problematic in terms of reliability, because as indicated before, the more items included in a test, the more reliable it becomes, because it elicits multiple performances on which decisions can be made. This in turn leads to a reduction of chance errors. It is therefore advisable to use as many tasks as possible within the limits of the available resources.

In order to support reliability, no fewer than six tasks were used for the purposes of this study. The combination of a high number of tasks and a high number of learners (n= 84) has led to a *large pool of writing performances* that suffices for the purposes of this study. In terms of the appropriateness of the discourse function and the ideational meaning constructed, 467 samples of writing were rated. As far as the appropriateness of different aspects of the language use is concerned, 365 samples were rated[1].

The results found on the basis of these samples are not used to make inferences about individual learners' competences, but the aim is to *identify tendencies occurring across the majority of learners' performances*. The results of the holistic rating concerned with the appropriateness of the discourse function and the analytic rating focusing on the completeness and correctness of the ideational meaning constructed are used in frequency distributions. These show on which level of performance most learners lie regarding the holistic rating and the score that most learners received in terms of the rating of ideational meaning constructed. In terms of the analytic rating of the textual realisation of meaning, frequency distributions are used to identify strengths and weaknesses in the learners' language use that occur across all tasks and all learners' performances.

Another aspect of performance assessment that poses challenges in terms of reliability is the fact that rating in performance assessments involves subjective judgements (e.g. Huot 1990: 250-257, McNamara 1996: 117-127). The reliability in the scoring of performances can be affected by *inter-rater variation*, when the scoring by two different raters show variation, or by *intra-rater variation*, when the rating behaviour of a single rater varies over the period of time during which the scoring is done (e.g.

[1] The total number of samples rated differs slightly from task to task, since not all 84 learners did all of the tasks. The numbers of learners who did not do a specific task are however small. Thirteen is the highest number of learners who did not do a task (task 6).
The difference in number between the samples used in rating the structure and linking of a text and the language use (textual realisation of meaning) and those used in the two other ratings (appropriateness of the discourse function and correctness and completeness of the ideational meaning constructed) are due to the fact that only ratable responses (e.g. Bachman and Palmer, 1996: 218-219 and Weigle, 2002: 131-134) were incorporated in the analytic scoring of the textual realisation of meaning. In other words, only responses that were long enough to make valid decisions about each of the seven aspects rated with these analytic scales were incorporated. Where responses were too short, for instance, no valid decisions could be made about aspects such as the effective organisation of structural units and the effective use of cohesive devices. These samples consequently had to be excluded from the analytic rating focusing on the structuring and linking of the texts and on the language use.

Alderson *et al.* 1995: 129, Davies *et al.* 1999: 88-89, 91-92, Weigle 2002: 134-135). As already pointed out in the general discussion on principles for quality assurance in tests, the subjectivity of the rating process can be reduced by means of a number of procedures, such as the use of rating scales by trained raters whose scores can be averaged to reach a final score.

However, as authors like McNamara (1996), Weigle (2002) and Alderson *et al*. (1995) point out, however, it is not possible to completely exclude variation in the rating of constructed responses, due to the complexity of the scoring process required by performance assessment. Where there is routine double marking (thus where all responses are marked by two raters), decisions should therefore be taken in terms of how much variation to allow between the scores given by different raters or the level of performance assigned to a specific response. Wiegle (2002: 128) for instance points out that it is common practice to tolerate ratings that are one score point apart on a six-point scale. Only when ratings are more than one scale point apart on such a scale, are they seen to be discrepant and to require a third rating of the response in question.

All three types of scoring that have been carried out in this study involved the use of *rating scales* that had been especially developed for the purposes of the study. These scales were developed in an extensive process that firstly involved designing the scales on the basis of the frameworks of expectations presented in chapter 3. The process secondly entailed applying preliminary scales to random selections of learners' texts in order to ensure that the scales cover the whole range of performances. Furthermore, these preliminary scales were presented to teachers who gave feedback on the relevance of the categories and criteria and on the feasibility of applying the scales. Lastly, the preliminary scales were also tested by other members of the project who then gave feedback on the extent to which the scale descriptors were clear and the extent to which they allowed all responses to be assigned to a specific level of performance described on the scale.

Different *measures were taken to support reliability in scoring the responses*. All raters received training in the application of the scales and benchmark texts were made available for the scoring of the discourse function. All of the learners' responses were double rated in terms of the ideational meaning constructed and raters' scores were

averaged to reach a final score per learner. Where the discrepancy between the scores given by the two raters was too big, a blind rating was done by a third rater and the two closest scores were then averaged. A blind rating entails the scoring of a performance without knowledge of the scores already given. As far as the holistic rating of the appropriateness of the discourse function is concerned, as well as in the case of the analytic rating of the textual realisation of meaning, random double marking of a proportion of the responses was carried out. High levels of inter-rater reliability have been achieved in terms of the holistic rating and the analytic scoring of the ideational meaning constructed. In terms of the analytic rating of the textual realisation of meaning, inter-rater reliability has not been consistently high. However, where differences between raters' scores occurred these were not more than one scale point. The levels of inter-rater reliability achieved is reported on in the discussion on each of the rating sessions.

4.3.3 Generalisability of the results

Generalisability is an aspect of reliability and validity. The variation in performances elicited and the smaller number of tasks used in performance assessments pose problems in terms of generalisability. *Performance has been proven to be highly task-specific* (e.g. Chalhoub-Deville 2001, Gipps 1994: 105-108, Huot 1990: 240-249, Linn *et al.* 1991: 18-19, Linn 1994: 9-11, Skehan 2001, Weigle 2002: 60-70). Thus, performance on tasks that appear to be similar, or on tasks from the same domain, can show considerable variation. Findings in terms of task-related variation in performance are consistent with findings about the context dependency of learning and cognition. This does not only affect reliability, which is concerned with consistency of performance and scoring, but also coverage, which is an aspect of validity.

Unless a domain is very narrowly defined, as is done in domain-referenced testing, it is difficult to achieve sufficient coverage of a domain, such as subject-specific task-based writing. Due to task-based variation in performance, the tasks used in a test would need to cover all possible tasks from the domain systematically in order to allow full generalisation to the domain. Although it is theoretically possible to use a grid design approach that aims to cover the critical dimensions of tasks from the domain in a structured way, defining these critical dimensions is problematic. Especially if the

domain that needs to be tested is broad and includes tasks in which different variables combine in ways that make it difficult to predict performance and to explain the factors influencing performance. To illustrate, more complex tasks from the domain of the subject geography include different types of input material that needs to be processed by applying a range of subject-specific procedures. This possible variation in the input material and procedures combined with possible variations in the subject matter the task focuses on and variation in the genres involved make it impossible to select tasks that cover all possible combinations and that can be used in the context of one single (or even numerous) assessment session(s). Performance assessment is therefore seen to be suitable for measuring in depth, but not necessarily for measuring in breadth.

In addressing this aspect of performance assessment, test theorists (e.g. Alderson *et al*. 1995, Gipps 1994, McNamara 1996) use a quality criterion referred to as *transferability*. Applying this criterion implies that detailed descriptions are given of the tasks used to elicit the performances. This then makes it possible to decide to what extent performance on a specific task can be generalised to performance on another.

In this study, detailed analyses of the tasks are used to theorise about the types of tasks to which the results of the *holistic rating* of learners' performances on each of the six tasks can be generalised. The results over all six tasks are thus not combined into a single score and used to draw inferences about learners' overall subject-specific task based discourse competence. Rather, the results are presented and discussed in detail per task and the levels of competence demonstrated are interpreted as specific for a particular type of task.

In terms of the *analytic rating of the textual realisation of meaning* on the other hand, generalisation is possible, because the same constructs were measured over all six tasks. Due to the variation in the tasks in terms of the input material, the thematic focus and the genres involved, the six tasks show a degree of variety that is sufficient to be able to draw inferences about specific areas of strengths and weaknesses in learners' subject-specific task-based writing.

Finally, in terms of the generalisability of the findings over the whole sample, care was taken to ensure a high level of consistency in the *administration of the test*. Learners all

wrote the test in school during a 90-minute period and the test was always administered by members of the project. Learners' all received the same test booklets which were collected at the end of each session and they all had to respond to the test tasks in standardised answer booklets. Also, the possibility that learners' gender or the schools they attend might have affected their scores was investigated. No correlations were found between learners' scores on the different ratings and their school or sex for any of the three ratings.

Each of the measures taken in order to maximise reliability and to support the validity of the assessment is now discussed in more detail. After a description of the sample, the process involved in developing and testing the tasks is explained. The tasks are also described. This is followed by a discussion on the administration sessions. Finally, attention is paid to the development and the application of the rating scales, including a comprehensive description of the scales.

4.4 The sample

The learners involved in the project in which this study is embedded all attend grammar schools in Lower Saxony where geography is taught in English in addition to the German-taught tracks. In all of these schools, learners have the possibility of following a so-called "bilingual track" from the 7^{th} up to the 10^{th} grade. During this time, the subject geography is taught in the 7^{th}, the 9^{th} and the 10^{th} grade. The total number of geography classes taught in German and English in these schools was taken as the basic population from which a sample had to be drawn by chance. After this procedure had been completed, contact was established with the respective schools in order to determine whether they were willing to co-operate. Seven schools responded positively and constituted the base sample of the project. In addition, two more schools were involved in the project for piloting purposes.

The writing responses used for the purposes of this study were produced by *84 10^{th} grade learners* following geography in English at four of the seven participating schools. In all four of these schools, history is presented as a bilingual track in addition to geography and most of the learners had followed bilingual education in both of these subjects. The sample consists of 36 (43%) boys and 48 (57%) girls. There are thus

slightly more girls than boys in the group, which is normal for the distribution of the genders in grammar schools in Germany if one looks at the gender constitution of groups reported about in large-scale studies such as PISA (e.g. Baumert *et al.* 2003: 229).

Information about the number of semesters of bilingual education in geography and in history the learners had followed was elicited by means of questionnaires. All except three learners indicated the number of semesters of bilingual education they had followed in each of the subjects. This information is presented in Table 4.1 below. The number of semesters indicated include the semester during which the test was administered. The group shows some variation in terms of the number of semesters that the participating learners had foreign language education in geography and in terms of the total number of semesters that learners had followed bilingual education. More than half of the learners (54%), however, had at least four semesters of bilingual education in geography and almost two thirds (63 %) had at least eight semesters of bilingual subject-specific education in total.

Number of semesters of bilingual training in geography	**Number of semesters of bilingual training in history**	**Number of learners**
-	4	1
1	4	5
2	2	1
2	4	1
2	6	28
3	4	1
4	-	1
4	2	3
4	4	23
6	2	2
6	4	15

Table 4.1: Number of semesters of bilingual subject-specific education the learners had followed in geography and history

The biggest group of learners (28) thus had two semesters of bilingual subject-specific education in geography and six in history, making a total of eight semesters of bilingual education. A slightly smaller group (23) had four semesters of bilingual education in geography and four in history, also leading to a total of eight semesters of bilingual

education. The third largest group (15) had ten semesters of bilingual subject-specific education of which six had been in geography and four in history.

It was assumed that a higher number of semesters of bilingual education in geography and of overall bilingual education would lead to higher scores on the test. It was therefore also assumed that in generalising about the levels of learners' subject-specific task-based written discourse competence and strengths and weaknesses in their writing, this differentiation between the learners would have to be accounted for. Consequently, three different groups were built for each of the two variables: number of semesters of bilingual education in geography and number of semesters of bilingual education in history. The percentage of learners per group identified in terms of these two variables are presented in the table below:

	Group 1 1-2 semesters	***Group 2*** 3-4 semesters	***Group 3*** 5-6 semesters
Geography	44%	35%	21%
History	7,5%	57,5%	35%

Table 4.2: Percentage of learners per group identified in terms of number of semesters of bilingual education followed in geography and history respectively

The means of these six groups' scores on the holistic rating of the appropriateness of the discourse function per task were then compared. The results were rather surprising, since the means of the different groups did not show much difference. Furthermore, where there were differences, these were not systematically related to the numbers of semesters of bilingual education in geography or history. The groups with the most semesters of bilingual education in geography or history, for instance sometimes had the lowest mean. In the table below the means of the different groups' scores in terms of the appropriateness of the discourse function on each of the six tasks are compared.

Number of semesters of bilingual education	**Means of group's scores on each of the tasks**					
Geography	**Task 1**	**Task 2**	**Task 3**	**Task 6**	**Task 7**	**Task 8**
1-2 semesters	2,29	2,56	2,24	2,50	2,54	2,36
3-4 semesters	2,36	2,79	1,50	2,60	2,50	2,33
5-6 semesters	1,82	2,47	2,54	2,50	2,27	2,14
History	**Task 1**	**Task 2**	**Task 3**	**Task 6**	**Task 7**	**Task 8**
1-2 semesters	2,33	2,67	1,83	3	2,17	1,83
3-4 semesters	2,15	2,67	2	2,49	2,45	2,45
5-6 semesters	2,25	2,52	2,07	2,54	2,50	2,25

Table 4.3: Comparison between means of groups' scores on holistic rating

The results suggest that subject-specific task-based written discourse competence, as assessed here, is a construct that is neither systematically developed in bilingual subject-specific education in geography or in history. Furthermore, the fact that the differences between the various groups' means are either small, or not systematically related to the number of semesters of bilingual education in geography or history, makes it possible to generalise the findings regarding the levels of competence shown and the strengths and weaknesses in the language use across the group. To further ensure that results could be generalised across the group, analyses of variance were carried out in order to determine whether the class a learner attended or learners' gender had an effect on performance. Again no significant differences were found between the performances of boys and girls or of learners attending different schools in terms of any of the ratings done.

Further aspects of the sample's profile, such as the family situation, the number of languages spoken, the socio-economic background and attitude towards the subject geography are discussed in Vollmer's (in preparation) detailed report on the findings of the project as a whole.

In the next section the tasks used to elicit learners' written responses are described.

4.5 The elicitation tasks

4.5.1 Development and refinement of a first draft for piloting purposes

In order to guarantee a high level of authenticity and content validity, the geography test used in the project was *developed in close co-operation with leading experts on geography education*[1]. The experts supported the project team in the development of a structural model of geographical competence, designed especially for the purposes of the project and discussed in chapters 2 and 3 (e.g. Vollmer 2007, in preparation). Furthermore, they assisted in the initial design and also in the finalisation of the test and at a later stage contributed to the development of the rating scales.

In developing the test, a decision was first taken about the *theme* the test should focus on in co-operation with the experts. Basic conceptualisations for possible tasks were also made. As far as the test theme is concerned, it was important to choose a theme that was not part of the 10th grade syllabus, but that had been dealt with in one of the earlier grades in order to make a comparison of the results between learners form different schools possible. Choosing a theme from the 10th grade syllabus might have caused a situation where some of the classes involved in the research had already dealt with the theme, while others had not.

Following the advice of the experts, *the tropical rain forest* was chosen as overall theme for the test. This theme was suggested, because of its importance for the subject community and its multi-dimensional nature. It is a central theme in the subject and is dealt with in the 8th, and in some federal states, also in the 11th grade. In addition to its curricular importance, the tropical rain forest serves to exemplify a number of more abstract content or thematic categories with which the community operates. This aspect together with its significance as a unique and fragile ecological system and its global effect on climate may explain its status in the curriculum. As a concrete example of a specific identifiable space, the tropical rain forest is studied in terms of aspects such as

[1] The main experts involved were Prof. Ingrid Hemmer from the Catholic University Eichstätt-Ingolstadt and Prof. Michael Hemmer from Münster University (*Westfälische Wilhelms-Universität Münster*), who are both members of the executive committee of the German Society for Geography (*Deutsche Gesellschaft für Geographie: DGfG)*. In addition to meetings with these two experts, the project team also had the opportunity to participate in a number of meetings including two further executive members of the *DGfG*, namely Prof. Gudrun Ringel from the Freiburg University of Education and Prof. Tilman Rohde-Jüchtern from the Friedrich Schiller University Jena. At the time, these four experts were in the process of developing a competence model for geography commissioned by the Ministry of Culture.

topography, climatology and ecology, as well as in terms of social, economic and ethical matters. It thus allows the use of a variety of tasks each focusing on a different aspect of the subject's more abstract thematic patterns and involving different types of texts as input material. The use of different types of input texts again requires learners to apply different subject-specific procedures in interacting with the material.

All of the tasks require the *integrated application of the competence areas* presented and explained in chapters 2 and 3, namely knowledge of subject matter, methodological competence, subject-specific communicative competence and volitional aspects.

Based on Vollmer's (e.g. 2007, in preparation) structural model of geographical competence and on the conceptual outlines for a number of tasks, the team set out to develop a first draft of the test. The geography experts provided the team with a large selection of material on the tropical rain forest from which task formulations and input texts could be taken. In addition, the team consulted various standard textbooks and collections of material used in teaching geography in the 10th grade. In developing the first draft, care was firstly taken to design tasks that integrated the different competence areas, instead of fragmenting them. In this way, more complex problem-solving behaviour could be stimulated. Secondly, attention was paid to the formal features of the tasks, namely to the formulation of the instruction and the presentation of the instruction and input material, as well as to their level of difficulty. This was done in order to ensure that the tasks also formally resembled tasks that learners had become accustomed to within the context of geography instruction. This is an important aspect in terms of so-called response validity, since tasks that are too unfamiliar to test takers do not allow them to demonstrate their true ability (e.g. Henning 1987: 96).

After the first draft of the test had been completed, it was given to the experts for feedback. The experts were to judge each of the tasks in terms of the following three aspects: firstly their authenticity, seen as the extent to which they correspond to tasks typically used in geography training in the 10th grade. Secondly, the experts were to judge the tasks in terms of the appropriateness of their perceived level of difficulty for 10th graders and thirdly in terms of the clarity of the instructions and appropriateness of the input material. Finally, the draft test as a whole was to be judged in terms of whether the tasks in total covered the competence areas for geography, referred to above, to a

sufficient extent. At the end of a continuous process during which feedback and suggestions of the experts were applied and a number of further drafts developed, a test was finalised for the purposes of pilot testing.

4.5.2 Pilot testing

The aim of piloting a test is to determine whether the test is appropriate for a particular group and to ascertain that the items lead to responses that can be used to make inferences about the construct to be measured (e.g. Alderson et al. 1995: 73-75, Henning 1987:10-11). Some tasks on a writing test for instance might unintentionally elicit only short written responses, so that it is not possible to make inferences about aspects such as effective organisation and linking of meaning and the use of a wide range of vocabulary. By involving a group of test takers that resemble those who are to write the final version of a test, it is further possible to determine whether the instructions are clear and whether the time allowed for completing the test is sufficient.

The 17 test tasks used in the project were *piloted in 3 classes at two different grammar schools in Lower Saxony*. The one class consisted of 10th grade learners following geography in English and one of 10th grade learners following it in German. In the third class, the geography teacher responsible alternated the teaching between the two languages, but learners chose a particular language at the beginning of the 8th grade and used this language in their participation in classroom activities. The tasks were tested together with other instruments used within the context of the larger project. The learners who had been selected to work on the tasks did so in the language in which they follow geography. In addition to completing the tasks, learners were also asked to comment on the tasks in writing after they had completed all of them. Furthermore, learners were invited to verbally report to the researchers any problems experienced in understanding and processing the tasks, which were then recorded in writing by the researchers. The respective teachers were also given copies of the task booklets and, like the experts in geography education, were requested to give feedback on the tasks in terms of their authenticity, their level of difficulty, the clarity of the instructions and the appropriateness of the input material.

Based on the feedback received during the pilot testing, *a final test with seventeen tasks and standardised answer booklets* were prepared for the administration of the test in the seven schools participating in the main data elicitation phase of the project. The test gives general instructions in an introductory part, which is then followed by sets of tasks organised into eight topic categories. Appendix B gives an overview of the eight task categories, the number of tasks per category and the input material used in each category.

Great care was taken to ascertain that the German and English versions of the test were equivalent. Furthermore, careful attention was paid to the layout and presentation of each of the test tasks in the test booklet so as to present the learners with clearly organised and formatted booklets that look attractive. In their discussion on different types of test items, Weirsma and Jurs (1990: 42-89), for example, point out that in addition to the wording of the general test instructions and that of each individual test item, the comprehensibility of test items also depends on test layout. Furthermore, they argue that a professional-looking test can have a positive effect on test takers' motivation to work on the test tasks. The different test tasks were therefore all visually clearly demarcated and except for one task, none of the tasks ran over two pages. Colour was also used in all of the visual input material.

In terms of this specific study, a decision had to be taken about which tasks to use for the assessment of learners' language use. Since the analytic rating of written responses is so time-consuming, it was not possible to use learners' responses to all seventeen tasks. Based on whether they led to the production of a longer coherent and cohesive written response and on the nature of the required discourse function, *six of the seventeen tasks were selected for the purposes of this study*. These are tasks 1c, 2a, 3, 6, 7 and 8[1]. The discussion on the exact nature of the tasks involved, which follows the description of the elicitation sessions below, is therefore only concerned with these six tasks. For further discussions on the different test tasks and the purposes for which they have been used in each of the three other studies embedded in the project, see Vollmer (2006a, 2006b, 2006c, in preparation), Heine (2006, 2007) and Troschke (2006, 2006, in preparation).

[1] The original task numbers are used in referring to them, in order to enable readers to draw relations between the different studies embedded within the project.

4.5.3 Data elicitation sessions

The data was elicited in the period of May to June 2005, thus during the final phase of the 10th grade. *Different measures were taken in order to minimise possible sources of inconsistencies in the administration of the tests*. The test was only administered by members of the project team. This was done in classrooms at each of the seven participating schools during a 90-minute session embedded in normal school hours. The learners who followed geography in English always wrote in one classroom and those who followed it in German in another. The researchers had clear instructions in terms of the procedures to be followed. They were to create a friendly test atmosphere that would limit feelings of anxiety, while keeping individual interaction with the test takers to the minimum and they were to remain in the classroom throughout the full 90 minutes that were given for the completion of the test.

The same sets of test booklets, 35 in English and 35 in German, were used at all the schools. Consequently, the presentation of the tasks was also kept consistent. The test takers were instructed not to write in the test booklets and to hand in the booklets after the test. The researchers then checked each booklet to make sure that learners had not written in them and eventually replaced pages as necessary. Furthermore, as already indicated above, a standardised answer booklet was used with all groups.

Learners were not expected to write their names on tests, but were all given a four-digit code. The first digit identified the school, the second indicated whether the learner followed geography in English or in German and the last two digits represented the identity number given to a learner within the context of the specific class involved.

4.5.4 The nature of the tasks

4.5.4.1 Importance of variation between the tasks in terms of the required discourse function

Although all six tasks used for the purposes of this study involve the integrated application of subject-matter knowledge, methodological competence and subject-specific communicative competence, the tasks clearly vary in terms of the discourse function they require learners' responses to realise[1]. This is firstly the result of variation

[1] The tasks are presented in Appendix A.

between the tasks in terms of their thematic focus or the aspect of the subject-specific thematic patterns they are concerned with. The six tasks thus require learners to apply subject-matter knowledge of a range of subject-specific thematic patterns. The variation in terms of the discourse function secondly stems from differences between the tasks related to the cognitive-semiotic activities they involve. The different tasks require learners to apply a variety of subject-specific procedures in constructing meaning on the basis of the texts that constitute the input material. Resulting from the differences in the required activities, the mini-genres that have to document these activities as end products also differ from task to task.

The fact that the tasks involve the application of knowledge of subject-specific thematic patterns and procedures and that they show considerable variation in the topics, input material and consequently in the required discourse functions, is *important in terms of validity*. Firstly, the authenticity of the tasks determine whether the responses they elicit can be used as a basis for making valid inferences about learners' subject-specific writing. Such inferences can only be made if the tasks can be regarded as authentic, in the sense that they involve the application of subject-specific competences. The variation in the tasks is important, because research on the use of writing tasks has shown that aspects such as the topic, the input material and the genre involved, have clear effects on the nature of the language used and on the levels of performance demonstrated (e.g. Breland et al. 1987, Huot 1990: 240-241, Skehan 2001: 170-174, Weigle 2002: 60-69). Findings suggest that differences in variables related to the discourse mode (here referred to as the discourse function) lead to variation in syntactic maturity, as measured by T-units, lexical variety and to observable differences in the levels of performance. Consequently, in order to be able to generalise findings about learners' subject-specific writing, the assessment had to involve as many tasks as possible with as much variation as possible in terms of the thematic focus, input material, required activities and required discourse function.

The differences between the tasks in terms of their requirements will now be disccused. The descriptions of the task requirements outlined below represent a consensus reached though a *comprehensive and subject-based analysis of the tasks and random selections of learners' responses*. The required thematic focus, activities, genres and discourse functions were identified through an intensive analysis of the tasks and of learners'

responses as part of developing the scales. This process entailed close co-operation with experts in geography education and with geography teachers. In the last phase of the process, the lists of requirements for each task were finalized in a further close analysis of the tasks together with a geography and linguistic expert.

4.5.4.2 Variation between the six tasks in terms of thematic focus

The range of themes covered by the first four tasks, task 1c, 2a, 3 and 6, includes the *climate* of the tropical rain forest, the *layering* of the tropical rain forest and the *living conditions* in each layer, the extent of the *international destruction* of the tropical rainforest presented in relative and absolute terms and the *principle of sustainability*. Tasks 7 and 8 represent a shift in the theme towards an examination of the *interrelatedness between, on the one hand, humans in general and the learners in particular, and on the other, the tropical rain forest*. The topical focus of task 7 is the international significance of the tropical rain forest and the way in which everybody on earth is affected by its destruction. Task 8 is concerned with measures the learner can personally take in order to contribute to the protection of the tropical rain forest.

All of the tasks involve *applying subject-matter knowledge* in solving the particular problem presented in the task, rather than the mere recalling of memorized information[1]. Although the tasks require the application of knowledge of a wide range of thematic patterns, the knowledge involved is limited to central aspects of subject matter learners are expected to have mastered by the end of the 9th grade.

The differences in the thematic focus of the tasks has the advantage that a *wide range of subject-specific terms and expressions* related to different thematic patterns are involved. Furthermore, this range is not only limited to geography, but also includes terms and expressions shared with other subjects. Task 1c allows insight into learners' use of subject-specific terms and expressions related to climatic aspects, such as type of climate, temperature and precipitation. Task 2a involves the use of terms and expressions associated with living conditions in different environments, which are also used in the discourses of biology. Task 3 requires the use of terms and expressions describing change and development in relative and absolute terms. These terms and

[1] See Heine (2007) for a cognitive model of the task processing that explains and illustrates the ways in which the tasks stimulate two identifiable cognitive processes, namely the recalling of information from memory and the construction of new knowledge during the processing.

expressions also occur in the discourses of other subjects such as economics, political and social studies. Furthermore, in responding to task 6 and constructing meaning about the principle of sustainability, notions associated with economic, social and ecological matters are entailed. The thematic pattern revolving around the international significance of the tropical rain forest merges a number of the notions referred to so far. Climatic, ecological, social and economic consequences and the associated thematic patterns are all involved. Task 8, on the other hand, does not necessarily require the use of any specific technical vocabulary, since it focuses on the scope for action that the learners themselves have to contribute to the protection[1]. It thus represents a wider focus on activities outside the context of the subject community.

4.5.4.3 Variation between the six tasks in terms of input material

The tasks all require learners to construct meaning on the basis of one or two input texts. The variety of input texts, together with variation in the tasks' activity requirements lead to the application of *a range of subject-specific procedures*. The texts belong to *text categories commonly used in geography and in other subject-specific contexts* and involve *different semiotic forms of presentation*. These include a climate graph (task 1c), a diagram (task 2a), a bar chart (where the bars are presented as trees) with information presented in percentages and a table with information given in km^2 (task 3), a cartoon (task 7), a 5-point scale (task 8) and three types of verbal texts, including two short statements, a definition and a suggestion (tasks 2, 6 and 7).

Whereas the climate graph is typical for the subject geography, the diagram, which depicts the characteristics of the tropical rain forest as an ecosystem, presents a text type that is also used in the context of biology. In addition to geography, history and political studies also make use of cartoons. Graphs and tables are text types applied by different subject communities in visual representations of processes of change and the different types of verbal texts are also used across subjects. The scale, on the other hand, is less conventional in subject-specific contexts, but represents a text widely used in society. Service providers, such as hotels, restaurants and airlines for instance use short questionnaires in order to monitor customer satisfaction. In this respect, as in terms of the use of subject-specific vocabulary, task 8 thus prompts learners to draw a relation

[1] Due to the its thematic focus, learners' responses to task 8 was not rated in terms of the sufficient and correct use of subject-specific terms and expressions in the analytic scoring of the textual presentation of meaning.

between a phenomenon studied by the subject community and the learners' experience outside the context of the subject.

Task 7 prepares this shift in focus from a subject-specific construction of experience, to a more individualised perspective, as it connects these two approaches. This is done by stimulating the application of subject-specific categories used in interpreting experience, e.g. climatology, ecology and economics in reflecting about ways in which all people on earth are affected. This increase in the extent of personal involvement is also reflected in the choice of personal pronouns used in the two statements serving as input material. Whereas the statement in task 7 reads "The rain forest concerns *us*[1] all.", the statement in task 8 reads: "*I* can personally contribute to the protection of the tropical rain forest".

The extensive use of a wide variety of text types is typical for the subject geography and so is the application of text types also used by other specific subject communities, such biology, history, economics and political studies. This results from the fact that geography is a subject which merges a number of approaches to an interpretation of the physical and biological world (e.g. Brunotte et al. 2002, Leser 2005). Economic matters, social aspects, ecological concerns and political issues are some of the themes focused upon in addition to those that are typical for geography, such as topography and climate.

The input material also *provides learners with the opportunity to apply a number of subject-specific conventions in the language used in their responses*. Detailed information about measurable phenomena expressed in exact numerical terms can for instance be extracted from the climate graph, the diagram of the living conditions in the tropical rain forest and the bar chart and table depicting the decline of the tropical rain forest. Using these numerical representations in the response contributes to the construction of meaning that is more exact. These input texts thus provide the test takers with considerable support that can lead to more appropriate language use.

Whereas the texts in the first three tasks present the test takers with support in the form of numerically encoded meaning, the definition and suggestion in task six present assistance in linguistically encoded meaning in the form of a number of terms that signal important abstract thematic patterns. References to ecological, social and

[1] The italics are used here for emphasis. The two personal pronouns were not italicized in the test booklets.

economic goals can be used as the basis for expressing networks of relationships between these more abstract notions and concrete examples of each.

Verbalising the processes, developments and relations depicted in the diagram, bar chart and cartoon potentially involve the use of more complex semantic patterns and terms and phrases expressing different types of logic relations.

Due to the variety and authenticity of the input texts and the range of subject-specific procedures involved in interacting with these texts, e.g. analysing a climate graph, the learners' responses to these tasks provide a comprehensive basis for the analysis of their subject-specific writing.

4.5.4.4 Variation between the six tasks in terms of the required discourse function

The particular combinations of thematic focus, input material and required activities result in six tasks that *vary greatly in terms of their required discourse functions*. As explained in section 3.5.1.2 in the previous chapter, a text realises a particular discourse function through its construction of ideational meaning with a specific thematic focus in a particular genre and register. The ideational meaning and genre realised by the text result from particular cognitive-semiotic activities, e.g. analysing a climate chart, identifying a type of climate, characterising and comparing living conditions in the different layers of the tropical rain forest etc.

Task 1c requires Kisangani's climate to be identified, as well as a detailed description of the climate on the basis of the climate graph. The thematic focus is thus a *particular type of climate and aspects that determine climate classification*, such as the average temperature, the total annual precipitation, fluctuations in terms of these two aspects over twelve months and humidity. In essence, the cognitive-semiotic activities involved revolve around *an analysis of the climate chart* that should lead to an *identification of the type of climate* depicted, as well as to *a description of this climate.* The nature of the instruction, however, makes it possible to realise the meaning in two ways. Since the instruction starts with a question: "What type of climate does Kisangani have?", followed by the requirement to describe the climate, it is firstly possible to identify the type of climate and then to present a descriptive explanation. In the descriptive explanation, the manifestation of aspects such as the temperature and total annual

precipitation should be described and the relations between aspects such as humidity, the amount of precipitation and rate of evaporation explained. Alternatively, a descriptive explanation can be presented that ends with a conclusion about the type of climate depicted.

Task 2a requires a *comparative characterisation* of *the living conditions in the three layers of the tropical rain forest* on the basis of *an analysis of a detailed diagram* depicting the features of the tropical rain forest as an ecosystem and contrasting the nutrient cycle of the tropical rain forest with that of deciduous forests. In this task, the thematic focus revolves around the factors that influence living conditions for animals and plants, namely the incidence of light, the maximum temperature, air humidity and the number of animals. In addition, the response should focus on the layering of the vegetation in an upper layer or canopy (tops of the giant trees), an intermediate layer (tree top layer with closed canopy) and a lower layer (bush and shrub layer). In analysing the diagram, the manifestation of the four factors that influence the living conditions in each of the three layers should be identified and compared.

Since the diagram has basically two dimensions each with a number of aspects: the three layers on the one hand and the four factors that influence living conditions on the other, the experts on geography education advised the project team to give the learners some orientation. It was therefore decided to identify a specific aspect and dimension as a starting point for the characterisation. Consequently, the instruction states that learners should start their characterisation with the giant trees. Support was thus given in terms of the overall structuring of the response, since the instruction suggests an organisation of meaning on the basis of a distinction between the three layers.

Task 3 involves a critical reflection about the influence that a particular semiotic representation can have on the meaning constructed. It requires learners to *comment critically* on a statement and a depiction of the international destruction of the tropical rain forest presented in percentages in a bar chart. This has to be done in the light of information about the destruction of the tropical rain forest measured in km^2 and presented in a table. Whereas the table provides information about the sizes of the areas covered by rain forests in America, Asia and Africa in the past, the bar chart provides information about the remaining rain forest in percentages, but does not give any

information about the original sizes of the areas covered by rain forests. Looking at the bar chart alone, it seems as if the most extensive destruction is taking place in Asia. The information in the table, however, makes it clear that in terms of the actual size of forest area that is being destroyed, the severest destruction is in America. In order to comment critically on *the way in which different semiotic representations construct different meanings*, the representation of *the destruction in the bar chart thus needs to be compared to the depiction of the destruction in the table*. The comparison should lead to a *criticism* of the statement and the bar chart *or a suggestion to improve the depiction of the destruction* on the three continents. In addition to a concern with the effect a particular semiotic representation of experience has on the meaning that is constructed, the thematic patterns involved are related to *the decline of the tropical rain forest on the different continents*.

In task 6, test takers are required *to apply a definition of sustainability in evaluating a suggestion about how the tropical rain forest can possibly be saved*. The activities involved here thus require two short written texts to be related to one another. This should result in a response in which the suggestion is evaluated and supporting arguments for the evaluation are presented. The thematic focus is on *the acceptability of an idea to protect parts of tropical rain forests by turning them into national parks as viewed against* the background of *the principle of sustainability*. This principle is concerned with *finding a balance* between the notions *natural resources*, *ecological*, *social* and *economic goals* and *rights and needs of future generations*.

In task 7, test takers are required to explain to what extent they believe that the statement "the rain forest concerns us all" is true. In addition to the short statement, the task also includes a cartoon depicting the rain forest as a hand that holds the earth and that is being cut off. Learners thus need to *express an opinion about the statement* which needs to be *supported by an explanation* that also includes *a reference to the cartoon*. Here, the thematic focus is thus related to the *international significance of the tropical rain forest*, which entails aspects such as *global warming* and the *greenhouse effect*, the notion of the tropical rain forest as *the "green lungs"* of the earth (which is related to the *concept of photosynthesis*) and the value of the rain forest as a *unique ecosystem*, *living space*, *source of income* and *medical resource*. The task thus requires learners *to identify complex chains of interrelationships over continental borders and to*

apply different categories in interpreting experience, e.g. climatic, ecological, social, economic and ethical concerns.

All consequences of destroying the forest are, however, not portrayed explicitly in the cartoon. As already indicated above, tasks 7 and 8 both differ from the other tasks in the sense that the input material provides clearly less support in terms of the ideational meaning to be constructed. Unlike the other tasks, where a subject-based analysis of the input material could generate most of the ideational meaning to be included in the response, the cartoon in task 7 only gives an indication of the tropical rain forest as a unique ecosystem renowned for its biodiversity. Further aspects of the international significance of the tropical rain forest, such as its global effect on climate, is only vaguely suggested by the fact that the rain forest is depicted as a hand holding the earth. The task thus requires learners to construct meaning based on subject-matter knowledge about the tropical rain forest.

Task 8 presents a statement: "I can personally contribute to the protection of the tropical rain forest" and a scale with the following options: absolutely true, true, partly true and partly not true, not true and absolutely not true. Learners need to *indicate their opinion on the scale and give verbal support for their opinion*. The thematic focus is concerned with *measures that the learners can personally take in order to contribute to the protection of the tropical rain forest*.

The three aspects of the required discourse functions of these six tasks: their thematic focus, the required activities and related mini-genres are summarised in Table 4.4. Since the mini-genres are the textual realisations of the activities, these two aspects are presented in one column. The input texts that were used are also listed, because they are central to the activities and directly influence them. Part of the ideational meaning listed for task 7 is italicised, in order to indicate that learners received no direct support from the material in terms of the thematic items involved. The aspect of biodiversity can, however, be inferred from the cartoon and is consequently not presented in italics. As far as the interpretation of the cartoon is concerned, only two of the most obvious interpretations are listed. A range of interpretations is naturally possible, but is too extensive to present in the table. The ideational meaning to be constructed in response to task 8 is also italicised, because the input material provided no support in this respect.

The possible thematic items are also not further indicated, because the ideational meaning to be constructed here is not based on a standardized subject-specific thematic pattern, but is concerned with learners' own interpretation of experience.

Task	**Types of semiotic representations presented as input texts**	**Aspects determining the required discourse function**	
		Required cognitive-semiotic activities and mini-genres to be realised[1]	**Required thematic focus of ideational meaning to be constructed**
1c	*Climate graph* for the Kisangani weather station	Identification / Conclusion	Type of climate: tropical
		Descriptive explanation	Specific manifestation of defining features of climate: temperature, precipitation, humidity, evaporation
2a	*Diagram* depicting the living conditions in the tropical rain forest and contrasting its nutrient cycle with that of deciduous forests	Comparative characterisation	Three layers of the tropical rain forest
			Factors that affect living conditions: incidence of light, maximum temperature, air humidity and number of animals
			Effects that manifestation of these factors have on living conditions for plants and animals
3	*Statement* that extent of destruction of tropical rain forest is highest in Asia *Bar chart* depicting the international destruction of the tropical rain forest in relative terms *Table* depicting the destruction in absolute terms as size of and change in forest area in km^2	Critical comment / Suggestion	Depiction of extent of international destruction of tropical rain forest
		Comment	Effect of type of representation on the meaning constructed about the extent of the destruction of the tropical rain forest on each continent
		Comparison	Depiction of the international destruction of the tropical rain forest in the bar chart and table Extent of the decline of the tropical rain forest in America, Asia and Africa
6	*Definition* of sustainability *Suggestion* to protect tropical rain forests by turning them into national parks	Critical comment	Acceptability of suggestion
		Assessment of suggestion	National park: exclusive concern with ecological goals Principle of sustainability: use of natural resources in which there is a balance between ecological, social and economic goals, rights and needs of future generations
7	*Cartoon* depicting the tropical rain forest as a hand holding the earth that is being cut off *Statement*: "The rain forest concerns us all."	Supporting comment	Extent to which statement is true
		Explanation	International significance of the tropical rain forest: *globally affects climate, destruction contributes to greenhouse effect* *"lungs of the earth", process of photosynthesis* *medical resource* *living space* *resource for developing countries* biodiversity*
		Interpretation (of the cartoon)	E.g. Tropical rain forest important for life on earth Value of tropical rain forest in terms of biodiversity
8	Statement: "I can personally contribute to the protection of the tropical rain forest" 5-point scale	Explanation (in support of opinion indicated on scale)	*Personal scope for action in terms of protection of the tropical rain forest*

Table 4.4: Summary of input material and aspects constituting the required discourse function

[1] The order in which the mini-genres are presented in the table is not prescriptive in terms of the order in which they must be realised in the text.

The 84 learners' responses to these six tasks that clearly show variation in terms of their required discourse functions ensures a broad basis for making inferences about bilingual geography learners' subject-specific writing. In the next section, the scales used to rate the performances are described in detail. After a general overview, the three types of scales are described and aspects of their development are discussed in detail.

4.6 Rating scales

4.6.1 General overview

4.6.1.1 Requirements the rating scales had to meet

Since the scales were used in the wider project, they had to be applicable to the rating of responses written in *English* and in *German*. This proved to be no problem, since the performances of learners' following bilingual subject-specific education showed no features that were not also present in the texts written in German. Furthermore, the texts written in these two languages showed the same degree of variation in terms of the levels of performance found. The only possible difference may be in the numbers of learners whose performances lie on a specific level. It would for instance be reasonable to expect that a larger number of performances produced by the bilingual learners lies on the lower levels of the language scales. However, this still needs to be confirmed through statistical analyses of the scores given to the performances written in German. These analyses are currently being undertaken by Vollmer and will be reported about in his (Vollmer: in preparation) forthcoming publication concerned with a comparison between the competence levels of learners following monolingual instruction as opposed to learners attending bilingual tracks.

A second requirement the scales had to meet was that they had to be based on categories and criteria that reflect appropriate *subject-specific* writing. Thus, they had to allow the scoring of performances in terms of the extent to which they adhere to the conventions followed by the subject community. This was achieved by doing social semiotic and systemic functional linguistic analyses of the contexts in which the learners' texts are embedded and the requirements in terms of the performances arising from these. These analyses of the subject-specific tasks, test-taking situation and broader school context

also entailed working closely with experts in geography education and geography teachers in order to verify the relevance of the categories and criteria included in the scales to members of the subject community.

Furthermore, the analytic language scales had to allow the identification of *specified features* of learners' writing that can be distinguished from one another in the written texts, so that the feedback given about the strengths and weaknesses in learners' writing could be as precise as possible. By giving exact indications in terms of the problematic aspects, these can by addressed in a more focused way in the subject-specific classroom. Thus, instead of having to operate with diffuse categories, such as "the overall structure" and the "the overall style", the aim was to break down these categories into smaller sub-categories that constitute these. This led to distinctions between categories such as the ordering of structural units in the text and the organisation of meaning into these structural units, or between the use of subject-specific terminology and the use of more general academic vocabulary.

4.6.1.2 Outline of the three types of rating scales used

As explained before, three types of rating scales were developed. The first category of scales consists of *six holistic scales* were used to rate the degree of task fulfilment, operationalised as the appropriateness of the discourse function of a response. The second category is a set of *seven analytic scales* that were developed in order to identify general strengths and weaknesses in learners' language use. The third category of scales consists of *six sets of dichotomous and three-point scales* that were used in assessing the ideational meaning constructed in learners' responses. All scales are presented in Appendix C.

The first two of these three types of scales are commonly used in writing assessment (e.g. Brown / Bailey 1984: 22-26, Weigle 2002: 109-121, Huot 1990: 238-239, Shohamy *et al.* 1992: 27-28). In *analytic scoring*, a written sample or performance is rated in terms of several aspects, for instance content, structure, cohesion, vocabulary, grammar and punctuation. Analytic rating can either be done in order to generate diagnostic information about the strengths and weaknesses in test takers' written performances, as is the case in this study, or the scores given in terms of each of the aspects can be tallied, to reach an overall score. In such cases different components of

the performances are often weighted, so that content and structure are for instance given more importance than punctuation.

With *holistic scoring*, a performance is given a single score, based on the overall impression that the performance makes on the rater. In most cases where holistic scoring is done, scoring guidelines usually outline the general characteristics of a performance on a specific level.

Both forms of scoring have a number of *advantages*, but also show some *disadvantages* (e.g. Weigle 2002: 109-121). The first advantage of using holistic scales is that they allow raters to score texts faster than analytic scales, because each text only needs to be read once and be assigned one score. When using analytic scales, on the other hand, a rater needs to read a performance several times, each time focusing on a different aspect of the text. Secondly, holistic scoring is also thought to focus readers' attention on what the writers do well, rather than to draw attention to what they do less well, which some test theorists believe to be a problem of analytic scales. Finally, holistic scoring is thought to be closer to readers' natural ways of interacting with and responding to texts as whole units.

However, holistic scoring does not give diagnostic information about writing performances, because only one single score is awarded. Furthermore, with holistic scales it is sometimes difficult to formulate guidelines that describe the characteristics of performances on different levels, because these characteristics may vary in dissimilar ways. To illustrate, if the characteristics outlined include aspects such as content, structure and language use, a performance may show good content and structure, but there may be problems in terms of grammar and vocabulary. It then becomes difficult to formulate guidelines that characterize this type of performance. Research also suggests that analytic scales lead to more reliable scoring, because multiple scores increase reliability just as additional items increase the reliability of a test.

The choice of whether to use analytic scoring or holistic scoring in a specific assessment situation, should in the end depend on the purpose of the assessment in question and on the type of information needed from the assessment, as well as on the resources available.

Due to the research interests of this study, it was decided to make use of both forms of rating. Analytic scoring was needed to gain insight into particular strengths and weaknesses of learners' writing and holistic scoring to rate the overall impression a response makes.

4.6.1.3 Challenges involved in developing the scales

The two biggest challenges in terms of developing *the analytic scales* for assessing learners' text structure and language use was firstly *to identify useful analytic categories* that allow a rater to make decisions about clearly distinguishable aspects of the textual realisation of meaning in a response. The second challenge was to identify criteria that allow a comprehensive assessment of the language use in a text and that reflect criteria used by members of the subject community. The co-operation with teachers in designing the scales showed that aspects such as correctness of grammar and vocabulary received very little if any attention. Decisions had to be made about whether the scales used in the context of this study had to include or exclude the criteria less valued by teachers.

The biggest challenge in terms of the development of *the holistic scales* was *to design scales that could cover and categorise the full range of performances*. The learners' responses namely showed considerable variation in terms of all three types of meanings constructed (ideational, interpersonal and textual). It was thus particularly challenging to identify a manageable number of scale levels that simultaneously made provision for the immense amount of differences between learners' responses. Aspects such as the comprehensiveness and correctness of the ideational meaning constructed on the one hand and the appropriateness of the mini-genre realised on the other, manifested on different levels in individual learners' texts. In one text, for instance, the ideational meaning constructed might be complete and correct, but inappropriate mini-genres might be realised. In another text, the appropriate mini-genres might be realised, but the ideational meaning might be incomplete and partly incorrect.

This variation shown in learners' responses is *typical for performance assessment*. As pointed out before, it is not possible to standardise responses to performance tasks, due to the number of task features that can affect performance in different ways. Designing

scales for the rating of performance assessments is thus always highly time-consuming. The advantage of performance assessment, on the other hand, is that it allows the assessment of higher order competences and provides insight into how learners deal with more complex problems.

The greatest challenge in designing *the dichotomous and three-point scales* used to rate the correctness and completeness of the ideational meaning was *identifying the required thematic items and semantic relations*. Each task required the construction of ideational meaning based on a thematic pattern used by the subject community. As pointed out in the previous chapter, these models of subject matter consist of collections of concepts and the logical relations between them. In order for an answer to be correct and comprehensive from a subject-specific point of view, certain thematic items thus had to be included and particular semantic relations had to be verbalised. Although these thematic patterns or models of subject matter are standardised, they are interpreted in different ways by members of the subject community. This required quite a number of meetings with members of the subject community in order to reach a consensus about each task's content requirements.

The way in which these challenges were dealt with is discussed in sections 4.6.2 to 4.6.4 describing the scale categories in detail and explaining the processes involved in their development.

4.6.1.4 General overview of development processes

The scales were all designed in an extensive development process based on an analysis and application of subject-specific conventions. The whole process involved *close co-operation with experts in geography education, geography teachers and graduates*, in order to ensure that the scales indeed reflected the expectations of members of the subject community. The development process firstly entailed creating *frameworks of expectations* that outline the requirements in terms of the meaning to be constructed in response to each task and the textual realisation of this meaning. *Preliminary scales* were then developed on the basis of these frameworks of expectations. This involved identifying criteria that could be used to differentiate between different levels of performance, e.g. completeness of content, correctness of content, extent and

correctness of the use of subject-specific terminology. The criteria formed the basis for the formulation of level descriptors.

The *preliminary scales were then applied to random selections of learners' performances* in order to refine them. This is a necessary step in scale development in performance assessment, since learners' performances on complex tasks show such variation (e.g. Gipps 1994: 91). Preliminary or concept scales therefore need to be developed and tested on actual performances before final rating can start, in order to ensure that the scales cover this variation. Especially the holistic scales used to rate the appropriateness of the discourse function and the dichotomous and three-point scales used to rate the ideational meaning constructed involved an extensive process of testing and refinement.

Adapting the holistic scales predominantly entailed adding criteria to the original level descriptors in order to account for variation in the appropriateness of the performances. In the case of the dichotomous and three-point scales used to rate ideational meaning, the modifications firstly involved adding new dichotomous scales. This was done in order to ensure that all possible and relevant thematic items and semantic relations that could be realised were indeed included in the scales. The changes secondly involved adding one more option, usually the option "partly", to a dichotomous scale, thus reflecting that a particular aspect of a thematic pattern was realised, but not in sufficient detail or not in an entirely correct way. In the scales designed for task 1c on the climate of the tropical rain forest, for instance, learners were awarded a full mark if they wrote that the climate is humid and half a mark if they described it as hot and damp or hot and wet.

The scales and the processes followed in developing them are now explained in detail. The structure of the discussion follows the order in which the different types of scales were developed and starts with the analytic scales before paying attention to the dichotomous and three-point scales and concluding with the holistic scales.

4.6.2 Analytic scales: Appropriateness of the textual realisation of meaning

4.6.2.1 Description of scales

These scales were *applied for diagnostic purposes* and *focus on the appropriate realisation of aspects of the subject-specific register*. Seven scales with five levels each were developed. The scales range from 4 (the highest score) to 0 (the lowest score). The scale presented as example below was used to rate sufficient and correct use of subject-specific terms and expressions.

Scale 2.1: Sufficient and correct use of subject-specific terms and expressions

Descriptor	**Score**
Subject-specific terms and expressions are consistently used where references are made to subject-related phenomena and are continuously used correctly.	4
Subject-specific terms and expressions are used to a large extent where references are made to subject-related phenomena and are used correctly in most instances. Only a limited number of subject-specific terms are lacking and/or a limited number of subject-specific terms and expressions are not used correctly. Incorrect terms might be used and/or collocations and spelling might be incorrect.	3
Subject-specific terms and expressions are used, but not to a sufficient extent and/or frequent errors are made in the application of subject-specific terms. About half of the references to subject-related phenomena are either made using terms and expressions that are not part of the subject-specific register and/or show errors regarding the terms used, collocations and spelling.	2
The use of subject-specific terms and expressions is limited and/or only a limited number of terms and expressions included are used correctly. In most instances, there are errors regarding the terms used, collocations and spelling.	1
No subject-specific terms or expressions are used.	0

These analytic scales are *divided into two categories*. The first category is concerned with *the effective organisation of meaning in a text* and includes three scales: the effective ordering of structural units in a text (scale 1.1), the effective organisation of meaning into different structural units (scale1.2) and the effective linking of structural units (scale 1.3).

The term "structural units" refers to sentence parts, sentences and paragraphs, thus structures conventionally used to organise meaning in written texts. *The effective ordering of structural units* is concerned with the way in which these structural units follow each other in a text. For example in a response to task 2a, where the living conditions of the tropical rain forest are to be characterised, one would expect paragraphs dealing with each of the layers to follow one another in a specific order, either realising meaning that describes the layers from top to bottom, or from bottom to

top. This scale also focuses on whether the same ideational meanings are unnecessarily repeated. Remaining with the example of characterising the living conditions in the tropical rain forest, *the effective organisation of meaning into different structural units* is illustrated by organising the meaning about the three layers into three different paragraphs. *The linking of structural units* is related to the use of cohesive devices. All three of these scales thus measure knowledge of discourse-semantics.

The four remaining scales measure *knowledge of the lexico-grammar*. These scales are: sufficient and correct use of subject-specific terms and expressions (scale 2.1), sufficient use of formal language and clear, succinct formulations (scale 2.2), sufficient use of general academic vocabulary and structures realising semantic relations (scale 2.3) and finally, correctness of grammar, vocabulary and punctuation (scale 2.4).

In applying scale 2.1, focusing on *the use of subject-specific terms and expressions*, possible terms and expressions that could be used in response to a specific task were listed, as well as terms and expressions that are incorrect. The list of terms and expressions that could be used for task 1c for instance includes expressions such as "Kisangani has a tropical climate", "the temperature remains constant at 25°C", " the precipitation/rainfall increases/decreases". Terms and expressions listed as inappropriate are for instance a reference to the climate as "hot and wet", or references to the seasons in describing the diurnal climate of the tropical rain forest.

Sufficient use of formal language and clear and succinct formulations are closely related and were therefore combined into one scale (scale 2.2). On the other hand, although the use of subject-specific terms and expressions makes language use more formal and formulations more succinct, these two aspects were put into different scales (scales 2.1 and 2.2). This was done in order to allow a rater to focus exclusively on the use of the subject-specific terms and expressions. In applying scale 2.2, the use of precise references to measurements also received attention. These include for instance references, presented in numerical terms, to the exact amount of total annual precipitation at task 1c, the height of the trees at task 2a and the size of rain forest destroyed presented in percentages and km^2 at task 3. References to specific points in time were also covered by these scales, e.g. an indication of the months in describing

fluctuation in the annual precipitation at task 1c and referring to the years 1990 and 2000 at task 3.

Scale 2.3 (*sufficient use of general academic vocabulary and structures realising semantic relations*) focuses on the use of different lexico-grammatical resources involved in indicating logic relations, e.g. "consequently", "as a result of", "in comparison to" and complex sentences in which related meanings are combined. Furthermore, the use of terms and expressions that are applied across different subject-specific contexts and that are typical for subject-specific discourses was rated under this point.

In rating the *accuracy of grammar, vocabulary and punctuation* (scale 2.4) a distinction was made between minor errors, such as incorrect spelling and punctuation and major errors, e.g. incorrect word order, concord and tense errors. The scales allowed a performance to be placed on a specific level, depending on the extent to which minor and major errors were made.

4.6.2.2 Process involved in developing the scales

Scale descriptors used in different writing assessments (Weigle, 2002, Bachman and Palmer, 1996) were first analysed, in order to *identify scale categories and criteria*. These categories and criteria were then interpreted against the background of the framework of expectations I had developed on the basis of a social view of task-based subject-specific writing. I then *designed preliminary scales and applied them* to bilingual learners' responses to 16 of the 17 test tasks. Five randomly selected responses to each task were rated for this purpose. One of the test tasks was not included, because it did not require an extended constructed response, but entailed completing a flow chart. On the basis of this first application of preliminary scales to learners' responses, I *selected six of the 17 tasks* for the purposes of this study. These tasks were selected because they elicited responses that enabled decisions about all aspects of learners' language use that were of interest to this study. These six tasks all involved more complex thematic patterns and led to longer responses. They thus provided the opportunity to rate the use of structuring and linking devices and of subject-specific terminology.

The preliminary scales were again applied to learners' texts during a *meeting held with four grammar school geography teachers*. This time the scales were also applied to texts produced in German. The teachers worked at three schools involved in the research project. Three teachers were from schools which participated in the main data elicitation phase and one from a school where piloting was done. By the time the data was collected, two of the teachers taught the bilingual track and two taught geography in German.

The aim of the meeting was firstly *to gain insight into the criteria the teachers apply* when they assess task-based subject-specific writing. The second aim was *to get feedback on the clarity of the scale descriptors and the manageability of the levels.* In this regard, the scales did not only have to be applied to responses produced in English to the six tasks used for the purposes of this study. Since other studies conducted within the context of the project also involved assessing these learners' writing competence and needed the scales, they had to be tested on all sixteen tasks leading to the constructed responses and on texts produced by learners who followed geography in German.

In a first rating session, each teacher was given a set of five responses to task 2a and a set of five responses to task 6. The responses were pre-selected, in order to assure a sufficient variety in the levels of performance. This was done by applying the preliminary scales to 20 learners' responses to these two tasks. Task 2a and task 6 were selected, because they elicited the longest responses from learners. The assumption was that more extensive performances would lead to more opportunities to identify the criteria the teachers used.

The teachers were given copies of responses produced in the languages in which they taught geography. The two teachers involved in bilingual geography education received copies of the same responses written in English and the two teaching in German received copies of the same responses written in German. The teachers were asked to score each text using the range of marks applied in German schools, because this is the scoring system with which they are familiar. They were also asked to list the strengths and weaknesses of each response in writing. Furthermore, they had to put the five responses to each of the two tasks in rank order from the best to the weakest performance. This was followed by a short discussion round during which the teachers

gave their general impressions of positive and negative aspects of the five responses. To conclude this first session, teachers were asked to list in writing the criteria they use when assessing learners' texts.

There was a *high level of consensus between all four of these teachers in terms of the rating criteria they apply*. These criteria can be summarised as follows: central to their assessments of learners' task-based subject-specific writing is, as can be expected, the extent to which the ideational meaning constructed and the mini-genres realised, meet the task requirements. Thus, the degree of overall task fulfilment was the primary concern for these four teachers. In terms of the mini-genres, the teachers pointed out that learners are trained how to respond to different instructional verbs. This, as the guidelines given on the production of different texts that were found in geography textbooks, illustrate the way in which the subject community standardises its meaning-making processes.

In terms of a focus on distinguishable components of learners' writing, the ideational meaning, the organisation and linking of meaning and the use of subject-specific terminology were seen to be central. Clarity and succinctness of expression and the correct linguistic realisation of logic relations were also singled out as important aspects in learners' subject-specific writing. On the other hand, it was indicated that the use of formal language and the use of correct vocabulary and grammar were less central.

In a second session, the teachers were requested to rate language use by applying the preliminary scales to all 16 tasks leading to constructed responses. After the scales had been briefly introduced to the teachers, they each set out to rate the responses of 16 out of the 17 tasks in a copy of a learner's task booklet. Again, the two teachers teaching geography in English each received a copy of the same test written in English and the two teaching in German received a copy of the same test written in German. The tests were randomly selected from the sample. The teachers had to write their scores into a scoring grid which was collected in order to determine the extent to which the teachers' scores corresponded. After they had completed the scoring, they were asked to verbally comment on the scales and to make suggestions for improving them in terms of their feasibility.

The teachers' feedback on the scales and input about the criteria they apply were used to *further refine the scales*. Although the teachers indicated that they did not pay a lot of attention to *formal language and correct vocabulary and grammar*, these aspects were nevertheless *kept as scale categories*, since they are features of the subject-specific register and of the register used in other formal educational and in professional contexts. They thus represent aspects of bilingually trained learners' language use that should receive attention, if the aim of this form of education is to prepare learners for effective participation in the discourses of educational and professional communities. Furthermore, the aim of the analytic rating was to generate diagnostic information and scores were not aggregated into a single score interpreted as a reflection of learners' overall writing proficiency. The use of formal language and correct grammar and vocabulary were therefore kept as categories, since they contribute to a richer description of learners' language use. Dropping scales concerned with the use of formal language and correct grammar and vocabulary would have meant losing potentially interesting information about strengths and weaknesses of learners' task-based subject-specific writing.

It is believed that *research is needed that systematically and representatively collects data about the categories and criteria bilingual teachers use when rating learners' performances* and about how these relate to their subjective theories about the aims of bilingual education. As Vollmer (2002c: 102) points out, very little is known about how assessments are carried out in the context of bilingual subject-specific education. More insight is needed, however, because the categories and criteria that teachers use reflect their views of what it is that bilingual subject-specific education should achieve. This again affects learning outcomes. The teacher's behaviour influences learners' understanding of what is valued in the context of the classroom and learners adapt their behaviour accordingly, for instance by focusing their attention on those aspects they know are assessed. Consequently, the norms and values the teachers operate with and the resulting rating criteria they apply should be made explicit as part of the ongoing attempt to define this form of education's profile and aims.

The *last adaptations to the scales* were made when they were being tested on a randomly selected sample of 20 learners' responses to the six tasks used in this study. 10 of the learners' responses were written in English and 10 in German. In this final phase,

the scales were applied by myself and two graduates: one in English and another in English and geography.

The process followed in developing these scales is summarised in the table 4.5.

Version of scales	**Development basis**	**Number an nature of samples involved**
Preliminary I	Analysing scales for writing assessment published in literature	-
	Developing framework of expectations in terms of task-based subject-specific writing	-
Preliminary II	Applying scales I to learners' responses	- Random selection of learners' responses - 5 alternating learners' English responses to 16 tasks
Preliminary III	Implementing 4 teachers' feedback on the rating criteria they use in their assessments	- Selected samples of 5 German and 5 English responses to tasks 2a and 6 - Selected on basis of application of preliminary scales II
	- Implementing teachers' feedback on preliminary scales II - Feedback based on application of scales to learners' responses	- 2 randomly selected tests - 1 English test and 1 German test - Responses to 16 tasks
Final	- Applying preliminary scales III to learners' responses - Done in co-operation with 2 graduates (English and geography)	- Random selection of 20 learners' responses to six tasks - 10 English and 10 German responses

Table 4.5 Summary of process involved in the development of the analytic scales used to rate the textual presentation of the meaning constructed

In explicating the criteria they apply when assessing learners' texts, the teachers also gave valuable input that was used in the development of the dichotomous and three-point scales used to rate the ideational meaning constructed. These scales and the process involved in designing them are now discussed.

4.6.3 Dichotomous and three-point scales: Completeness and correctness of meaning constructed

4.6.3.1 Description of scales

The six sets of dichotomous and three-point scales and the holistic scales are similar in two ways. They are firstly *specified per task*, unlike the analytic scales that were designed to be applied across all tasks. This was done because of the second feature these scales have in common. They focus on the aspects which are unique to a required response to a specific task, namely *the meaning that should be constructed and the mini-genres that should be realised*. A relatively high correlation was also found between the means of the scores received on these two ratings over all six tasks (r=.853).

The differences between the sets of dichotomous and three-point scales on the one hand and the holistic scales on the other are firstly that the sets of dichotomous and three-point scales *do not draw attention to the structuring and linking of a text or to the register used*. As will be explained later, however, the dichotomous and three-point scales do take the use of subject-specific terms into consideration in cases where a specific phenomenon has to be identified or where the terms are provided in the input material.

The second difference between these two different types of scales is that the dichotomous and three-point scales *allow a closer analysis of the correctness and completeness of the meaning constructed in a response and simplify the judging process*. This is achieved by the fact that these sets of scales break down the meaning to be constructed into single propositions formulated as questions. As the following extract from the collection of scales used to rate the ideational meaning constructed in response to task 1c illustrates, the questions then require a rater to choose between two or three options: yes/no, or yes/partly/no.

2	***CORRECT IDENTIFICATION of the type of climate***			
2.1	Is Kisangani's climate identified as a *tropical climate*?	Yes 0,5		No 0
2.2	Is Kisangani's climate identified as a *type of tropical climate*?[1]	Yes 0,5[2]		No 0
2.3	Is it indicated that it is humid?	Yes 1	Partly[3] 0,5	No 0
2.4	Is it indicated that the precipitation is higher than the evaporation?	Yes 1		No 0

Consequently, the collections of scales guide a rater through a learner's text in a step by step way, with each scale focusing the rater's attention on a particular aspect of the meaning constructed and requiring a simple decision to be made about whether a specific proposition is realised or not. The scales thus provide a means of dealing more reliably with the significant variation in the ideational meaning constructed in response to performance test tasks.

The yes/partly/no-option serves as a means to award marks where a particular aspect of the meaning required to be constructed is realised, but in a way that is not fully acceptable in a subject-specific context. This applies to the following instances, firstly where a task requires a specific phenomenon to be identified and described, as is the case with tasks 1c and 2a and *the correct subject-specific terms are not used*. As the extract from the set of scales above illustrates, responses to task 1c had to include the terms "tropical climate" and "humid" in order to be awarded full marks on the scales dealing with these aspects of the thematic pattern. At task 2a, responses were also required to include the correct subject-specific terms in referring to the layers of the tropical rain forest, because the terms are provided in the input material. In this instance the use of the correct terms was thus taken to indicate methodological competence in interacting with the input material.

Learners' responses could also score half of the possible marks on a specific scale if *certain meanings were implied, but not explicated*. To illustrate, if a response to task 7 did not explicitly include references to the cartoon in the argument, but referred to the destruction of the tropical rain forest, half of the possible marks were awarded on one of the scales.

[1] For instance if the term "tropical climate" is not directly used, but it is indicated that the climate is similar to the one in the tropical rain forest.
[2] Award this mark if the term *tropical climate* is used. In other words, if a mark was awarded at 1.1, a mark should automatically be awarded for 1.2.
[3] Award 0,5 if the climate is described as hot and damp/wet.

These first two examples of instances where learners' responses were awarded half of the possible marks on a scale illustrate the interrelation between content and language highlighted by social semiotic theorists and systemic functional linguists discussed in section 3.5.1.2. In developing the dichotomous and three-point scales and applying them to learners' responses it became clear that specific aspects traditionally related to the language rather than the content domain can simply not be excluded in rating the content or in social semiotic terms, the meaning to be constructed. The use of subject-specific terms and explicit and precise formulations naturally directly affect the meaning constructed in a text. This again points to the importance of paying focused attention to language aspects in the subject-specific classroom. Subject-specific content does not exist separately from the social semiotic systems through which it is constructed.

Learners' responses were also not awarded full marks *if the meaning constructed was not comprehensive or detailed enough*. At task 2a for instance, learners' responses only scored half of the possible marks on one of the scales when the influences of the prevailing living conditions on plants were only discussed in terms of one of the layers. In order to score full marks on this particular scale, the responses had to refer to the effects on at least two of the three layers. Depending on how detailed the arguments were, learners' responses to tasks 7 and 8 were for instance also given full or half marks on the scales focusing on this aspect of the meaning to be constructed.

The last example of instances where half of the possible marks were awarded is the case where *broader categories are exemplified*, *but the examples are not directly relevant to the task*. This had to be applied in the case of task 6. The principle of sustainability focused upon in this task includes economic, ecological and social goals. Where responses included examples of these categories, but the example itself is not in alignment with the principle, half of the marks were awarded on the scale dealing with this aspect. In referring to economic aspects for example, some learners wrote that it is expensive to establish and manage a national park. This is of course an example of economic considerations. However, economic considerations in the sense of the

principle of sustainability refer to the concerns of groups that lose their basis of income, because they are not allowed to exploit natural resources anymore.

The extract from the set of scales above also illustrates *the weighting of scores* on different scales. As will be explained in the next section, the weighting was done in co-operation with experts on geography education and geography teachers. It reflects the perceived level of difficulty involved in realising the specific meaning the scale deals with and its centrality to the thematic pattern involved. Two is the highest mark that a scale makes provision for and is for instance awarded for a detailed explanation of one interrelationship between humans and the tropical rain forest in responding to task 7. Where a scale makes provision for a yes-/partly-/no-distinction, the partial inclusion of an element is either awarded a 0.5, if the full inclusion is awarded a 1, or the partial inclusion is awarded a 1, if 2 is the highest mark possible.

The scales divide into *two main categories*: "Meaning to be constructed" and "Meaning that can additionally be constructed". The propositions listed under the latter category are not required by the model answer or the framework of expectations, but are related to the thematic pattern the task deals with.

Learners' *scores on these scales were tallied to reach a total per task*. The sum of all the marks that can be awarded under “meaning to be constructed” was taken as a task's total. In order to calculate a learner’s mark for a specific task, the marks that had been awarded under “meaning to be constructed” and “meaning that can additionally be constructed” were added. It is thus theoretically possible that a learner’s mark for a specific task is higher than the task’s total. In practice, however, this never was the case. Rather, the marks that had been awarded to a learner under “meaning that can additionally be constructed” served to compensate for marks not gained under the first category.

Depending on the nature of the task, each of the six collections of scales *operationalises a particular combination of the following criteria*:

Completeness of meaning constructed

- completeness of the meaning constructed in terms of the main aspects to be included (all six tasks),
- completeness of the meaning constructed in terms of details, including information about measurable phenomena presented in the input material (task 1c, task 2a, task 3),
- completeness of the criteria extracted from the material as support for view presented (task 6).
- number and completeness of explanations and arguments presented as support for view taken (task 7, task 8).

Correctness and appropriateness of meaning constructed

- correctness of information included (task 1c, task 2a),
- correctness of view presented (task 3, task 6),
- appropriateness of view presented (task 7, task 8),
- quality of explanations and arguments (measured in terms of extent to which they are reflective) presented as support for view taken (task 7, task 8).

4.6.3.2 Process involved in developing the scales

The project team *developed model answers to all 17 tasks* and refined these during meetings with one of the experts in geography education. This involved analysing the tasks and creating outlines of the ideational meaning realised by a subject-based processing of the tasks. A subject-based processing of a task entails correctly applying knowledge of the standardised subject-based thematic pattern involved and following subject-based procedures in constructing meaning on the basis of the instruction and the input material. In developing the outlines, the project team naturally also took the level of knowledge required of 10^{th} grade learners into consideration. The model answers thus did not include references to subject matter not dealt with in school before the 10^{th} grade.

The *model answers were then broken down into propositions*, thus into the basic semantic relations between the thematic items involved in the thematic pattern. Four of the propositions identified in the model answer to task 1c on the climate of the tropical rain forest are for instance:

- It is a tropical climate.
- It is humid.

- The precipitation is higher than the evaporation.
- The fact that the precipitation is higher than the evaporation leads to humidity.

Scores were then assigned to each of the propositions. As already pointed out in the previous section, the scores reflect expert opinion on the level of difficulty involved in realising a specific proposition, as well as the relative importance of the proposition within the overall thematic pattern.

The *model answers were then compared to a random selection of 20 bilingual learners' answers* to the 17 tasks. Questions that arose about the acceptability of particular meanings realised in learners' texts were dealt with in discussions with the expert on geography education.

One of the results of comparing the model answers to learners' texts was that *propositions were identified that went beyond the requirements set in terms of the meaning to be constructed*. Consequently, the category "Meaning that can additionally be constructed" was created in the scales. As briefly explained before, this category includes all thematic items and semantic relations that are not necessarily required for a response to a task to be regarded adequate from a subject-based point of view, but that are factually correct and related to the thematic pattern the task deals with. An example from the model answer to task 1c is a reference to the fact that the climate of the tropical rain forest is diurnal and not seasonal.

Following the workshop with the four teachers referred to in the discussion on the development of the analytic scales, the model answers were further developed, predominantly by expanding on two aspects. As a result of the teachers' feedback, *propositions were firstly added to the scales so that they reflect more comprehensively the information that had to be extracted from the input material and included in the answer in order to make it more precise*. For example, in the model answer to task 1c, an exact indication of the temperature in degrees Celsius and total annual precipitation in millimetres was included and details about the exact height of each of the layers in the vegetation in the tropical rain forest were included in the model answer to task 2a. The model answers were thus elaborated to reflect meaning resulting from a subject-based interaction with the input material, which is related to procedural competence. Furthermore, *references to the mini-genres to be realised were added to the scales*. For

example, it was explicated that the living conditions in the three layers must be compared, or that there must be an opinion expressed and support provided at tasks 3 to 8.

The *model answers were then changed into simple dichotomous and three-point scales and applied to 20 bilingual learners' responses* to the six tasks. Questions about the acceptability of the meanings constructed in the responses that arose from this second application of the scales were treated in a *second meeting with teachers*. Two teachers teaching geography in English and German respectively participated in the meeting. One taught at a school that was involved in the main data elicitation phase of the project and one at a school where piloting was done. In addition to responding to concerns I voiced about problems experienced in rating the 20 learners' texts, the teachers were also given copies of learners' responses to the tasks and asked to rate these, using the preliminary scales. Since the focus was not on the language use, both teachers received three copies of learners' responses to the six tasks written in German and three written in English. The responses were randomly selected. In applying the preliminary scales to these responses, the teachers evaluated the scales in terms of whether:

1. all required thematic items and semantic relations were covered by the scales,
2. the scales reflected knowledge of thematic patterns that can be expected from 10th grade geography learners,
3. the mark allocation per scale was acceptable and
4. all the scales were clearly formulated and easy to apply.

Furthermore, in rating the learners' texts, further propositions that could be added to the category "Meaning that can additionally be constructed" were also identified.

In a final phase, the *scales were applied to another random selection of 20 learners' responses* (10 produced in English and 10 written in German) by myself, as well as one English and one geography graduate. This was done in order to further improve the clarity of the formulations used in the scales and to ensure that the scales captured the range of ideational meanings constructed in learners' responses.

The process followed in developing these scales is summarised in Table 4.6.

Version of model answers (MA) / scales (S)	**Development basis**	**Number and nature of samples involved**
MA I	- Developing model answers in co-operation with expert on geography education - Entailed analysing requirements of 17 tasks and identifying subject-matter knowledge required of 10th grade learners	-
MA II	- Comparing model answers I and learners' responses	20 randomly selected English tests
	- Addressing problems identified in comparing model answers to learners' texts - Done in co-operation with expert in geography education	
MA III	Elaborating model answers II by implementing 4 teachers' feedback on rating criteria they use in their assessments	- Selected samples of 5 German and 5 English responses to tasks 2a and 6 - Selection on basis of application of preliminary language scales II
S I	Changing model answers III to 6 of 17 tasks into scales	-
	- Applying scales to learners' responses	20 randomly selected English responses to 6 tasks
S II	- Addressing problems identified in applying scales I to learners' responses - Done in co-operation with 2 geography teachers	-
	- Implementing teachers' feedback on scales I - Feedback based on application of scales to learners' responses	- Randomly selected responses to 6 tasks - 3 alternating learners' English responses to six tasks and 3 alternating learners' German responses to six tasks
Final	- Applying scales III to learners' responses - Done in co-operation with 2 graduates (English and geography)	- 20 randomly selected responses to six tasks - 10 English and 10 German responses

Table 4.6 Summary of process involved in the development of the scales used to rate the ideational meaning constructed

The preparatory work done in designing the rating scales for language use and the ideational meaning constructed provided a solid basis for the development of the holistic scales. The scales and process involved in designing theme are discussed below.

4.6.4 Holistic scales: rating the appropriateness of the discourse function / the degree of task fulfilment

4.6.4.1 Description of scales

The holistic scales were *specified per task* and *seven levels of performance* were defined per scale. The scores range from 6 for the highest level of performance to 0 for the lowest.

That *criteria that underlie the holistic scales divide into two categories*. The first pertains to criteria related to the ideational meaning required to be constructed and the second to the appropriate textual realisation of this meaning. The *required ideational meaning* results from:

- An interpretation of the task that is contextually appropriate (subject-specific methodological competence).
- The application of knowledge of the relevant subject specific thematic pattern (subject-matter knowledge and methodological competence which facilitates the activation and combination of knowledge of a thematic pattern relevant to the task).
- The application of knowledge of the conventional uses of social semiotic systems (subject-specific communicative competence) in constructing meaning on the basis of input texts by following subject-specific procedures (methodological competence).
- The willingness to adhere to the conventions of the subject's discursive practices (volition and motivation).

The *appropriate textual realisation of this meaning* results from:

- The application of knowledge of the text structures (mini-genres) that conventionally realise the activities the task requires one to engage in (subject-specific communicative competence and methodological

competence which facilitates the activation of knowledge of text structures relevant to the requirements of the task).

- The application of knowledge of the subject's register (subject-specific communicative competence).

The criteria are as follows:

In terms of the ideational meaning required to be constructed:

- The extent to which the meaning realised is *complete*, *correct* and *relevant to the task*.
- The extent to which the meaning constructed is *consistent* in regard to the central position taken in an argument (relevant for tasks 3, 6, 7 and 8).

In terms of the textual realisation of the meaning constructed:

- The extent to which the *mini-genres* required by the task are realised.
- The extent to which the meaning constructed is *logically organised* and *linked* to form a text.
- The extent to which a *subject-specific register* is used appropriately.

The *level descriptors characterise performances in which these criteria are met to a more or a lesser extent*. Each scale consists of *a general description of a performance on that level*, which is then followed by *a more detailed characterisation* of a typical performance on that level. The detailed descriptions of performances were added, because, as Bachman and Palmer (1996: 221) point out, when level descriptors are too general raters can develop their own criteria, which adversely affects reliability. These detailed descriptions operationalise the criteria. In the case of comprehensiveness of content for instance, the content to be included was defined. This is illustrated by the following level descriptors taken from the holistic scale used in rating learners' responses to task 6. As already explained before, this task required learners to evaluate a suggestion to turn the tropical rain forest into a national park against the background of the principle of sustainability.

In these two scale descriptors, the criteria "*comprehensive* support" from level 6 and "*sufficient* support" from level 4 are for instance operationalised in terms of the number of goals of the principle of sustainability that are correctly applied. In the case of level 6 all four goals are correctly applied and in the case of level four at least 2 goals are correctly applied.

Level 6

There is a succinct and *balanced* evaluation of the suggestion. This is preceded or followed by succinct and *comprehensive* support that is fully consistent with the evaluation and that is based on a *complete* and a *correct* application of the principle of sustainability in evaluating the suggestion. The argumentation is well structured.

- The evaluation *criticises* the suggestion for the fact that it does not make equal provision for the different goals of the principle of sustainability. The support is based on the *correct* application of *all four* of the following goals of the principle of sustainability in evaluating the suggestion: protecting the environment (ecological goal), taking the rights and needs of future generations into consideration, making provision for social goals and paying attention to economic goals. *Only relevant content* is included.
- The text is *consistently* logically structured and effectively linked.
- A subject-specific register is *consistently* used.

Level 4

There is a *balanced* evaluation of the suggestion. The evaluation *may be* clearly formulated. There is *sufficient* support that *may be* clearly formulated, but that may show a *limited degree of inconsistency*. The support is based on an application of the principle of sustainability in evaluating the suggestion. The application may *not* be *entirely correct*. The argumentation *may be* generally well structured.

- The evaluation *criticises*, or *does not fully support* the suggestion. It *may be* pointed out that the suggestion does not make equal provision for the different goals of the principle of sustainability. The support is based on the *correct* application of *at least two* of the goals of the principle of sustainability in evaluating the suggestion. *Some content* may be included that *is* related to the topic, but *not central* to the content requirements of the task.
- The text may *not* be *consistently* logically structured and sufficiently linked.
- A subject-specific register is *mostly* used.

Due to the variation in learners' performances, which as already stated earlier is a typical feature of performance assessment, the biggest challenge in terms of the development of the holistic scales was to *find a balance between the variation in learners' responses and a practical number of scale levels*. The possibility of operating with a plus and a minus per level, thus for instance to distinguish between a 3+, a 3 and

a 3- was also investigated. This, however, led to rather low inter-rater reliability. It also proved to adversely affect intra-rater reliability, because it becomes impossible to handle a large number of scale levels consistently.

A first measure taken to address this problem was to formulate level descriptors using *may* and *or*. The level descriptor above for instance states: "There is *sufficient* support that *may be* clearly formulated, but that may show a *limited degree of inconsistency*." Formulating level descriptors in this way allows learners' texts to be assigned a specific level, in spite of some variation still shown in terms of a number of aspects. In the case of task 6 the scales for instance allow variation regarding the clarity of the formulation and the extent to which the support given is consistent with the statement made.

Although the scales do allow some variation in the performances on a particular level, *the performances all have at least one critical feature in common that distinguishes them from performances one level above and below*. For example, in the case of the descriptors for task 6, the overall descriptor for level 3 states:

> **Level 3**
> There *may be a balanced* evaluation that *may be* clearly formulated. There is support that *may be sufficient*, but that may show *some degree of inconsistency*. The support is *to some extent* based on a *correct* application of the principle of sustainability in evaluating the suggestion. The argumentation *may be* generally well structured.

Where the evaluation of the suggestion must be balanced in order for a performance to score a 4, in the sense that it does not fully support the suggestion, a performance on level 3 does not need to be balanced. In other words a performance on level 3 can also only criticise or fully support the suggestion. In cases where a balanced evaluation is included in a response, a distinction between a performance on level 4 and one on level 3 is made on the basis of the number of goals that are applied correctly. In a performance on level 3, only one goal needs to be applied correctly, whereas at least two goals must be applied correctly in order for a response to score a 4. This is illustrated by the outline of the criteria on level 3:

- The evaluation *may criticise*, or *may not fully support* the suggestion. The support is based on the *correct* application of *at least one* of the goals of the principle of

> sustainability in evaluating the suggestion. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.

Level 1 of the scale descriptors for task 6 serves as an example of a descriptor that makes use of *the "or-option"*. Here, 2 different types of performances are grouped on one level, because they deviate from the expectations to a large extent.

> **Level 1**
>
> There is *no evaluation* of the suggestion. There is only a discussion of Agenda 21 and/or the principle of sustainability that may show *major deficiencies* in terms of logical structuring and formulation.
>
> **Or**, there is an *evaluation* that *may be balanced*, but there is *no or hardly any support*, **or** the support is *not based* on an application of the principle of sustainability at all. The evaluation may show *major deficiencies* in terms of logical structuring and formulation.

A second measure taken to support raters in assigning learners' performances a specific level in spite of the variation was to *identify benchmark texts* that illustrate typical performances on the different levels. The process involved in formulating the scale descriptors and identifying the benchmark texts is explained below.

4.6.4.2 Process involved in developing the scales

In essence, designing these scales firstly entailed *combining the frameworks of expectations* in terms of the meaning to be constructed on the one hand and the textual realisation of this meaning on the other. It secondly involved *identifying the criteria* that could be used to differentiate between performances on different levels in a holistic scoring. Initial *level descriptors were then formulated* that focused attention on the effect of the text as a whole in terms of the ideational meaning constructed, the mini-genres realised and the language used.

These preliminary scales were applied to 10 responses written in English and 10 in German to the six tasks and the scale descriptors were further refined. The process was undertaken in co-operation with a graduate who studied English and geography and another project member who is a linguistics expert. In order to make the level descriptors clearer, *the detailed guidelines that explicate expectations indicated in the*

overall level descriptor were added. As illustrated in the section above, this involved specifying criteria such as comprehensiveness of meaning constructed, used in the overall level descriptor, by listing the propositions that had to be included.

A second measure undertaken to make the level descriptors as clear as possible and to support inter-rater reliability in the holistic scoring was to identify benchmark texts. As Weigle (2002: 112) for instance explains, benchmark texts or examples of performances on the different scale levels serve to support reliability in the holistic scoring of texts as they further illustrate the scale criteria. We identified benchmark texts by applying the final version of the scales to the 20 learners' responses. If a specific level of performance was not illustrated by any of the performances in this random sample, further learners' responses were rated until benchmark texts were found. In the whole sample, no benchmarks texts could however be found for level 6 on any of the tasks.

The process involved in developing the holistic scales is summarised in Table 4.7.

Version of scales	**Development basis**	**Number and nature of samples involved**
Preliminary I	Combining model answers and final scales used to rate the meaning constructed and the textual realisation of this meaning	-
	Identifying criteria that can be used to distinguish between performances, seen on the level of the text as a whole	
Final	- Applying scales I to learners' responses - Done in co-operation with a graduate in geography and English and another project member (linguist)	- 20 randomly selected responses to the 6 tasks - 10 English and 10 German responses
	Identifying benchmark texts	- Primarily 20 responses referred to above - Responses from wider sample used to identify benchmark texts not found in the original selection

Table 4.7 Summary of process involved in the development of the holistic scales used to rate the appropriateness of the discourse function

In the concluding section of this chapter, an overview is given of the way in which these scales were applied in rating sessions.

4.7 Rating

4.7.1 Rating of the completeness and correctness of the meaning constructed

The completeness and correctness of the meaning constructed was rated in *August 2006*. After a one-day training session, a group of five raters set out to rate the responses from the English and German samples to the six tasks. All responses were *double rated*.

Six responses, three written in English and three in German were used in *training* the raters and were not included in the final sample rated. The training involved explaining each of the scales to the group of raters and rating each of the six responses using the scales. After the rating of each response, the options (yes/no or yes/partly/no) ticked for each of the scales were compared. In cases where there were differences between the raters' decisions, consensus was reached about the interpretation and application of the particular scale in question.

The raters were also instructed in the *rating procedures*. Since there had to be an independent double rating of all responses, the rating and scoring had to be done on special rating grids. The rating grids enabled raters to indicate their decision per scale and to calculate the total score a learner received per task. The rating of all responses to a specific task also had to be done in a single rating session to avoid possible adverse affects that time lapses in the rating process may have had on intra-rater reliability.

The rating proceeded in *six rating sessions over six consecutive days*. The rating proceeded under controlled conditions as all raters sat together in one room. Each rating session focused on one specific task and started with a trouble-shooting forum and discussion session. These feedback and discussion sessions created the opportunity to find solutions to unanticipated problems and to remind the raters of the rating standards agreed upon during the training. After each rating session, learners' tests were re-arranged. This was done, so that the position of a learner's test in the pile changed and it

could be avoided that a text preceding a particular sample could systematically influence the rating of the sample. The English responses to the six tasks were mainly rated by two geography graduates. The raters indicated their rating decisions in separate rating grids, so that they could not see each other's rating decisions.

Inter-rater reliability was computed using the Pearson Product-Moment Correlation. Acceptable to high levels of agreement, which range from r.855 to r.934, were reached on all tasks, except in the case of task 7 where the correlation was r.627. It is not entirely clear what caused the differences in scores given to performances on this task by the different raters. It might be related to the fact that the complexity of the arguments presented in the responses had to be assessed. It is possible that there might have been differences in the raters' views in terms of the level of complexity of the arguments. Other aspects that might have contributed to differences between the scores the two raters gave on the different tasks include fatigue on the part of the rater, ambivalent or vaguely formulated responses that could have led to different interpretations of the content, or a rater misinterpreting parts of a scale, without it being identified during the training or feedback sessions.

I then compared the two raters' scores on each of the tasks and carried out a *third blind rating* in all cases where the raters' scores differed by more than one point, except at task 2a. A variation of 1,5 points was allowed for this task, because the total was so much higher than the totals of the other tasks[1]. A blind rating involves rating a text again without looking at the original rating decisions. In all cases of the third, blind rating, the score given differed less than one point from one of the initial ratings. The *two closest scores were then added and divided by two* in order to calculate a final mark. This procedure supports reliability, since the final result incorporates two ratings and not only one single rating. This final mark was used as the basis for the frequency distributions showing the range and number of marks on different levels for all six tasks. These frequency distributions are presented and discussed in the next chapter.

[1] The totals of the tasks are as follows:
task 1c: 8, task 2a: 19,5, task 3: 10, task 6: 7, task 7: 7 and task 8: 7.

4.7.2 Analytic rating of the textual realisation of meaning

Random sample double rating (Alderson *et al.* 1995: 144, McNamara 1996: 118) preceded the main rating phase and was done by myself and another rater. However, due to the time it takes to do an analytic rating, the sample had to be limited to 8 texts per task and only included the responses to 5 of the tasks. No double rating was carried out for task 2a. The scores given per task on each of the seven scales were written into separate scoring grids. Computation of *inter-rater reliability* showed varying results, ranging from acceptable levels of reliability higher than r=.80 to levels showing little agreement. However, the scores never differed by more than one scale point.

The low level of inter-rater reliability on some of the analytic scales can firstly be explained by the fact that the second rater could not be trained extensively in using these scales, due to time constraints. It could secondly be attributed to the low number of samples involved in calculating inter-rater reliability and can thirdly be related to the fact that inter-rater reliability was calculated in terms of a specific scale per task and not on the basis of the average score given on a specific scale over all six tasks. For the purposes of this study, the latter is of greater importance, however. Since the aim is to get an overview of strengths and weaknesses in learners' writing and tasks have such a profound effect on performance, the averages achieved on each scale over all six tasks are used as the basis for the frequency distributions presented in Chapter 5.

The *main rating phase* followed the one-day double rating session and was carried out in *August 2006*. The rating *proceeded over a period of 6 days* and involved the analytic rating of *365 samples*. As already pointed out in section 4.6.7, not all of the responses were included in the analytic rating. No rating was carried out with any of the seven analytic scales in cases where responses were too short to allow valid and reliable decisions about the consistency with which structural units are ordered and linked effectively and formal language and clear and succinct formulations are used. Responses that were clearly not written to the end, for example responses that stopped in the middle of a sentence, were also not rated. As Bachmann and Palmer (1996: 218-219) point out: "In order for extended responses to be rated reliably and validly, the language produced has to constitute what is called a 'ratable sample' of language. One characteristic of a ratable sample of language is that there is opportunity for the full range of components that are to be rated to occur. A second characteristic is that a

ratable sample provides evidence of the test taker's highest level of ability on these components.".

As with the rating of the meaning constructed, each rating session focused on one specific task. The tests were also re-arranged after each rating session in order to preclude the possibility that rating was systematically influenced by a test's position in the pile.

4.7.3 Holistic rating of the appropriateness of the discourse function

I carried out the holistic rating together with another project member who rated the responses written in German. The rating took place over *a period of 6 days in January 2007*.

Random sample double marking (Alderson *et al.* 1995: 144, McNamara 1996: 118) was also carried out here to determine the level of agreement between the two ratings. Twenty texts written in English and 20 written in German were double rated for this purpose. Inter-rater reliability was again computed using the Pearson Product-Moment Correlation. High levels of agreement were reached on all tasks, ranging from r.879 for task 7 to r.973 for task 6.

Each rating session also focused on a specific task and the piles of tests were also re-arranged before each rating session.

4.8 Conclusion

As illustrated in this chapter, great care was taken to enhance the reliability and validity of the assessment. The measures taken in support of validity include the use of authentic tasks and rating scales designed in co-operation with members of the subject community. The challenges posed by performance assessment in terms of reliability were addressed by means of the inclusion of a large number of tasks and written responses and the application of rating scales. This provides a solid empirical basis for assertions about this group of learners' subject-specific writing presented in the next chapter.

5
Scoring results: analysis and discussion

In this chapter, I present the results of the study and use examples of learners' responses to illustrate general features of the learners' subject-specific task-based writing. The results clearly indicate a developing subject-specific written discourse competence and point to the need for further training in terms of their knowledge of the tropical rain forest and subject-specific vocabulary, as well as their ability to analyse input material and their ability to write formal English. In the discussion, attention is first paid to the results of the holistic scoring of the appropriateness of the discourse function. This is followed by a discussion on problems in terms of the meaning constructed in learners' texts. Finally, the strengths and weaknesses of their learners' textual presentation of meaning are outlined.

5.1 Approach followed in data processing and interpretation

All statistical processing of the data was done with SPSS and was carried out by a psychology major who analysed all data elicited in the larger project as part of his own thesis.

For the purposes of my study, scores on all three types of rating were organised according to their frequency of occurrence. These *frequency distributions* (e.g. Bachman 2004: 43, Henning, 1987: 36) are used to make inferences about general features occurring in the subject-specific task-based writing of the specific group of learners involved in the project. This is done by identifying the range of scores, as well as the level of performance demonstrated by the largest number of learners on a specific scale. The identification of general strengths and weaknesses in this group of learners' written performances creates a basis for further research into bilingually-taught learners' task-based subject-specific writing. Further studies can for instance explore the extent to which particular strengths and weaknesses identified in the context of this study also occur in the writing of other bilingually-taught groups.

In the following sections, the distributions of the scores on all three types of rating are presented and discussed. The discussion starts with an analysis of the results of the

holistic rating. The results of the content (meaning constructed) rating are then discussed, since these results show a number of similarities to the results of the holistic scoring. Finally, attention is paid to the results of the analytic scoring of the textual presentation of meaning.

5.2 Holistic scoring of the appropriateness of the discourse function: presentation and analysis of the results

Generally speaking, learners' performances clearly point to *a developing subject-specific written discourse competence*. For all tasks, the mode (e.g. Bachman 2005: 55, Henning 1987: 40) or the score that was given most often is a 3. At some tasks, quite a large number of performances were also assigned lower levels of performance. At task 3 for instance, no less than 25 % of the performances were awarded a 0. At task 1c, 31% of the performances scored a 1. This is as high as the number of performances scoring a 3 on this task. At task 7, 28% of the performances were awarded a 2. A score of 5 on the other hand was only given to 3% of the responses to task 2a, 4% of the responses to task 3 and 4% of the responses to task 8. A 6 was never given. This can be explained by the fact that the level descriptor used on level 6 reflects the framework of expectations used as the basis for designing the scales. It thus describes a performance in which all conventions are followed. Consequently, it can be argued that it is not expected that learners in the 10th grade are able to perform on this level yet.

The performances *do not show significant problems in terms of the realisation of mini-genres*. Most learners realised the mini-genres required by the tasks. The only exception occurred at task 3, which could possibly be attributed to the instructional verb used. Task 3 required learners to "comment critically" on information about the extent of the destruction of tropical rain forest on the different continents. It is assumed that the critical comment is a genre with which these learners were not familiar. The fact that learners' responses do not show particular problems in terms of realising the structures of the mini-genres corresponds to their scores on the analytic scales used to rate the ordering of structural units and the organisation of meaning into different structural units. The means of the scores over all six tasks in terms of the rating done with these analytic scales are clearly higher than the means for the sufficient and appropriate use of

subject-specific terminology and the sufficient and appropriate use of formal language and succinct and clear expressions.

The fact that the learners' performances do not show major problems in terms of the realisation of the mini-genres, is ascribed to the following: firstly, these structures show similarities across different, but related languages such as German and English. Thus, these two languages clearly vary less from one another in terms of the discourse semantics, than in terms of the lexico-grammar. Secondly, as I pointed out a number of times in chapters 3 and 4, learners receive instruction on producing a number of mini-genres. This is for instance demonstrated by the extracts from the geography textbooks presented in Appendix D. The teachers who attended the workshops also indicated that they give learners explicit guidance on the appropriate realisation of these structures.

The *realisation of ideational meaning in the learners' responses, however, is rather problematic*. This became clear in both the holistic rating, where the incorrectness and incompleteness of the ideational meaning constructed negatively affected the overall score and in the analytic rating of the content carried out with the dichotomous and three-point scales. As explained in chapters 3 and 4, constructing ideational meaning involves applying knowledge of the subject-specific thematic patterns, as well as applying knowledge of social semiotic systems in order to construct meaning on the basis of the instruction and the input material. In addition to the fact that the ideational meaning constructed in the responses was often incorrect and incomplete, the responses tended to be imprecise and to lack details. Furthermore, it seems as if learners experienced problems in combining meaning from different input texts.

The first two tasks supported the construction of the ideational meaning required. The climate graph used in Task 1c and the diagram used in 2a both provided learners with detailed information in the form of numerical values (e.g. the amount of precipitation, the height of the trees) that they could use in their own responses in order to construct meaning in a more precise way. However, both at task 1c and 2a, *most learners failed to incorporate these details*. At task 2a, terms were also provided in the material that could be used to refer to the three layers of the tropical rain forest and to the factors that affect the living conditions in each layer. Many of the learners either did not use these, or

copied them incorrectly from the input texts, referring to the "brush and scrub layer" for instance, instead of to the "bush and shrub layer".

Task 3 and 6 required learners to *combine meaning from two different texts to construct new meaning*. Eleven learners did not attempt task 3 and thirteen did not do task 6. These numbers are clearly higher than the numbers of learners who did not attempt the other tasks. This is taken to indicate that the learners perceived the tasks to be too difficult. As pointed out before, a relatively large number of learners who did attempt task 3, received a 0 on their performances. The difficulties caused by this task are firstly ascribed to the use of the instructional verb "comment critically" (discussed above) and secondly by the fact that meaning from different sources had to be combined. At task 6, learners also clearly showed problems in extracting all relevant information from the definition of sustainability and in applying it to assess the suggestion. Most learners for instance did not apply more than 2 of the 4 goals of sustainability presented in the material. Furthermore, terms used in the definition that belong to a more formal register and that represent items on a higher level of abstraction of the thematic pattern, such as "ecological, social and economic goals" were rarely used by learners. They did give examples of these categories in their responses, for example by referring to the right of the Yanomami to live in the forest, but failed to connect an example such as this one to the category "social goals".

Learners' responses to task 7 and 8, where less support in terms of input material was given, was *also characterised by the incorrect and incomplete realisation of thematic patterns*. At task 7, for instance, where learners had to explain the nature of the relation between humans and the tropical rain forest, many learners could only give one explanation. For both these tasks, the meanings constructed also tended to remain on a very general level. Again taking task 7 as an example, learners would for instance refer to the importance of the tropical rain forest for the climate, but would then fail to explain how the tropical rain forest affects climate.

Problems in terms of the construction of ideational meaning are further analysed in the discussion on the results of the analytic scoring involving the dichotomous and three-point scales. In the following sections the distribution of performances on the holistic rating is discussed per task and the most striking characteristics of learners'

performances on each task are pointed out. Each discussion is preceded by a table that summarises information about the distribution. The table firstly reports the range of performances demonstrated, as well as the mode. Furthermore, the number of learners who did not attempt the task is indicated. As pointed out above, if the number of learners who did not attempt to do a specific task is clearly higher than for other tasks, it is taken as an indication of the task's perceived level of difficulty. It can also be related to a loss of concentration or to fatigue, especially if the task appeared towards the end of the test. Finally, the percentage of performances on each level is given. In this regard, only the valid percentage is reported. Thus, the percentages given are based on the number of learners who did the task. The table is followed by a bar chart that presents the frequency distributions visually.

It must be pointed out here that the *frequency distributions themselves should not be interpreted as reflecting the difficulty levels of the different tasks*. The fact that a larger number of learners' performances received a higher score on a specific task, does not mean that the task was easier, because a specific level of performance, say a 3, does not reflect a comparable level of performance over all tasks. This results from the fact that tasks showing variation in terms of their required discourse functions were used in support of generalizability. This naturally led to significant variance in the types of performances elicited by the different tasks. Consequently, it was not possible to design scale descriptors that allow all of the performances on a specific task to be categorised and that are simultaneously comparable across different tasks.

5.2.1 Task 1c

Range	0-4
Mode	1, 3
Number of learners who did not attempt the task	-
Percentage of learners on each level	
0	5%
1	31%
2	20%
3	31%
4	13%

Table 5.1: Task 1c: Summary of data – appropriateness of the discourse function

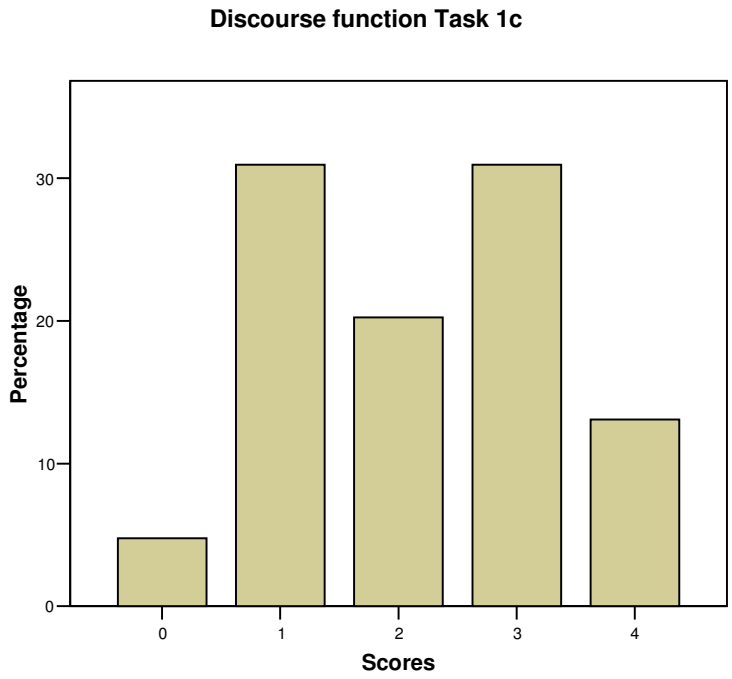


Figure 5.1: Task 1c: Distribution of scores - appropriateness of the discourse function

Task 1c required learners to identify and describe the climate of Kisangani on the basis of a climate graph. The scores on this task show a bimodal distribution with an equal number of learners scoring a 1 (31%) and a 3 (31%). A relatively large number of learners also scored a 2 (20%). The range of performances (0 – 4) on this task and the distribution of performances *give reason for concern* in terms of the learners' knowledge of a thematic pattern and a text type that are typical for the subject geography. Climate is one of the basic concepts in geography and therefore a concept one would expect 10th grade learners to have mastered. The same applies to the climate graph, which unlike the other input texts used in the six tasks, is specific to the subject. One therefore would expect 10th grade learners to be able to extract basic information from it.

No learner scored a 5. On the other hand, no less than 31% of the learners scored a 1. In performances on level 1, *the type of climate is either not identified*, or the *classification is incorrect*. Furthermore, the *descriptive explanation usually focuses either only on temperature or on precipitation*. The following two texts illustrate these typical problems occurring in performances on level 1. In the first example, the climate is not identified correctly. The description of the climate, however, although vague, refers to

both temperature and precipitation. In the second example, the climate is identified as a subtropical climate and the description only focuses on the precipitation. The learner also fails to indicate millimetre as the unit of measurement used in indicating the amount of precipitation. Furthermore, it is incorrect to describe the precipitation in the tropical rain forest as "low" and imprecise to describe it as "fairly high", since the amount of precipitation in this climatic zone is high in comparison to the precipitation in other types of climate. Naturally, it is correct to state that the amount of precipitation is lower during certain months of the year.

Task 1c, 3001

> *Kisangani has a maritime climate due to the even rate of temperature. In addition to this the amount of precipitation is fairly high which indicates a maritime climate as well.*

Tack 1c, 3002

> *I think that this city is located near the equator so that you can recognize a subtropical climate. In January the precipitation is low, in April and Mai the precipitation is higher, in June and July the precipitation is low and then it rise from 90 – 235 till November, then it falls again.*

Due to the fact that this task involves a thematic pattern and input material that is so specific to the subject geography, the *possible effect the number of semesters of bilingual subject-specific education in geography might have had on learners' performance* on the task was checked. The results taken from the multivariate analysis of variance, however, suggest that this variable had no effect on the performance. Firstly, there was no significant difference between the mean of the scores of the group who had one to two semesters of bilingual subject-specific training in geography and that of the group who had three to four semesters of bilingual instruction in geography. The mean for the first group was 2,29 and the mean for the second group was 2,36. The mean of the group who had five to six semesters of bilingual training in geography was even lower than that of the first group, namely 1,82.

The performances on levels 1 and 2 usually also *did not include the detailed and exact information about temperature and precipitation* presented in the form of numerical

values in the input material. This points to a lack of understanding of the importance of accuracy and precision in constructing meaning in subject-specific contexts and an inability to extract important information from input texts.

Another striking feature of learners' performances on task 1c was the fact that the climate was often identified as humid, or as hot and wet. Thus, *commonsense categories used in the classification of climate types were applied, rather than a subject-based one*. This was the case for all performances that scored a 2 and for some that scored a 3. It was still possible to get a 3 though, even if the climate was identified as humid, if the descriptive explanation included *references to both temperature and precipitation*. Thus, the main difference between a performance scoring a 2 and one scoring a 3 was the comprehensiveness of the description of the climate. The main difference between a performance scoring a 3 and one scoring a 4 was whether the *climate was correctly identified*. A performance did not only have to include a number of detailed references to both the temperature and precipitation to score a 4, but also had to point out that Kisangani has a tropical climate. A performance in which the climate is correctly identified but the description incomplete, was awarded a 3.

Below are three examples of performances on each of these three levels: levels 2, 3 and 4. Like many other performances awarded a 2, the first text (2003) identifies a "humid" instead of a "tropical climate" and focuses almost exclusively on precipitation, while only briefly and vaguely referring to temperature. No reference is made to the fact that the average annual temperature is 25°C or to the fact that it is constant throughout the year. The second performance (5006), which was awarded a 3, also identifies the climate as "humid", but the description is more comprehensive, including references to both precipitation and temperature. In the last example (3017), the climate is identified correctly and the description refers to temperature and precipitation. The description is also more detailed and precise, because of the inclusion of the temperature in degrees Celsius and amount of precipitation in millimetres. This performance was consequently awarded a 4.

Task 1c, 2003

> *Kisangani has a very humid climate. The rainfall is much higher than the averag anual temperature, that's what is showing in figure 2. At the beginning of the*

year the rainfall is rather low, but in april the rain starts to fall. In july it is like in January. There is a rainfall about 100 mm a day. Between october and november there is the highest rainfall. They noticed about 230 mm a day. In the end of the year the rainfall decreases.

Task 1c, 5006

The climate in Kisangani is very humid throughout the whole year. The precipitation line is almost never under 100 mm. The climax of the precipitation line lies above 220 mm. The temperature is constantly 25°C throughout the year.

Task 1c, 3017

The climate doesn't really change. The only thing is the precipitation, which variates between 100 mm in january and 230 in september. The annual amount of precipitation is about 1804 mm and the temperature is about 25°C during the whole year. By seeing at all these facts, I can say, that it's a tropical climate near the equator.

The holistic rating not only focused on the meaning constructed, but also on the textual presentation of this meaning. As already pointed out in the introductory section on the holistic rating, learners' responses were significantly less problematic in terms of the realisation of the required mini-genres than in the construction of the required ideational meaning. However, *problems in terms of the textual presentation of the meaning constructed* often led to responses being awarded a lower score. In the first of the following two examples (2009), the text contains all information to be awarded a 3. However, since the text is insufficiently linked, it was placed on level 2. The second example (2005) scored a 3 instead of a 4, although the climate is correctly identified and the description refers to both temperature and precipitation. This results from the fact that these references are not presented in a structured way, but that the text jumps between references to the type of climate, temperature and precipitation and repeats itself in terms of references to temperature.

Task 1c, 2009

Kisangani has a very hot climate and the climate is also humid. The climate is warm, because Kisangani is in a savanna. The climate doesn't change in the

months, it has always the same temperatures. The temperature is always 25°C. The precipitation is in July very low and in October very high.

Task 1c, 2005

Kisangani's climate is very tropical. It has nearly the same temperature of 25°C all through the year. It is a very humid climate and most rainfall comes down from August to November, most in October. Another very hot and wet period from March to June, where most rain falls in April. The average annual temperature is 25°C and the average annual precipitation is 1804 mm.

5.2.2 Task 2a

Range	0-5
Mode	3
Number of learners who did not attempt the task	1
Percentage of learners on each level	
0	4%
1	13%
2	23%
3	42%
4	15%
5	3%

Table 5.2: Task 2a: Summary of data – appropriateness of the discourse function

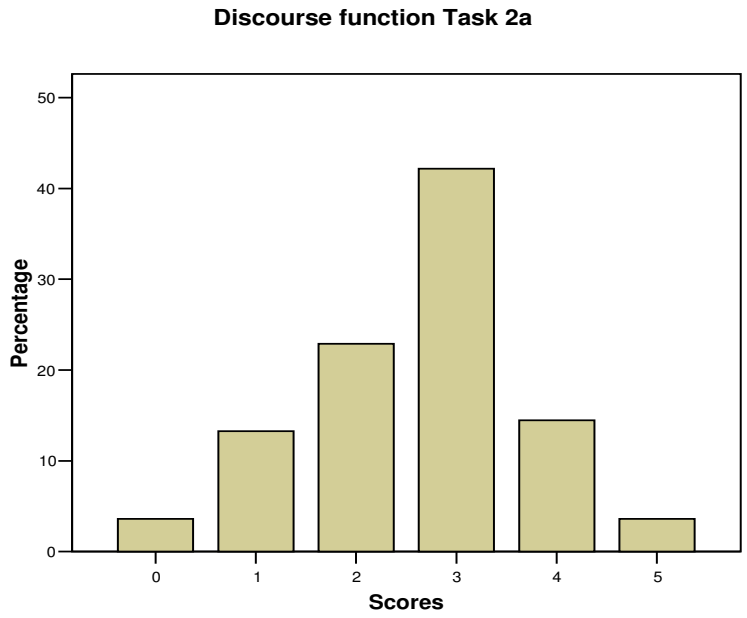


Figure 5.2: Task 2a: Distribution of scores - appropriateness of the discourse function

Task 2a required learners to characterise the living conditions in the three layers of the tropical rain forest on the basis of a detailed diagram. The mode on this task is 3 with 42% of the learners receiving this score. A small percentage (3%) of learners also managed to score a 5 on this task.

One of the criteria used to distinguish between the different levels was the extent to which the characterisation was done on the basis of *a comparison between the living conditions in each of the layers*. Performances on levels 6, 5 and 4 had to include comparisons, whereas performances on the lower levels did not have to include comparisons. This aspect was included in the framework of expectations used in designing the scales for this task after the first workshop with the four geography teachers. All four teachers stated that they would expect learners to compare the living conditions in the three layers, although this is not explicitly indicated by the instruction. They argued that the three layers are identified on the basis of the differences in the living conditions and that a comparison between the three layers thus reflects a subject-based construction of meaning. They also pointed out that they regard knowledge about the differences between the living conditions in the three layers as part of the body of basic knowledge about the tropical rain forest that learners should have mastered by the 10^{th} grade.

The extent to which the living conditions in the different layers are compared is one of the distinguishing factors between a performance on level 5 and one on level 4. Furthermore, a performance had to include *the correct terms for the different layers* to be awarded a 5 and *all four factors: incidence of light, maximum temperature, air humidity and number of animals had to be included consistently* in the characterisation of the living conditions in each of the three layers. Also, performances on level 5 included more *explanations about the interaction between the different factors* playing a role in the living conditions in the different layers.

The first of the following two texts (2019) includes a number of these explanations, e.g. it explains why fewer animals live in the tops of the giant trees, or why the temperature is the lowest in the bush and shrub layer. This text was also awarded a 5, because of the number of comparisons between the layers included and because of the fact that

reference is constantly made to all four of the factors affecting the living conditions. One of the problems illustrated by this response, however, is the *lack of detailed information*, for example an indication of temperature in degrees Celsius and of incidence of light in percentages, as well as a consistent reference to the height of the different layers in meters. The second performance (5006) consistently includes detailed information about temperature and incidence of light, but only once refers to the height of a layer in meters. Neither does it include explanations about the interactions between the factors affecting the living conditions. Furthermore, in the description of the bush and shrub layer no reference is made to the number of animals. This performance was consequently awarded a 4.

Task 2a, 2019

> *In the giant trees of up to 60 m height, there is most of the sunlight and also the highest temperature of the layers. The air humidity is quite low there, because the rain falls through these trees and makes the ground more humid. Due to these factors, in this layers are less animals than in the others. In the second layer with a height of 30 to 40 meters, the tree top layer with closed canopy there is medium light and a medium temperature. The air humidity is higher than in the giant trees, so there are also living more animals. In the bush and shrub layer, which is the lowest, there is hardly any sunlight falling in and so the temperature is the lowest of the three layers, but the precipitation is the highest and there are living many animals as well.*

Tas 2a, 5006

> *The giant trees are about 60m high. The temperature up there lies at 35°C. There is less air humidity and only a few animals can live there. The incidence of light is nearly 100%.*
>
> *The next layer is the tree top layer. While the incidence of light and the temperature decrease the air humidity and the number of animals increase. In this layer you can also find some nutrients.*
>
> *Finally we have the bush and shrub layer. On the ground there's only 10% of light incidence. The temperature has fallen under 25°C while the air humidity has almost doubled after the tree top layer. In the ground you can find many nutrients.*

Apart from the fact that a performance on level 3 did not have to include any comparison between the living conditions, a performance was also awarded a 3 rather than a 4 if it *consistently excluded one of the four factors*: incidence of light, maximum temperature, air humidity and number of animals, or if the *characterisation of the living conditions in one of the levels was limited*. Often the difference between a performance on level 4 and one on level 3 was also determined by the amount of details included and by factors related to the textual presentation of meaning. Although the following text includes some comparisons between the living conditions in the four layers, as well as two references to the interrelations between the living conditions, it was only awarded a 3. The text shows a number of *language errors* and except for the reference to the height of the tree top layer with closed canopy, contains no detailed information, causing it to be rather vague in its characterisation of the living conditions.

Task 2a, 4006

> *The highest tries get the most light, so that it can grow fast. But they reach also the hottest area and the air humidity is very low so high above the soil. So animals are not to be seen so often, too.*
> *The other trees at the hight till 40 meters get also a lot of sunlight, but not as much as the giant trees do. Also the heat is lower and the air humidity quite high. The number of animals is very high there.*
> *The bushe and shrub layer mostly gets any lights, not much heat, but very much humidity and have a big amount of animals.*

Incomplete content, for example the exclusion of one of the factors affecting the living conditions or even of one of the three layers reflect *problems in terms of following subject-based procedures in constructing meaning on the basis of the input material*. All three layers and all relevant factors to be included in the characterisation were presented in the input material. A lack of details also points to problems in terms of the ability to construct meaning on the basis of input material, as information about the exact height of the temperature, trees and humidity in each layer could be extracted from the input material. Furthermore, the correct terms for the different labels were also presented in the input material, but as the example above shows, were not always used in the written texts.

Level 2 was the level the second biggest group of learners achieved with their performances. The main distinguishing factor between a 3 and a 2 was *the number of layers identified* and *the comprehensiveness of the characterisation of the living conditions.* All three layers had to be identified for a performance to be awarded a 3 and consistently had to include references to two or three of the factors determining the living conditions. On the other hand, a performance could still score a 2, if only two layers were identified or if the description of the living conditions in each layer included only two or three factors.

In the following text, three layers are identified, but not correctly named. Furthermore, no details are included and in the description of the bush and shrub layer, no information about the incidence of light and air humidity is included. The text also shows a number of language errors and weaknesses in terms of structuring and linking. Whereas the description of the living conditions in the tops of the giant trees is presented in a separate paragraph, the living conditions in the two remaining layers are described in one paragraph. All three descriptions are linked using the additive conjunction "and", rather than a contrastive conjunction, such as "however". The text was consequently awarded a 2.

Task 2a, 2013

> *On the top of the giant trees there is too hot, because the son is shining there all the time and there is no shade.*
> *And if you go under the giant trees there isn't so hot because there is more shade and that make something colder so more animals can live there then in giant trees. And in the ground at the tree "Stammen" there is much colder where the most animals can live. But the soil there are the nutrients.*

The last text presented as an example here illustrates typical problems occurring in performances on level 1. No distinction is made between three different layers and the description of the living conditions is incomplete and vague.

Task 2a, 5007

> *The living conditions in the rain forest are very hard, because there is only very little incidence of light and temperature. I can imaging that it is exhausting if there's such a high air humidity.*

The layering of the vegetation in the tropical rain forest is one of its unique features and knowledge about this layering is central to an understanding of the tropical rain forest. Furthermore, most information required from the responses to score at least a 4 was included in the material. Consequently, the fact that the three layers were often not correctly identified and all relevant factors not included in describing the living conditions firstly suggests that an important aspect of the thematic pattern revolving around the tropical rain forest has not been mastered. It secondly points to problems in terms of the close analysis of input material.

5.2.3 Task 3

Range	0-5
Mode	3
Number of learners who did not attempt the task	11
Percentage of learners on each level	
0	25%
1	11%
2	19%
3	30%
4	11%
5	4%

Table 5.3: Task 3: Summary of data – appropriateness of the discourse function

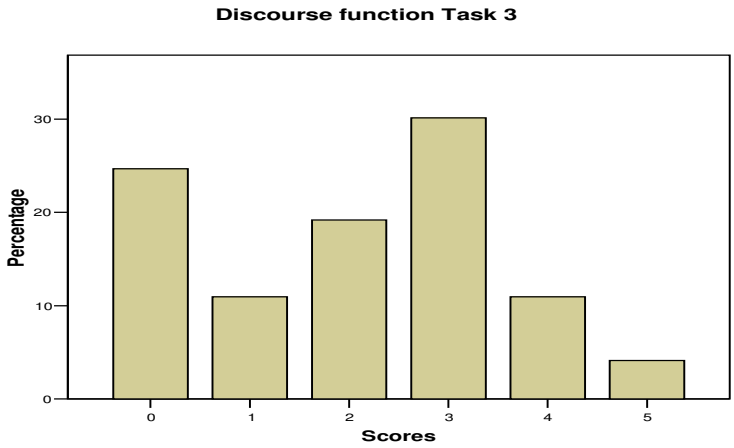


Figure 5.3: Task 3: Distribution of scores - appropriateness of the discourse function

Task 3 required learners to reflect about the effect that different semiotic presentations depicting the international destruction of the tropical rain forest have on the meaning constructed. A graph showing the remains of tropical rain forest in percentages by the end of the 80s and by 2010 had to be compared to a table in which the sizes of the forest area, measured in km^2, destroyed between 1990 and 2000 are presented. The two texts construct different meanings about the extent of the destruction. In proportional terms, the destruction is highest is Asia, but regarding the actual size of the area of rain forest destroyed, more forest area is lost in America. The results of the scoring of task 3 and the number of learners who did not attempt the task, suggest that *the task caused the learners some difficulty*.

Although the mode is 3 and 30% of the learners achieved this level, 11 of the learners did not attempt the task. It is the *second highest number of learners that did not attempt a specific task*. The only other task that more learners did not do is task 6. Furthermore, 25% of the learners' performances scored a 0. There was no *other task where a 0 was awarded this often*. A 0 indicates that none of the required mini-genres was realised and that the content of a response is not relevant to the demands of the task.

The difficulties caused by this task can probably be ascribed to two aspects. The first is, as has already been pointed out earlier, *the instructional verb used in the task*. The task requires learners to "comment critically". The teachers involved in the workshops pointed out that this is not an instructional verb commonly used in the context of subject-specific training in geography. It is also assumed that this instructional verb is not regularly used in other subjects. Furthermore, it is possible that learners had *difficulty in constructing meaning on the basis of the two different semiotic presentations and in comparing the meaning*s constructed.

Most performances that were given a 0 either *explained why the tropical rain forest is destroyed*, as is the case in the first response (5012) below, or *explained what the consequences of the destruction are*, illustrated by the second performance (5005). In addition to the fact that responses on this level neither met the requirements of the tasks in terms of the mini-genre realised, nor in terms of the meaning constructed, the performances often showed *problems in terms of the textual presentation of meaning*.

The first response below, for instance, shows clear deficiencies in terms of clarity. It is also problematic in terms of structuring and linking. In the second response, the word "increases" is used incorrectly and the last sentence remains vague.

Task 3, 5012

> *Maybe the protection of the tropical rain forest isn't very good. So they can destruct them without any interests of people. The rain forest in Asia is found on isles. Maybe that has a reason, too.*

Task 3, 5005

> *The destruction of the rainforest is very critical for our climate. The rain forests are producing most of the world's air and it absorbs the CO_2, which is polluting the atmosphere. If the size of the rain forest increases the remaining trees are not able to absorb enough CO_2. So destruction supports global warming. Asia really should decrease.*

Performances that did not criticize or support the depiction of the destruction of the tropical rain forest in the graph, but that used the information from the table and graph to *describe the destruction and to make predictions* about the extent of the destruction in future were awarded a 1. The text below illustrates a typical performance on this level. It also shows that the contradiction in the depiction of the extent of the destruction in the two texts was not identified.

Task 3, 3016

> *Figure 4 shows the huge decline of the original tropical rain forests. So in America there still will stand only 39% of the original tropical rain forest in 2010 in Africa 40% and in Asia only 19%. If the destruction continue this way, soon there won't be any tropical rain forests. This shows also the example of Brazil. In 1990, there had been 5669 980 km^2 forest are, in 2000 it decreased by 230 900 km^2 forest.*

Performances on level 2 either *did not explicitly criticise the depiction of the destruction in the graph or they fully supported the statement* that the graph makes about the destruction, namely that the most severe destruction takes place in Asia. The two texts

below illustrate these characteristics of performances on level 2. The first text below (3013) contradicts itself, in the sense that it first supports the presentation in the graph, but then in the last sentence suggests that the destruction may be more severe in South America. It thus remains vague in its criticism of the graph. The second text (3013) on the other hand, clearly supports the view of the destruction presented in the graph. As is the case with all performances on this level, none of these texts refer to the fact that the graph and table construct different meanings about the extent of the destruction on the three continents. Both responses also focus on information from only one of the input texts. In the case of the first response below, the focus is on information as presented in the table. In the case of the second response, the information presented is only drawn from the graph.

Task 3, 3013

> *The destruction of the tropical rain forest should be the highest in Asia. I think this statement is right. For exampel in Indonesia, there is a annual change of -13 120 km^2. So there is a high reduction of rain forest between the years 1990 and 2000. But in America, especially in Brazil is a higher decline of forest area. The area of rain forest in Brazil was reduced of 23 090 km^2 between the years 1990 and 2000. As a conclusion, the destruction of tropical rain forest is nearly the same in Asia like in South America. Perhaps it is a bit higher in South America, especially Brazil.*

Task 3, 2028

> *The table and Figure 4 show us that the percentage of the remains of the original rain forest declines. And that very fast. In American "only" 18% were destroit in the time between the 80's and 2010. In Africa were 17% destroit.*
> *There is no real difference. But in Asia the people destroit 21% of the rain forests, so it there is the highest destruction.*

In order to be awarded a 3 or more, a *response had to criticise the way in which the graph depicts the destruction*. The most important distinguishing features between performances that were awarded a 3 and those that were awarded a 4, were firstly *the degree of explicitness in the argumentation* and secondly the *extent of the support given*. In the first example (3008) below, which scored a 3, the statement made by the graph is

criticised and it is pointed out that the actual size of the area of rain forest destroyed is bigger in Brazil than in Asia. The second example (5011), which was awarded a 4, takes the argumentation one step further by pointing out that there is a difference between the depiction of the destruction in percentages and one in square kilometres. There is thus an explicit reference to the units of measurement used and the way in which they influence the meaning constructed. Furthermore, in the second response below the comparison is drawn on the level of the continents - Asia and America - and Brazil is identified as a country on the American continent. Response 3008 fails to point this out.

Task 3, 3008

> *This statement isn't completely right. In Brazil the most rain forest was destructed and will have been destructed. The reason is that Brazil has the biggest rain forest, but Asia has not. If you compare them now, you can see that Asia will have the smallest forest but that Brazil loses more forest.*

Task 3, 5011

> *The statement that the destruction of the tropical rain forest is highest in Asia is only partly true. The decline shown in percent doesn't point out the real amount of forest area which has been destroyed. It is true that Asia has destroyed the highest percentage of its rain forest but the highest amount of km^2 was destroyed in America or to be more specific in Brazil. To get a fairly exact picture of the destruction one has to look at the size of the areas itself as the tropical rain forest is wider spread in America than in Asia.*

Neither of these two responses includes detailed information about the exact percentage of original rain forest left on the three continents or of the size of the areas that were destroyed. Neither do they include references to the periods of time involved: the end of the 80s, by the year 2010 and between 1990 and 2000. Surprisingly, this is a feature of quite a number of the better responses to task 3. In other words, quite a number of responses that did criticise the graph for its depiction of the destruction and that did identify the contradiction in the two depictions failed to include detailed information.

The first of the two responses below, which both scored a 5, does not include any detailed information. However, it is *more explicit in its criticism* of Figure 4 and the

way in which it depicts the international destruction than response 5011 above, although it shows more problems in terms of language use. By explaining that the size of forest area destroyed in Brazil is bigger than the size of forest area destroyed in the whole of Asia, the extent of the destruction in Brazil is also made clearer. The second response (4016) below also shows problems in terms of the textual presentation of meaning, especially when compared to response 5011 above. The style is not always appropriate and there are major grammar errors. However, it does explain the effect of the depiction in percentages and in square kilometres and *uses exact references extracted from the input material* in its argumentation. It is also more precise in its comparison of the extent of the destruction in the different countries than 5011.

Task 3, 4014

> *The problem in figure 4 is, that the changes are shown only in percent but not with concrete numbers.*
> *In table 1 you can see the real annual changes of forest area in km^2. You can read that in Brazil more forest is destroyed than in whole Asia. But in percent, the amount of forest destroyed in America is lower than in Asia because in America is more forest at all.*
> *Consequently it is difficult to draw the green trees ("100%-trees") in the same hight for all of the three continents.*

Task 3, 4016

> *The numbers in Fig 4 are percent numbers. If a country has 10 tropical trees, and 9 are destroyed, it would be 90%, but it wouldn't be so bad.*
> *But it isn't a lie to say that Asia has a very high loss of tropical rain forest areas. In Indonesia for example have been 1 181 100 km^2 forest. Only 10 years later, 13 120 km^2 were destroyed. An other example from America: Brazil! In 1990 there have been 5 669 980 km^2. Nearly 4 times as much as Indonesia. 10 years later 23 090 km^2 were destroyed. "Only" twice as much as in Indonesia. There are more examples and in numbers, there is more forest destroyed in America than in Asia. But the percent is truely higher in Asia.*

The fact that responses on level 2 and lower did not criticise the statement made by the graph or identify the contradiction in the depiction of the destruction by the graph and

table suggests that a large number of learners had difficulty relating information from the two different input texts presented in this task. It also suggests that the learners had difficulty in reflecting about the effect that the form of depiction has on the meaning constructed. A further problem in terms of interaction with input material illustrated by performances on this task – also those on higher levels – is the ability to extract detailed information from input texts. Seeing that competence in analysing different text types, such as pictures, graphs, tables, and verbal texts is such a central skill in geography, the learners' responses to task 3 give reason for concern and point to the need for further and more detailed studies into bilingual learners' interaction with input material in the geography classroom.

Task 6

Range	0-4
Mode	3
Number of learners who did not attempt the task	13
Percentage of learners on each level	
0	1%
1	20%
2	22%
3	37%
4	20%

Table 5.4: Task 6: Summary of data – appropriateness of the discourse function

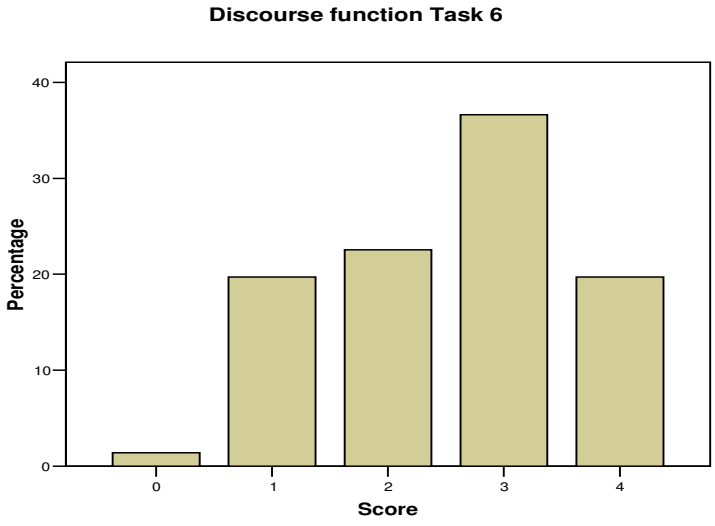


Figure 5.4: Task 6: Distribution of scores - appropriateness of the discourse function

As in the case of task 3, task 6 also required learners to combine meaning from two texts. Learners namely had to evaluate a suggestion to turn the tropical rain forest into a national park against the background of a definition of the principle of sustainability. This is the task that the highest number of learners (13) did not attempt to do. In general, learners' performances on this task suggest that they *experienced difficulties in terms of extracting relevant information from the definition and applying this information to evaluate the suggestion*. The definition of sustainability presented in the task states that regarding the use of natural resources, a balance must be found between ecological, social and economic goals, while the rights and needs of future generations should also be taken into consideration. There are thus four types of goals that need to be taken into account. Most learners only referred to one of these goals in their evaluation of the suggestion.

Learners who scored a 4 or 3 *evaluated the suggestion and correctly applied the principle of sustainability*, in the sense that they showed understanding of the fact that the principle *requires different goals to be balanced*. A performance was awarded a 3 if *one of these goals was applied correctly* and a 4 if *two of these goals were applied correctly*. The fact that a relatively high number of learners (20%) scored a 4 here, should therefore not be taken to mean that learners demonstrated a higher level of competence on task 6 than on the other tasks. The requirements that needed to be met for a performance to be awarded a 4 were not particularly high in comparison to the framework of expectations that demands that all four goals be applied correctly in a response to task 6.

Performances were awarded a 2, if they reduced the principle of sustainability to its ecological concerns or its concerns with the rights and needs of future generations. Most learners whose performances scored a 2 *fully supported the suggestion* to turn the rain forest into a national park, whereas learners whose performances scored a 3 or a 4 illustrated an understanding that the suggestion is problematic when related to the principle of sustainability. Learners whose performances scored a 2 motivated their support for the suggestion by *arguing incorrectly that it reflected the goals of the principle of sustainability*.

The first text (3018) below illustrates a performance on level 4. It does not fully support the suggestion, but implies that although the solution serves ecological goals, it has negative effects on an economic level. The second text (3014) below is an example of a performance on level 3. As in the case of the first text, the suggestion is not simply supported fully. However, only one of the goals of the principle of sustainability is correctly applied. Against the background of the principle, the reference to the rights and needs of future generations is correct. The argument that people will not be willing to leave the area, because they are only concerned with profit, is based on an incorrect interpretation of the principle. The principle namely also makes provision for taking the economic needs of people into consideration. The response thus does not illustrate an understanding that people also have the right to use natural resources in order to generate money. The third text (5020) below illustrates the central features of performances on level 2. The suggestion is supported fully and attention is only paid to one of the goals, in the case of this example to an ecological goal.

Task 6, 3018

> *This suggestion for a solution is very good. If there are such national parks there is no extinction of any trees or animals. The tropical rain forest will be protect. Another advantage of this solution is that there aren't any people who can destroy the tropical rain forest.*
>
> *But there is a disadvantage, too. The states which have debts can't sell them back because of missing money. They can't sell their trees of tropical rain forest anymore because they are I national parks.*

Task 6, 3014

> *I would support the suggestion for a solution because the tropical rain forest should be remained for our next generations. But it would be a problem to finish this possible solution, too, because people living in that area would not leave their territories on their own. They would not care about the problems and only concentrate on profitable work. Therefore, it would be difficult to finish such a plan, but it would be good.*

Task 6, 5020

> *I think it is good. The nature and the animals have their own living area and nobody is allowed to touch anything of this. It is a possible way to protect the tropical rain forest.*

All performances that scored a 1 either *provided very limited support* for their evaluation of the suggestion, or *provided support that was not based on an application of the principle of sustainability*. In the first of the two texts (3009) below the suggestion is evaluated, but limited and vague support is given. In the second text (5012) the suggestion is also evaluated, but the support is not related to the principle of sustainability.

Task 6, 3009

> *It's a good idea that I can support, because national parks are a fine thing and can be used everywhere.*

Task 6, 5012

> *I don't think that this is a very good idea. Because of the human beings it won't work! Everyone must have the opinion, that you can protect the rain forest. They don't have to be forced to think that, but when there is a National Park they will be forced to protecting and human beings don't want to be forced.*

The principle of sustainability is a central construct in geography. However, the fact that not one learner referred to more than two of the goals of the principle and that many learners were not able to identify the problems related to the suggestion, indicate that these learners do not fully understand the principle of sustainability. Problems in terms of learners' performances on this task also suggest that they have difficulty in analysing a written text, extracting relevant information from it and combining meaning from two different texts. The definition clearly referred to all four of the goals. This means that even with limited knowledge of the principle, it was possible to identify them. The fact that the goals were not always correctly applied in evaluating the suggestion is ascribed to the fact that the learners might have had difficulty in understanding the abstract categories: ecological, social and economic goals, as they are used in relation to the principle of sustainability.

5.2.5 Task 7

Range	0-4
Mode	3
Number of learners who did not attempt the task	5
Percentage of learners on each level	
0	1%
1	18%
2	28%
3	43%
4	10%

Table 5.5: Task 7: Summary of data – appropriateness of the discourse function

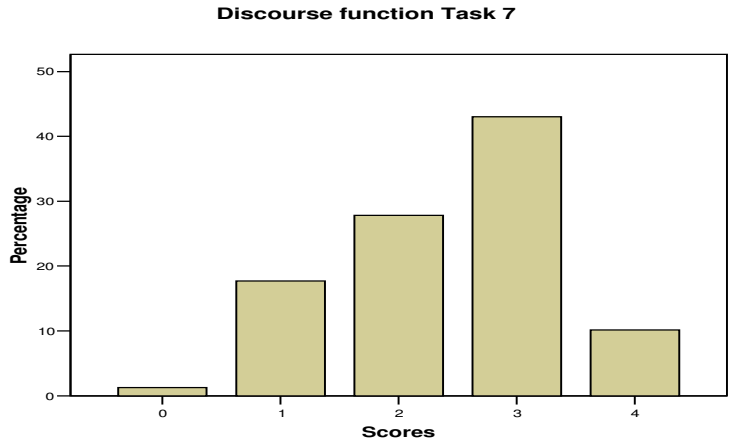


Figure 5.5: Task 7: Distribution of scores - appropriateness of the discourse function

Unlike the first four tasks, task 7 provided learners with no verbal support in terms of the meaning to be constructed in their responses. The task required learners to explain in how far it is true that the rain forest concerns all humans. The only input material presented was a cartoon depicting the rain forest as a hand that is holding the earth and that is being cut off. As in the case of the first four tasks, however, the mode for the responses to task 7 is again a 3.

Generally speaking, learners' argumentation on this task showed a *lack of knowledge of the international significance of the tropical rain forest*. Most learners could only refer to *one interrelationship between humans and the tropical rain forest and this reference was usually presented in vague terms*. This is illustrated by the following response

(2006) in which reference is made to the fact that the tropical rain forest affects the world's climate, but it is not explained why the tropical rain forest has an effect on the climate.

Task 7, 2026:

> *I think this statement is true because we need the rain forest. Our climate depend on the rainforest. We don't know how terrible it would be when there were no rain forests.*

This is an example of a response on level 3. A 3 was awarded when the *statement was supported* and reference was made to *at least one interrelationship* between humans and the tropical rain forest. Some responses on level 3 also included a description of the cartoon. In order to be awarded a 4, the statement had to be supported and reference had to be made to *at least two ways in which the destruction of the tropical rain forest adversely affects the world population*. A response on level 4 also had to *include a correct description or plausible interpretation of the cartoon*.

The first (3010) of the two responses below was also awarded a 3. The statement is supported, but the international significance of the tropical rain forest is only explained in terms of the fact that it functions as the earth's so-called "green lungs". The process by which the tropical rain forest absorbs CO_2 and produces oxygen through the process of photosynthesis is, however, not explained. The second text (3024) was awarded a 4. It contains a description and interpretation of the cartoon and expresses support for the statement that the rain forest concerns all. It also provides two references to the interrelationship between humans and the tropical rain forests, namely the fact that the destruction of tropical rain forests contributes to the greenhouse effect and the fact that the tropical rain forest is the living space for human beings and a habitat for animals. Again, however, the explanations remain vague. There is no detailed explanation of the processes involved in the greenhouse effect and the way in which it possibly influences the climate. There is also no detailed explanation of the ways in which mining and agricultural activities threaten the traditional lifestyle of groups of Indians living in the tropical rain forest, such as the Yanomami, or destroy the habitat of animal and plant species that are only found in the tropical rain forest.

Task 7, 3010

> *The statement "The rain forest concerns us all" is true. If the rainforests are destroyed more and more, the nature will be destroyed, too. And if the nature is destroyed, our life will be much harder or even impossible. So the rain forests concerns us all because the trees give us fresh air and when we destroy the rain forest, we will not have / get fresh air and this follows that we cannot life without the rainforests. We depend on the forest and the trees.*

Task 7, 3024

> *The statement that the rain forest concerns us all is true because without the rain forest there will be an increase in the greenhouse effect and the living space for many animals and people will be destroyed.*
> *In the picture that fact is shown in that way, that the rain forest has a hand and with this hand the forest holds the world. On the world there is a man who wants to cut the hand and the forest but if he do so he will fall down with the whole world.*
> *The picture says that we and the world are addicted to the rain forest and the rain forest are in some ways addicted to us.*
> *If we don't cut the rain forests they'll help our world.*

A 2 was awarded if the statement was supported, but the *explanation of the interrelationship between humans and the tropical rain forest was not plausible* from a subject-specific point of view, as is the case with response 3016 below. Furthermore, a 2 was awarded to a response that did not explain an interrelationship between humans and the tropical rain forest, but that merely described and interpreted the cartoon, as in the case of response 2008 below. A 2 was also awarded if, as in the case of response 3009 below, no clear opinion about the extent to which the statement is true was expressed and no clear reference was made to the cartoon, but an explanation of the international significance of the tropical rain forest was given.

Task 7, 3016

> *This statement is true, as the trees hold the soil together, where we live on. Consequently we destroy ourselves, if we destroy the trees.*

Task 7, 2008

> *In the picture we can see a person who cuts down the rain forest. But the rain forest is a hand, which carries the world, on which he is sitting. Maybe he doesn't see that fact because he is too excited about cutting down the forest. So, I think the statement is true and the protection of the rain forest is important for everybody.*

Task 7, 3009

> *If the rain forest is destroyed, the ecosystem is brought out of order. You say also the „green lungs of the earth". By destroying the rainforest we destroy our oxygen support.*

Most responses that were awarded a 1, either contained only *a brief description of the cartoon*, as in the case of response 4006 below, or presented *an implausible explanation of the international significance* of the tropical rain forest, as in the case of response 2007.

Task 7, 4006

> *In this picture the roots of the rainforest holds the world (earth) in the hand and the cutter would kill the whole mankind if he cut the forest.*

Task 7, 2007

> *Cutting down the rainforest means threatening habitats of animals and our own lifes. If there is no rainforest on earth any longer we can't survive, because we need the money which could be gotten out of it and also the animals maybe only to eat.*

The small number of explanations of the international significance of the tropical rain forest and the lack of details in these explanations, in the responses to task 7, suggest that this group of learners need further training in their ability to present detailed arguments. It also indicates that there are deficiencies in this group of learners' knowledge about the international significance of the tropical rain forest.

5.2.6 Task 8

Range	0-5
Mode	3
Number of learners who Did not attempt the task	7
Percentage of learners on each level	
0	12%
1	18%
2	22%
3	32%
4	12%
5	4%

Table 5.6: Task 8: Summary of data – appropriateness of the discourse function

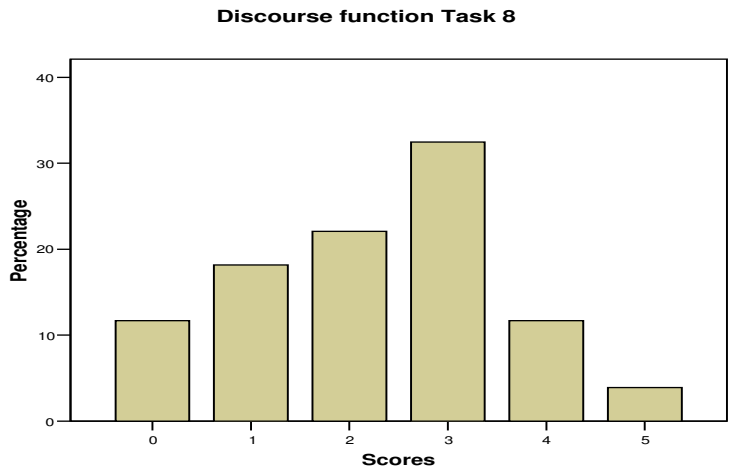


Figure 5.6: Task 8: Distribution of scores - appropriateness of the discourse function

In chapter 3, section 3.5.4.2, reference was made to the fact that pedagogical discourse aims to develop specific subject positions in which learners accept the values that regulate the discourse of the subject community and apply these in their own contributions to the classroom discourse. Task 8 focused on the *subject positions* reflected in the learners' written responses and *assessed the extent to which certain subject-specific values seemed to have been internalised* by the learners. The task required learners to react to the statement: "I can personally contribute to the protection of the tropical rain forest" by indicating their opinion on a five-point scale ranging from "absolutely true" to "absolutely not true". Learners were also required to support their opinion. The educational experts and the teachers made clear that they expect geography learners to show a willingness to contribute to the protection, but to also

demonstrate an understanding of the ways in which an individual living in Europe is restricted in terms of the extent to which he or she can contribute. This specific subject position was used as the basis in developing the holistic scale applied in assessing the responses to task 8.

The *criteria that underlie the differentiation between the levels* identified in the scale are firstly whether or not the option ticked on the scale is acceptable from a subject-specific point of view. The framework of expectations developed in co-operation with the educational experts and teachers for this task determined that one of the options: "absolutely true", "true" or "partly true, partly not true" had to be chosen. Secondly, the identification of the seven different levels of performance was based on the number of arguments presented in support of the opinion expressed and the extent to which these arguments were detailed. Furthermore, distinctions between the levels of performance were based on whether the arguments focused on one's personal scope for action, rather than on what people in general can do to protect the tropical rain forest. Finally, attention was paid to whether the response reflected an understanding of the fact that one is also limited in one's personal capacity to contribute to the alleviation of this global problem.

Level 3 again represents the mode. A relatively large number of performances scored a 0 (12%) on this task. Following task 3, where 25% of the performances scored a 0, this is the task with the second highest number of performances receiving no marks. These performances mainly scored a 0, because only an opinion was indicated on the scale. No verbal support for the opinion was given. This might be because the task was at the end of the test, so that learners did not have enough time to respond, or they might have chosen not to respond, due to fatigue and loss of concentration.

The first text (3017) below is one of the few that scored a 5. A response was awarded a 5, if *at least two arguments were presented in sufficient detail* in support of the opinion indicated on the scale. The arguments *had to focus on opportunities* the learner personally has to contribute to the protection of the rain forest, *but also on limitations* in terms of the personal scope for action. The learner ticked the option "partly true, partly not true" on the scale. In the support for this opinion, the way in which consumers' behaviour can affect the activities of companies responsible for the destruction of the

tropical rain forest is explained. Attention is also drawn to projects to which individuals can donate money that is used to buy and protect parts of the tropical rain forest. The lack of direct political influence is, however, also pointed out. Text 4015 below lists four ways in which one can personally contribute to the protection. It also points out that one single person's influence remains limited. The learner ticked the option "true" on the scale. Although more arguments are presented than in response 3017, this second text is an example of a response that scored a 4, since the *single arguments do not clearly explain the chain of events involved in each of these actions*. It is for instance not explained how organisations try to save the tropical rain forest, or how the decision not to buy certain products contributes to its protection; neither is it made clear why the protection needs the involvement of larger groups of people.

Task 8, 3017

> *I can contribute to the protection by buying no wood of the rainforest or exercise books without wood of the tropical rain forests. If everyone does this, the people selling tropical wood can't sell their trees any longer. In addition to this, I can buy smaller parts of the rainforests and than leave them as they are.*
> *But it's difficult to get involved in the politics of a country like Brazilia, because I am not brazilian, so I have no right to change the laws there. This is the part only the inhabitants can do.*

Task 8, 4015

> *I think it is possible for everybody to do something for the rainforests. Not much, not only one person can save the whole rainforest, but their are organisation which try to save the rainforest and you can pay money to them, or take part in the demonstration against the destruction of the rainforests. Or you can stop buy products with wood, for example Hefte. This are only small steps, but if everybody helps, it is possible to save the rainforests.*

The main characteristic of performances on level 3 is the fact that the arguments firstly *do not reflect the complexity of the different aspects involved in the protection of the tropical rain forest*. Responses that scored a 3 mostly only focused on the possibilities individuals have to contribute to the protection, without taking limitations in terms of personal capacity into consideration. A second characteristic is that the arguments did

not explain in enough detail how the actions of an individual contribute to the protection of the tropical rain forest. In response 5012 to task 8 below, for instance, the learner ticked the option "absolutely true". In the supporting argument, attention is only paid to two ways in which an individual can contribute to the protection, namely by donating money to Greenpeace and buy buying a specific type of beer. No reference is made to aspects that restrict individuals' influence. Furthermore, the ways in which Greenpeace can contribute to the protection or the effect that buying this specific brand of beer has on attempts to save the tropical rain forest are not explained.

Task 8, 5012

> *To help the rain forest we can spent money to Greenpeace or we can buy fe. Krombacher because the animals mustn't die and the forests hold the world so we can't will it.*

Level 2 performances were firstly distinguished from level 3 performances based on the amount of support given for the opinion indicated on the scale. In a level 3 performance, at least two supporting arguments had to be given. If *only one supporting argument* was given, a response scored a 2. Furthermore, a performance received a 2, if *it argued that no contribution could be made* and it only focused on aspects limiting the possibilities individuals have to help protect the tropical rain forest. In the first (2028) of the two responses below, the learner ticked the option "true" and presents only one argument in support of this opinion indicated on the scale. The argument itself does not explain in sufficient detail how the fact that one uses recycled paper contributes to the protection of the tropical rain forest. Furthermore, the statement that most "exercise books" are made from tropical wood is not true. Wood from the tropical rain forest is mainly used to manufacture furniture and not to produce paper. In the second example (3015), the learner ticked the option "not true" and argues that only the companies involved in activities that destroy tropical rain forests and the governments of countries allowing this destruction of their forests can be expected to take action.

Task 8, 2028

> *We can start with a lot of small things which would help. The exercise book's for example are mostly made of this rain forests. So we can buy recicled paper.*

Task 8, 3015

> *In my opinion it is not true, because I can do nothing, only the firms and their chefs can stop the destruction, if they had bought it from the also guilty government.*

Finally, performances that scored a 1, either expressed the opinion that one can personally contribute to the protection, but *did not explain how one can contribute personally*, or they argued that *one cannot contribute and offered limited support* for this opinion. In response 2020 below, for instance, the option "true" was ticked. However, the written text itself does not contain any suggestions in terms of how one can contribute. In terms of the second text below (3009), the option "absolutely not true" was chosen. The argument given in support of this opinion is, however, more vague than the argument presented in response 3015 above.

Task 8, 2020

> *I think so, because there are not just trees destroyed. The living area of animals gets destroyed and I'm sure nobody would like diggers demolishing his house. If the rain forest gets destroyed the whole climate would change dramatically.*

Task8, 3009

> *Because I can't do anything to save the rainforest myself and so as a single person you can't provide anything for this.*

As in the case of task 7, the responses to this task also point to the need to further develop learners' ability to argue in a detailed way. The responses on a level 3 and lower also indicate that learners' ability to present sophisticated arguments on complex matters need more attention, since some of the arguments reflected a simplistic view of the issues involved in protecting the tropical rain forest.

Few responses expressed the opinion that one cannot contribute to the protection of the tropical rain forest personally, suggesting that the geography training the learners received had been rather successful in terms of developing a particular subject position. Of course, the concern for the protection of the tropical rain forest has also been promoted by the media and even by advertising campaigns run by certain

manufacturers, such as the beer company some learners referred to in their text. The development of this specific subject position can therefore probably also be ascribed to factors other than merely the pedagogical discourse around the tropical rain forest in the geography classroom.

In the following section, the results of the scoring of the completeness and correctness of the meaning constructed are discussed. Since typical features in terms of the meaning constructed on each of the individual tasks have already been explained in this section, an outline of general features of the meaning constructed in learners' responses is given in section 5.3.

5.3 Completeness and correctness of the meaning constructed: presentation and analysis of the results

The table below summarises the results of the scoring of the correctness and completeness of the meaning constructed. These results were obtained through the application of the dichotomous and three-point scales. In calculating the frequency distributions, *the mean of the two closest scores* for each of the responses were used. As already explained in chapter 4, all responses were double rated. In some cases, the score obtained through a third, blind rating replaced one of the initial scores. The blind rating was carried out in cases where the first two raters' scores differed more than one full mark. This applied to all tasks, except task 2a, where a difference of one and a half marks was allowed, because raters had to use a much higher number of scales in rating this task than the other tasks. There were therefore more opportunities for variation in their decisions than for the other tasks. The percentages presented in the table were calculated using the number of learners who did the task. The totals of the respective tasks are also included. The highest score on task 7 and on task 8, namely 7,5 is higher than the respective totals (7) of each of these two tasks. As explained in the previous chapter, learners could score a mark that is higher than the total of the task, because marks were also given for content that was correct and relevant to the task, but that went beyond the minimum requirements of the task. The totals themselves, however, were only based on the minimum requirements.

Task	Lowest score	Percentage of performances receiving this score	Highest score	Percentage of performances receiving this score	Percentage of performances scoring half or less of the possible marks
1c	1 / 8	3,6%	6,75 / 8	1,2%	46,4%
2a	0 / 19,5	2,4%	14,5 / 19,5	1,2%	45,8%
3	0 / 10	30,1%	7 / 10	1,4%	86,3%
6	0 / 7	1,4%	6,75 / 7	2,4%	53,5%
7	1 / 7	1,3%	7,5 / 7	2,6%	30,3%
8	1,5 / 7	32,5%	7,5 / 7	1,3%	57,1%

Table 5.7: Completeness and correctness of the meaning constructed: summary of the results

In terms of *tasks 1c and 2a*, slightly fewer than half of the performances scored half or less than half of the possible marks. The difficulties posed by task 3 are again, as in the case of the results of the holistic scoring, reflected in the results of the rating focusing on the meaning constructed. A third of the performances scored no marks and slightly more than 85% of the performances scored half or less of the allotted marks.

The rating results for *task 7*, however, show that only one third of the learners scored half or less than half of the total mark. This clear difference to the results on the other tasks is ascribed to the fact that the dichotomous and three-point scales used to rate task 7 required that only one detailed and one basic explanation of the interrelationship between humans and the rain forest are presented. Alternatively, three basic explanations could be given. Quite a number of learners managed to do this.

Slightly more than half of the performances on *tasks 6 and 8* scored half or less then half of the total mark. The fact that almost one third of the performances on task 8 only

scored 1,5 marks, also indicates that learners had difficulty with this task. This is also reflected by the fact that 12% of the performances on this task scored a 0 in the holistic rating. The problems learners had with this task are ascribed to the fact that the task was at the end of a 90-minute test. It is assumed that learners might have suffered from fatigue, causing a loss of concentration. Many learners for instance only ticked an option on the scale, without giving any verbal support. In rating the correctness and completeness of the content, half a mark was given if learners simply ticked an option on the scale and a full mark was then also given if they ticked one of the options "absolutely true", "true" "partly true, partly not true". It was thus possible to score 1,5 on this task without giving any verbal support.

The fact that slightly fewer or slightly more than half of the performances on four of the six tasks scored half or less of the possible marks and that only 15% of the performances on task 3 scored more than half of the marks point to *problems in terms of the correctness and completeness of the meaning constructed in learners' responses*. In the previous section, these problems in terms of content have already been identified as they occurred per task. In this section, a general overview will be given of deficiencies in terms of firstly the completeness of meaning and secondly the correctness of meaning constructed in the learners' responses. A selection of responses to the tasks are presented in order to illustrate the problems.

5.3.1 Incompleteness of the meaning constructed

Central aspects, as well as *supporting details were often lacking*. Using task 1c as an example to illustrate problems in this respect, the main aspects to be included in a response to this task were an identification of the type of climate and a description of the temperature and the precipitation. In identifying the climate as tropical and as humid, there also had to be a reference to the fact that the precipitation is higher than the evaporation. The descriptions of the temperature and precipitation had to include the following details: it had to be stated that Kisangani has a high average temperature of 25°C that remains constant throughout the year. In terms of the precipitation, it had to be pointed out that the total annual rainfall is high, namely 1804 mm and that there are two peaks in the rainfall.

A comparison between two responses to task 1c illustrates problems in terms of the comprehensiveness of the meaning constructed. The content of the first response (2005) below, which scored 6,75 out of 8 on task 1c, is comprehensive and detailed. The response did not get full marks, because of the identification of the climate as "very tropical" and "very humid". Furthermore, it is not explained that the precipitation is higher than the evaporation. In comparison to the second response (3004) below, the comprehensiveness of the meaning constructed in the first response becomes clear, though. In addition to the fact that the climate is identified incorrectly in the second response, reference is only made to temperature, whereas no information about precipitation is included. Furthermore, the reference to the temperature only explains that it is constant, but not that it is high. This second response scored 1.

Task 1c, 2005

> *Kisangani's climate is very tropical. It has nearly the same temperature of 25°C all through the year. It is a very humid climate and most rainfall comes down from August to November, most in October. Another very wet period from March to June, where most rain falls in April.*
>
> *The avarage annual temperature is 25°C and the avarage annual precipitation is 1804mm.*

Task 1c, 3004

> *Kisangani has got a subtropical climate, because there is no unregularity in the heat all the year.*

The incomprehensiveness of the content in many cases probably resulted from *problems in terms of interaction with the input material*. In the case of tasks 1c, 2a and 6 the input material contained references to central aspects that had to be included. The climate graph presented information about temperature and precipitation, the diagram depicting the layering and living conditions in the ecosystem included references to all three layers, as well as to all four factors influencing the living conditions and the definition

of sustainability referred to all four types of goals that had to be applied in evaluating the suggestion. Furthermore, the input material presented in tasks 1c, 2a and 3 includes references to measurable phenomena, such as the temperature in degrees Celsius, the height of the different layers of the vegetation in meters or the size of rain forest destroyed in km^2. The inclusion of these details leads to a higher level of precision in the meaning constructed, which is a feature of the subject-specific discourse of geography that reflects the nature of one of its central activities, namely exact measurement.

The omission of central aspects and supporting details is an aspect of responses to all of the tasks. In the responses to tasks 7 and 8, where explanations and arguments had to be presented, the responses showed features particular to these two tasks. Many responses only included a *limited number of explanations and arguments* as support for the view taken and the *explanations and arguments themselves often showed little detail*. Whereas the first response (2006) below, which scored no less than 7,5 on task 7, includes four references to the international significance of the tropical rain forest, the second response (2009) only points out that the destruction will have consequences. However, it is not explained what these consequences are. This response scored 1,5 out of 7. The third response (2014), which scored 3,5, does refer to a specific consequence of the destruction of the tropical rain forest, namely natural catastrophes. However, it is not explained how the destruction leads to natural catastrophes and only this one consequence is listed.

Task 7, 2006

> *I think that this picture is true, because if the people destroy the rain forest the whole climate will change and that concerns us all. The greenhouse effect will grow. The rain forest is very important, for the climate and many special species live there. There are also many people who depend on the forest and can't exist without it. Without the forest our live will change, too and so the picture shows the truth like the Germans says: Wir sägen an unserem eigenen Ast.*

Task 7, 2009

> *If the rain forest would be destroyed the whole world must live with the consequences. It's hardly possible to live with the consequences.*

Task 7, 2014

> *The most problems of the rain forest come from people. They destroyed the environment and if we people don't stop destroying and ignoring the warn signals, we will have a world with much natural catastrophs.*

In the first response (3024) to task 8 below, which scored full marks (7 out of 7), there is a rather detailed explanation of how one can contribute to the protection of the tropical rain forest. It is pointed out that one can join an environmentalist group and an example of such a group is given. Furthermore, examples are given of the actions such groups can take. It is, however, also explained that one is limited in terms of one's ability to contribute to the protection of the tropical rain forest. In the second response (2009), on the other hand, it is only explained that one can help by donating money. There is no reference to whom one can donate the money and no explanation of how this money can be used to protect the tropical rain forest. This response scored 3 (half a mark was given for the fact that an opinion was indicated on the scale in the answering booklet, a further half a mark was given for the fact that some form of support was provided, one mark was given, because one of the options "absolutely true", "true" or "partly true, partly not true" was ticked and one mark was given for the inclusion of one basic argument).

Task 8, 3024

> *I think that this statement is on the one side true because if we engaged us for helping the rain forests we can do this in a group like "WWF" or others. With these groups we can demonstrate for the support of the government to the forests and we can collect money to support the people and animals in the forests to give them new homes or something like that. On the other side I don't think that the statement is true, because we can see on the statistics and the tables given on the pages before that the groups and organisations haven't a great effect on reaching their aims to support the not cutting down the forests.*
> *But a little help is more than nothing.*

Task 8, 2009

> *I think everybody can do something for the rain forest, e.g. spend money.*

In addition to the lack of details, *the argumentation in most learners' texts also showed a tendency towards oversimplification*. Response 2009 to task 8 above for instance illustrates this lack of sophistication resulting from an exclusive focus on the ways in which a person can contribute to the protection of the tropical rain forest, without referring to the ways in which individuals are limited in their scope of action.

5.3.2 Incorrectness of the meaning constructed

The *information included in responses to tasks 1c and 2a were often incorrect*. In response 3011 to task 1c below for instance, there is a confusion between climate and temperature, there are incorrect references to seasons, the precipitation from May to July is incorrectly described as "very low" and the unit of measurement for precipitation is incorrectly indicated as meters.

Task 1c, 3011

> *The climate graph for Kisangani is/remain constant. Kisangani has a average climate of 25°C all over the year. It is near the equator and located in the tropical rain forest in the Mid of Africa. Kisangani has a relief about 460m and the amount of precipitation is in winter and summer very low (about 100m) and in Spring and autumn exspecially very high (about 220 m at the highest point).*

In the following two responses to task 2a, the description of the living conditions in the tropical rain forest are incorrect from a subject-specific point of view. In the first response (5007), the tropical rain forest seems to be evaluated in terms of its suitability as a living space for humans. Rather than describing the layering and explaining how the temperature, humidity and incidence of light in each layer affect the vegetation and number of animals, the tropical rain forest is presented as an almost hostile living environment. The response does not reflect an understanding that there are different layers in the vegetation in the tropical rain forest and that different species need different living conditions. This response received no marks. In the second response (4005) below, the living conditions in each layer are reduced to the amount of nutrients in the layer and the affect this has on the vegetation. The response does not reflect an

understanding that different factors constitute the living conditions and that not only plant, but also animal life is affected by it. Furthermore, the view of the nutrient cycle presented is not correct from a subject-specific point of view. The nutrients are presented as independent entities which regulate their own release and uptake in the ecosystem. It is not explained that they are stored in plants. This response scored 3 out of 19,5 possible marks, due to the correct identification of the three layers.

Task 2a, 5007

> *The living conditions in the rain forest are very hard, because there are only very little incidence of light and temperature. I can imaging that it is exhausting if there's always such a high air humidity.*

Task 2a, 4005

> *The Giant trees have the biggest roots of all trees. So they take the nutrients out of the soil. The living conditions aren't the best because the nutrients fall down after some time in the lower tree layers. In the Tree top layer the living conditions are better than at the giant trees. A reason for this are the downfalling nutrients. But after time the nutrients will drop to the soil so at the bush and shrub layer the living conditions are the best. In summary you can say that the living conditions rise from the layer of the giant trees to the layer of bushes and shrubs because there are all the nutrients.*

The *views presented in responses to tasks 3 and 6 were often incorrect from a subject-based point of view*. This is ascribed to *problems in terms of the interpretation of the input material*. Whereas the first response (5014) below reflects a correct interpretation of the input material presented in task 3, the second (3005) reflects an incorrect view of the extent of the destruction of the tropical rain forest on the different continents. In addition to incorrectly stating that the destruction of the tropical rain forest is highest in Asia, the response includes content that is not relevant to the task set. Response 5014 scored 7 out of 10. Response 3005 was not awarded any marks.

Task 3, 5014

> *It's righ that Asia has cut off the highest percentage of the rain forest, but in America they cut twice the amount of wood as in Asia. So in the number of trees America destroy more wood as Asia.*
> *But Asia had less trees than America, so if the cut the same amount of wood the American forest stay longer there.*

Task 3, 3005

> *The destruction of the rain forest is too high in every part of the world, but in Asia it's the highest one. This cannot go on like this, because the tropical rainforests produce a main part of the oxigyn of the world and we cannot survive without oxigyn. The destruction also causes the green house effect, because burning down the trees causes a lot of smoke.*

Many of the responses to task 6 failed to explain that the suggestion to turn the remaining forest areas into national parks is problematic and instead tended to focus on only one or two of the goals identified at the world summit of Rio de Janeiro. It is assumed that the problems occurring in learners' responses to this task result from the fact that they did not correctly apply the definition of sustainability presented in the task in evaluating the suggestion. All four goals are mentioned in the definition and the definition does point out that these goals need to be balanced. The first response (5011) below scored 6,75 out of 7, because it reflects an understanding that the suggestion is not optimal and refers to three of the four goals included in the concept of sustainability (no reference is made to social goals). The second response (2025), on the other hand, evaluates the suggestion positively. Reference is made to ecological and economic issues, but economic matters are presented as a purely negative aspect. "People who have concerns" in the third sentence, refers to company owners ("Konzern" in German refers to a group of companies). The response thus does not reflect an understanding that economic goals should also be taken into consideration and that a balance should be found between the different goals. This response scored a 2 on task 6.

Task 6, 5011

> *Basically, I like the suggestion of conservation areas but according to Agenda 21 ecological and economic goals should be treated as equal. But the suggestion*

> *would favour the ecological goals as the economic goals would be completely abandoned. But I think big parts of the rain forest have to be conserved because future generation will have to depend on the tropical rain forest. Maybe conservation areas are the best way to do so.*

Task 6, 2025

> *I think it's important to care about the Umwelt. So this idea in this text is a great one. Because when the rainforests can be left to itselfs again, so they can't be destroyed by people who have concerns. This would be natural and everything would be okay.*

Finally, *one aspect in which the majority of responses followed the conventions of the subject community is in terms of views presented in tasks 7 and 8*. Although only a limited number of supporting arguments were presented in most cases and the arguments tended to remain vague, the majority of responses to task 7 expressed the opinion that the tropical rain forest is internationally significant. At task 8, most responses expressed the opinion that one can, at least to some extent, contribute to the protection of the tropical rain forest. As already explained in the section on the results of the holistic rating, task 8 is concerned with assessing the extent to which the responses reflect certain subject-specific values, namely that the tropical rain forest is important and that one has a personal responsibility to contribute to its protection. The majority of responses reflected this stance. In addition to the effect the pedagogical discourse probably had in developing these subject positions, it is assumed that the attention the tropical rain forest has been enjoying in the international media over that past few years also contributed to developing these views on the value of the rain forest.

The results of the scoring of the correctness and completeness of the meaning constructed, as the results of the holistic scoring of the appropriateness of the discourse function indicate that the majority of learners have not yet mastered the subject-specific thematic patterns concerned with the tropical rain forest. The results also point to the need to pay attention to the development of these learners' ability to argue in a detailed and sophisticated way. Finally, the results suggest that these learners need more training in analysing input material.

In the final section, the results of the analytic rating of the textual presentation of meaning is discussed. The results on the scale concerned with the sufficient and correct use of subject-specific vocabulary also point to the fact that these learners are still in the process of developing their knowledge of the subject's thematic patterns.

5.4 Presentation and analysis of results of analytic scoring: appropriateness of the textual realisation of meaning

In calculating the frequency distributions presented in the table below, the means of learners' performances on each of the scales over all six tasks were used. The table is followed by a list with the seven scale titles.

Scale	**Lowest score**	**Percentage of learners whose overall performance received this score**	**Highest score**	**Percentage of learners whose overall performance received this score**	**Mode(s)**	**Percentage of learners whose overall performance received this score**
1.1	2,5	1,2%	4	17,9%	3,67 4	17,9%
1.2	2,8	1,2%	4	11,9%	3,5	16,7%
1.3	2,33	1,2%	4	14,3%	3,5 3,67 4	14,3%
2.1	1	2,4%	2,8	1,2%	2	22,6%
2.2	1,17	1,2%	3,5	1,2%	2	21,4%
2.3	2	13,1%	4	1,2%	2,33	14,3%
2.4	2	4,8%	3,5	1,2%	3	26,7%

Table 5.8: Appropriateness of the textual realisation of meaning: summary of the results

Effectiveness of the textual organisation of meaning:

Scale 1.1: Effective ordering of structural units in the text

Scale 1.2: Effective organisation of meaning into different structural units

Scale 1.3: Effective linking of structural units

Appropriateness of the language use:

Scale 2.1: Sufficient and correct use of subject-specific terms and expressions

Scale 2.2: Sufficient use of formal language and clear, succinct formulations

Scale 2.3: Sufficient use of general academic vocabulary and structural units realising semantic relations

Scale 2.4: Correctness of grammar, vocabulary and punctuation

The results show that *more problems occur in terms of the appropriateness of the language use than in the area of structuring and linking*. Whereas the lowest scores on the first three scales are all higher than 2, the lowest scores on scales 2.1 and 2.2 are 1 and 1,17 respectively and 2 on the two remaining scales used to assess the appropriateness of the language use. Furthermore, the highest score of 4 was achieved on all three of the first scales and on scale 2.3. The highest score on the scale used to rate the sufficient and correct use of subject-specific terms and expressions (2.1) is clearly lower (2,8). The highest score on each of the two remaining scales (2.2 and 2.4) is 3,5. The percentages of learners receiving the highest score for their performances on each of the scales are also clearly higher on the first three scales, than on the last four. Finally, the modes of the scores on the three scales used to rate aspects related to structuring and linking are higher that the modes of the scores on the four scales used to rate aspects related to the appropriateness of the language use. These score distributions can be ascribed to the fact that the competences needed to structure and link a text effectively are to a large extent also developed in the L1. The appropriateness of the language use is, however, solely dependent on the L2 competence.

The most common problems in terms of the textual realisation of meaning found in the performances will now be discussed in more detail and illustrated by means of examples from learners' responses.

5.4.1 Problems in terms of structuring and linking

The mini-genres the tasks require learners to produce and the thematic patterns each of the tasks involves, create the basis for *making decisions about how effectively meaning is organized and linked in a response*. By organizing meaning into different structural units, e.g. sentence parts, sentences and paragraphs and by presenting the structural units in a particular order, the *different defining parts of a specific mini-genre can be demarcated* in conventional ways and *semantic relations between thematic items can be reflected*.

In an argument, for example, it is conventional for the opinion to be presented either at the beginning of the text, followed by support, or at the end of the text, as a conclusion reached on the basis of the evidence provided. Presenting the opinion in one sentence and each argument in another is one possible way of distinguishing the different stages of the genre.

Furthermore, the particular way in which meaning is organized in a text can reflect the standardized semantic relations between different items in a thematic pattern. Taking the example of the thematic items climate and its defining features, temperature, precipitation and humidity as an example, the semantic relations can ideally be realised in one of two ways. Since climate is the overall and abstract category and the result of a particular combination of temperature, precipitation and humidity, meaning about the type of climate can be presented first, followed by detailed references to how high the temperature, precipitation and humidity are. Alternatively, the aspects of a specific climate, thus the temperature, precipitation and humidity can be described first, followed by a reference to the type of climate.

A written response to task 1c that presents meaning about temperature in the first sentence, then refers to the type of climate in the second, before presenting information about the precipitation in the third, fails to use the potential of ordering sentences in a specific way in order to realise the nature of the semantic relations between these thematic items. The organization of meaning in this order does not reflect the hierarchical nature of the relation between the type of climate, temperature and precipitation.

Another example of the way in which the ordering of structural units in a text contributes to reflecting the relations between the thematic items involved, is a possible organization of meaning about the layers of the tropical rain forest: the giant trees, tree top layer with closed canopy and bush and shrub layer. The differences between these layers and their physical relation to each other can for instance be reflected in a text by organizing meaning about each of the layers into a different paragraph and ordering the paragraphs in a way that reflects the vertical position of the layers: at the top, in the middle or at the bottom. By starting with the giant trees, then describing the tree top layer with closed canopy and then the bush and shrub layer, the physical organization of these layers is reflected in the ordering of the paragraphs in the text.

The *use of cohesive devices* serve to link ideational meanings constructed in different structural units and to indicate how the meanings constructed are organized as text. The production of texts that realise more complex thematic patterns involving a number of thematic items and semantic relations, as well as the production of texts that realise mini-genres with more than one defining part, requires the use of linking devices between different structural units. Texts in which cohesive devices are not used to a sufficient extent fail to connect the meanings realised in the different structural units. Consequently, these texts do not realise all of the semantic relations between items in the thematic pattern and become choppy.

The most common problems in terms of structuring and linking in the learners' responses can be summarized as follows: the *organization of sentences did not always reflect the standardised thematic patterns* between the items of the thematic patterns with which the tasks were concerned. This was firstly caused by the position of a particular sentence in the text as a whole and secondly by the fact that closely interrelated meanings were either presented in different sentences, or logically distinguishable meanings were presented in the same sentence. Secondly, *meanings were sometimes repeated* in two different sentences. Finally, in some texts *a lack of cohesive devices* caused the texts to be choppy and not to fully reflect the logic of the thematic pattern underlying the meaning realised in the text.

In the response (2017) to task 1c below, the text switches between references to temperature and precipitation, rather than to first describe temperature and then precipitation. Furthermore, cohesive devices are not used sufficiently to realise the semantic relations between the meanings constructed. The text starts with an identification of the type of climate. This is followed by two sentences, each giving information about the average annual temperature and precipitation respectively. The following sentence then provides information about the constancy of the temperature, before the variation in the precipitation is described in the last two sentences. The description of the variation in the precipitation also switches between references to months during which the precipitation is higher and months during which it is lower.

This specific organization of meaning might have been chosen in order to first present information about the annual values and then to indicate that the temperature remains constant whereas the precipitation varies. However, the meaning constructed about the annual values are presented in two separate, unlinked sentences. Furthermore, the contrast between the constancy of the temperature and the variation in the precipitation is not explicitly expressed. The possible logic underlying the text is thus not explicated.

Task 1c, 2017

> *Kisangani has a very humid and hot climate. The average annual temperature is about 25°C. The average annual rainfall is about 1804 mm per m^2. As you can see in the climate graph the temperature is nearly constant 25°C. In July there is the smallest rainfall (100 mm), in October the most (230 mm). Another very humid month is March, and an other quite dry month is January.*

The following response (2009) to task 1c is not only rather choppy, because cohesive devices are not used to a sufficient extent, but meaning about the constancy of the temperature is also repeated. This problem is further aggravated by the fact that it seems as if the concepts climate and temperature are used interchangeably, which is an error in terms of the correct use of subject-specific terminology. By realising meaning about the so-called "constancy of the climate" and constancy of the temperature in one sentence, a close interrelationship between these two concepts is suggested. Meaning about the constancy of the climate is then repeated in the following sentence, which also includes

more precise information about the temperature in degrees Celsius. The last sentence then presents information about the precipitation, without indicating any relation between the concepts climate and temperature referred to in the previous sentences.

Task 1c, 2009

> *Kisangani has a very hot climate and the climate is also humid. The climate is warm because Kisangani is in a savanna. The climate doesn't change in the months, it has always the same temperatures. The temperature is always 25°C. The precipitation is in July very low and in October very high.*

In the following response (2006) to task 8, in which the option "true" was ticked on the scale, the opinion that one can personally contribute to the protection of the tropical rain forest is presented twice. This is done first at the beginning and then again in the middle of the text. Furthermore, an argument explaining in which way one is limited in contributing as an individual, is presented between two arguments explaining how one can contribute. This organization of meaning does not fully make use of the potential of structuring and linking the text in such a way that the opinion held is highlighted and the arguments for this opinion are grouped together, before those that relativise them are presented. Furthermore, due to this organization of meaning the relation between the two arguments supporting the opinion does not become clear. The first argument that one can contribute by joining an environmental organization should have been followed directly by the argument that one can donate money, should one not have time to join such an organization. Furthermore, the argument that one can donate money is presented as one long sentence, although there are two clearly distinguishable parts. The sentence starts with an explanation of what one can do and then also includes an evaluation of this form of contribution. Presenting the explanation in one sentence and the evaluation in another would have demarcated these two functions more clearly.

Task 8, 2006

> *I think it is true, because it is important that everybody do something for the forest. If I begin to do something maybe others follow me and we can really move something in the protection. There are many organizations you can be part in, which want to help the rain forest. So I can personally contribute to the protection. But I can only do that, if I don't do that alone, so I really need the*

> *support of other people. I can also spend money to do that, when I don't have the time, and this is of course better than just to sit at home and say: If I can't be part in a help organization, because I don't have the time, I can't do anything.*

Although some problems in terms of the structuring and linking of the texts could be found, the learners' language use showed clearly more need for focused attention. The problems in terms of the language use occurring in the learners' performances will now be discussed and illustrated.

5.4.1 Problems in terms of language use

The results clearly show that in terms of the appropriateness of the language use, *the sufficient use of subject-specific terms and expressions is the most problematic aspect*. Problems in this respect are firstly the *use of commonsense terms* rather than subject-specific terms, for example using the expression "hot and wet" to refer to Kisangani's climate rather than the subject-specific term "tropical". The second type of problem is the *incorrect use of subject-specific terms*, for instance using the word "climate" to refer to temperature, or using the terms "seasons", "summer" and "winter" in describing the diurnal climate of the tropical rain forest. Furthermore, the learners often made *collocation errors*, *misspelled subject-specific terms* or used the *German*, rather than the English subject-specific term.

The second aspect of learners' use of the register that clearly shows problems is *the sufficient use of formal language and clear, succinct formulations*. Problems in this respect can firstly be related to the fact that *subject-specific terms*, which distil meaning, were *not used to a sufficient extent*.

Secondly, problems here are also connected to the fact that learners *did not make use of details in the form of information about measurable aspects* presented in the input material. Whereas references to temperature and amount of precipitation for instance often remained vague in responses to task 1c, references to temperature, incidence of light, and the height of the trees in responses to task 2a often lacked clarity. Many responses to task 3 also failed to be explicit and clear in terms of references to a point in time and to the decline in forest cover on the different continents. Furthermore,

comparisons often remained vague, because the different aspects that were compared to each other were not always explicitly mentioned. This for instance occurred were comparisons were drawn between the living conditions in the different layers of the tropical rain forest or between the decline in forest cover on the different continents. Also, an array of forms associated with an informal, spoken register, such as s*hort forms* (isn't, doesn't, I'm etc.), *phrasal verbs* and colloquialisms were used. In addition, the extensive use of the active, rather than the passive voice negatively affected the realisation of a subject-specific register.

In terms of *the use of general academic language*, some responses suffered from a lack of variation in the range of vocabulary and grammar structures used. In a number of responses only simple sentences and words belonging to a basic vocabulary were constantly used. Logic relations were often also not explicated or only explained in the simplest of terms. Furthermore, many of the responses would include references to information extracted from the input material, without referring to the sources of information.

Finally, *errors in terms of grammar, vocabulary, spelling and punctuation* occurred in almost all texts. These include a wide range of minor and major errors. However, there was only a limited number of cases in which the errors were so severe that meaning was obscured.

In the following response (3010) to task 1c, there are a number of problems in terms of the use of subject-specific terminology. Firstly, the subject-specific term "climate" is used incorrectly to refer to temperature. There are also references to seasons, which is incorrect in terms of describing the climate of the tropical rain forest and the word "coordinates" is used incorrectly to refer to Kisangani's latitude and longitude. Thirdly, the subject-specific term "precipitation" is misspelled once and "degree" is used incorrectly in the singular form. Furthermore, collocations containing subject-specific terms are not always used correctly, e.g. the use of the verb "to be" in explaining the variation in the amount of precipitation and the reference to "precipitation of all month (sic.)" instead of "total annual precipitation". In addition, the text shows a number of vocabulary errors, such as the use of the adjective "continuous", instead of the adverb in describing the adjective "high", the incorrect use of "also" to introduce the third

sentence and the incorrect expression "is on a high at" in referring to Kisangani's altitude. "October" is also misspelled.

Task 1c, 3010

> *The climate of Kisangani is continuous high (about 26 degree) from January to December. The precepitation is much higher, so all months are humid. Most of the rain is on Octobre (about 230mm) and in the summer months (e.g. July) there is not so much precipitation (about 100 mm). Also in January, the precipitation is about 100 mm. The annual temperature is 25°C and the precipitation over all month is 1804mm, which is a lot. Kisangani is on a high at about 460m and has the coordinates 1°N/25°E.*

The following response (3007) to task 2a below, shows a number of problems in terms of the use of subject-specific terms. Firstly, the reference to the highest layer in the vegetation, namely the giant trees, is used incorrectly in the singular form. Secondly, the unidiomatic expression "takes the most light" is used in referring to the incidence of light occurring in this layer. Thirdly, "hot temperature", instead of the collocation "high temperature" is used. Furthermore, incorrect verbs are constantly used in describing the nutrient cycle. It is incorrectly stated that the leaves of the giant trees "get many nutrients" instead of stating that they "store many nutrients". It is also incorrectly stated that the tree top layer with closed canopy and the bush and shrub layer "get these nutrients" and that the nutrients "go in" the soil. The two other layers identified in the vegetation are not correctly named and are only referred to as "the lower trees". The term "ground" is used, instead of the term "soil". The meaning constructed also remains rather vague, because no detailed information about the height of the trees in metres, the temperature in degrees Celsius and the incidence of light in percentages is included. There are also a number of vocabulary errors. The personal pronoun "he" is used incorrectly to refer to "the giant tree" and the singular pronoun "it" is used incorrectly to as a substitute for the plural noun "nutrients". Furthermore, the phrasal verb "fall down" is used incorrectly. Finally, there is a concord error in the last sentence.

Task 2, 3007

> *The giant tree is the highest tree in the rain forest. In figure 3 you can see that he takes the most light. So he can grow very well. Because of the low air*

> *humidity and the hot temperature, the leaves get many nutrients. If these leaves fall down the animals and the lower trees get these nutrients. Later the nutrients go in the ground and the giant tree use it again.*

In response 3011 to task 3 below, the meaning constructed also remains particularly vague. This is firstly the result of the fact that information about the decline of the tropical rain forest presented in percentages and in square kilometers in the input material is not included. There are also no references to the periods of time involved. This is secondly caused by the use of the word "reason" without any clear reference to the aspect that this so-called reason should explain. This vagueness is caused by the use of the pronoun "this", as well as by the determiner "one" and the adjective "other" in referring to the noun "reason". Furthermore, the lack of clarity in the text is caused by errors in terms of the vocabulary used, such as the incorrect use of the noun "proportion", the verb "participated" and the conjunction "because (of)". The meaning of the word "useful", which is misspelled, also remains obscured within the context of the text. The fact that no commas are used in the second sentence and that no comma is used before the phrase "comparing Asia to Africa and America" in the third sentence, also adversely affects the text's clarity. Finally, the last sentence also contains a concord error.

Task3, 3011

> *As we can see in table 1, the decline of tropical rain forest has increased. To see it in the proportion Asia has not participated in the decline because e.g. deforestation. This is one reason and the other reason is comparing Asia to Africa and America, Asia have more higher trees which get more light and which are in general more usefull.*

The following response (2014) to task 3 does not adhere to a number of the conventions of subject-specific language use either. Through the generalised and direct references to e.g. "The Asians", the use of the active voice and the personal pronoun "they", the text realises a stereotyped perception of Asians and Africans. In terms of describing the decline of the tropical rain forest in Asia, the focus is on a group of people seen to be responsible for the destruction of the tropical rain forest, rather than on the extent of the destruction itself. Thus, there is a concern with those who act, rather than with the

results of the actions. Furthermore, short forms are used and except for the fourth sentence, no complex sentences are used. This use of simple sentences and the lack of cohesive devices cause the text to be choppy. In other words, the text also shows problems in the area of structuring and linking. There are also a number of vocabulary errors. The word "condition" is used incorrectly in the plural form and the adjective "industrial" is used instead of the adverb in combination with the adjective "interested". The phrase "industrially interested" is also unidiomatic in English. In addition, the comparison between Asia and Africa remains vague, because it contains no explicit reference to Asia. This constitutes a problem in terms of the use of clear formulations. Finally, the personal pronoun "they" instead of the possessive pronoun "their" is used and the nominal phrase "the way of" is unidiomatic.

Task3, 2014

> *The Asians use the wood for their industry. They don't care very much about the conditions of the rain forest. They only think about money. They ignore the danger of natural catastrophes. Africa isn't so industrial interested and they care about they nature. That's much better than the way of the Asians.*

In the following response (3023) to task 7, subject-specific terms are used in German. Furthermore, the expression "the greenhouse effect rises up" contains a phrasal verb that does not collocate with the subject-specific term "greenhouse effect". The preposition "of" in the first sentence does not collocate with "the trees" and "the rain forest". In terms of the use of the phrasal verb "to make out" in the second sentence, the reference to oxygen incorrectly follows the reference to carbon dioxide, so that the position of the prepositional phrase "out of" is incorrect. There is also a concord error in the second sentence. The phrase "The modern human being" in the third sentence is not idiomatic and is followed by the verb "produce", leading to a concord error. Furthermore, the position of the adverbial phrase "every day" is incorrect; the verb "to have" is used incorrectly in combination with "ecological catastrophe"; the article "a" instead of "an" is used before the adjective "ecological" and the preposition "at" is used incorrectly in combination with the noun "end". This last sentence also reflects problems in terms of structuring and linking, since it includes references to both the problematic behaviour of individuals and companies and the possible results of this behaviour. The logic relations between these different aspects are, however, not explicated.

Task 7, 3023

> *The statement is true because the trees of the rain forests make out of Kohlenstoffdioxid Sauerstoff. But when the forest are burned down the greenhouse effect rises up. The modern human being produce so much Kohlenstoffdioxid with driving cars and having big firms and the rain forests get every day smaller it will have a ecological catastrophe at the end.*

The problems found in terms of the textual presentation of meaning indicate that more focused attention should have been paid to the development of these learners' ability to use the subject-specific register. In the following chapter, general suggestions for the development of learners' subject-specific written discourse competence in the bilingual classroom are outlined.

5.5 Conclusion

In interpreting the results it should naturally be taken into account that these learners were in the tenth grade when they wrote the geography competence tests. One can therefore expect that their subject-specific competences are still developing. The results point to areas in which the development of their subject-specific competences should be guided in a focused way. Firstly, the results clearly show that they have not mastered the thematic patterns related to the tropical rain forest that learners are expected to have mastered by the end of the tenth grade. This is reflected by the fact that the meaning constructed in the responses was often incorrect and incomplete.

The incorrectness and incompleteness of the meaning constructed is also ascribed to problems in terms of interacting with input material. In this regard, process research that can cast light on the exact nature of the problems learners are experiencing in extracting information from input texts and in combining meaning from different texts is needed.

The learners' argumentation skills also reveal the need for focused attention. In this regard, the ability to argue in a reflective way, taking different points of view into account, as well as the ability to argue in a detailed way, supporting arguments with

examples and making the different phases of the argument clear to the reader, should be addressed.

The results of the assessment of the textual presentation of meaning point to the need for more focused attention on the correct and sufficient use of subject-specific vocabulary and formal language. In this regard, studies that observe and analyse teachers' approaches towards developing the learners' subject-specific discourse competence are needed. More detailed and systematic studies on different aspects of the learners' writing, for instance on inappropriate stylistic features and on the vocabulary and grammar errors made, are also needed. Should the results of such studies point to systematic errors, corrective approaches could be designed and their effectiveness tested.

In the last chapter, an outline is given of ways in which the deficiencies found in the learners' subject-specific discourse competence in the context of this study can be addressed in the classroom. Furthermore, suggestions for further research into the task-based written discourse competence of bilingually trained learners are made.

6
Conclusion and perspectives

In this final chapter, I first present general suggestions about the development of learners' written discourse competence. These are followed by suggestions specifically directed towards the bilingually taught classroom. The suggestions are based on the insights gained through a social semiotic interpretation of the dynamics of subject-specific discourse and the results of the study. I then outline further research needs in terms of learners' subject-specific discourse competence. In the last part of this chapter, I give an overview of the study's intended theoretical and methodological contribution to the discourse on bilingual subject-specific training.

6.1 Developing learners' task-based subject-specific written discourse competence: suggestions for the classroom

A social semiotic and systemic functional linguistic approach has been followed in exploring the nature of subject-specific discourses in the context of this study. Applying social views about the nature of language and human meaning-making processes in subject-specific contexts has firstly made it possible to develop a definition of a task's required discourse function that integrates aspects of content and form into one concept. Secondly, operating with meaning and the processes and systems involved in its construction as central notions has made it possible to explain how knowledge of subject matter, methodological competence and subject-specific communicative competence interact in the process of constructing a response to a subject-specific task. Thirdly, this approach has enabled the development of rating instruments that can be used to identify aspects of learners' task-based subject-specific writing that do not adhere to the conventions of the subject's discourse. In a final step, insights gained from a social view of the nature of subject-specific discourses and the results of the assessment carried out in the context of this study are used to make suggestions for the classroom. In the first part, three general principles are identified that should guide all forms of subject-specific education. In the second part, specific suggestions are made for the bilingual geography classroom.

6.1.1 Demystify the discourse

Subject communities use language in ways that are different from the language of family life, so that many learners find the subject-specific register difficult to master. As explained in chapter 2, subject-specific registers do however present ways of using language that are central to self-regulatory participation in educational, professional, public-administrative and political discourses in industrialised cultures. Learners are social beings that need to be empowered to reach personal aims through successful participation in these discourses. The subject-specific teacher thus has a responsibility to consciously and actively develop learners' subject-specific discourse competence.

The first dimension of this responsibility is to *explicate the conventions of subject-specific language use*, as well as the *functions* these conventions have and the *values* they reflect. To illustrate: *the need for verbal explicitness and precision* in subject-specific discourses must be pointed out to learners. They must also be brought to understand that adhering to this convention firstly facilitates mutual meaning-making processes that are not dependent on shared knowledge that is confined to specific spatial, cultural and linguistic contexts. The second function of this convention, namely to reflect the nature of the subject's standardised meaning-making activities, in which specific procedures are used in measuring phenomena, should be explained. In other words, it must be made clear to learners that temperature for instance is measured in the same way by members of the subject community all over the world and that they report about their observations in a way that makes it possible to share information about these observations among themselves. In explaining the function of the need to be as verbally explicit and precise as possible, a central aim of the subject community's meaning-making processes thus becomes clear, namely to develop an understanding of experience by means of standardised forms of observation and to share this understanding with other members of the subject community.

Furthermore, the consequent *standardised nature of a subject-community's genres* as text structures that serve to realise specific functions in the communication between members of the subject community should be made clear. This means that simply teaching learners a standardised way of structuring a description, an explanation or an evaluation is not enough. On the contrary, it might be harmful to merely focus on the form, because these structures are then presented as ready-made containers into which

meaning is always poured in the same way. This is problematic, since it firstly presents an over-simplified view of the ways in which genres are used in discourse. It does not, for instance, acknowledge that different mini-genres, such as definitions, descriptions, explanations and evaluations are combined to form longer and more complex genres in a subject-community's discourse. Secondly, only teaching learners to mimic standardised and simplified text structures keeps them from developing the ability to effectively and creatively use these genres to achieve personal goals in subject-specific discourse, because these structures are made to seem inflexible.

The standardised nature of the subject's thematic patterns should also be brought to learners' attention. This implies that in teaching subject matter, learners should be given the opportunity to reflect about the ways in which the subject community organises experience differently from the models with which learners might be operating. In this respect, the thematic pattern should be broken down into discrete meanings that can be identified and analysed in terms of the thematic items involved in the pattern and the semantic relations between them. These networks should be presented to learners in different social semiotic forms of representation. The verbal explanation of the interrelations between humans and the rain forest for instance should also be realised in the form of a diagram. Learners should also be given the opportunity to construct thematic patterns in different social semiotic forms, verbalising a diagram, or creating a schematic representation of the teacher's explanation of a particular phenomenon for example.

Learners also need to understand *the role of subject-specific terminology in constructing meaning about a subject's thematic patterns*. Learners need to understand that subject-specific terminology serves as a way of distilling meaning and that it consequently allows effective exchange between members of the community. They also need to know that these terms reflect particular taxonomies or ways of organising experience that are accepted as true by the community. This means that subject-specific terms should not be seen and taught in isolation, but should be presented in networks that realise the semantic relations that exist between the individual terms. Learners should also be given opportunities to create such terminological networks themselves, so that they can reflect about the semantic relations between these thematic items. In this way, learners do not

only learn to use subject-specific terms correctly, but also explore the subject's thematic patterns and thus develop their knowledge of subject-specific matter.

A subject community's discourse requires learners to operate with models of experience that might be different from or more complex than those with which they have become familiar in other contexts. The community's use of social semiotic systems might also differ clearly from the ways in which these systems are used in other contexts. In this respect, the subject-specific teacher becomes a mediator who guides the learner in his or her initiation into the subject's activities. Consequently, *the mechanics of the subject's discourse need to be made clear* to the learner. This firstly means *drawing attention to the conventions* and giving learners the opportunity to reflect about the ways in which the subject's discourse differs from other discourses. However, learners can only be empowered to participate in a subject community's discourse in an effective and appropriate way if they also know and understand *the functions and values that underlie the subject community's activities* – including its use of social semiotic systems. A lack of understanding of these functions and values implies that a learner is left to merely mimic the conventions, without having the opportunity to steer his or her participation in the discourse. The learner can thus not decide on a specific course of action and execute the plan by following the conventions and can therefore not participate in the subject's discourse in a self-regulatory way.

In order to be able to explicate the conventions and to make the functions and values that underlie them clear to learners, *teachers should of course personally understand the dynamics of the discourse of the subject*. This requires that training programs for teachers also acknowledge the role of the teacher as mediator whose responsibility it is to demystify the discourse of the subject.

6.1.2 Give learners ample opportunity to construct meaning

Learners need to be given *ample opportunity to independently realise the subject's thematic patterns in spoken interaction, spoken production and in writing*. Learners thus need to be given sufficient practice in using language as a social semiotic system in order to construct meaning in a subject-specific context. In this regard, not only frequency, but also quality is important. Learners should be required to produce

extended stretches of speech and writing in the subject-specific classroom, so that they can firstly explore the subject's thematic patterns and secondly the devices a language's discourse-semantics and lexico-grammar offer to realise these patterns. Tasks that require one-word answers, or that only require learners to focus on a limited number of thematic items and on simple semantic relations, firstly do not suffice as a way of mastering the subject's thematic patterns in their full dimension. Secondly, such tasks also cannot effectively develop learners' ability to structure and link meaning, since they only require a limited number of devices from a language's discourse-semantics and lexico-grammar to be used.

In requiring learners to produce extended oral and written responses, *learners should naturally be given ample opportunity to explore the thematic pattern involved and to plan their responses*. By for instance implementing tasks that require learners to create mind maps on a specific theme that has been dealt with in class and then to present their interpretations orally to the class, learners are given the opportunity to examine the thematic patterns involved. In giving detailed feedback on the interpretations realised by the learners, the teacher creates an opportunity for learners to compare their own conceptualisations to the standardised thematic pattern. In a third step, learners can then be required to produce a written response. Learners are thus given the opportunity to review their understanding of the thematic pattern yet again. Allowing learners time to plan and revise oral and written contributions involving more complex semantic networks permits a type of interaction with the subject's thematic patterns that is qualitatively different from constructing an immediate oral response to a question posed by the teacher.

Furthermore, by *relating the production of an extended oral or written response to a specific activity structure*, learners' ability to produce texts that effectively realise genres is developed. An example of such a task would be to require learners to analyse a climate graph and to give a detailed description of the climate. This combination of constructing ideational meaning, as the result of enacting a specific activity structure (analysing a climate graph) and realising the meaning in a particular genre (a description), facilitates the development of the ability to produce texts that appropriately

effectuate discourse functions. It also develops learners' understanding of the ways in which language is used to reflect the subject community's activities.

In order to master the subject-specific register, learners need to be given practice in using it. In order to develop this ability effectively, learners need regular practice in using the subject-specific register and need to construct extended responses involving more complex thematic patterns.

6.1.3 Give learners effective feedback

In supporting learners' ability to use the subject-specific register, teachers need to give learners clear feedback on the appropriateness of their spoken and written responses to subject-specific tasks. Naturally, time limitations make it impossible to give detailed feedback on the discourse-semantics and lexico-grammar of each spoken or written performance. Rather, *specific phases during a weekly or monthly training cycle should be reserved for a focus on the effectiveness and appropriateness of their speaking and writing*.

Giving learners feedback on the ways in which their texts are structured and linked and on their language use requires that subject-specific teachers view the ability to use different social semiotic systems - including language - appropriately as a central subject-specific competence. The development of this competence should therefore not be regarded as an add-on that can only receive attention once sufficient attention has been paid to the development of subject-specific knowledge or subject-specific methodological competence. Rather, *the role of language and other social semiotic systems in the construction of meaning about subject-specific phenomena and in the reflection of the nature of subject-specific activities needs to be acknowledged*. In other words, subject-specific teachers need to acknowledge that exchange about subject-specific content can only take place through the use of social semiotic systems and that using them to construct meaning is a central activity of subject communities.

In order to give effective feedback, subject-specific teachers need to *be critically aware of the learner's language use*. This entails that teachers select certain phases of classroom activity or written responses to specific tasks and react to instances of

language use that do not adhere to the conventions. The phases of classroom activity during which there will be a focus on the appropriateness of the language use and the written products that will be rated with a stronger focus on the use of the lexico-grammar and discourse semantics need to be announced. Learners thus need to be informed about the fact that there will be a more directed focus on language use. Learners themselves also need to be brought to an understanding of the importance of language use in subject-specific contexts and to view it as a central subject-specific competence.

In his or her reactions to learners' written and spoken responses, the subject-specific teacher needs to *give clear and focused feedback*. This entails that he or she identifies the specific aspects of language use that deviates from the conventions. Using clearly identified criteria, as for instance those used in the analytic scales applied for the purposes of this study, is helpful in identifying the exact nature of problems occurring in learners' use of the subject-specific register. Operating with a framework consisting of distinguishable concepts such as structuring and linking, appropriate use of subject-specific terminology, appropriateness of style and correctness of grammar and vocabulary simplifies the process of analysing learners' responses and precisely identifying the deviations. Learners should also be familiar with the criteria used, since they reflect general conventions of the subject's discourse. In knowing the criteria, learners can direct their own use of the lexico-grammar and discourse semantics. Thus, learners themselves are given tools with which they can reflect about their language use during the speaking and writing process, when aiming to develop their ability to use the subject-specific register.

In identifying aspects of the textual presentation of meaning that are not appropriate, subject-specific teachers should also explain why the way in which meaning has been realised is not acceptable in a subject-specific contexts. In other words, *form should be viewed as functional* and there should be a focus on the structure, as well as its use for specific purposes.

Effectively developing learners' ability to use the subject-specific register also requires that *attention is paid to the realisation of all three types of meanings: the ideational, the social and the textual*. All three of these types of meanings are constructed

simultaneously and as explained in the discussion on register in chapter 3, the inappropriate realisation of one of these three type of meanings negatively affects a text's coherence. Due to the perceived importance of subject matter in subject-specific contexts, there tends to be an exclusive focus on the appropriate realisation of ideational meaning, whereas the construction of textual and interpersonal meaning is neglected. However, as argued in chapters 2 and 3, the subject-specific classroom should prepare learners for participation in discourses in other institutionalised contexts, such as other educational, as well as professional and public-administrative contexts. In these contexts the ideational meaning constructed is usually clearly different from that of the subject-specific classroom, but there are similarities in terms of the expected textual and interpersonal meanings.

Paying attention to *textual meaning* implies reacting to the use of the lexico-grammar and discourse-semantics in a way that heavily relies on the possibility of feedback. This, for instance, includes the constant use of reference items without making the referent or participant clear, or the ordering of information that does not take a distinction between presumed new and familiar information into account. Focusing on the construction of textual meaning also entails challenging learners not to base the choices they make during the writing process on the assumption: "the teacher knows what I mean", but to be as explicit as possible. Using tasks in which a written text must be produced for a reader other than the teacher can facilitate the development of learners' abilities in this regard. An example of such a task would involve expecting learners to produce a response to a task that should serve to explain a particular phenomenon to readers of a popular science magazine.

The nature of the interpersonal relationships involved in interaction situated within the activities of institutions require the application of specific structures. Learners therefore need to receive feedback on language use that realises an orientation towards the reader that is inappropriate in contexts where the power relations, frequency of contact and nature of the affective involvement require the use of formal language. In other words, teachers need to make learners aware of the effects of attitudinal rather than neutral lexis and of the modal resources available for realising a specific orientation towards the meaning one is making.

The teacher's role as mediator who initiates learners into the discourse of the subject community and in so doing, lays the foundation for their ability to participate in discourses in other institutionalised contexts, requires the teacher to give learners clear, structured and comprehensive feedback on their written and spoken responses. Explicating the nature of the subject's discourse, giving learners ample opportunity to produce extended spoken and written products and giving feedback are important in developing learners' ability to use the subject-specific register in all subject-specific contexts.

In the next section, the focus will be on aspects that are seen as requiring special attention in the context of bilingual subject-specific training.

6.1.4 Suggestions for the bilingual geography classroom

The results of the assessment carried out in the context of this study and the inferences made about deficiencies in these learners' subject-specific task-based written discourse competence form the basis for the following suggestions: learners firstly need to be guided towards *careful analyses of tasks*. One way of approaching this analysis is to focus on the requirements of the task in terms of the meaning to be constructed and the textual presentation of this meaning, while also paying attention to the support the task offers one in terms of constructing this meaning. The following questions can guide a learner in identifying the requirements and support: (a) What exactly should the answer be about? (What is the thematic focus of the ideational meaning to be constructed?) (b) What form of support does the task give in terms of the content to be included? (Which input texts are used and in which way can they contribute to the construction of meaning?) (c) Which details are presented in the input material that can be used to make an answer more exact? (d) Which text type is required? (Which mini-genre is the most appropriate in order to realise the activities in semiotic form?)

Learners secondly need to learn how to *use the input material as a resource*. This involves identifying the support provided by the input material in terms of the ideational meaning to be constructed. The input material firstly provides support regarding the

thematic items to be included and semantic relations to be realised. The climate graph used in task 1c for instance, provides information about temperature and precipitation. This serves as an indication that information about both these concepts should be included in the answer. In task 2a, the diagram includes information on three layers and four factors involved in the living conditions, indicating that the meaning constructed in a response to the task should include all these thematic items. The same applies to task 6, where the definition of sustainability includes references to four aspects that need to be taken into account and that need to be balanced. Furthermore, the input material provides support in terms of detailed information that can be extracted and that used in the written response, in order to make the meaning constructed more exact. In addition, the input material may include subject-specific terms, as was the case with task 2a, which should be used in the response.

Thirdly, learners *need practice in constructing meaning by combining information from two different sources*, as for example required by task 3 and 6. In this respect, learners should also practice combining meaning on the basis of texts that are realised by different social semiotic systems, for example a verbal text and a bar chart. In constructing meaning on the basis of different texts, learners' ability to critically analyse the effect the different social semiotic presentations have on the meaning constructed should be developed. In task 3 for instance, the way in which the international destruction of the tropical rain forest was presented led to the construction of different meanings about where the extent of the destruction is the most severe.

Furthermore, *attention needs to be paid to the development of learners' ability to use subject-specific terms and expressions correctly*. This firstly entails that production tasks are preceded by preparatory exercises in which learners are given the opportunity to identify the different thematic items and analyse the semantic relations between them, for example by creating word webs. This secondly involves that learners are expected to use these terms and expressions in actual discourse where they have to be embedded in sentences.

The ability to use subject-specific terms correctly is a component of the ability to use the subject-specific register. Enabling learners' to use this register also involves *developing their ability to apply other resources from the discourse-semantics and*

lexico-grammar in order to produce a text in formal English. Learners should thus not only be expected to construct ideational meaning correctly, but also to construct interpersonal meaning appropriately. This involves that learners are not only expected to produce unprepared oral answers, but that they are also given the opportunity to produce planned oral and written responses in which they have time to reflect about the resources available to them in realising a subject-specific register. In addition to giving learners opportunity to practice using the subject-specific register, the bilingual subject-specific teacher naturally needs to give learners clear feedback on their attempts to use this register.

Finally, learners *need to be challenged to fully develop their arguments*. In other words, it is important that learners are expected to produce detailed arguments that do not remain on a vague and general level.

These suggestions in terms of the development of bilingual learners' task-based subject-specific written discourse competence are based on the results of this specific study. In the next section, suggestions for further research into subject-specific discourse competence are given.

6.2 Perspectives for further research

As it has been pointed out in the first and second chapter, research on bilingually trained learners' use of the subject-specific register is limited. In the past, studies tended to focus on the extent to which bilingual subject-specific training contributed to the development of learners' general competence in the foreign language. Furthermore, little is known about bilingually trained learners' subject-specific writing. Insights into these aspects are however needed in order to effectively develop learners' ability to use the subject-specific register and in doing so, to prepare them for participation in other discourses taking place in institutionalised contexts. Consequently, there is a need for further studies focusing on different aspects of bilingually trained learners subject-specific discourse competence and their task-based writing. A number of suggestions for possible research projects are listed below.

This study involved the written performances of 84 10th grade learners following geography in English in Lower Saxony. The results of this study firstly need to be compared to the results of studies involving 10th grade learners following geography in *other federal states* or the results of a study conducted on a *national level*. Secondly, further studies are needed on the task-based subject-specific writing of learners in *the different grades* and studies on the task-based subject-specific written discourse competence of learners following bilingual subject-specific education in *other subjects* are needed.

This study broadly identified a range of problematic aspects in a group of bilingual learners' subject-specific task-based writing. In order to effectively address these, *more detailed analyses of learners' texts and the exact nature of each of the problematic areas are needed*. For example, problems in terms of learners' use of subject-specific terminology need to be studied more closely, the types of problems involved need to be categorised, and focused teaching solutions for these problems need to be developed and tested.

Furthermore, this study focused on written products. In order to develop a better understanding of the nature of the problems involved in task processing, *process research is needed* that looks at the procedures learners follow in constructing meaning on the basis of the task instruction and input material. Understanding the processes better can lead to a more precise identification of the methods needed in order to address the deficiencies manifesting in the written products.

Also, *detailed comparisons between the written performances of bilingually trained learners and those produced by learners following the subject in German are needed*. Such comparisons can be used to determine whether the problems found are specific to the written responses of learners trained in a foreign language, or whether some or all of them also occur in the written texts of learners following a subject in the L1. Preliminary analyses of this nature, carried out in the context of the project in which this study was embedded, suggest that the texts produced by 10th grade geography learners following the subject in German show similar problems to the texts produced by the bilingually trained learners.

Finally, seeing that teachers are the mediators responsible for initiating learners into the discourse practices of the subject, more information is needed on *how bilingual subject-specific teachers view their role* in terms of the development of learners' discourse competences and literacy. It should be established to what extent teachers accept that they play a central role in the development of learners' ability to use the subject-specific register, which is the basis for participation in institutionalised discourses beyond the school. Teachers can only effectively support learners' initiation into the subject community, if they accept the role as mediator and take their responsibility in this regard seriously. It is also important to determine *how teachers, as members of the subject community, define the relevant norms and conventions of subject-specific discourse*. If teachers are to guide learners towards adhering to the particular conventions, the teachers themselves need a critical awareness of these conventions. Related to this would be research on the rating criteria implicitly applied by teachers when they are assessing learners' subject-specific writing.

6.3 Closing remarks

This study aims to contribute to an understanding of subject-specific discourse competence and its assessment in a number of ways. Firstly, it has identified areas of strengths and weaknesses in a group of bilingually trained geography learners' subject-specific writing and has presented examples of instruments that can be used to rate learners' writing. Through these results and rating scales, the study intends to draw attention to problematic aspects of a group of bilingually trained learners' use of the subject-specific register and to present a methodological basis that can be applied in other studies on bilingually trained learners' task-based writing. In this respect, the study aspires *to contribute to the research on bilingual subject-specific education in Germany by investigating learners' task-based subject-specific writing - an aspect that has not received much attention up until now - and by indicating the need for more research on this aspect*.

The processes that were involved in developing and applying the rating scales in the context of this study made the challenges involved in measuring high order performances on complex tasks clear. The significant variation in learners' responses and the difficulty involved in identifying levels and formulating level descriptors that

firstly allowed the performances to be reliably and appropriately categorised, but that were also still feasible in their application, pointed to complexities that are not always openly discussed by test developers. In order to move the field forward, it is believed that test developers need to share their experiences and the insights gained in the processes of developing, administering and rating performance assessments. Only by means of open discussion, can the methodologies involved be optimised. This study therefore aspires to contribute to *an understanding of the practical aspects involved in promoting the reliability and validity of assessment initiative*s.

The study thirdly attempts to *contribute to the demystification of subject-specific discourse*. This has firstly been done by organising the expectations in terms of written responses to subject-specific tasks into three categories: the ideational meaning to be constructed, the procedures to be followed in constructing this meaning and the textual presentation of this meaning. It has secondly been done by explaining the dynamics of the discourse through identifying ways in which a subject-community's values, its activities and the different functions texts play in the community's activities influence the conventions regulating the subject-specific discourse.

Furthermore, the study aspires to contribute to *a more comprehensive understanding of the concept of a task's discourse function and of the areas involved in subject-specific written discourse competence*. Through a reinterpretation based on social views of existing definitions of these concepts, the study attempts to provide a conceptualisation of a task's discourse function that integrates aspects of content and form. Also, using the concept of pedagogical research and its role in developing subject positions, the definition of discourse competence presented in this study integrates volitional aspects in addition to subject knowledge, methodological competence and subject-specific communicative competence.

Finally, this study aspires to *contribute to a greater awareness of the role the subject-specific classroom and teacher have in developing learners' ability to use the subject-specific register*. In this respect, the study has explained the importance of the ability to use social semiotic systems appropriately for effective participation in the subject community's activities. It has also explained the central role that mastery of the subject-specific register plays in the ability to participate in discourses in other educational,

professional and public contexts. In other words, one of the central intentions of the study is to point out that language use can not be seen as an aspect that is of little concern to the subject-specific teacher, since the appropriate and effective use of language for subject-specific purposes is a central subject-specific competence and contributes to the development of literacy.

The bilingual subject-specific classroom provides unique opportunities to acknowledge the role the ability to use language appropriately and effectively has for participation in subject-specific activities. Since learners are using a foreign language, the focus naturally does not only fall on familiarising learners with the subject matter and the subject's activities, but also on the language used. These opportunities can however only be exploited fully, if teachers support these familiarisation processes in a structured way. This implies that the conventions the subject community follows in terms of the models of experience with which it operates, the procedures it follows in constructing meaning and the social semiotic systems it uses to present meaning are made clear to learners. Through research, problems learners experience in following these conventions and that hamper their initiation into the subject community can be identified and solutions found. By paying attention to learners' use of the subject-specific register in the classroom and through research, the potential bilingual subject-specific training harbours for preparing learners for contexts beyond the classroom can be optimised.

Appendix A

Elicitation tasks

- Task 1c: Kisangani's climate
- Task 2a: Living conditions in the three layers of the tropical rain forest
- Task 3: International destruction of the tropical rain forest
- Task 6: Sustainability
- Task 7: International significance of the tropical rain forest
- Task 8: Personal scope for action

1C. CLIMATE

c) What type of climate does Kinsangani have? Describe Kisangani's climate in detail with the help of the chart.

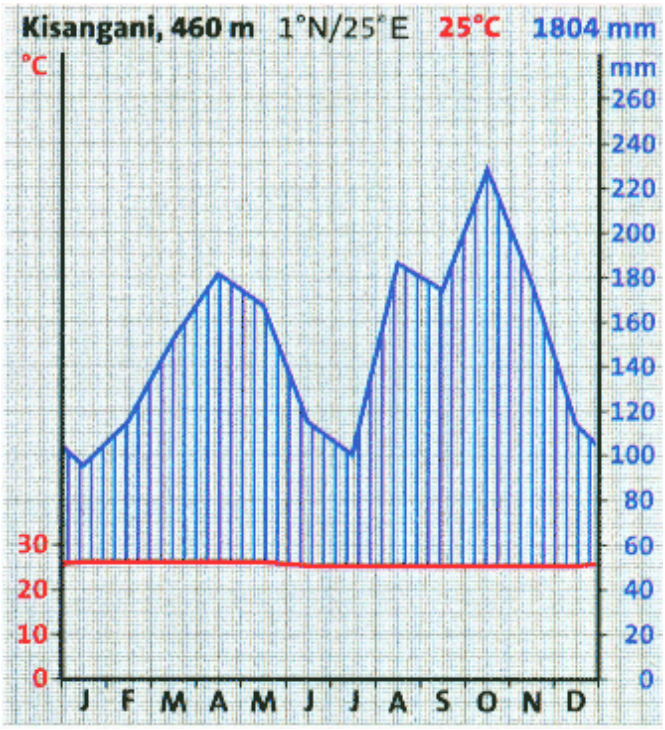


Fig. 2

2. THE ECOSYSTEM IN THE TROPICAL RAINFOREST –NUTRIENT CYCLE

a) Characterize the living conditions in the tree layers of the tropical rain forest with the help of Figure 3. Start with the giant trees.

b) The tropical rain forest does not grow out of the soil but lives on it. Explain this sentence with the help of Figure 3 and compare the nutrient cycle of the tropical rain forest with that of the deciduous forests in our regions.

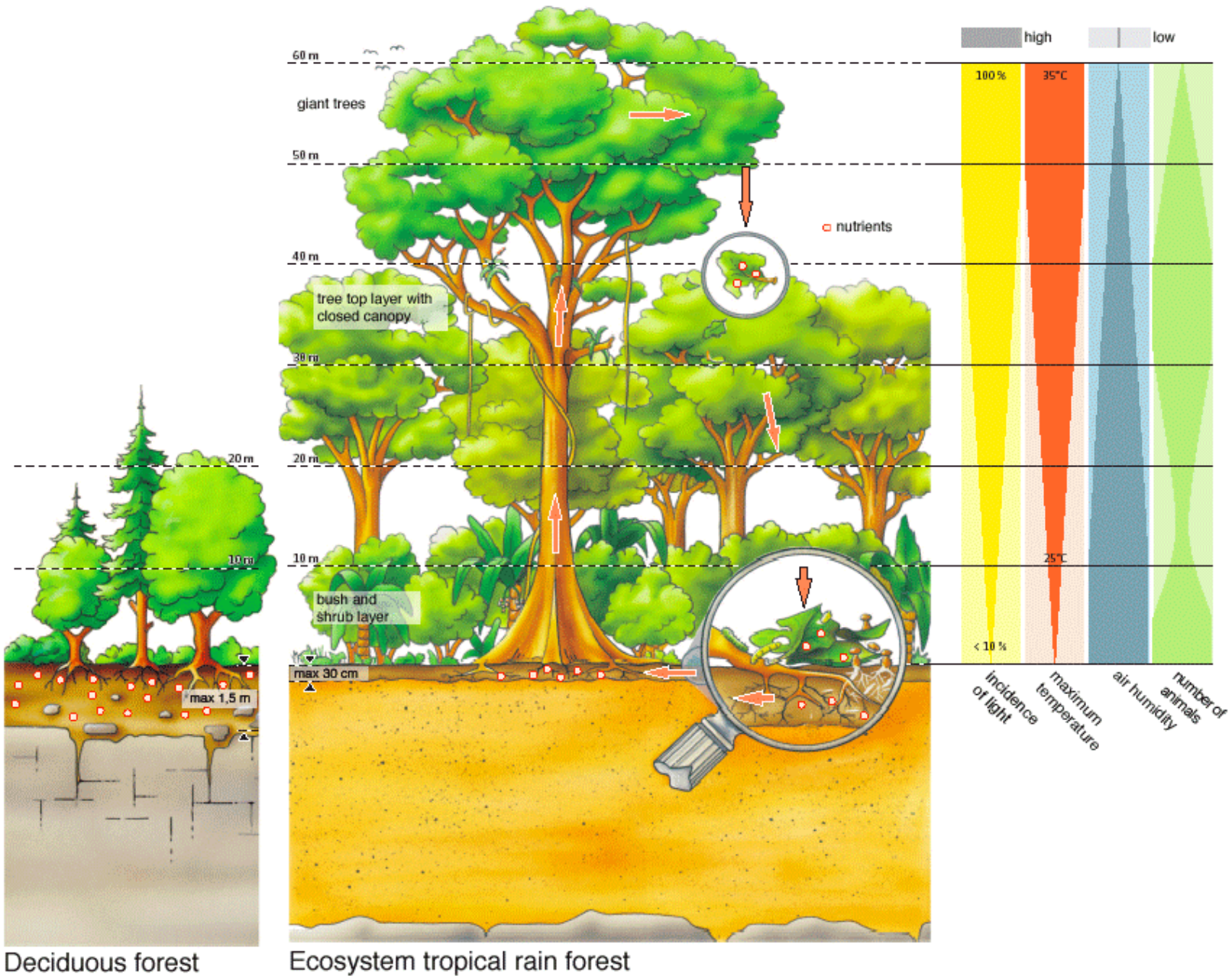

60 m
giant trees
50 m
nutrients
40 m
tree top layer with closed canopy
30 m
20 m
10 m
bush and shrub layer
max 30 cm
high
low
100 %
35°C
25°C
< 10 %
incidence of light
maximum temperature
air humidity
number of animals
20 m
10 m
max 1,5 m
Deciduous forest
Ecosystem tropical rain forest

Fig. 3

3. DESTRUCTION OF THE TROPICAL RAIN FOREST

In recent decades the destruction of the tropical rain forest has increased continuously. The following two figures show the extent of this destruction.
According two Figure 4, the destruction of the tropical rain forest is worst in Asia. Comment critically on this statement by relating it to the information presented in Table 1.

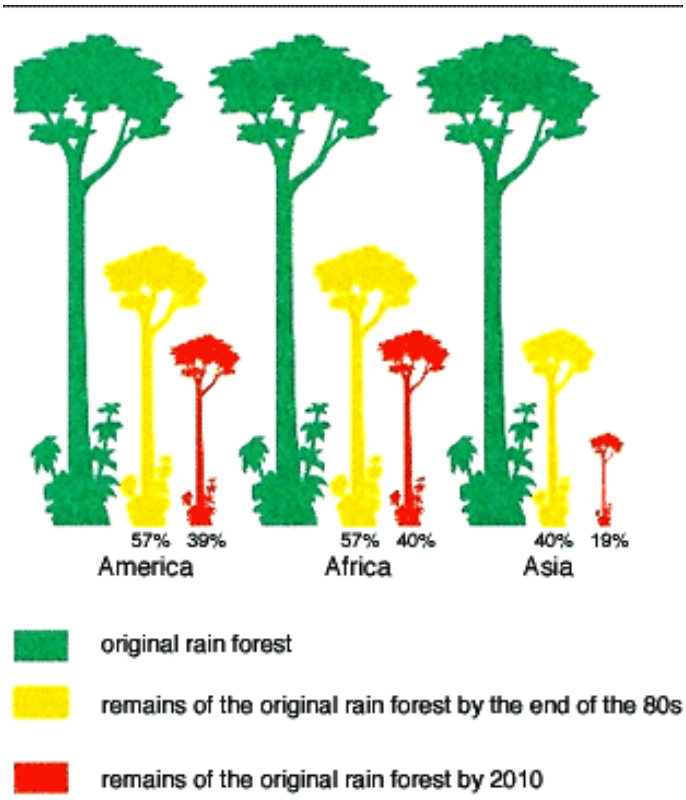


Fig. 4: Decline of the original tropical rain forest

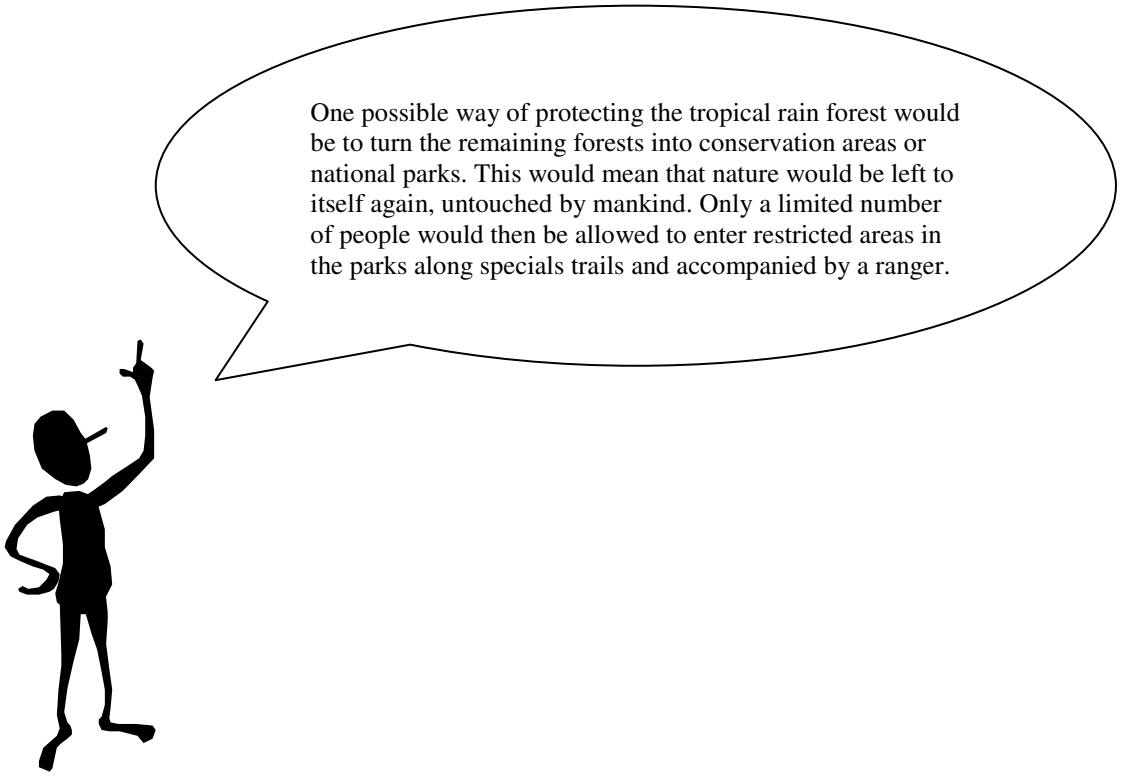
Countries	Forest area in km²: 1990	Forest area in km²: 2000	Annual changes of forest area in km²
Angola (Africa)	709 980	697 560	-1 240
Bolivia (America)	546 790	530 680	-1 610
Brazil (America)	5 669 980	5 439 050	-23 090
Dem. Rep. Congo (Africa)	1 405 310	1 352 070	-5 320
Indonesia (Asia)	1 181 100	1 049 860	-13 120
Columbia (America)	515 060	496 010	-1 900
Malaysia (Asia)	216 610	192 920	-2 370
Mexico (America)	615 110	552 050	-6 310
Myanmar (formerly Birma) (Asia)	395 880	344 190	-5 170
Peru (America)	679 030	652 150	-2 690
Zambia (Africa)	397 755	312 460	-8 510
Venezuela (America)	516 810	495 060	-2 180

Table 1: Decline in forest cover in countries with large forest areas 1990–2000

6. SUSTAINABILITY

Since the world summit of Rio de Janeiro in 1992 (Agenda 21) the principle of sustainability has been globally accepted. This principle means that in terms of the use of natural resources, ecological, social and economic goals should be treated as equally important. At the same time, the rights and needs of future generations should be respected, so that they are not disadvantaged in any way through the exploitation of the resources.

a) Against the background of the aims of Agenda 21, how do you evaluate the following suggestion for a solution? Support your answer.



7. HUMANS AND THE RAIN FOREST

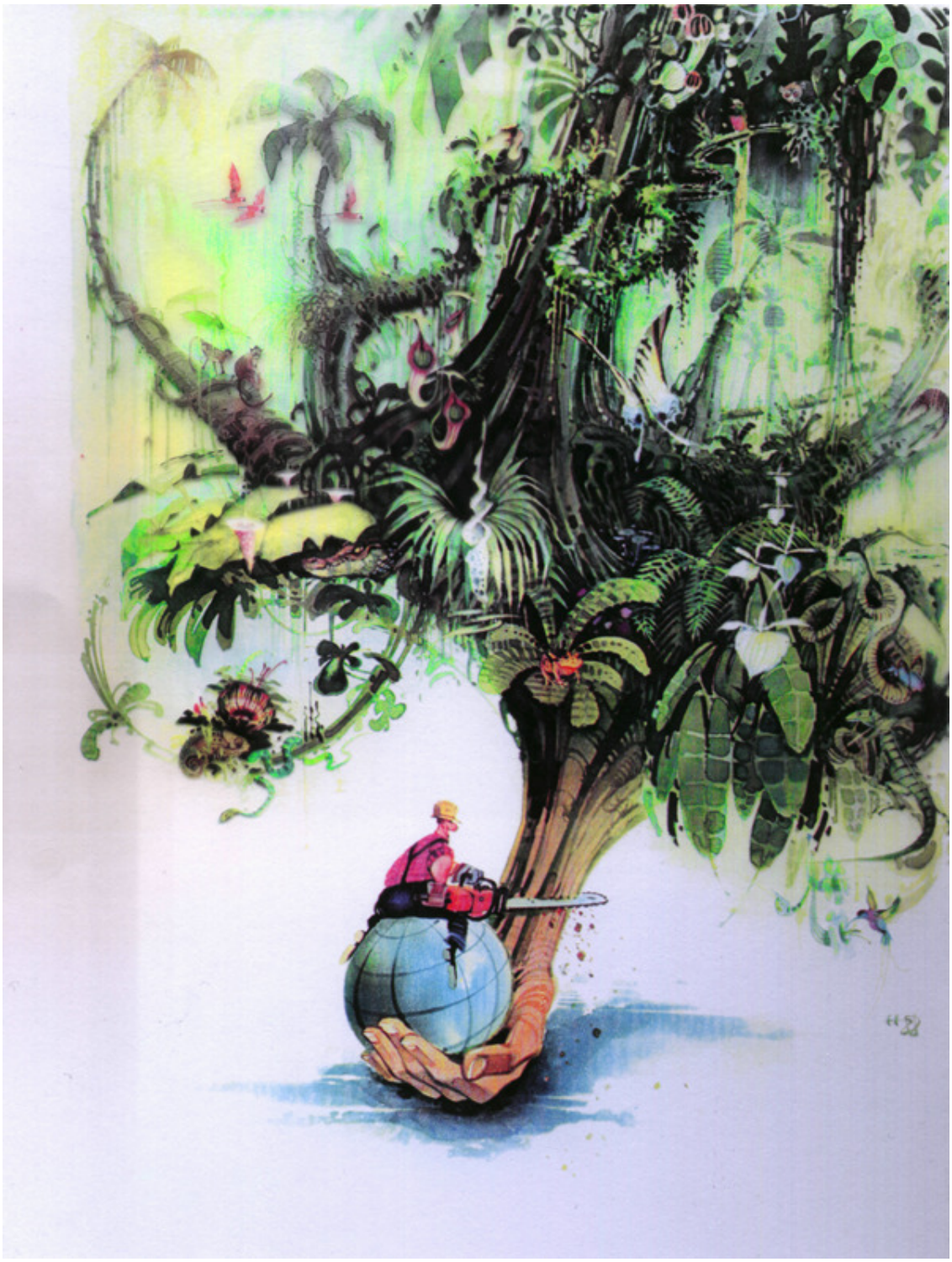


Fig. 5

Take a look at Figure 5 and explain in how far the following statement is true.

"The rain forest concerns us all".

8. THE EXTENT TO WHICH ONE CAN CONTRIBUTE PERSONALLY

This part deals with the question of what you can do personally to contribute to the protection of the tropical rain forest. First, give your opinion about the following statement by **indicating in your answering sheets on page 8** the extent to which it is true for you. Then, support your decision.

I can personally contribute to the protection of the tropical rain forest.

Appendix B

Summary of task categories and input material

Topical focus of task categories	**Number of tasks**	**Input material**
1) The distribution and climate of the tropical rain forest	5	1) A map showing the international distribution of rain forests. 2) A climate graph of the Kisangani weather station.
2) The ecosystem in the tropical rain forest and the nature of the nutrient cycle	2	1) A rather detailed diagram that depicts the ecosystem and nutrient cycle of the tropical rain forest, as well as the nutrient cycle found in deciduous forests. 2) A sentence that reads: "The tropical rain forest does not grow out of the soil but lives on it".
3) The international destruction of the tropical rain forest	1	1) A diagram portraying the decline of the tropical rain forests in America, Africa and Asia by means of images of differently sized trees accompanied by percentages of the original rain forests that were left by 1980 and that will be left by 2010 according to forecasts. 2) A table with information about the decline in the areas covered by rain forests in different countries between 1990 and 2000. The information is presented in absolute numbers of forest area in km^2 and of annual changes of forest area in km^2.
4) Causes and effects of the destruction of the tropical rain forest	3	1) A list of causes and effects of the destruction. 2) An uncompleted flow chart categorising the causes and effects of the destruction.
5) Opinions about the use of the rain forest	3	1) Two short texts, the one representing the view of a representative of the Yanomamis and the second representing the view of a government official.
6) The principle of sustainability applied to the tropical rain forest	1	1) A short definition of sustainability. 2) A suggestion about how the rain forest could be saved, but which only pays attention to ecological matters while ignoring social and economic aspects.
7) The international significance of the tropical rain forest	1	1) A cartoon depicting the rain forest as a hand that holds the earth and that is being cut off. 2) A statement that reads: "The rain forests concerns us all."
8) The extent to which individuals can contribute to the protection of the tropical rain forest	1	1) A statement stating that one can personally contribute to the protection of the rain forest. 2) A scale with 5 options ranging from absolute disagreement to absolute agreement on which a choice should be indicated by ticking one of the options.

Summary of task categories and input material used in the geography competence test

Appendix C

Rating scales

Holistic scales:

- Degree of task fulfilment: Appropriateness of the discourse function

Analytic scales:

- Appropriateness of the textual realisation of meaning
- Completeness and correctness of meaning constructed

Holistic scales:

Degree of task fulfilment – appropriateness of the discourse function

Task 1C

Level 6

There is a well structured, succinct, correct and *comprehensive* descriptive explanation that precedes or follows the correct identification of the climate as a *tropical* climate, or a conclusion that the climate is tropical.

- The descriptive explanation includes correct information about temperature and precipitation.
  - In terms of temperature, it is correctly stated that the average temperature is 25°C (it is not simply stated that it is warm, but the correct temperature is given in degrees Celsius) and it is pointed out that the temperature is constant throughout the year.
  - In terms of precipitation, it is correctly and precisely indicated that the precipitation is constantly high: namely that it is hardly under 100mm per month. It is stated, however, that the monthly amount of precipitation varies over the year and that there are two peaks in the precipitation over the year. It is further explained that the precipitation is higher than the evaporation and that it is therefore humid.

  Only relevant content is included.
- The text is *consistently* logically structured and effectively linked.
- A subject-specific register is *consistently* used.

Level 5

There is a *generally* well structured and clearly formulated, correct and *adequate* descriptive explanation that precedes or follows the correct identification of the climate as a *tropical* climate, or a conclusion that the climate is tropical.

- The descriptive explanation includes correct information about temperature and precipitation, but *one of the sub-points under precipitation* may be *lacking*, or may *not* be *expressed numerically*. It is explained that the precipitation is higher than the evaporation and that it is therefore humid. A *limited amount of content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text is *for the most part* logically structured and effectively linked.
- A subject-specific register is *consistently* used.

Level 4

There is a descriptive explanation that is *sufficient*, but that may show a *limited number of errors* and that precedes or follows the correct identification of the climate as a *tropical* climate, or a conclusion that the climate is tropical. The descriptive explanation and identification, or conclusion *may be* generally well structured and clearly formulated.

- The description includes information about temperature and precipitation, but a *limited number of sub-points under precipitation* may be *lacking*, or may be *incorrect*, or may *not* be *expressed numerically*. It *may be* pointed out that it is humid (or that it is hot and wet, or hot and damp), but an *explanation* about the relation between the amount of precipitation and degree of evaporation *may be lacking*. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may for the most part* be logically structured and sufficiently linked.
- A subject-specific register is *mostly* used.

Level 3

The type of climate *may be* identified as *tropical*, **or** may be classified as *humid*. Or it may be concluded that the climate is tropical or humid. This is preceded or followed by a descriptive explanation that is *still acceptable for the particular classification*, but that may show a *limited number of errors*. The descriptive explanation and identification, or conclusion *may be* generally well structured and clearly formulated.

- The description includes information about temperature and precipitation, but a *limited number of sub-points under temperature or precipitation* may be *lacking*, or may be *incorrect*. Exact references *expressed numerically* are *probably not included*. It *may be* pointed out that it is humid, but an *explanation* about the relation between the amount of precipitation and degree of evaporation may be *lacking*. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may for the most part* be logically structured and sufficiently linked.
- A subject-specific register is *mostly* used.

Level 2

The type of climate is not identified as tropical, or it is not concluded that the climate is tropical. The climate is *classified as humid*, **or** *related paraphrases* are given. This is preceded or followed by a descriptive explanation that *may still be acceptable*, but that may show *some errors*. The descriptive explanation and identification, or conclusion may show *major deficiencies* in terms of logical structuring and formulation.

- The description includes *limited information* about temperature and precipitation. *Some sub-points under temperature or precipitation* may be *lacking*, or may be *incorrect*. Exact references *expressed numerically* are *probably not included*. It *may be pointed out* that it is humid, but an *explanation* about the relation between the amount of precipitation and degree of evaporation is *lacking*. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may not* be logically structured and sufficiently linked.
- A subject-specific register may *not always* be used.

Level 1

There is *no identification* of, or conclusion about a type of climate, **or** the classification is *incorrect*. There is a *limited* description of the climate that may show *some errors* and that may show *major deficiencies* in terms of logical structuring and formulation.

- The description *may focus* either on temperature *or* on precipitation. *Some sub-points under temperature or precipitation* may be *lacking*, or may be *incorrect*. Exact references *expressed numerically* are *probably not included.* It *may be pointed out* that it is humid, but an *explanation* about the relation between the amount of precipitation and degree of evaporation is *lacking*. *Some content* may be included that is only *peripherally related* to the topic.
- The text *may not* be logically structured and sufficiently linked.
- A subject-specific register may *not always* be used.

Level 0

None of the required mini-genres is realised. The *type of climate is not identified*, or no conclusion is reached and there is *no description* of the climate. The text is *severely deficient* in terms of logical structuring and formulation.

- There is *no* or *hardly any reference related* to the *climate graph*.
- The text is *unstructured* and *fragmentary*.
- A subject-specific register *is not* used.

Task 2A

Level 6

All three layers are *correctly* named. There is a well structured, succinct, correct and *comprehensive* characterisation of the living conditions in each of the three layers. The living conditions in the layers are *effectively* compared.

- The characterisation of the living conditions in each of the three layers *consistently includes* references to the manifestation of *all four* of the following factors: incidence of light, maximum temperature, air humidity and number of animals. These references are *well-supported* by detailed information in the form of numerical values taken from Figure 3. The characterisation also includes *numerous* explanations about the interrelationships between these four factors and the living conditions in the three layers are *consistently* compared. *Only relevant content* is included.
- The text is *consistently* logically structured and effectively linked.
- A subject-specific register is *consistently* used.

Level 5

All three layers are *correctly* named. There is a *generally* well structured and clearly formulated, correct and *adequate* characterisation of the living conditions in each of the three layers. The living conditions in the layers are *sufficiently* compared.

- The characterisation of the living conditions in each of the three layers *consistently includes* references to the manifestation of *all four* relevant factors. These references may *to some extent* be *supported* by detailed information in the form of numerical values taken from Figure 3. The characterisation also includes *some* explanations about the interrelationships between these four factors and the living conditions in the three layers are *frequently* compared. A *limited amount of content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text is *for the most part* logically structured and effectively linked.
- A subject-specific register is *consistently* used.

Level 4

All *three layers* are referred to, but the *correct label* may *not always* be *used*. There is a *sufficient* characterisation of the living conditions in each of the three layers that may show a *limited number of errors*. The living conditions in the layers are compared *to some extent*. The comparative characterisation *may be* generally well structured and clearly formulated.

- The characterisation of the living conditions in each of the three layers may *not consistently include* references to the manifestation of *all four* relevant factors. References to the manifestation of the factors may *to a limited extent* be *supported* by detailed information in the form of numerical values taken from Figure 3. The characterisation includes *some* explanations about the interrelationships between these four factors. There is *some* comparison between the living conditions in the three layers. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may for the most part* be logically structured and sufficiently linked.
- A subject-specific register is *mostly* used.

Level 3

All *three layers* are referred to, but the *correct label* may *not always* be *used*. The characterisation of the living conditions in each of the three layers is *still sufficient*, but may show a *number of errors*. The living conditions in the layers *may be* compared *to some extent*. The comparative characterisation *may be* generally well structured and clearly formulated.

- The characterisation of the living conditions in each of the three layers may *consistently exclude one* or *two* factors, **or** the *characterisation* of the living conditions in *one of the* three *layers* may *be limited*. References to the manifestation of the factors may *to a limited extent* be *supported* by detailed information in the form of numerical values taken from Figure 3. The characterisation includes a *limited number* of explanations about the interrelationships between the factors mentioned. There *may be some* comparison between the living conditions in the three layers. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may for the most part* be logically structured and sufficiently linked.
- A subject-specific register is *mostly* used.

Level 2

Two to three layers are referred to, but the *terms used* to refer to the layers may be *somewhat vague*. There is an *insufficient* characterisation of the living conditions in the layers referred to that may show a *number of errors*. There *may be a limited degree* of comparison between the living conditions in the layers mentioned. The comparative characterisation may show *major deficiencies* in terms of logical structuring and formulation.

- The characterisation of the living conditions in the *layers* referred to may include only *two* or *three* factors. References to the manifestation of the factors are *probably not supported* by details in the form of numerical values taken from Figure 3. The characterisation *may include* a *limited number* of explanations about the interrelationships between the factors mentioned. There *may be little* comparison between the living conditions in the three layers. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may not* be logically structured and sufficiently linked.
- A subject-specific register may *not always* be used.

Level 1

One to two layers are referred to, but the *terms used* to refer to the layers may be *somewhat vague*. There is a *highly limited* characterisation of the living conditions in the layers referred to that may show a *number of errors*. There *may be hardly any or no* comparison between the living conditions in the layers mentioned. **Or**, *three layers* are named, but the *terms used* to refer to the layers may be *somewhat vague*. There is an *irrelevant* characterisation of the living conditions in the three layers that may show a *number of errors.* The living conditions in the layers *may be* compared *to some extent*. The comparative characterisation may show *major deficiencies* in terms of logical structuring and formulation.

- The characterisation of the living conditions in the one or two layers referred to may include *only one* of the factors. References to the manifestation of the factors are *probably not supported* by details in the form of numerical values, taken from Figure 3. There *may be hardly any* comparison between the living conditions in the three layers. **Or,** the characterisation of the living conditions in the *three layers* may include only *factors* that are *irrelevant* for a distinction between the living conditions in the three layers. There may, for example be references to nutrients. References to the manifestation of the factors are *not supported* by details in the form of numerical values taken from Figure 3. If more than one factor is included, the characterisation *may include* a *limited number* of explanations about the interrelationships between the factors mentioned. There *may be some* comparison between the living conditions in the three layers. *Some content* may be included that is only *peripherally related* to the topic.
- The text *may not* be logically structured and sufficiently linked.
- A subject-specific register may *not always* be used.

Level 0

None of the required mini-genres is realised. *None of the three layers* is *named* and there is *no comparative characterisation* of the living conditions in the three layers. The text is *severely deficient* in terms of logical structuring and formulation.

- There is *no* or *hardly any reference* that *relates to Figure 3* and to the *living conditions* in the three *layers* of the tropical rain forest.
- The text is *unstructured* and *fragmentary*.
- A subject-specific register *is not* used.

Task 3

Level 6

There is a *succinct and correct* criticism of the idea put forward in Figure 4, **and/or** a suggestion to improve the statement. There *is* a succinct comment on the difference in the presentation of data in the two sources of information presented in the task. Support *is* given that is succinct and fully consistent with the statement made. The argumentation is well structured.

- It is pointed out that in terms of the size of the area of rain forest that is destroyed (or in absolute terms) the destruction is the biggest in America and it *is added* that in terms of the proportional decrease (or in percentage terms), the destruction is the biggest in Asia. Additionally, the difference between the depiction of the international destruction of the tropical rain forest in Figure 4 and Table1 *is reflected upon*. The support includes references to the sizes of the areas covered by tropical rain forest and the extent of changes in km^2. These references are presented in the form of numerical values taken from Table 1. *Only relevant content* is included.
- The text is *consistently* logically structured and effectively linked.
- A subject-specific register is *consistently* used.

Level 5

There is a *clearly formulated and correct* criticism of the idea put forward in Figure 4 **and/or** a suggestion to improve the statement. There *may be* a clearly formulated comment on the difference in the presentation of data in the two sources of information presented in the task. There *is some* support given that is clearly formulated and consistent with the statement made. The argumentation is *generally* well structured.

- It is pointed out that in terms of the size of the area of rain forest that is destroyed (or in absolute terms) the destruction is the biggest in America and it *is added* that in terms of the proportional decrease (or in percentage terms), the destruction is the biggest in Asia. Additionally, the difference between the depiction of the international destruction of the tropical rain forest in Figure 4 and Table 1 *may be reflected upon*. The support includes *some* references to the sizes of the areas covered by tropical rain forest and the extent of changes in km^2. These references are presented in the form of numerical values taken from Table 1. A *limited amount of content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text is *for the most part* logically structured and effectively linked.
- A subject-specific register is *consistently* used.

Level 4

There is a *correct* criticism of the idea put forward in Figure 4 **and/or** a suggestion to improve the statement. These *may be* clearly formulated. There *may be* a comment on the difference in the presentation of data in the two sources of information presented in the task that *may be* clearly formulated. *Some support may be* given. The support *may be* clearly formulated, but may show a *limited degree of inconsistency*. The argumentation *may be* generally well structured.

- It is pointed out that in terms of the size of the area of rain forest that is destroyed (or in absolute terms) the destruction is the biggest in America and it *is added* that in terms of the proportional decrease (or in percentage terms), the destruction is the biggest in Asia. Additionally, the difference between the depiction of the international destruction of the tropical rain forest in Figure 4 and Table1 *may be reflected upon*. The support *may include some* references to the sizes of the areas covered by tropical rain forest and the extent of changes in km^2. These references are presented in the form of numerical values taken from Table 1. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may for the most part* be logically structured and sufficiently linked.
- A subject-specific register is *mostly* used.

Level 3

There is a criticism of the idea put forward in Figure 4 **and/or** a suggestion to improve the statement. However, these *may be vaguely formulated*. There is *probably no* comment on the difference in the presentation of data in the two sources of information presented in the task. *Limited support may be* given. The support *may be vaguely formulated* and may show a *limited degree of inconsistency*. The argumentation *may be* generally well structured.

- It is pointed out that in terms of the size of the area of rain forest that is destroyed (or in absolute terms) the destruction is the biggest in America. It *may be added* that in terms of the proportional decrease (or in percentage terms), the destruction is the biggest in Asia. The difference between the depiction of the international destruction of the tropical rain forest in Figure 4 and Table1 is *probably not reflected upon*. The support *may include a limited number of* references to the sizes of the areas covered by tropical rain forest and the extent of changes in km^2. These references are presented in the form of numerical values taken from Table 1. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may for the most part* be logically structured and sufficiently linked.
- A subject-specific register is *mostly* used.

Level 2

There is *no distinguishable* criticism of the idea put forward in Figure 4 **and/or** a suggestion to improve the statement. The text as a whole suggests that the idea is not fully accepted. **Or**, there is an *unrestricted confirmation* of the idea put forward in Figure 4. There may be *some* support that *may be consistent* with the statement made. The argumentation may show *major deficiencies* in terms of logical structuring and formulation.

- If the text as a whole suggests that the idea is not fully accepted, there *may be* a reference to the fact that the total area covered by tropical rain forest is bigger in America than in Asia *or* a reference to the fact that the total changes in forest area per year is higher in America than in Asia. There are *probably no* references to the sizes of the areas covered by tropical rain forest and the extent of changes in km^2. **Or**, if there is an unrestricted confirmation of the idea put forward in Figure 4, references to the destruction of the tropical rain forest in percentage terms *and* references to the destruction of the tropical rain forest in absolute terms taken from Figure 4 and Table 1 *may be* included. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may not* be logically structured and sufficiently linked.
- A subject-specific register may *not always* be used.

Level 1

There is *no criticism, or confirmation of* the idea put forward in Figure 4 and *no suggestion* to improve the statement. There is only a description of and /or a prediction about the international destruction of the tropical rain forest that may be correct. The argumentation may show *major deficiencies* in terms of logical structuring and formulation.

- The description / and or prediction may include references to the destruction of the tropical rain forest in percentage terms *and* references to the destruction of the tropical rain forest in absolute terms taken from Figure 4 and Table 1. *Some content* may be included that is only *peripherally related* to the topic.
- The text *may not* be logically structured and sufficiently linked.
- A subject-specific register may *not always* be used.

Level 0

None of the required mini-genres is realised. There is *no criticism of* the idea put forward in Figure 4, or a *suggestion* to improve the statement and *no support*. The text is *severely deficient* in terms of logical structuring and formulation.

- There is *no comment on* the idea put forward in Figure 4 and there is *no* or *hardly any reference* that *relates to Figure 4* and *Table 1*.
- The text is *unstructured* and *fragmentary*.
- A subject-specific register *is not* used.

Task 6

Level 6

There is a succinct and *balanced* evaluation of the suggestion. This is preceded or followed by succinct and *comprehensive* support that is fully consistent with the evaluation and that is based on a *complete* and a *correct* application of the principle of sustainability in evaluating the suggestion. The argumentation is well structured.

- The evaluation *criticises* the suggestion for the fact that it does not make equal provision for the different goals of the principle of sustainability. The support is based on the *correct* application of *all four* of the following goals of the principle of sustainability in evaluating the suggestion: protecting the environment (ecological goal), taking the rights and needs of future generations into consideration, making provision for social goals and paying attention to economic goals. *Only relevant content* is included.
- The text is *consistently* logically structured and effectively linked.
- A subject-specific register is *consistently* used.

Level 5

There is a clearly formulated and *balanced* evaluation of the suggestion. This is preceded or followed by clearly formulated and *adequate* support that is fully consistent with the evaluation and that is based on a *correct* application of the principle of sustainability in evaluating the suggestion. The argumentation is *generally* well structured.

- The evaluation *criticises*, or *does not fully support* the suggestion. It is pointed out that the suggestion does not make equal provision for the different goals of the principle of sustainability. The support is based on the *correct* application of *at least three* of the goals of the principle of sustainability in evaluating the suggestion. A *limited amount of content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text is *for the most part* logically structured and effectively linked.
- A subject-specific register is *consistently* used.

Level 4

There is a *balanced* evaluation of the suggestion. The evaluation *may be* clearly formulated. There is *sufficient* support that *may be* clearly formulated, but that may show a *limited degree of inconsistency*. The support is based on an application of the principle of sustainability in evaluating the suggestion. The application may *not* be *entirely correct*. The argumentation *may be* generally well structured.

- The evaluation *criticises*, or *does not fully support* the suggestion. It *may be* pointed out that the suggestion does not make equal provision for the different goals of the principle of sustainability. The support is based on the *correct* application of *at least two* of the goals of the principle of sustainability in evaluating the suggestion. *Some content* may be included that *is* related to the topic, but *not central* to the content requirements of the task.
- The text *may for the most part* be logically structured and sufficiently linked.
- A subject-specific register is *mostly* used.

Level 3

There *may be a balanced* evaluation that *may be* clearly formulated. There is support that *may be sufficient*, but that may show *some degree of inconsistency*. The support is *to some extent* based on a *correct* application of the principle of sustainability in evaluating the suggestion. The argumentation *may be* generally well structured.

- The evaluation *may criticise*, or *may not fully support* the suggestion. The support is based on the *correct* application of *at least one* of the goals of the principle of sustainability in evaluating the suggestion. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may for the most part* be logically structured and sufficiently linked.
- A subject-specific register is *mostly* used.

Level 2

There is an evaluation of the suggestion that is *unbalanced*. The evaluation *may be* clearly formulated. There is support that *may be sufficient*, *clearly formulated* and *consistent*. However, the support is based on an *incorrect* application of the principle of sustainability in evaluating the suggestion. The argumentation may show *major deficiencies* in terms of logical structuring and formulation.

- The evaluation *fully supports* or *fully rejects* the suggestion. The support is based on a *reduction* of the principle of sustainability to its concern for ecological goals and / or the rights and needs of future generations. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may not* be logically structured and sufficiently linked.
- A subject-specific register may *not always* be used.

Level 1

There is *no evaluation* of the suggestion. There is only a discussion of Agenda 21 and/or the principle of sustainability that may show *major deficiencies* in terms of logical structuring and formulation.

Or, there is an *evaluation* that *may be balanced*, but there is *no or hardly any support*, **or** the support is *not based* on an application of the principle of sustainability at all. The evaluation may show *major deficiencies* in terms of logical structuring and formulation.

- If there is no evaluation, but only a discussion, it *may* include references to the goals of the principle of sustainability. **Or**, if there is an evaluation, it may *fully support* the suggestion, or *criticise* the suggestion. If there is support, it may focus on *aspects* that are *not directly related* to the principle of sustainability. It may, for instance, be pointed out that the suggestion would be too expensive to implement, or that it would not get enough political support etc. *Some content* may be included that is only *peripherally related* to the topic.
- The text *may not* be logically structured and sufficiently linked.
- A subject-specific register may *not always* be used.

Level 0

None of the required mini-genres is realised. There is *no evaluation* of the suggestion and *no support*. The text is *severely deficient* in terms of logical structuring and formulation.

- There is *no criticism* of or *support for* the suggestion and there are *no references* that *relate to* the principle of *sustainability*.
- The text is *unstructured* and *fragmentary*.
- A subject-specific register *is not* used.

Task 7

Level 6

There is a succinct and *acceptable* opinion about the extent to which the statement is true. This is preceded or followed by a succinct, *comprehensive* and plausible explanation about the international significance of the tropical rain forest that fully supports the opinion expressed. A brief description and valid interpretation of the cartoon *is* included. The argumentation is well structured.

- The opinion expressed *supports* the statement. The explanation includes correct references to *at least three* interrelationships between humans and the tropical rain forest. The nature of the interrelationships is explained *in full detail*. The cartoon *is* described as depicting the tropical rain forest as a hand that holds the earth and that is being cut off and interpreted to suggest that the tropical rain forest is a resource that humans need, but that that they are destroying. *Only relevant content* is included.
- The text is *consistently* logically structured and effectively linked.
- A subject-specific register is *consistently* used.

Level 5

There is a clearly formulated and *acceptable* opinion about the extent to which the statement is true. This is preceded or followed by a clearly formulated, *adequate* and plausible explanation about the international significance of the tropical rain forest that fully supports the opinion expressed. A brief description and valid interpretation of the cartoon *is* included. The argumentation is *generally* well structured.

- The opinion expressed *supports* the statement. The explanation includes correct references to *at least two* interrelationships between humans and the tropical rain forest. The nature of the interrelationships is explained *in some detail*. The cartoon *is* described as depicting the tropical rain forest as a hand that holds the earth and that is being cut off and interpreted to suggest that the tropical rain forest is a resource that humans need, but that they are destroying. A *limited amount of content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text is *for the most part* logically structured and effectively linked.
- A subject-specific register is *consistently* used.

Level 4

There is an *acceptable* opinion about the extent to which the statement is true. The opinion *may be* clearly formulated. There is a *sufficient* explanation about the international significance of the tropical rain forest. The explanation may, however, *not* be *entirely plausible* and may *not* be *fully consistent* with the opinion expressed. The explanation *may be* clearly formulated. A brief description and valid interpretation of the cartoon *is* included. The argumentation *may be* generally well structured.

- The opinion expressed *supports* the statement. The explanation includes references to *at least two* interrelationships between humans and the tropical rain forest. A limited number of references *may not be entirely correct*. The references to the interrelationships may remain on a general level and *only few details may be* included. The cartoon *is* described as depicting the tropical rain forest as a hand that holds the earth and that is being cut off and interpreted to suggest that the tropical rain forest is a resource that humans need, but that they are destroying. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may for the most part* be logically structured and sufficiently linked.
- A subject-specific register is *mostly* used.

Level 3

There is an *acceptable* opinion about the extent to which the statement is true. The opinion *may be* clearly formulated. However, there is a *somewhat limited* explanation about the international significance of the tropical rain forest that may *not* be *entirely plausible*. The explanation *may be* clearly formulated. A brief description and valid interpretation of the cartoon *may be* included. The argumentation *may be* generally well structured.

- The opinion expressed *supports* the statement. The explanation includes a reference to *at least one* interrelationship between humans and the tropical rain forest. The references *may not be entirely correct* and may remain on a *very general level*. The cartoon *may be* described as depicting the tropical rain forest as a hand that holds the earth and that is being cut off and interpreted to suggest that the tropical rain forest is a resource that humans need, but that they are destroying. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may for the most part* be logically structured and sufficiently linked.
- A subject-specific register is *mostly* used.

Level 2

There is an *acceptable* opinion about the extent to which the statement is true, but there is an *inadequate* explanation about the international significance of the tropical rain forest that *is not entirely plausible*. A brief description and valid interpretation of the cartoon is *probably not included*. The argumentation may show *major deficiencies* in terms of logical structuring and formulation. **Or**, there may be *no explicit opinion* expressed about the extent to which the statement is true. There is either a description *and / or* valid interpretation of the cartoon, **or** there is only an explanation about the international significance of the tropical rain forest that is *somewhat limited* and that may *not* be *entirely plausible.* The description/interpretation or explanation may show *major deficiencies* in terms of logical structuring and formulation.

- If an *opinion* is expressed, it *supports* the statement. If there is an *explanation* about the international significance of the tropical rain forest, it includes a *possibly incorrect* reference to *only one* interrelationship between humans and the tropical rain forest that remains on a *general level*. **Or**, if there is a *description and interpretation* of the *cartoon*, it *is* pointed out that the cartoon depicts the tropical rain forest as a hand that holds the earth and that is being cut off. The cartoon is interpreted to suggest that the tropical rain forest is a resource that humans need, but that they are destroying. **Or**, if there is an *explanation* about the international significance of the tropical rain forest, it includes a reference to *at least one* interrelationship between humans and the tropical rain forest. The reference *may not be entirely correct* and may remain on a *general level*. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may not* be logically structured and sufficiently linked.
- A subject-specific register may *not always* be used.

Level 1

No opinion is expressed about the extent to which the statement is true. There is either only a *brief* description of the cartoon that may be correct, **or** there is only an *inadequate* explanation about the international significance of the tropical rain forest that is *implausible and/or vague*. The description or explanation may show *major deficiencies* in terms of logical structuring and formulation. **Or**, there is an *unacceptable* opinion about the extent to which the statement is true and there is an *implausible* explanation about the tropical rain forest's lack of international significance. The argumentation may show *major deficiencies* in terms of logical structuring and formulation.

- If the cartoon is *described*, it *may be* pointed out that the cartoon depicts the tropical rain forest as a hand that holds the earth and that is being cut off. **Or**, if there is an *explanation* about the international significance of the tropical rain forest, it includes a reference to *only one* interrelationship between humans and the tropical rain forest that is either *incorrect* and/or that remains on a *general level*. **Or**, if an *opinion* is expressed, it *does not support* the statement. *Some content* may be included that is only *peripherally related* to the topic.
- The text *may not* be logically structured and sufficiently linked.
- A subject-specific register may *not always* be used.

Level 0

None of the required mini-genres is realised. *No opinion* is expressed about the extent to which the statement is true. There is *no explanation* about the international significance of the tropical rain forest and there is *no comprehensible description or valid interpretation* of the cartoon. The text is *severely deficient* in terms of logical structuring and formulation.

- There is *no support* for or *criticism of* the statement. There is *no explanation or description* that is *related to* the *international significance* of the tropical rain forest or the *cartoon*.
- The text is *unstructured* and *fragmentary*.
- A subject-specific register *is not* used.

Task 8

Level 6

An *acceptable* opinion is indicated on the scale. Succinct and *comprehensive* support is presented. The support is oriented towards the writer's *personal* scope for action and illustrates a *sophisticated* understanding of the complexity of the issues involved. The support is fully consistent with the opinion expressed. The argumentation is well structured and succinct.

- One of the options *absolutely true*, *true*, or *partly true and partly not true* is ticked on the scale. *At least three detailed* supporting arguments are included. These focus on *possibilities*, but *also on limitations* in terms of the extent to which the writer can *personally* contribute to the protection of the tropical rain forest. *Only relevant content* is included.
- The text is *consistently* logically structured and effectively linked.
- A subject-specific register is *consistently* used.

Level 5

An *acceptable* opinion is indicated on the scale. Clearly formulated and *adequate* support is presented. The support is oriented towards the writer's *personal* scope for action and illustrates a *sufficient* understanding of the complexity of the issues involved. The support is fully consistent with the opinion expressed. The argumentation is *generally* well structured and clearly formulated.

- One of the options *absolutely true, true*, or *partly true and partly not true* is ticked on the scale. *At least two* supporting arguments are included. The arguments are presented in *sufficient detail* and focus on *possibilities*, but *also on limitations* in terms of the extent to which the writer can *personally* contribute to the protection of the tropical rain forest. A *limited amount of content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text is *for the most part* logically structured and effectively linked.
- A subject-specific register is *consistently* used.

Level 4

An *acceptable* opinion is indicated on the scale. *Sufficient* support is presented. The support is oriented towards the writer's *personal* scope for action and illustrates a *basic* understanding of the complexity of the issues involved. There may be a *limited degree of inconsistency* in the support. The argumentation *may be* generally well structured and clearly formulated.

- One of the options *absolutely true, true*, or *partly true and partly not true* is ticked on the scale. *At least two* supporting arguments are included. The arguments are presented in *very little, if any detail*. The arguments focus on *possibilities*, but *also on limitations* in terms of the extent to which the writer can *personally* contribute to the protection of the tropical rain forest. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may for the most part* be logically structured and sufficiently linked.
- A subject-specific register is *mostly* used.

Level 3

An *acceptable* opinion is indicated on the scale. There is *some* support for the opinion expressed that may be *inconsistent.* The support is *more or less oriented* towards the writer's *personal* scope for action, but illustrates a *limited* understanding of the complexity of the issues involved. The argumentation *may be* generally well structured and clearly formulated.

- One of the options *absolutely true, true*, or *partly true and partly not true* is ticked on the scale. *At least two* arguments are presented. The arguments are presented in *very little, if any detail* and focus either exclusively on *possibilities or on limitations* in terms of the extent to which the writer*, personally, and* possibly also *people in general,* can contribute to the protection of the tropical rain forest. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may for the most part* be logically structured and sufficiently linked.
- A subject-specific register is *mostly* used.

Level 2

An *acceptable* opinion is indicated on the scale. However, there is *limited* support for the opinion expressed and the support may be *inconsistent*. The support is *more or less oriented* towards the writer's *personal* scope for action, but illustrates a *limited* understanding of the complexity of the issues involved. The argumentation may show *major deficiencies* in terms of logical structuring and formulation. **Or**, an *unacceptable* opinion is indicated on the scale. *Sufficient* support is presented that *may be consistent*. The support is *more or less oriented* towards the writer's *personal* scope for action. However, the support illustrates a *limited* understanding of the complexity of the issues involved. The argumentation may show *major deficiencies* in terms of logical structuring and formulation.

- In the case of an *acceptable opinion* with *limited support*, one of the options *absolutely true, true*, or *partly true and partly not true* is ticked on the scale. *Only one* argument is presented. The argument *may not be presented* in *any detail*. Furthermore, the argument focuses either exclusively on a *possibility*, *or on a limitation* in terms of the extent to which the writer*, personally, and* possibly also *people in general,* can contribute to the protection of the tropical rain forest. **Or**, in the case of an *unacceptable opinion* with *sufficient support*, one of the options *not true, or absolutely not true* is ticked on the scale. *At least two* arguments are presented. The arguments *may* be presented in *some detail*. However, the arguments focus exclusively on *limitations* in terms of the extent to which the writer*, personally, and* possibly also *people in general,* can contribute to the protection of the tropical rain forest. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may not be* logically structured and sufficiently linked.
- A subject-specific register may *not always* be used.

Level 1

An *acceptable* opinion is indicated on the scale. There *may be sufficient* support for the opinion expressed that may be *consistent*. However, the support *is not oriented* towards the writer's *personal* scope for action at all and may show a *limited* understanding of the complexity of the issues involved. The argumentation may show *major deficiencies* in terms of logical structuring and formulation. **Or**, an *unacceptable* opinion is indicated on the scale. There is *inadequate* support that may be *inconsistent*. The support *may be more or less oriented* towards the writer's *personal* scope for action. However, the support illustrates a *limited* understanding of the complexity of the issues involved. The argumentation may show *major deficiencies* in terms of logical structuring and formulation.

- In the case of an *acceptable opinion* with *sufficient support* that is *not oriented* towards the writer's *personal scope for action*, one of the options *absolutely true, true*, or *partly true and partly not true* is ticked on the scale. *More than one* argument *may be* included. The arguments *may be* presented *in some detail* and may focus either exclusively on *possibilities*, *or on limitations* in terms of the extent to which *people in general* can contribute to the protection of the tropical rain forest. **Or**, in the case of an *unacceptable opinion* with *inadequate support*, one of the options *not true, or absolutely not true* is ticked on the scale. *Only one* supporting argument is presented. The argument is *not presented* in *any detail*. Furthermore, the argument focuses exclusively on *a limitation* in terms of the extent to which the writer*, personally, and* possibly also *people in general,* can contribute to the protection of the tropical rain forest. *Some content* may be included that is only *peripherally related* to the topic.
- The text *may not* be logically structured and sufficiently linked.
- A subject-specific register may *not always* be used.

Level 0

An *acceptable opinion*, **or** an *unacceptable* opinion is indicated on the scale. However, there is *no support* for the opinion expressed. No inferences can be made about any aspects of the learner's writing.

- One of the options *absolutely true*, *true*, or *partly true and partly not true* is ticked on the scale. **O**r, one of the options *not true* or *absolutely not true*, is ticked on the scale.

Completeness and correctness of meaning constructed:

TASK 1C

CODE	ASPECTS TO BE RATED	SCORES		
A	**Meaning to be constructed**			
1	***MAIN ASPECTS to be included: Reference to the TYPE OF CLIMATE and inclusion of BOTH CLIMATIC ASPECTS: temperature and precipitation***			
1.1	Is there an identification of a particular *type of climate*?	Yes 0,5		No 0
1.2	Is there a description of *temperature*?	Yes 0,5		No 0
1.3	Is there a description of *precipitation / rainfall / amount of rain*?	Yes 0,5		No 0
2	***CORRECT IDENTIFICATION of the type of climate***			
2.1	Is Kisangani's climate identified as a *tropical climate*?	Yes 0,5		No 0
2.2	Is Kisangani's climate identified as a *type of tropical climate*?[1]	Yes 0,5[2]		No 0
2.3	Is it indicated that it is humid?	Yes 1	Partly[3] 0,5	No 0
2.4	Is it indicated that the precipitation is higher than the evaporation?	Yes 1		No 0
3	***DETAILS to be included in the description of temperature***			
3.1	Is it indicated that the temperature is *high*, or that it is about *25°C*?[4]	Yes 0,5		No 0
3.2	Is it indicated that the temperature remains (almost) *constant* throughout the year?	Yes 0,5		No 0
4	***DETAILS to be included in the description of precipitation***			
4.1	Is it indicated that the annual precipitation is *high*, or that the total rainfall per year is *1804 mm*?[5] / Is it indicated that it never rains less than 100 mm per month?	Yes 0,5		No 0
4.2	Is it indicated that the precipitation shows *variation*?	Yes 0,5		No 0
4.3	Is it indicated that the variation in the precipitation shows *two peaks*?[6]	Yes 0,5		No 0
5	***REFERENCES TO MEASURABLE PHENOMENA to be included in the descriptions of temperature and precipitation***			
5.1	Is information about the *average temperature* correctly included? (25°C)	Yes 0,5		No 0
5.2	Is information about the *total annual precipitation* correctly included? (1804 mm)	Yes 0,5		No 0
B	**Meaning that can additionally be constructed**			
6	***FURTHER DESCRIPTION of the CLIMATE***			
6.1	Is it indicated that there is no seasonal climate, but a diurnal climate?[7]	Yes 0,5		No 0
6.2	Is it indicated that this type of climate is normal along the equator?	Yes 1		No 0
6.3	Is it indicated that the high precipitation and high temperature lead to high humidity?	Yes 0,5		No 0
6.4	Are the two peaks in rainfall related to inter-tropical convergence?	Yes 0,5		No 0
6.5	Is it indicated that the vegetation period lasts the whole year?	Yes 0,5		No 0
6.6	Is there a comparison between the total annual precipitation in Kisangani and Osnabrück?	Yes 0,5		No 0

Sub-total for meaning constructed: 8+

[1] For instance if the term "tropical climate" is not directly used, but it is indicated that the climate is similar to the one in the tropical rain forest.
[2] Award this mark if the term *tropical climate* is used. In other words, if a mark was awarded at 1.1, a mark should automatically be awarded for 1.2.
[3] Award 0,5 if the climate is described as hot and damp/wet.
[4] It is technically incorrect to say that it is hot or warm. There must be a reference to the temperature in degrees Celsius.
[5] It is technically incorrect to say that it is wet or damp. There must be a reference to the amount of precipitation in millimetres.
[6] Award a mark, even if the learner does not explicitly refer to two peaks, but if it becomes clear from the description that he or she identifies two and not three periods during which the rainfall is at a maximum.
[7] It is also acceptable if a learner writes day climate or daytime climate.

TASK 2A

CODE	ASPECTS TO BE RATED	SCORES		
A	**Meaning to be constructed**			
1	***MAIN ASPECTS to be included: REFERENCES to ALL THREE LAYERS of the tropical rain forest***			
1.1	Is there any reference going beyond the mere naming of the layer to *the crowns of the giant trees*?	Yes [1] 1	Partly[2] 0,5	No 0
1.2	Is there any reference going beyond the mere naming of the layer to *the tree top layer with closed canopy*?	Yes [3] 1	Partly[4] 0,5	No 0
1.3	Is there any reference going beyond the mere naming of the layer to *the bush and shrub layer*?	Yes [5] 1	Partly[6] 0,5	No 0
2	***MAIN ASPECTS to be included: Reference to EFFECTS of incidence of light, temperature and air humidity ON PLANTS AND ANIMALS***			
2.1	Is there any reference made to the possible influences of the prevailing living conditions *on plants*?	Yes[7] 1	Partly[8] 0,5	No 0
2.2	Is there any reference made to the possible influences of the prevailing living conditions *on animals*?	Yes 1	Partly 0,5	No 0
3	***DETAILS to be included in the description of the living conditions in the layer formed by the crowns of the giant trees***			
3.1	Is it indicated that the *incidence of light* is high/higher in this layer / that it is 100%?	Yes 0,5	No 0	
3.2	Is it indicated that the *temperature* is high/higher in this layer / that it is 35°C?	Yes 0,5	No 0	
3.3	Is it indicated that the *air humidity* is low/lower/the lowest in this layer.	Yes 0,5	No 0	
3.4	Is there a reference to the lower *number of animals* in this layer?	Yes 0,5	No 0	
4	***DETAILS to be included in the description of the living conditions in the tree top layer with a closed canopy***			
4.1	Is it indicated that the *incidence of light* is lower than in the layer formed by the crowns of the giant trees / higher than in the bush and shrub layer / that it is between 40% and 70%?	Yes 0,5	No 0	
4.2	Is it indicated that the *temperature* is lower than in the layer formed by the crowns of the giant trees / higher than in the bush and shrub layer / that it is about 30°C?	Yes 0,5	No 0	
4.3	Is it indicated that the *air humidity* is higher than in the layer formed by the crowns of the giant trees / lower than in the bush and shrub layer?	Yes 0,5	No 0	
4.4	Is there a reference to the higher *number of animals* in this layer?	Yes 0,5	No 0	
5	***DETAILS to be included in the description of the living conditions in the bush and shrub layer***			
5.1	Is it indicated that the *incidence of light* is low/lower in this layer / is about 15% - 10%?	Yes 0,5	No 0	
5.2	Is it indicated that the *temperature* is low/lower/the lowest in this layer / is 25°C and lower?	Yes 0,5	No 0	
5.3	Is there a reference to the higher *air humidity* in this layer?	Yes 0,5	No 0	
5.4	Is there a reference to the higher *number of animals* in this layer?	Yes 0,5	No 0	

6 continues on next page.

[1] Award 1 if one of the correct terms: "crowns of the giant trees"/"giant trees" is used
[2] Award 0,5 if there is a reference to the layer, but the correct term is not used.
[3] Award 1 if one of the correct terms: "crowns of the trees with a closed canopy"/"tree top layer with closed canopy" is used.
[4] Award 0,5 if there is a reference to the layer, but the correct term is not used.
[5] Award 1 if the correct terms: "bush and shrub layer" is used.
[6] Award 0,5 if there is a reference to the layer, but the correct term is not used.
[7] The reference is made in terms of two to three layers. The same applies to scale 2.2.
[8] The reference is only made in terms of one layer. The same applies to scale 2.2.

6	***REFERENCES TO MEASURABLE PHENOMENA to be included in the descriptions of all three layers***			
6.1	Is information about the height of the giant trees correctly included? (40m – 60m / 60m and higher)	Yes 0,5	No 0	
6.2	Is information about the incidence of light in the layer formed by the tops of the giant trees correctly included? (100%)	Yes 0,5	No 0	
6.3	Is information about the maximum temperature in the layer formed by the tops of the giant trees correctly included? (35°C)	Yes 0,5	No 0	
6.4	Is information about the height of the tree top layer with a closed canopy correctly included? (20m – 40m)	Yes 0,5	No 0	
6.5	Is information about the incidence of light in the tree top layer with a closed canopy correctly included? (between 40% and 70%)	Yes 0,5	No 0	
6.6	Is information about the maximum temperature in the tree top layer with a closed canopy correctly included? (about 30°C)	Yes 0,5	No 0	
6.7	Is information about the height of the bush and shrub layer correctly included? (0-15 m / 10m)	Yes 0,5	No 0	
6.8	Is information about the incidence of light in the bush and shrub layer correctly included? (about 15% - 10%)	Yes 0,5	No 0	
6.9	Is information about the maximum temperature in the bush and shrub layer correctly included? (25°C and lower)	Yes 0,5	No 0	
7	***SHOWING RELATIONS: Comparison of layers and reference to relations between different aspects playing a role in the living conditions in each layer***			
7.1	Is there an attempt to compare the living conditions in the layers to one another?	Yes[1] 2	Partly 1[2]	No 0
7.2	Is there an attempt made to draw causal relations between the aspects playing a role in the living conditions?	Yes, explicitly[3] 2	Yes, implicitly[4] 1	No 0
B	**Meaning that can additionally be constructed**			
8.1	Is there a reference to the fact that the tropical rain forest is a particularly fragile ecosystem?	Yes……0,5	No…….0	

Sub-total for meaning constructed: 19,5+

[1] The comparison relates to all three layers.
[2] The comparison relates to two layers.
[3] The causal relation is named.
[4] The causal relation is suggested through the ordering of ideas.

TASK 3

CODE	ASPECTS TO BE RATED	SCORES		
A	**Meaning to be constructed**			
1	***MAIN ASPECTS to be included: CRITICAL COMMENT and SUPPORT***			
1.1	Is there any *critical comment* made that relativises or criticises the statement made by Figure 4?	Yes 1		No 0
1.2	Is this critical comment/suggestion *supported* in any way?[1]	Yes 1		No 0
2	***ACCEPTABILITY of COMMENT***			
2.1	Is it indicated that regarding the *proportional decrease*, or *in percentage terms*, or in terms of *the size of the original rain forest*, the destruction is the biggest in *Asia*?	Yes 1	Partly[2] 0,5	No 0
2.2	Is it indicated that regarding *the size of areas of tropical rain forest* that are destroyed, or in *absolute terms* the destruction is the biggest in *America*?	Yes 1	Partly[3] 0,5	No 0
3	***SUPPORT in the form of MEASURABLE PHENOMENA***			
3.1	Is any of the following pieces of information about the annual changes of forest area in km^2 in America correctly included? (all American countries together: -37 780 km^2 / Brazil: -23 090 km^2 / an approximation of the extent of the changes)	Yes 1		No 0
3.2	Is information about the annual changes of forest area in km^2 in Asia correctly included? (-20 660 km^2)	Yes 1		No 0
4	***Further SUPPORT for critical commentary***			
4.1	Is it pointed out that the compound score of the *annual changes of forest area in km^2* in the six *American countries* is *higher* than the compound score of the annual changes in km^2 in the three Asian countries. / Is it pointed out that the *annual changes of forest area in km^2* in *Brazil* alone is higher than the *annual changes of forest area in km^2* in the three Asian countries put together?[4]	Yes 1		No 0
4.2	Is it pointed out that the *total area* covered by forest is *bigger in / the biggest in America/Brazil*?	Yes 1		No 0
5	***Presenting SUGGESTION to improve the statement***			
5.1	Is there any *suggestion* made *to improve the statement* presented in the task (by distinguishing between proportional and absolute terms)?	Yes 1		No 0
6	***REFLECTION about DEPICTION of DATA***			
6.1	Is there any reflection about the differences between the forms in which the data is presented in the two sources?	Yes 1		No 0

Sub-total for meaning constructed: 10

[1] It is not possible to get a mark at 1.2, if a mark was not awarded at 1.1, because the points raised must support the critical point of view taken in terms of the acceptability of the statement.

[2] Award 0,5 if the statement is relativised, for instance if a statement such as: "The destruction is not necessarily the biggest in Asia." is made.

[3] Award 0,5 if the statement is relativised, for instance if a statement such as: "The destruction could be higher in America." is made.

[4] Award the mark, even if the learner does not refer to Asia specifically, but if a comparison is expressed, for instance, if the learner writes: "The decrease in forest area is the biggest in America".

TASK 6

CODE	ASPECTS TO BE RATED	SCORES		
A	**Meaning to be constructed**			
1	***MAIN ASPECTS to be included: Expression of OPINION about suggestion and APPLICATION of assessment CRITERIA***			
1.1	Is an *opinion* about the suggestion expressed?	Yes 0,5	No 0	
1.2	Is any form of *support* presented for the particular opinion expressed through the application of some or other *assessment criteria*?	Yes 0,5	No 0	
2	***ACCEPTABILITY of OPINION expressed***			
2.1	Is it clear from the assessment as a whole that the suggestion is regarded as *not fully acceptable or appropriate*?	Yes 1	No 0	
3	***CRITERIA to be applied in assessing the suggestion***[1]			
3.1	Is it indicated that *ecological goals* should be taken into account, e.g. that the tropical rain forest should be protected?	Yes[2] 1	Partly[3] 0,5	No 0
3.2	Is it indicated that the *rights and needs of future generations* should be taken into account, e.g. that future generations must also have the opportunity to use the tropical rain forest?	Yes 1	Partly 0,5	No 0
3.3	Is it indicated that *economic goals* should be taken into account, e.g. that it must still be possible to use the rain forest as a source of income?	Yes 1	Partly 0,5	No 0
3.4	Is it indicated that *social goals* should be taken into account, e.g. that it must still be possible for the indigenous people to live in the rain forest?	Yes 1	Partly 0,5	No 0
3.5	Is it indicated that a balance must be found between ecological goals, rights and needs of future generations, social goals and economic goals? / Is it indicated that the suggestion does not make equal provision for the four goals of the principle of sustainability?	Yes 1	Partly 0,5[4]	No 0
B	**Meaning that can additionally be constructed**			
4	***SHOWING RELATIONS: Linking abstract categories to concrete examples***			
4.1	Is there an attempt made to give examples and to link them to any of the following four categories: ecological goals, rights and needs of future generations, social goals and economic goals?	Two or more 2	At least one 1	No 0
5	***Presenting ALTERNATIVE suggestions***			
5.1	Is there an attempt made to offer one or more alternative suggestions?	Yes 1	No 0	
5.2	Is at least one of the alternative suggestions in line with the principles of sustainability?	Yes 1	No 0	
6	***Applying ADDITIONAL CRITERIA***			
6.1	Additional reference(s) to an *ecological* goal /aspect.	Yes[5] 1	Partly[6] 0,5	No 0
6.2	Additional reference(s) to *the rights and needs of future generations*.	Yes 1	Partly 0,5	No 0
6.3	Additional reference(s) to a *social* goal /aspect.	Yes 1	Partly 0,5	No 0
6.4	Additional reference(s) to an *economic* goal /aspect.	Yes 1	Partly 0,5	No 0

Sub-total for meaning constructed: 7+

[1] The minimum requirement is one reference to each of the four types of goals: ecological goals, the protection of the rights and needs of future generations, social goals and economic goals. This implies that more than one reference to any one of the four types of goals should be rated under "Meaning that can additionally be constructed".
[2] The criteria used relate to the aims of Agenda 21 or the principles of sustainability.
[3] Although the criteria used fall under one of the categories: ecological aspects, rights and needs of future generations, social aspects and economic aspects, the criteria do not directly relate to the aims of Agenda 21 or the principles of sustainability. To illustrate, there might be a reference to an economic aspect, such as the cost of running a national park.
[4] Award 0,5 if two to three aspects are incorporated.
[5] The criteria used relate to the aims of Agenda 21 or the principles of sustainability.
[6] Although the criteria used fall under one of the categories: ecological aspects, rights and needs of future generations, social aspects and economic aspects, the criteria do not directly relate to the aims of Agenda 21 or the principles of sustainability. To illustrate, there might be a reference to an economic aspect, such as the cost of running a national park.

TASK 7

CODE	ASPECTS TO BE RATED	SCORES		
A	**Meaning to be constructed**			
1	***MAIN ASPECTS to be included: Expression of OPINION about statement and inclusion of SUPPORT***			
1.1	Is there an *opinion* expressed about the truth of the statement?	Yes 0,5	No 0	
1.2	Are any points raised that *support* this opinion?	Yes 0,5	No 0	
2	***ACCEPATBILITY of OPINION expressed***			
2.1	Is the statement evaluated as *true*?	Yes 1	No 0	
3	***QUALITY AND NUMBER OF ARGUMENTS presented in evaluation***[1]			
The following arguments are all acceptable.				
3.1	Is it pointed out that the tropical rain forest *globally affects climate* and that the destruction of the tropical rain forest thus has consequences for everybody on earth? / Is there a reference to the relation between the destruction of the tropical rain forest and the *greenhouse effect*?	Yes,[2] detailed 2	Yes,[3] basic 1	No 0
3.2	Is it pointed out that the tropical rain forest is the *"lungs of the earth"* and is needed for *balance on earth*? / Is there a reference to the process whereby *CO2 is changed into oxygen* through the process of photosynthesis?	Yes, detailed 2	Yes, basic 1	No 0
3.3	Is it pointed out that the tropical rain forest is a *unique ecosystem* and that this is a value in its own right?	Yes, detailed 2	Yes, basic 1	No 0
3.4	Is it pointed out that the tropical rain forest ensures *biodiversity* and that this is a value in its own right? / Is it pointed out that a *large number of animals* live in the tropical rain forest and that there is a *large variety of species*?	Yes, detailed 2	Yes, basic 1	No 0
3.5	Is it pointed out that the tropical rain forest is valuable as a *medical resource* ("the chemistry of the future") that humans can benefit from?	Yes, detailed 2	Yes, basic 1	No 0
3.6	Is it pointed out that the tropical rain forest provides *wood* that is used all over the world.	Yes, detailed 2	Yes, basic 1	No 0
3.7	Is it pointed out that there are groups of people for whom the tropical rain forest is a *living space* and that due to humanitarian reasons it concerns us?	Yes, detailed 2	Yes, basic 1	No 0
3.8	Is it pointed out that the tropical rain forest is a *resource for developing countries* and that, if destroyed, developed countries would have to provide even more aid to these developing countries in the long run?	Yes, detailed 2	Yes, basic 1	No 0
3.9[4]	Is it pointed out that if we destroy nature, e.g. the tropical rain forest, we destroy the basis of our existence on earth?	Yes 1	No 0	
4	***INVOLVING the CARTOON***			
4.1	Is there a reference to the cartoon? / Is the cartoon used as a basis for an argument/the arguments?	Yes, explicitly 1	Yes,[5] implicitly 0,5	No 0
4.2	If there is an interpretation of the cartoon, is it valid?	Yes 1	No 0	
B	**Meaning that can additionally be constructed**			
5	***SHOWING RELATIONS: Linking abstract categories to concrete examples***			
5.1	Is there any attempt made to present the arguments as concrete examples of abstract categories, e.g. humanitarian, ecological, economic, medical etc. considerations?	Yes 1	No 0	

Sub-total for meaning constructed: 7+

[1] The maximum that can be achieved here is 3. In other words, the minimum requirement is either one detailed and one basic argument, or three basic arguments. All marks above 3 are to be calculated under "Meaning that can additionally be constructed".

[2] An example of detailed argument would be: "The tropical rain forest absorbs CO_2 and produces oxygen through the process of photosynthesis. This is why it is called 'the green lungs of the earth'. The destruction of the tropical rain forest will disturb this process, which will mean that there will be more CO_2 and less oxygen. This will have effects for all humans, since we need oxygen to survive.".

[3] An example of a basic argument is: "The tropical rain forest is the green lungs of the earth. We need it for balance on earth.".

[4] A mark is only awarded at 3.9, if a learner was not given a mark under any of the previous categories.

[5] Although there is no direct or explicit reference to the cartoon, there are references to the destruction of the tropical rain forest and / or deforestation.

TASK 8

CODE	ASPECTS TO BE RATED	SCORES		
A	**Meaning to be constructed**			
1	***MAIN ASPECTS to be included: Indication of OPINION and presentation of SUPPORT***			
1.1	Is an *opinion* indicated on the scale?	Yes 0,5		No 0
1.2	Is any form of *support* presented for the particular opinion indicated?	Yes 0,5		No 0
2	***ACCEPATBILITY of OPINION indicated***			
2.1	Is any of the following three options selected: - absolutely true - true - partly true, partly not true?	Yes 1		No 0
3	***QUALITY AND NUMBER OF ARGUMENTS presented to support the evaluation***[1]			
The following arguments are all acceptable.				
3.1	Is it pointed out that one can contribute by *not buying* or *consuming* products that have been won in ways that are harmful to the tropical rain forest?	Yes,[2] detailed 2	Yes,[3] basic 1	No 0
3.2	Is it pointed out that one can contribute by *informing oneself* about issues related to the tropical rain forest and by *developing an understanding* of the problems involved?	Yes, detailed 2	Yes, basic 1	No 0
3.3	Is it pointed out that one can contribute by *making people* in one's immediate environment *aware* of problems related to the tropical rain forest?	Yes, detailed 2	Yes, basic 1	No 0
3.4	Is it pointed out that one can contribute *by joining environmental organizations* involved in the protection of the tropical rain forest?	Yes, detailed 2	Yes, basic 1	No 0
3.5	Is it pointed out that one can make a contribution, for instance by *donating money* to aid agencies, or by *supporting certain projects* where money spent on products is used to protect the tropical rain forest?	Yes, detailed 2	Yes, basic 1	No 0
3.6	Is it pointed out that one can *influence decision-makers*, for instance through protests, as well as political involvement and action?	Yes, detailed 2	Yes, basic 1	No 0
3.7	Is it pointed out that as an individual, one's *power* to have an effect is *limited*, because of the global dimension of the problem?	Yes, detailed 2	Yes, basic 1	No 0
3.8	Is it pointed out that it is difficult to do something about the problem, because of the fact that there are *powerful groups* which have a vested interest in the tropical rain forest as a source of income?	Yes, detailed 2	Yes, basic 1	No 0
4	**Presenting a *BALANCED VIEW***			
4.1	Are there references to *possibilities and limitations* in terms of one's personal scope for action?	Yes, compre-hensive[4] 2	Yes, adequate[5] 1	No 0
B	**Meaning that can additionally be constructed**			
5	**DRAWING RELATIONS: Linking abstract categories to concrete examples**			
5.1	Is there any attempt made to present the arguments as concrete examples of abstract categories, e.g. contributions on a humanitarian, an ecological, an economic etc. level?	Yes 1		No 0

Sub-total for meaning constructed: 7+

[1] The maximum that can be achieved here is 3. In other words, the minimum requirement is either one detailed and one basic argument, or three basic arguments. All marks above 3 are to be calculated under "Meaning that can additionally be constructed".

[2] An example of a detailed argument would be: "I can contribute by not buying goods of which the production is harmful to the tropical rain forest, e.g. furniture made from tropical wood."

[3] An example of a basic argument is: "I should not buy things from the tropical rain forest."

[4] There are at least two references to possibilities (3.1 – 3.6) and two references to limitations (3.7 and 3.8).

[5] There is at least one reference to a possibility and there is one reference to a limitation/ There is at least one reference to a possibility and there are two references to limitations.

Appropriateness of the textual realisation of meaning

1) Effectiveness of textual organisation of meaning

Scale 1.1: Effective ordering of structural units in the text

Descriptor	Score
The order in which sentence parts, sentences (and paragraphs) follow each is consistently logical and there is no repetition in the meaning constructed.	4
The order in which sentence parts and sentences follow each other is for the most part logical. Seen in relation to the text's length, there are only a limited number of instances in which the logic underlying the position of a sentence part, or sentence is not clear, or there is limited repetition in the meaning constructed. If there are paragraphs, they all follow each other in a logical order.	3
There is some logic evident in the order in which sentence parts and sentences follow each other to form a text. However, seen in relation to the text's length, there are a number of instances in which the organisation of the structural units is not logical. The order in which a number of sentence parts and sentences are presented in the text may not be logical and/or there is some repetition in the meaning constructed. If there are paragraphs, the position of a particular paragraph is not logical in terms of the structure of the text as a whole.	2
There is little clearly identifiable logic to the order in which sentences follow each other and/or there is frequent repetition in the meaning constructed. If there are paragraphs, the logic underlying the ordering of the paragraphs is not clear.	1
There is no identifiable logic to the order in which sentences (and paragraphs) follow each other and/or there is constant repetition in the meaning constructed.	0

Scale 1.2: Effective organisation of meaning constructed into different structural units

Descriptor	Score
The structural presentation of meaning in sentences (and paragraphs) is consistently logical.	4
The structural presentation of meaning in sentences (and paragraphs) is for the most part logical. Seen in relation to the text's length, there are only a limited number of instances in which, for instance, too many thematic items and semantic relations are presented in one sentence (paragraph), or semantically closely interrelated thematic items are divided between different sentences (paragraphs).	3
The structural presentation of meaning in sentences (and paragraphs) is to some extent logical. However, seen in relation to the text's length, there are a number of instances in which, for instance, too many thematic items and semantic relations are presented in one sentence (paragraph), or semantically closely interrelated thematic items are divided between different sentences (paragraphs).	2
There is little clearly identifiable logic to the structural presentation of meaning in sentences (and paragraphs).	1
There is no identifiable logic to the structural presentation of meaning in sentences (and paragraphs).	0

Scale 1.3: Effective linking of structural units

Descriptor	Score
Sentences and sentence parts are consistently well linked, producing a clear flow of ideas.	4
Sentences and sentence parts are for the most part well linked. Seen in relation to the text's length, there are only a limited number of instances in which the linking could have been more effective.	3
Sentences and sentence parts are to some extent linked. However, seen in relation to the text's length, there are a number of instances of disconnectedness, so that parts of the text might be fragmentary or choppy.	2
Sentences and sentence parts are linked to a limited degree, causing parts of the text to be so fragmentary that a flow of ideas is obstructed.	1
There is no identifiable linking of sentences and sentence parts.	0

2) Appropriateness of language use

Scale 2.1: Sufficient and correct use of subject-specific terms and expressions

Descriptor	Score
Subject-specific terms and expressions are consistently used where references are made to subject-related phenomena and are continuously used correctly.	4
Subject-specific terms and expressions are used to a large extent where references are made to subject-related phenomena and are used correctly in most instances. Only a limited number of subject-specific terms are lacking and/or a limited number of subject-specific terms and expressions are not used correctly. Incorrect terms might be used and/or collocations and spelling might be incorrect.	3
Subject-specific terms and expressions are used, but not to a sufficient extent and/or frequent errors are made in the application of subject-specific terms. About half of the references to subject-related phenomena are either made using terms and expressions that are not part of the subject-specific register and/or show errors regarding the terms used, collocations and spelling.	2
The use of subject-specific terms and expressions is limited and/or only a limited number of terms and expressions included are used correctly. In most instances, there are errors regarding the terms used, collocations and spelling.	1
No subject-specific terms or expressions are used.	0

Scale 2.2: Sufficient use of formal language and clear, succinct formulations

Descriptor	Score
Formal language and clear, succinct formulations are used throughout the text.	4
Formal language and clear, succinct formulations are used most of the time. Seen in relation to the text's length, there are only a limited number of instances where a colloquial form, rather than the more formal variant is used and/or where a formulation could have been clearer, or more succinct.	3
Formal language and clear, succinct formulations are used to some extent. However, seen in relation to the text's length, there are a number of instances in which the language use is not formal and/or in which formulations are vague, or could have been more succinct.	2
The use of formal language and clear and succinct formulations is limited. The style is predominantly informal and/or most formulations are vague.	1
The language use is consistently informal and/or vague.	0

Scale 2.3: Sufficient use of general academic vocabulary and structures realising semantic relations

Descriptor	Score
There is considerable variation in the vocabulary used and words and expressions belonging to a general academic register are included. A number of structures realising complex semantic relations are included.	4
There is sufficient variation in the vocabulary used. Some words and expressions belonging to a general academic register might be included. Some structures realising complex semantic relations might be included.	3
There is some variation in the vocabulary used. A limited number of words and expressions belonging to a general academic register might be included. A limited number of structures realising complex semantic relations might be included.	2
There is limited variation in the vocabulary used. No words and expressions belonging to a general academic register are included. There might be repetition in the vocabulary and expressions used. No structures realising complex semantic relations are included.	1
There is constant repetition in the grammar structures and vocabulary used. Only words belonging to a basic vocabulary and structures realising simple semantic relations are used.	

Scale 2.4: Correctness of grammar, vocabulary and punctuation

Descriptor	Score
Language is used with a very high level of accuracy. There might be one or two minor errors, e.g. a word might be misspelled or a punctuation mark might be absent.	4
Language is used with a relatively high level of accuracy. Most of the errors that do occur are minor errors, such as spelling and punctuation errors and the use of the incorrect part of speech. None of the errors obscures the meaning constructed on sentence level.	3
Language is used with some degree of accuracy. Major errors, such as tense, concord, sentence structure and word choice errors, as well as minor errors, occur. In a limited number of cases the errors might obscure the meaning constructed on sentence level.	2
Language is used with a low level of accuracy. There is a large number of major and minor errors that might cause some of the sentences to be incomprehensible.	1
Language is used with such a low level of accuracy that the meaning constructed is obscured.	0

Appendix D

Guidelines for text production and analyses taken from geography textbooks

Examples of guidelines for the production of texts with particular rhetorical structures (mini-genres) and for the analysis of texts, taken from *Diercke: Erdkunde Klasse 11*, 2001, published by Westermann, *Diercke: Erdkunde für Gymnasien in Niedersachsen, Klasse 11*, 2006, published by Westermann and *Diercke: Erdkunde für Gymnasien in Niedersachsen, Klasse 7/8*, 2005, published by Westermann.

- Understanding test instructions from Diercke (2001: 189)
- Understanding instructions from Diercke (2006: 157)
- Analysing a climate graph from Diercke (2005: 16-17)

Methode: Fragestellungen – die Klausur

Klausuren sind Hürden, die übersprungen werden müssen. Ihre Bearbeitung fällt aber leichter, wenn Sie folgende Tipps beherzigen.

1. Die Vorbereitung der Klausur:
Bereiten Sie sich inhaltlich und methodisch gut auf die Klausur vor, indem Sie den Stoff aus dem vorangegangenen Unterricht gründlich wiederholen (nicht erst am Tag vor der Klausur). Zusammenfassungen auf Karteikärtchen sind oftmals sehr hilfreich.

2. Das Vorgehen in der Klausur:
- Lesen Sie alle Aufgaben (siehe hierzu auch Punkt 3) genau durch und schauen Sie sich das angegebene Arbeitsmaterial an.
- Ordnen Sie das Arbeitsmaterial den Aufgaben zu, soweit das vom Lehrer noch nicht geschehen ist.
- Erstellen Sie nach dieser ersten Übersicht einen Zeitplan für Ihre Bearbeitung.
- Werten Sie die Materialien aus! Marker, Farbstifte und Randbemerkungen helfen beim Kennzeichnen bestimmter Sachverhalte und Kernaussagen. Achten Sie auf Querverbindungen und Verknüpfungen bei den Materialien.
- Legen Sie zu jeder Aufgabe in Stichworten eine übersichtliche Gliederung an, die als roter Faden für die Reinschrift dienen soll.
- Fertigen Sie die Reinschrift an (siehe hierzu auch Punkt 4)! Beachten Sie hierbei, dass Ihre Aussagen immer einen klaren Materialbezug haben.

3. Die Arbeitsanweisungen:
Die Klausuraufgaben enthalten, genauso wie die Arbeitsaufträge im Unterricht, eine genaue Anweisung, was von Ihnen in der entsprechenden Aufgabe verlangt und erwartet wird. Hierzu ein Überblick:
- **nennen:** ohne Erläuterungen auflisten/aufzählen
- **wiedergeben:** vorgegebene/bekannte Inhalte in eigenen Worten wiederholen/zusammenfassen
- **skizzieren:** in groben Zügen das Wichtigste verdeutlichen
- **beschreiben:** über einen Sachverhalt durch umfassende Angaben berichten
- **darstellen:** einen Sachverhalt mit Text, Diagramm, Tabelle und/oder Zeichnung ausführlich wiedergeben
- **gliedern:** nach vorgegebenen oder selbstgewählten Merkmalen ordnen/unterscheiden
- **kennzeichnen:** Typisches/Auffälliges herausstellen und ggf. veranschaulichen
- **untersuchen:** an einen Gegenstand gezielte Arbeitsfragen stellen und gewonnene Erkenntnisse darlegen
- **vergleichen:** zu vorgegebenen oder selbst gewählten Gesichtspunkten Gemeinsamkeiten und Unterschiede feststellen
- **übertragen:** Vorgegebenes/Bekanntes auf ein anderes (Raum-)Beispiel übertagen
- **erklären:** das Verstehen von Erscheinungen/Entwicklungen/Zusammenhängen/ Ursachen ermöglichen
- **erläutern:** Sachverhalte beschreiben und deren Beziehungen verdeutlichen
- **analysieren:** Strukturmerkmale und deren Zusammenhänge (veranschaulichend) herausarbeiten
- **erarbeiten:** ein neues Konzept in Ansätzen entwickeln
- **ableiten:** auf der Grundlage vorhandener/bekannter Ergebnisse eigene Schlussfolgerungen ziehen
- **(über-)prüfen:** eine Hypothese (begründete Vermutung) an neuen Beobachtungen oder ihrer inneren Logik messen
- **diskutieren/erörtern:** einen Sachverhalt im Hinblick auf positive und negative Erscheinungen/Wirkungen ausführlich untersuchen, die Ergebnisse darstellen und bewerten
- **Stellung nehmen:** zu einem Sachverhalt/einer Behauptung eine eigene, begründete und bewertende Meinung äußern
- **beurteilen, bewerten:** ein eigenes Urteil über Richtigkeit/Wahrscheinlichkeit/Zumutbarkeit/Angemessenheit/Anwendbarkeit eines Sachverhaltes/einer Behauptung fällen und dieses Urteil stichhaltig begründen

(zusammengestellt nach: Brameier, U. und W. Fraedrich: Klausuren: Tpps für den Erfolg. - In: „Praxis Geographie" 10/1993, S. 10)

4. Die Reinschrift:
- Lassen Sie einen ausreichenden Rand für die Korrektur.
- Fassen Sie Ihre Darstellung gemäß den Arbeitsanweisungen klar, verständlich und sprachlich einwandfrei ab. (Achten Sie auf Rechtschreibung, Zeichensetzung und Grammatik.)
- Schreiben Sie Ihre Antworten in ausformulierten Sätzen!
- Gliedern Sie Ihre Darstellung formal durch Absätze. Es erleichtert Ihnen die Übersicht über Ihre Bearbeitung.
- Kennzeichnen Sie die wörtliche Übernahme von Textpassagen als Zitat.

5. Der Schluss:
- Lesen Sie Ihre Bearbeitung abschließend unter Beachtung der Punkte 3 und 4 noch einmal durch. (Diese Zeit sollten Sie von vornherein mit einplanen!)

Entschlüsselung von Arbeitsanweisungen – Wie sollen Inhalte präsentiert werden?

Arbeitsanweisungen: Was wird von Ihnen erwartet?	
Anforderungsbereich I	
(be-)nennen	Sachverhalte ohne Erläuterung angeben
wiedergeben	bekannte Sachverhalte oder einem Material entnommene Informationen mit eigenen Worten unkommentiert zusammenfassen
darlegen/darstellen	Sachverhalte detailliert und fachsprachlich angemessen aufzeigen
beschreiben	gesetzmäßige und raumspezifische Sachverhalte aus Materialien strukturiert darlegen
gliedern	einen Raum nach selbst gewählten oder vorgegebenen Kriterien systematisierend ordnen
Anforderungsbereich II	
einordnen/zuordnen	Sachverhalte in einen systematischen Zusammenhang einfügen
charakterisieren	geographische Sachverhalte in ihren Eigenarten beschreiben und typische Merkmale kennzeichnen
analysieren	ein Ganzes (z.B. einen Raum) nach bekannten Ordnungsmerkmalen aufgliedern und systematisch untersuchen
erklären	Sachverhalte so darstellen, dass Bedingungen, Ursachen und Gesetzmäßigkeiten verständlich werden
erläutern	Sachverhalte in ihren komplexen Beziehungen verdeutlichen (auf der Grundlage von Kenntnissen bzw. Materialanalyse)
vergleichen	Gemeinsamkeiten und Unterschiede von geographischen Sachverhalten erkennen und darlegen
Anforderungsbereich III	
beurteilen	begründete Aussagen über die Richtigkeit, Wahrscheinlichkeit, Angemessenheit bzw. Anwendbarkeit eines Sachverhalts machen, ohne persönlich Stellung zu nehmen
Stellung nehmen	zu einem Sachverhalt bzw. einer Behauptung differenziert argumentierend eine eigene Meinung äußern
erörtern	einen Sachverhalt oder eine vorgegebene Aussage eingehend von verschiedenen Seiten, das Für und Wider abwägend betrachten und zu einer abschließenden Einschätzung kommen
entwickeln	Vorschläge, Einschätzungen, Maßnahmen darlegen, die zu einer inhaltlich weiterführenden und zukunftsorientierten Betrachtung führen

Quelle: www.nibis.de/nli1/gohrgs/operatoren/operatoren_erdkunde.pdf

16 METHODE • METHODE • METHOD

Klimadiagramme zeichnen und auswerten

Deine nächste Urlaubsfahrt ist etwas besonderes und soll zusammen mit deinen Eltern nach Tunis in Tunesien führen. Du möchtest wissen, wie das Wetter an deinem Urlaubsort sein wird. Kannst du die Badehose oder den Badeanzug in den Koffer packen oder ist es dann in Tunis zu kalt oder zu regnerisch zum Baden? Der aktuelle Wetterbericht gibt darüber für ein paar Tage Auskunft. Da du die Reise aber erst in einigen Monaten antreten willst, hilft dir ein Klimadiagramm, das das Wetter an einem bestimmten Ort (hier Tunis) über einen längeren Zeitraum darstellt. Darin werden Durchschnittswerte von Temperatur und Niederschlag über einen Zeitraum von mindestens 30 Jahren ermittelt und in das Klimadiagramm eingezeichnet. Mit den Niederschlags- und Temperaturwerten von Tunis (M1) kannst du das Klima in einem Klimadiagramm besonders gut auf Millimeterpapier darstellen. Beachte die folgenden Schritte, um das Klimadiagramm richtig zu zeichnen und später auswerten zu können.

		J	*F*	*M*	*A*	*M*	*J*	*J*	*A*	*S*	*O*	*N*	*D*	*Jahr*
Tunis/Tunesien, 4 m ü. M.	T in °C	11,3	11,9	13,1	15,6	19,2	23,0	26,3	26,6	24,1	20,3	15,8	12,4	18,3
36°50'N/10°14'O	N in mm	60	54	46	37	23	10	3	7	35	67	58	67	467

M1 *Temperaturen und Niederschläge in Tunis/Tunesien*

Beachte (1. Schritt):
Der Wert auf der rechten Achse ist immer doppelt so groß wie der auf der linken Achse (bei einem Monatsmittel von 10 °C verdunsten 20 mm des Monatsniederschlags, bei 20 °C=40 mm usw.).

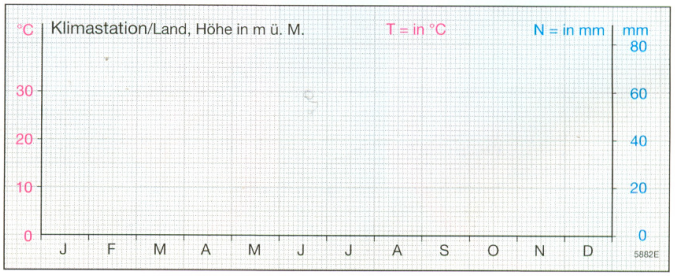
Info

Wetter und Klima
Als **Wetter** bezeichnet man das Zusammenwirken von Temperatur, Luftdruck, Wind, Bewölkung und Niederschlag zu einem bestimmten Zeitpunkt an einem bestimmten Ort. Als **Klima** wird der durchschnittliche Ablauf des Wettergeschehens an einem bestimmten Ort bezeichnet, gemessen über einen Zeitraum von mindestens 30

1. Schritt: Die Achsen

Grundlage für das Zeichnen eines Klimadiagramms ist ein Grundgerüst, das für alle Klimadiagramme gleich ist.

a) Zeichne auf der waagerechten Jahresachse die Monate von Januar (J) bis Dezember (D) im Abstand von 1 cm ein.
b) Auf der linken senkrechten Achse wird die Skala für die Temperaturwerte (in °C) eingetragen (1 cm entspricht 10 °C).
c) Auf der rechten senkrechten Achse wird die Skala für die Niederschlagswerte (in mm) eingetragen (1 cm entspricht 20 mm Niederschlag).



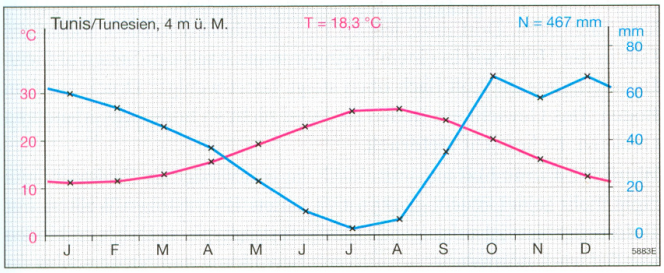
M2 *Die Achsen (1. Schritt)*

2. Schritt: Die Kurven

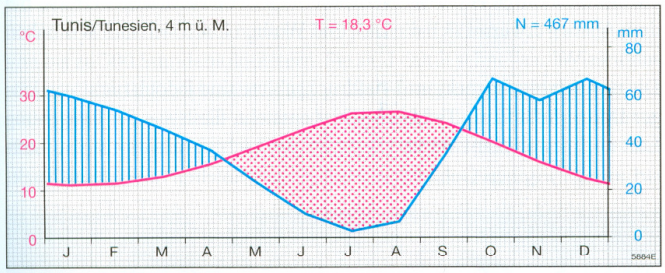
a) Zeichne die Temperatur- und die Niederschlagswerte als kleine Kreuze in die Monatsmitte (gerundete Werte).

METHODE • METHODE • METHODE 17

b) Die Temperaturwerte werden mit einer roten Linie verbunden. Die Niederschlagswerte werden als blaue Linie gezeichnet. Beide Linien werden bis zu den senkrechten Achsen verlängert.
c) Zeichne blaue senkrechte Linien zwischen beide Kurven, wenn die Niederschlags- über der Temperaturkurve liegt (humide Monate) (M4).
d) Zeichne rote Punkte zwischen beide Kurven, wenn die Temperaturkurve über der Niederschlagskurve liegt (aride Monate) (M4).
e) Bei Stationen mit besonders hohen Niederschlägen wird die Einteilung für die Niederschlagsmenge über 100 mm gestaucht, d. h. statt 20 mm-Schritte erfolgen 100 mm-Schritte. Nur so können die hohen Werte in das Diagramm übernommen werden.



M3 *Die Kurven (2. Schritt)*



M4 *Das Klimadiagramm von Tunis*

3. Schritt: Die Auswertung (Beispiel Tunis)

a) Lage der Klimastation mit Jahresdurchschnittstemperatur und Jahresniederschlag: Tunis (36°50'N/10°14'O) liegt 4 m über dem Meer (ü. M.), Jahresdurchschnittstemperatur 18,3 °C, Jahresniederschlag 467 mm.
b) Wärmster Monat (Temperaturmaximum): August (26,6 °C).
c) Kältester Monat im Jahr (Temperaturminimum): Januar (11,3 °C).
d) Jahresamplitude (Unterschied kältester und wärmster Monat): 15,3 °C.
e) Höchster Niederschlag: Oktober und Dezember (jeweils 67 mm).
f) Geringster Niederschlag: Juli (3 mm).
g) Humide Monate: Oktober bis April, aride Monate: Mai bis September.
h) *Zusammenfassung:* Tunis hat milde, regenreiche Winter und heiße, trockene Sommer. Die ariden Monate erschweren das Pflanzenwachstum. Die Klimastation liegt in den Subtropen (s. Seiten 24/25).

Info

Humid und arid

Der Begriff humid stammt vom lateinischen Wort „humidus" ab und bedeutet „feucht".
In humiden Monaten fällt mehr Niederschlag als verdunsten kann. Die Niederschlagskurve liegt über der Temperaturkurve.
Der Begriff arid stammt vom lateinischen Wort „aridus" ab und bedeutet trocken.
In ariden Monaten fällt weniger Niederschlag als verdunsten kann. Die Temperaturkurve liegt über der Niederschlagskurve.

Beachte (3. Schritt):

- Zeitraum mit den höchsten/niedrigsten Niederschlägen = Niederschlagsmaximum/Niederschlagsminimum
- Pflanzenwachstum = **Vegetationsperiode (Vegetationszeit)**

1 Werte in gleicher Weise das Klimadiagramm von El Obeid/Sudan (Seite 118, M1) aus.

2 Zeichne mithilfe der Klimadaten auf Seite 231 das Klimadiagramm von Bangkok/Thailand und werte es aus.

Bibliography

Abendroth-Timmer, Dagmar / Bonnet, Andreas / Breidbach, Stephan / Hoffmann, Reinhard / Kricher, Ernst / Lutz, Küster / Rymarczyk, Jutta / Vollmer, Helmut J. / Zydatiß,Wolfgang (2004): Didaktiken im Dialog – für eine integrative Didaktik des bilingualen Unterrichts. In: Bonnet, Andreas / Breidbach, Stephan (Eds.): *Didaktiken im Dialog. Konzepte des Lehrens und Wege des Lernens im bilingualen Sachfachunterricht*. Frankfurt am Main: Peter Lang. 13-27.

Alderson J. Charles / Clapham, Caroline / Wall, Dianne (1995): *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.

Anderson, Lorin W. / Krathwohl, David R. (Eds.) (2001): *A Taxonomy for Learning, Teaching, and Assessing*. New York: Longman.

Bach, Gerhard (2002): Bilingualer Unterricht: Lernen – Lehren – Forschen. In: Bach, Gerhard / Niemeyer, Susanne (Eds.): *Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven.* (2., überarb. und erw. Aufl.) Frankfurt am Main: Peter Lang. 11-23.

Bachman, Lyle F. (1990): *Fundamental Considerations in Language Testing*. Oxford: University Press.

Bachman, Lyle F. / Palmer, Adrian S. (1996): *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: University Press.

Bachman, Lyle F. / Palmer, Adrian S. (2004): *Statistical Analyses for Language Assessment*. Oxford: University Press.

Baker, Eva L. / O'Neil Jr., Harold F. / Linn, Robert L. (1993): Policy and Validity for Performance-Based Assessment. *American Psychologist* 48/12. 1210-1218.

Baumert, Jürgen / Artelt, Cordula / Klieme, Eckhard / Neubrand, Michael / Prenzel, Manfred / Schiefele, Ulrich / Schneider, Wolfgang / Tillmann, Klaus-Jürgen / Weiß, Manfred (Eds.) (2003): *PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*. Opladen: Leske + Budrich.

Bazerman, Charles (1998): Emerging perspectives on the many dimensions of scientific discourse. In: Martin, James, R. / Veel, Robert (Eds.): *Reading Science: Critical and functional perspectives on discourses of science*. London: Routledge. 15-28.

Berkenkotter, Carol / Huckin, Thomas (1995): *Genre Knowledge in Disciplinary Communication: Cognition/Culture/Power*. Hillsdale NJ: Lawrence Erlbaum Associates.

Bernstein, Basil B. (1990): *The structuring of Pedagogic Discourse. Class, codes and control vol. IV*. London: Routledge.

Bhabha, Homi K. / Rutherford, Jonathan (1990): The Third Space. Interview with Homi Bhabha. In: Rutherford, Jonathan (Eds): *Identity: Community, Culture, Difference*. London: Lawrence & Wishart. 207-221.

Biederstädt, Wolfgang (2002): Möglichkeiten und Grenzen des Englischen als Arbeitssprache im Geographieunterricht der Klassen 7 – 10. In: Bach, Gerhard / Niemeyer, Susanne (Eds.): *Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven.* (2., überarb. und erw. Aufl.) Frankfurt am Main: Peter Lang. 127-135.

Böhme, Hartmut / Matussek, Peter / Müller, Lothar (2000): *Orientierung Kulturwissenschaft. Was sie kann, was sie will*. Reinbek: Rohwolt.

Breidbach, Stephan (2001): Wissenschaftskompetenz: Zur Bedeutung reflexiver Wissenschaftstheorien für die Ausbildung von Lehrerinnen und Lehrer im bilingualen Sachfachunterricht. In: Abendroth-Timmer, Dagmar / Bach, Gerhard (Eds.): *Mehrsprachiges Europa*. Festschrift für Michael Wendt zum 60. Gerburtstag. Tübingen: Narr. 161-175.

Breidbach, Stephan (2003): Transkulturalität – Paradigma für den Bilingualen Sachfachunterricht. In: Eckerth, Johannes / Wendt, Michael (Eds.): *Interkulturelles und transkulturelles Lernen im Fremdsprachenunterricht*. Frankfurt am Main: Peter Lang. 219-234.

Breidbach, Stephan (2004): Bildung, Kultur, Wissenschaft: Grundzüge einer reflexiven Wissenschaftsdidaktik für den bilingualen Sachfachunterricht. In: Bonnet, Andreas / Breidbach, Stephan (Eds.): *Didaktiken im Dialog. Konzepte des Lehrens und Wege des Lernens im bilingualen Sachfachunterricht*. Franfurt am Main: Peter Lang. 153-164.

Breland, Hunter M. / Camp, Roberta / Jones, Robert J. / Morris, Margaret M. / Rock, Donald A. (1987): *Assessing writing skill* (Research Monograph No. 11). New York: College Entrance Examination Board.

Brinton, Donna M. / Snow, Marguerite Ann / Wesche, Marjorie B. (1989): *Content-Based Second Language Instruction*. Boston: Heinle & Heinle Publishers.

Brown, James D. / Bailey, Kathleen M. (1984): A Categorical Instrument for Scoring Second Language Writing Skills. *Language Learning* 34/4. 21-42.

Brown, James D. / Hudson, Thom (2002): *Criterion-referenced Language Testing*. Cambridge: Cambridge University Press.

Brunotte, Ernst / Gebhardt, Hans / Meurer, Manfred / Meusberger, Peter / Nipper, Josef (Eds.) (2002): *Lexikon der Geographie*. Heidelberg: Spektrum.

Chalhoub-Deville, Micheline (2001): Task-based assessments: Characteristics and validity evidence. In: Bygate, Martin / Skehan, Peter / Swain, Merrill (Eds.): *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*. 210-228.

Christie, Frances (Ed.) (1984): *Children writing: reader*. Geelong, Vic.: Deakin University Press.

Christie, Frances (1985): Language and schooling. In: Tchudi, Stephen N. (Ed.): *Language, schooling, and society*. Upper Montclair, NJ: Boynton/Cook.

Chirstie, Frances (1998): Science and apprenticeship: The pedagogic discourse. In: Martin, James R. / Veel, Robert (Eds.): *Reading Science: Critical and functional perspectives on discourses of science*. London: Routledge. 152-177.

Cope, Bill / Kalantzis, Mary (1993): The power of literacy and the literacy of power. In: Cope, Bill / Kalantzis, Mary (Eds.): *The powers of literacy: A genre approach to teaching writing*. Pittsburgh: University Press. 63-89.

Council of Europe (2001): *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: University Press.

Cummins, James (1979): Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism* 19. 121-129.

Cummins, James (1981): The role of primary language development in promoting educational success for minority students. In California State Department of Education (Ed.), *Schooling and language minority students: a theoretical*

framework. Evaluation, Dissemination and assessment Center, California State University, Los Angeles.

Davies, Alan / Brown, Annie / Elder, Cathie / Hill, Kathryn / Lumley, Tom / McNamara, Tim (1999): *Dictionary of language testing*. Cambridge: Cambridge University Press.

Deutsche Gesellschaft für Geographie (DGfG) (2006): *Bildungsstandards im Fach Geographie für den Mittleren Schulabschluss*. Berlin: DGfG.

Eggins, Suzanne (1994): *An Introduction to Systemic Functional Linguistics*. London: Pinter Publishers.

Ernst, Manfred (1995): Fehlerkorrektur und Leistungsbewertung im bilingualen Sachfachunterricht. *Praxis des neusprachlichen Unterrichts* 42/3. 258-264.

Gee, James P. (2002): Literacies, identities and discourses. In: Schleppergrell, Mary J. / Colombi, Cecilia M. (Eds.): *Developing advanced literacy in first and second languages: Meaning with power*. Mahwah, N.J.: Lawrence Erlbaum. 159-175.

Giese, Bettina (1994): Bewertung und Korrektur der Sprechfertigkeit im Unterricht und im mündlichen Prüfungen. *Fremdsprache und Hochschule* 40/1994. 135.

Gipps, Caroline V. (1994): *Beyond Testing: Towards a Theory of Educational Assessment*. London: The Falmer Press.

Gnutzman, Claus (1995): Sprachbewußtsein ("Language Awareness") und integrativer Grammatikunterricht. In: Gnutzman, Claus / Königs, Frank G. (Eds.): *Perspektiven des Grammatikunterrichts*. Tübingen: Narr. 267-284.

Gnutzman, Claus (1997): Normen im Englischunterricht: Zur Rolle des Sprachbewußtseins ("Language Awareness") und seiner kognitiven, sozialen und affektiven Vernetzung. *Zeitschrift für Fremdsprachenforschung* 8/12. 213-226.

Hallet, Wolfgang (1998): The Bilingual Triangle. Überlegungen zu einer Didaktik des bilingualen Sachfachunterrichts. *Praxis des neusprachlichen Unterrichts* 45/2: 115-125.

Hallet, Wolfgang (2002): *Fremdsprachenunterricht als Spiel der Texte und Kulturen. Intertextualität als Paradigma einer kulturwissenschaftlichen Didaktik*. Trier: WVT.

Hallet, Wolfgang (2004): Bilingualer Sachfachunterricht als interkultureller Diskursraum. In: Bonnet, Andreas / Breidbach, Stephan (Eds.): *Didaktiken im Dialog. Konzepte des Lehrens und Wege des Lernens im bilingualen Sachfachunterricht*. Frankfurt am Main: Peter Lang. 141-152.

Halliday, Michael A.K. (1978): *Language as a social semiotic*. London: Edward Arnold.

Halliday, Michael A.K. (1985): *Introduction to Functional Grammar*. Edward Arnold: London.

Halliday, Michael A.K. (1989) (2nd edition): Part A of *Language, context and text: aspects of language in a social semiotic perspective*. Oxford: University Press.

Halliday, Michael A.K. (1998): Things and relations. Regrammaticising experience as technical knowledge. In: Martin, James R. / Veel, Robert (Eds.): *Reading Science: Critical and functional perspectives on discourses of science*. London: Routledge. 185-235.

Halliday, Michael A.K. / Martin, James R. (1993): *Writing Science: Literacy and Discursive Power*. London: Falmer Press.

Hasan, Ruqaiya (1985): *Linguistics, Language and Verbal Art*. Geelong, Vic.: Deakin University Press. (Republished by Oxford University Press 1989)

Hasan, Ruqaiya (1989) (2nd edition): Part B of *Language, context and text: aspects of language in a social semiotic perspective*. Oxford: University Press.

Heine, Lena (2006): Wissensbasierte Lösungsprozesse von Fachaufgaben. In: In: Timm, Johannes-Peter (Ed.): *Fremdsprachenlernen und Fremdsprachenforschung: Kompetenzen, Standards, Lernformen, Evaluation*. Tübingen: Narr. 267-282.

Heine, Lena (2007): *Kognitive Prozesse bilingualer Lerner bei der fremdsprachlichen Verarbeitung von Fachinhalten*. Osnabrück: unpublished PhD thesis.

Henning, Grant (1987): *A Guide to Language Testing: Development, Evaluation, Research*. Cambridge: Newbury House Publishers.

Hoffmann, Reinhard (2004): Geographie als bilinguales Sachfach: Fachdidaktische Grundsatzüberlegungen. In: Bonnet, Andreas / Breidbach, Stephan (Eds.): *Didaktiken im Dialog. Konzepte des Lehrens und Wege des Lernens im bilingualen Sachfachunterricht*. Frankfurt am Main: Peter Lang. 205-219.

Hughes, Arthur (1989): *Testing for Language Teachers*. Cambridge: University Press.

Huot, Brian (1990): The Literature of Direct Writing Assessment: Major Concerns and Prevailing Trends. *Review of Educational Research* 60/2. 237-263.

Johns, Ann M. (1997): *Text, Role, Context. Developing Academic Literacies*. Cambridge: University Press.

Klieme, Eckhard / Avenarius, Hermann / Blum, Werner / Döbrich, Peter / Gruber, Hans / Prenzel, Manfred / Reiss, Kristina / Riquarts, Kurt / Rost, Jürgen / Tenorth, Heinz-Elmar / Vollmer, Helmut J. (2004). *The Development of National Educational Standards. An Expertise*. Bonn: Bildungsministerium für Bildung und Forschung.

Lemke, Jay (1990): *Talking Science: Language, Learning, and Values*. Norwood: Ablex Publishing Corporation.

Lemke, Jay (1998): Multiplying meaning. Visual and verbal semiotics in scientific text. In: Martin, James R. / Veel, Robert (Eds.): *Reading Science: Critical and functional perspectives on discourses of science*. London: Routledge. 87-113.

Lenz, Thomas (2004): Schriftliche Leistungsüberprüfung im bilingualen Geographieunterricht. In: Bonnet, Andreas / Breidbach, Stephan (Eds.): *Didaktiken im Dialog. Konzepte des Lehrens und Wege des Lernens im bilingualen Sachfachunterricht*. Frankfurt am Main: Peter Lang. 103-114.

Leser, Hartmut (Ed.) (2005): *Wörterbuch allgemeine Geographie 13., völlig überarb. Aufl.* Braunschweig, München: Westermann, DTV.

Lindfors, Judith W. (1981): *Children's Language and Learning*. New York: Prentice Hall.

Linn, Robert L. (1994): Performance Assessment: Policy Promises and Technical Measurement Standards. *Educational Researcher* 23/9. 4-14.

Linn, Robert L. / Baker, Eva L. / Dunbar, Stephen (1991): Complex, Performance-Based Assessment: Expectations and Validation Criteria. *Educational Researcher* 20/8. 15-21.

Martin, James R. (1984): Language, Register and Genre. In: Christie, Frances (Ed.): *Children writing: reader*. Geelong, Vic.: Deakin University Press. 21-29.

Martin, James R. (1985): *Factual writing: exploring and challenging social reality*. Geelong, Vic.: Deakin University Press. (Republished by Oxford University Press 1989).

Martin, James R. (1992): *English Text: system and structure*. Amsterdam: Benjamins.

Martin, James R. / Veel, Robert (Eds.) (1998): *Reading Science: Critical and functional perspectives on discourses of science*. London: Routledge.

McNamara, Timothy F. (1996): *Measuring Second Language Performance*. London: Longman.

Mohan, Bernard (1986): *Language and Content*. Reading, Mass.: Addison-Wesley.

Mohan, Bernard (1998): Knowledge structures in oral proficiency interviews for international teaching assistants. In: Young, Richard / Weiyun, Agnes (Eds.): *Talking and Testing. Discourse approaches to the assessment of oral proficiency*. Amsterdam: John Benjamins. 173-204.

Messick, Samuel (1989): Validity. In: Robert L. Linn (Ed.): *Educational Measurement* (3rd ed.). New York: American Council on Education, Macmillan Publishing Company. 13-103.

Meyer, Sheree L. (1993): Refusing to play the confidence game: Illusion of mastery in the reading/writing of texts. *College English* 55/11. 46-63.

Pellegrini, Anthony / Galda, Lee (1998): *The Development of School-Based Literacy*. A social ecological perspective. London: Routledge.

Portmann-Tselikas, Paul R. (2002): Textkompetenz und unterrichtlicher Spracherwerb. In: Portmann-Tselikas, Paul R. / Schmölzer-Eibinger, Sabine (Eds.): *Textkompetenz. Neue Perspektiven für das Lehren und Lernen*. Innsbruck: Studien Verlag. 13-43.

Poynton, Cate (1985): *Language and Gender: making the difference*. Geelong, Vic.: Deakin University Press.

Rischke, Dimo M. (1997): Erdkundeunterricht 'in English' macht viel mehr Spaß! *Geographie heute* 18/156. 44-45.

Shohamy, Elana / Gordon, Claire M. / Kraemer, Roberta (1992): The Effect of Raters' Background and Training on the Reliability of Direct Writing Tests. *The Modern Language Journal* 76. 27-33.

Shohamy, Elana (2006): Rethinking Assessment for Advanced Language Proficiency. In: Byrnes, Heidi / Weger-Gunthrap, Heather / Sprang, Katherine A. (Eds.): *Educating for Advanced Foreign Language Capacities. Construct, Curriculum, Instruction, Assessment*. Washinton, D.C.: Georgetown University Press.188-208.

Simpson, Paul (1990): Modality in Literary-Critical Discourse. In: Nash, Walter (Ed.): *The Writing Scholar: Studies in Academic Discourse*. Newbury Park: Sage Publications. 63-94.

Skehan, Peter (2001): Tasks and language performance assessment. In: Bygate, Martin / Skehan, Peter / Swain, Merrill (Eds.): *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*. 167-185.

Swain, Merrill (1985): Communicative competence: some roles of comprehensible input and comprehensible output in its development. In: Gass, Susan M. / Madden Carolyn G. (Eds.): *Input in Second Language Acquisition*. Rowley, MA: Newbury House. 235-253.

Swain, Merrill (1993): The output hypothesis: Just speaking and writing aren't enough. *Canadian Modern Language Review* 50: 158-164.

Troschke, Randi C. (2006): Kollaboratives Problemlösen: inhaltliche, sprachliche und soziale Bedeutungsaushandlung bilingualer Lerner. In: Timm, Johannes-Peter (Ed.): *Fremdsprachenlernen und Fremdsprachenforschung: Kompetenzen, Standards, Lernformen, Evaluation*. Tübingen: Narr. 231-248.

Troschke, Randi C. (2007): Analyse interaktiver Gesprächsprotokolle: inhaltliche, sprachliche und soziale Aushandlungsprozesse beim kollaborativen Problemlösen. In: Vollmer, Helmut J. (Ed.): *Synergieeffekte in der Fremdsprachenforschung. Empirische Zugänge, Probleme, Ergebnisse*. Frankfurt am Main: Peter Lang. 237-255.

Troschke, Randi C. (in preparation): *Kooperatives Problemlösen – gemeinsame Wissenskonstruktion bilingualer Lerner bei der Bearbeitung fachlicher Aufgaben*. Osnabrück: PhD thesis.

Ventola, Eija (1987): *The Structure of Social Interaction: a systemic approach to the semiotics of service encounters*. London: Pinter.

Ventola, Eija (Ed.) (1991): *Functional and Systemic Linguistics: Approaches and Uses*. Berlin: Mouton de Gruyter.

Ventola, Eija (1997): Modulization: Probability – an exploration into its role in academic writing. In: Duszak, Anna (Ed.): *Culture and Styles of Academic Discourse*. Berlin: Mouton de Gruyter. 157-179.

Vollmer, Helmut J. (2002a): Förderung des Spracherwerbs im bilingualen Sachfacunterricht. In: Bach, Gerhard / Niemeyer, Susanne (Eds.): *Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven.* (2., überarb. und erw. Aufl.) Frankfurt am Main: Peter Lang. 139-158.

Vollmer, Helmut J. (2002b): Bilingualer Sachfachunterricht als Inhalts- und als Sprachlernen. In: Bach, Gerhard / Niemeyer, Susanne (Eds.): *Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven.* (2., überarb. und erw. Aufl.) Frankfurt am Main: Peter Lang. 51-73.

Vollmer, Helmut Johannes (2002c): Leistungsfeststellung und Leistungsbewertung im bilingualen Sachfachunterricht: Ein Desideratum. In: Breidbach, Stephan / Bach, Gerhard / Wolff, Dieter (Eds.): *Bilingualer Sachfachunterricht. Didaktik, Lehrer-/Lernerforschung und Bildungspolitik zwischen Theorie und Empirie*. Frankfurt am Main: Peter Lang. 101-121.

Vollmer, Helmut Johannes (2006a): Kompetenzaufgaben als Forschungs- und als Evaluationsinstrument. In: Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. / Krumm, Hans-Jürgen (Eds.): *Aufgabenorientierung als Aufgabe. Beiträge zur 26. Frühjarhrskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Narr. 244-255.

Vollmer, Helmut Johannes (2006b): Fachlernen und Sprachlichkeit: Aufgabenbasierte Kognition, Kooperation und Kommunikation bei bilingualen und monolingualen Lernern. In: Hahn, Angela / Klippel, Friederike (Eds.): *Sprachen schaffen Chancen. Dokumentation des 21. Kongresses für Fremdsprachendidaktik.* München: Oldenbourg. 145-159.

Vollmer, Helmut Johannes (2006c): Fachlichkeit und Sprachlichkeit. *Zeitschrift für Fremdsprachenforschung* 17/2. 201-244.

Vollmer, Helmut J. (2007): Zur Modellierung und empirischen Erfassung von Fachkompetenz am Beispiel der Geographie. In: Vollmer, Helmut J. (Ed.):

Synergieeffekte in der Fremdsprachenforschung. Empirische Zugänge, Probleme, Ergebnisse. Frankfurt am Main: Peter Lang. 279-298.

Vollmer, Helmut Johannes (in preparation). *Geographische Kompetenzen von monolingualen und bilingualen Lernern am Ende der Klasse 10 des Gymnasiums. Ergebnisse einer empirischen Untersuchung*. Osnabrück: Universität Osnabrück (Monograph).

Vygotski, Lev S. (1978): *Mind in Society*. Cambridge, Mass.: Harvard University Press.

Weigle, Sarah C. (2002): *Assessing Writing*. Cambridge: Cambridge University Press.

Weinert, Franz E. (2001): Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In: Weinert, Franz E. (Ed.): *Leistungsmessung in Schulen*. Weinheim, Basel: Beltz Verlag. 17-31.

Wiersma, William / Jurs, Stephen G. (1990): *Educational Measurement and Testing*. 2nd Edition. Boston: Allyn and Bacon.

Wignell, Peter / Martin, James R. / Eggins, Suzanne (1990): The Discourse of Geography: Ordering and Explaining the Experiential World. *Linguistics and Education* 1/4. 359-392.

Wode, Henning (1995): *Lernen in der Fremdsprache. Grundzüge von Immersion und blingualem Unterricht*. München: Hueber.

Zydatiß, Wolfgang (2000): *Bilingualer Unterricht in der Grundschule. Entwurf eines Spracherwerbskonzepts für zweisprachige Immersionsprogramme*. München: Hueber.

Zydatiß, Wolfgang (2002): Konzeptuellen Grundlagen einer eigenständigen Didaktik des bilingualen Sachfachunterrichts: Forschungstand und Forschungsprogramm. In: Breidbach, Stephan / Bach, Gerhard / Wolff, Dieter (Eds.): *Bilingualer Sachfachunterricht. Didaktik, Lehrer-/Lernerforschung und Bildungspolitik zwischen Theorie und Empirie*. Frankfurt am Main: Peter Lang. 31-61.

Zydatiß, Wolfgang (2004a): Einleitung: Sachfachlicher Kompetenzerwerb im bilingualen Sachfachunterricht. In: Bonnet, Andreas / Breidbach, Stephan (Eds.): *Didaktiken im Dialog. Konzepte des Lehrens und Wege des Lernens im bilingualen Sachfachunterricht*. Frankfurt am Main: Peter Lang. 89-90.

Zydatiß, Wolfgang (2004b): Überlegungen zur fächerübergreifenden Evaluation des bilingualen Unterrichts: Textkompetenz als Schlüsselqualifikation fremdsprachigen Sachfachlernens. In: Bonnet, Andreas / Breidbach, Stephan (Eds.): *Didaktiken im Dialog. Konzepte des Lehrens und Wege des Lernens im bilingualen Sachfachunterricht*. Frankfurt am Main: Peter Lang. 91-102.

Zydatiß, Wolfgang (2005): Diskursfunktionen in einem analytisch curricularen Zugriff auf Textvarietäten und Aufgaben des bilingualen Sachfachunterrichts. *Fremdsprachen Lehren und Lernen* 34. 156-173.

Zydatiß, Wolfgang (forthcoming): *Deutsch-Englische Züge in Berlin (DEZIBEL). Methoden und Ergebnisse des Evaluationsprojekts zum bilingualen Sachfachunterricht an Gymnasien*. Frankfurt am Main: Peter Lang.