Universität Bielefeld

Fakultät für Linguistik und Literaturwissenschaft

Masterarbeit
im Studiengang `Master of Education´
im Fach `British and American Studies´

zum Thema:

Just, Systematic but still Motivating?
- Towards a Method of Assessing In-Class Participation

vorgelegt von
Jens Peter Fischer

Erstgutachter/in: Claudia Brüning
Zweitgutachter/in: Dr. Patricia Skorge

Bielefeld, September 2009

Table of Contents:

I	**Introduction**		**1**
	Methodology		2
II	**Three Dimensions of Requirements**		**5**
	II.1	Institutional Requirements	6
	II.2	Motivational Requirements	11
	II.3	Scientific Requirements	16
		II.3.1 Objectivity	17
		II.3.2 Reliability	21
		II.3.3 Validity	23
	II.4	Summary of Key Issues	25
		Concurrences and Areas of Conflict	27
III	**The Requirements and the Real World – The Study**		**29**
	III.1	Study design	29
	III.2	Results	40
		III.2.1 Quality Criteria	40
		III.2.2 Method of Analysis	42
		III.2.3 Description of the Results	42
		III.2.4 Interpretation of the Results	56
IV	**Synthesis and Conclusion**		**64**
V	**Bibliography**		**67**

Appendices:

A oFb Homepage
B Pre-Test for Study
C Results of Pre-Test
D Full Catalogue of Question Items
E Hypotheses Tableau
F Table of Questions
G Questionnaire
H Letter to Principals
I Flyer for English Teachers
J Presentation of Raw Data
K Summary of Raw Data
L Explanation of Statistical Terms
M Contingency Tables

Acknowledgements

I would like to thank Claudia Brüning and Patricia Skorge for their patience with me and invaluable advice. Further thanks go to the "oFb – der onlineFragebogen" project for their free academic online questionnaire service that helped me carry out my research work and to the StatBeCe for their helpful suggestions. Finally, my thanks to go the many teachers who participated in my study, without whom the research would not have had a point.

1 - Introduction

I - Introduction

> *Unabweisbar ist mit der Berufsrolle von Lehrerinnen und Lehrern die Pflicht zur Bewertung von Schülerinnen- und Schülerleistungen verbunden.* (Jürgens, 2005, p. 41)

This quotation from Jürgens puts a general truth about the professional practice of teachers into words. Whenever teachers want to help their students understand the learning matter, support them with the work they have to do, and make them enjoy the learning process, they also have a number of very different tasks to perform at the same time: They have to evaluate the students' learning progress and, at regular intervals, carry out a binding assessment that summarizes the complexity of the students' individual learner characteristics into a set of numbers, i.e. grades.

Grading is the part of the work of teachers that many describe as their least favorite one; the teachers can hardly be sure that their assessment is completely justified. It is always necessary to generalize and weigh some aspects of a student's performance against others. They are also aware that this assessment may play a decisive role in determining the perspectives their students have in their later lives. There is a need to assess well and thoroughly; at the same time, a proper methodological framework is necessary to organize the assessment process, and finally assessment often takes place under immense time pressure.

One area in which these problems are especially significant is that of in-class participation, that is, assessment of how the students perform in all aspects of daily classroom life. Even more than in other situations where they have to assess, such as written exams, assessing in-class participation puts the teachers into a situation where much is uncertain: It is probably not enough to evaluate students' performance only before the end of term, because the impression would not be representative, but how often does it have to be done? How often can it be done? What is really part of in-class assessment? The quality of what a student says? The quantity? Probably both, but to which extent? And what about social behavior in class? Can the teachers be sure that they evaluate students justly? Will they mainly see good things about the students they expect to be good, and bad things about the weaker ones? Is there anything they can do about this?

These are only some of the questions that can be asked about assessment of in-class participation. Kirk (2003), who touched on the subject in her book *Beurteilung mündlicher Leistungen*, notes:

> *Die in Kapitel 1 dargestellten Erfahrungen einzelner Lehrer/innen und die dabei geschilderten Probleme verdeutlichen große Unsicherheit in diesem Bereich.* (p. 27)

At least some of this insecurity can be traced back to the fact that there is hardly any scientific literature on the topic that could help teachers. Of course, there is a lot on grading in general, and also on *Mündliche Leistungen*, but this usually deals with oral exams in school. The specific consideration and problems about grading in-class participation have, to the author's knowledge, not yet been investigated and there is no theoretical framework that teachers could use as a basis of their assessment practice. Both the teachers and the students suffer from the effects of this deficiency: Teachers are wondering whether they are assessing the students the right way and asking the above questions; the students suffer, in the worst case, by receiving an unjustified grade, as Jürgens (2005) explains:

> *Allerdings verlassen sich die Lehrerinnen und Lehrer zumeist auf allgemeine (Gelegenheits-)Beobachtungen, die sie beispielsweise vor Zeugnisterminen zu Globalurteilen unzulässigerweise zusammenfassen […].* (p. 87)

The purpose of this paper is to develop a theoretical framework that provides criteria that a method of assessing in-class participation should fulfill. As a second step, a study will examine how this framework relates to the way in which in-class participation is assessed by teachers, in the hope that this investigation can help to reduce the absence of methodologies in this area of grading.

Methodology

In order to develop a list of criteria for methods of assessing in-class participation, the aspects that have to be considered are both of theoretical and practical nature. The distinction between purely theoretical considerations on the one hand and their practical consequences on the other hand will be maintained in the course of the investigation, with the aim of providing both theoretical ideals as well as a practical perspective.

A lot of scientific research has been done on the subject of assessment. It covers, for example, how grades influence student motivation and how grading has to be carried out to match scientific standards. The results of this research provide a reasonable

theoretical foundation for the investigation. There are also a number of factors in assessing in-class participation that are given by the German education authorities. These aspects lead to three groups of requirements, institutional, motivational and scientific ones. They will be discussed in the first, theoretical part of this thesis. The result of that section will be a list of criteria for the assessment of in-class participation.

This purely theoretical list of criteria will form the basis of the second part of the paper that is connected to a study carried out in the context of this investigation. Teachers working in the German school system were questioned about their grading practices and attitudes towards grading. The questions used were created on the basis of the findings of the theoretical section. Finally, it will be attempted to reconcile the results of the theoretical and practical sections of this paper and find a synthesis between the two.

Before the investigation can be started, its scope should be defined exactly. How narrow and precise is the context in which we want to investigate the assessing of in-class participation? The scope has already been ‘reduced’ to teaching within the German public school system. However, different types of schools and different subjects also have unique issues to be dealt with. For example, when one considers motivational issues arising with the grading of oral participation, one has to give special attention to problems that are unique to foreign language teaching, such as foreign language anxiety. In order for this investigation to render reasonable results that do not have the limitation that they only apply in certain cases, it seems legitimate to narrow down the scope considerably. Therefore, we will deal with the context of assessing in-class participation in *Gymnasien* in North Rhine-Westphalia in the subject of English.

To describe the subject of investigation, the term *method of assessing in-class participation* will be used throughout this thesis. A constant term to denote this is certainly reasonable; it also means that, before other concerns, it should be explained why this term is chosen. The term *in-class participation* is chosen because it is very general and can be best used to contain the various aspects of *Sonstige Leistungen*, the term used in German schools and school legislature. It is chosen because it is the best term found, although what is understood by *Sonstige Leistungen* is not completely containted in the translation. For example, as will be seen, homework is,

by North Rhine-Westphalian education legislature, also a part of *Sonstige Leistungen* and thus encompassed by *in-class participation*, although it is, strictly speaking, not done in class. The term is used as describing everything the students do that is related to the work in class, as opposed to exams.

The concept that we are concerned with is a method based on the distinction between approach, method and technique as put forward by Edward M. Anthony (1963, pp. 63-67). An approach is a set of beliefs about a certain area of teaching; a method is a way of putting these beliefs into practice; techniques are the actual implementations of a method. Based on these distinctions, the design of this thesis can also be described in the following way: Existing approaches towards different aspects of grading are examined and applied to the field of assessing in-class participation. The aim is to compile a set of rules that a method of assessing it has to fulfill. In the second section, based on a study conducted, possible techniques that can be deduced from these rules are evaluated by teachers and compared to their own assessment practice in order to highlight what the practical demands on a method are.

2 – Three Dimensions of Requirements

II Three Dimensions of Requirements

First, it should be examined what, in general, a method of assessing in-class participation could look like. For this, we will use theories on the assessment procedure in general, so as to identify the overall structure of such a method. Jürgens (2005, p. 45) suggests that there are two different stages to any assessment of student achievements, namely *Leistungsmessung und –beurteilung*, as he explains in the following way:

> *Die Begriffe Leistungsmessung und –beurteilung werden häufig bedeutungsgleich verwendet […] obwohl die Bewertung als interpretativer Vorgang dem Verfahren der Erhebung nachgeordnet ist, soll das Beurteilungsergebnis nicht schon vorab feststehen.* (Jürgens, 2005, p. 45)

Jürgens argues that the surveying of an achievement, for example a teacher's observation in class, has to be separated chronologically from the act of interpreting what has been surveyed. Applied to in-class participation, this means that the interpretation of student behavior should not occur at the same time that the teacher observes it but afterwards, so that the interpretation can be done without it being influenced by the situation of observation. However, when one transfers this simple model that describes the measurement and interpretation of one single student achievement to the situation that a method of grading in-class participation in general is in, one finds that some adaptation is necessary. Firstly, when it comes to in-class participation, we are dealing with a (probably large) number of individual student actions that are observed in one way or the other. Also, it is difficult for the teachers to record their observations at the time they make them if this is in class. These issues are better accounted for by the model that Kirk (2003) suggests:

> *Der Weg von der Informationsgewinnung zur Beurteilung sollte dabei in den drei Schritten Beobachtung […]/ Beschreibung /Bewertung erfolgen.* (p. 34)

According to Kirk, the observation of student behavior, the recording of the observation (this is what she means by *Beschreibung*), and the interpretation should be the different steps of the procedure. If we relate this to the situation we are likely to encounter in schools, this can be interpreted as that in step one, teachers make (any number of) observations during the lessons. In the second step, these have to be recorded; this is hard to do during the lesson, so it has to happen afterwards. As it is impossible to keep many details (such as a large number of observations) in mind for a long time, the recording should take place regularly. The third step then takes place

at the end of the school term or whenever a grade based on the information available is given.

Having established a rough framework of how a method of assessing in-class participation should look, the next step is to identify the groups of criteria that this method has to fulfill. These are firstly the requirements put forward by the school as an institution, that is, the regulations set up by legislation and the school authorities. This thesis will not engage in a debate of these regulations, as they are binding laws that the teachers have to follow. These will in the following be called *institutional requirements*. Secondly, it will be argued that a method of assessing students' in-class performance that inevitably involves grading needs to be able to deal with the motivational issues that assessment and especially grades create. As a consequence, we will identify what we will call *motivational requirements*. The third group that will be considered are the scientific criteria that a method has to fulfill to be reliable and objective, in the following called *scientific requirements*.

II.1 Institutional Requirements

There are three layers of educational legislature that form the legal framework for teaching English at *Gymnasien* in North Rhine-Westphalia, and thus also for the grading carried out in this context. Firstly, the most general is the *Schulgesetz für das Land Nordrhein-Westfalen* in its current version of Febuary 2005. Article 48 is concerned with the principles of assessing student performance. It states the intention and the form of the assessment:

> *Die Leistungsbewertung soll über den Stand des Lernprozesses der Schülerin oder des Schülers Aufschluss geben[…]. Die Leistungen werden durch Noten bewertet. […] Grundlage der Leistungsbewertung sind alle von der Schülerin oder dem Schüler im Beurteilungsbereich „Schriftliche Arbeiten" und im Beurteilungsbereich „Sonstige Leistungen im Unterricht" erbrachten Leistungen.* (Ministerium, 2009a, p. 10)

This establishes the assessment of student achievements as an obligation; a later article (§57) states that the task of assessing is to be taken over by the teachers. It also establishes the practice of using grades for representing the level of students' performance. We learn that there are two main fields of assessment, *Schriftliche Arbeiten* and *Sonstige Leistungen im Unterricht*. Based on these vague labels, we can expect that oral participation is part of the second of those two fields.

The second layer of educational legislature contains the *Ausbildungs- und Prüfungsordnungen* for the two sections of the Gymnasium, *Sekundarstufe I* and *II* (*APO S1* and *APOGOSt*, the first covering grades five to nine/ten and the second covering grades ten/eleven to twelve/thirteen). For the *Sekundarstufe I*, the term *Sonstige Leistungen* that the *Schulgesetz* introduces is specified in the following way:

> *Zum Beurteilungsbereich „Sonstige Leistungen" gehören alle im Zusammenhang mit dem Unterricht erbrachten mündlichen und praktischen Leistungen sowie gelegentliche kurze schriftliche Übungen in allen Fächern. Die Leistungen bei der Mitarbeit im Unterricht sind bei der Beurteilung ebenso zu berücksichtigen wie die übrigen Leistungen.* (Ministerium, 2007a, p. 2)

These regulations provide a more precise definition of what *Sonstige Leistungen* are. The term refers to oral participation in class and also to the other work that is done in class, which is not specified. As a consequence, in this paper, we have to deal with a very wide definition of in-class performance, one that includes oral participation as well as performance at written tasks, group work ect. In the quotation given above, the legislator also stipulates that in-class participation should have the same significance for the grade that a student is given as other factors, once again stressing the importance of this aspect of student achievement.

The *Ausbildungs- und Prüfungsordnung* for the *Sekundarstufe II* gives a similar definition:

> *Zum Beurteilungsbereich „Sonstige Mitarbeit" gehören alle im Zusammenhang mit dem Unterricht erbrachten schriftlichen, mündlichen und praktischen Leistungen mit Ausnahme der Klausuren und der Facharbeit gemäß § 14 Abs. 3. Die Formen der „Sonstigen Mitarbeit" richten sich nach den Richtlinien und Lehrplänen für den Unterricht in der gymnasialen Oberstufe.* (Ministerium, 1999, p. 4)

While the definition of *Sonstige Leistungen* from the *APO S1* is more or less repeated in a briefer form, we are also referred to the *Richtlinien und Lehrpläne* for further details.

The third layer of educational legislature is the most concrete one: The curricula for English at *Gymnasien* for *Sekundarstufe I* and *II* .They give more details about *Sonstige Leistungen*. The *Kernlehrplan G8* for the subject English is the most recent curriculum for the *Sekundarstufe I*, and will therefore be considered here. Again, it stresses that the two fields of assessment, *Schriftliche Arbeiten* and *Sonstige*

Leistungen are to have exactly the same weight for the grade (Ministerium, 2007b, p. 46). It also gives a list of what is to be part of *Sonstige Leistungen*. These are:

> *[…] verstehende Teilnahme am Unterrichtsgeschehen sowie kommunikatives Handeln und Sprachproduktion schriftlich wie vor allem mündlich […], wobei individuelle Beiträge zum Unterrichtsgespräch sowie kooperative Leistungen im Rahmen von Team- und Gruppenarbeit zu beachten sind […] einzelne […] Kompetenzen in fest umrissenen Bereichen des Faches (u. a. kurze schriftliche Übungen, Wortschatzkontrolle, Überprüfungen des Hör- und Leseverstehens, vorgetragene Hausaufgaben oder Protokolle einer Einzel- oder Gruppenarbeitsphase), […] längerfristig gestellte komplexere Aufgaben, die von den Schülerinnen und Schülern einzeln oder in der Gruppe mit einem hohen Anteil der Selbstständigkeit bearbeitet werden […].* (Ministerium, 2007b, p. 46)

This detailed list of *Sonstige Leistungen* shows that a wide range of different aspects has to be considered. It includes oral performance in class as well as performance in written work in class, group work, and different forms of in-class tests and homework. The *Richtlinien und Lehrpläne für die Sekundarstufe II* offer a similar list that adds further points:

> *[…] Beiträge zum Unterrichtsgespräch, die Leistungen in Hausaufgaben, Referaten, Protokollen, sonstigen Präsentationsleistungen, die Mitarbeit in Projekten und Arbeitsbeiträge […* und *] die schriftliche Übung, die benotet wird.* (Ministerium, 1999, p. 100)

A slight shift of focus can be seen here, as class participation forms that require independent student work such as presentations and project work are mentioned more explicitly than in the regulations for the *Sekundarstufe I*. In the pages that follow the enumeration given above, one finds detailed explanations on how to grade those different types of performances. But this time, the curriculum also contains a statement on how exactly class participation should be evaluated:

> *Qualität, Intensität und Selbstständigkeit der Mitarbeit im Unterricht sind entscheidende Grundlagen der Beurteilung im Bereich „Sonstige Mitarbeit". Im Sinne der Entwicklung und Bewertung der kommunikativen Kompetenz kommt dabei der mündlichen Kommunikationsfähigkeit eine besondere Rolle zu […].* (Ministerium, 1999, p. 100)

Again, we can see that the curriculum places an emphasis on independent student work. While written work is mentioned in the list of aspects that should be graded, it is made clear that the ability to communicate orally is of special significance. Quality, intensity (probably mainly referring to how often a student participates) and independence constitute one set of criteria that should be used for evaluating students. Subsequently, another one is given; students are supposed to show

"Sprachkompetenz [...] Sach- und Problemkompetenz [und] Sozialkompetenz" (Ministerium, 1999, p. 101). This raises the question how exactly those two sets of criteria are to be combined. *Sprachkompetenz* and *Sach- und Problemkompetenz* are both measures that refer to the quality of student participation and can therefore be assigned to the maxim of *Qualität* from the first set. *Sozialkompetenz* is connected to the ability to work and participate independently, but goes beyond that, as it also includes the capacity for teamwork and constructive work habits (Ministerium, 1999, p. 101). Intensity, which we assume to refer to the quantity of participations from one student, remains as an additional factor.

The *Richtlinien und Lehrpläne für die Sekundarstufe 2* also introduce the following requirement:

> *Leistungsbewertung setzt voraus, dass die Schülerinnen und Schüler im Unterricht Gelegenheit hatten, die entsprechenden Anforderungen in Umfang und Anspruch kennen zu lernen und sich auf diese vorzubereiten. […] Die Bewertung ihrer Leistungen muss den Schülerinnen und Schülern auch im Vergleich mit den Mitschülerinnen und Mitschülern transparent sein.* (Ministerium, 1999, p. 88)

Transparency is established as an important property of a method of assessing in-class participation. Not only are the teachers asked to give the students a clear idea of what they expect of them, the students are even entitled to information about why they were graded in a certain way.

Connecting the different regulations considered above, one obtains a list of what is legally subject of grading in-class participation based on enumerations of items to be graded. They include presentations, project work, written tasks, homework, and, with special significance, oral participation. On the other hand, they do not explain how exactly the different factors are supposed to be weighed, how frequently and how regularly *Sonstige Leistungen* should be evaluated, and, most importantly, what method or procedure should to be used. These questions, however, are far from easy to answer and leave the teachers without the knowledge and skills to carry out the assessment of in-class participation properly.

It has been seen already that the teachers have the responsibility to evaluate students and assign grades. In a later article of the *Schulgesetz*, the functions of the *Fachkonferenz* or *Bildungsgangkonferenz*, the group of teachers at a school who teach in a certain subject or in a certain program, is introduced:

> *Die Fachkonferenz entscheidet in ihrem Fach insbesondere über […] Grundsätze zur Leistungsbewertung.* (Ministerium, 2009a, p. 15)

As can be seen here, the school authorities delegate decisions about the conditions and methods of evaluating in-class participation to the schools and the teachers. Whereas what is to be evaluated is rather clearly defined, the most difficult questions regarding the evaluation are left unclear. On the one hand, this may be due to the fact that education authorities want to leave teachers a certain pedagogic latitude (Jürgens, 2005, p. 112). On the other hand, as Kirk (2003, p.27) notes, there is a great insecurity about this area of grading and therefore it is also possible that school legislature tries to avoid having to be the vanguard by setting methodological standards in an area where there is yet no common agreement.

Kopfnoten

Realizing the wide social responsibility of education, the new North Rhine-Westphalian *Schulgesetz* defines

> *[…] den Auftrag der Schule klar an einem Verständnis von ganzheitlicher Persönlichkeitsbildung […]. Deshalb ist es zeitgemäß, die Entwicklung des Arbeitsverhaltens und Sozialverhaltens bewusst in den Blick zu nehmen und zu fördern. […] Schülerinnen und Schüler wie ihre Eltern haben Anspruch auf klare und verständliche Rückmeldungen und Bewertungen auch zu diesem Entwicklungsbereich.* (Ministerium, 2008a, p. 2)

This feedback is given in the form of the so-called *Kopfnoten*, grades that reflect the students' personal and social behavior. They include firstly the number of classes that a student missed without excuse on the report card, and secondly additional grades for working habits and social habits that are also included on the report card (Ministerium, 2009a, p. 10). The *Ausbildungs- und Prüfungsordnungen* for the *Sekundarstufe I* and *II* specify that the following grades must be given:

> *Noten für das Arbeitsverhalten in den Teilbereichen Leistungsbereitschaft, Zuverlässigkeit/Sorgfalt und Selbstständigkeit und Noten für das Sozialverhalten in den Teilbereichen Verantwortungsbereitschaft, Konfliktverhalten und Kooperationsfähigkeit.* (Ministerium, 2007a, p. 2*;* Ministerium, 1999, p. 2)

The primary time and place in which teachers can observe the behavior assessed by the *Kopfnoten* is undoubtedly in class. Therefore, an up-to-date method for assessing in-class participation also has to evaluate these aspects. However, it should be kept in mind that the *Kopfnoten* are viewed very controversially and that there also are fundamental doubts of their accuracy. One example for this is that, depending on the

way teachers grade *Sonstige Leistungen*, *Kopfnoten* may introduce a double grading of aspects of in-class participation. If a teacher considers the degree to which a student works independently, as suggested by the *Kernlehrplan G8* (Ministerium, 2007b, p. 46), the *Kopfnote Selbstständigkeit* is likely to assess the same behavior again, resulting in a double grading.

II.2 Motivational Requirements

Student Motivation

Zoltán Dörnyei (2001) argues that "motivation is one of the key issues in language learning" (p. 1). Pintrich and Schrauben (1992) describe one of the problems with neglecting motivation when they note the following about cognition-only models, that is, models that do not consider emotional factors related to motivation:

> *Cognition-only models have difficulty explaining why students who seem to have the requisite prior knowledge and relevant cognitive strategies do not activate them for many school tasks.* (p. 149)

While there already are a number of points to be made in favor of that statement in general, it is especially true when looking at the relationship between grading and motivation. Therefore, in the following, the effects that grading in general and grading in-class participation in particular can have on student motivation will be examined, with the aim of formulating a number of motivational requirements that a method of grading in-class participation has to fulfill.

The relationship between grading and motivation has often been characterized in a very negative way. Dörnyei (2001) summarizes that "for motivational psychologists 'grade' is definitely a 'four-letter word'" (p. 131). On the other hand, Kirk (2003) notes:

> *Darüber hinaus ist der pädagogischen Funktion [von Noten] in besonderem Maße Beachtung zu schenken. Die Hervorhebung von Stärken in Teilbereichen, das gezielte Aufzeigen von Schwächen […] kann die Lern- und Leistungsmotivation stärken. […] Im Gegensatz zu einer fachbezogenen Gesamtnote können Beurteilungen mündlicher Leistungen […] diesen Funktionen in besonderer Weise gerecht werden.* (p. 35)

What is implied by Kirk's explanations is that a more neutral approach to grading that is aware of the many possible problems with grades but also recognizes their potential functions is more reasonable than condemning them. It has to be kept in mind that for students, grades are "extrinsic considerations" (Pintrich & Schrauben,

1992, p. 156), meaning that they represent an external end and not something students do for themselves. This is opposed to intrinsic motivation, which means that a student approaches tasks with "a focus on learning and mastery" (Pintrich & Schrauben, 1992, p. 156). While there is evidence that intrinsic motivation may lead to a more intense involvement with a subject matter, being "high in both intrinsic and extrinsic orientation" yields the most positive effects (Pintrich & Schrauben, 1992, p. 157). Furthermore, there is a general tendency for the effects of extrinsic motivation, like those of grades, to increase with age (Pintrich & Schrauben, 1992, p. 102). Therefore, let us now try to give a reasonable account of the possible problems with grading and assessment and the remedies for it.

Grades can influence students' self-perception in a significant manner. These, in turn, influence students' "academic performance, persistence, and choice of different tasks". (Wigfield & Harold, 1992, p. 100) Therefore, the ways in which grades influence students' self-perception have to be identified and, if possible, controlled.

As Dörnyei (2001) tells us, "Grades tend to focus students' attention on ability rather than effort" (p. 131). This means that students may wrongly assume that the grade they get for their in-class participation refers to their language skills, such as fluency, only. Especially with weaker students, this could mean that they attribute the poor grade they get to their lack of language skills, which then increases anxiety and reduces the motivation to try to do better because ability cannot be increased as easily as effort.

Another factor related to students' self-perception has already been introduced implicitly: foreign language anxiety. In addition to the general anxiety that is caused by the knowledge that one is being assessed, the fear of speaking or writing in a foreign language also often keeps students from participating because they feel insecure about the language they are using. Foreign language anxiety is especially strong when it comes to speaking, as this is the most direct form of communicating in a language and mistakes cannot be hidden once a statement has been uttered. Speaking, however, is the most common form of in-class participation. Thus we have two different but related forms of anxiety that are both at work when it comes to grading in-class participation in a foreign language class. As a remedy to these problems, Ames (1992) suggests that students should always be given opportunities

to improve their grades, because these "suggest […] to students that mistakes and errors are part of the learning process and not indicative of failure to learn" (p. 341).

Kirk (2003, p. 23) makes the observation that students' performance in school is made up of the actions of the teachers and of the school as well as of those of the students themselves. This means that the teachers' behavior towards the students is relevant to their achievement. This happens in two different ways. The first one is that the expectations teachers have affect their own judgment; this will be dealt with in the subsection of scientific requirements. The other one is that the teachers' ideas of what skills a certain student has are communicated to the students, possibly without the teacher noticing. This way, teacher expectations can have a direct effect on student motivation. Wigfield and Harold (1992) state:

> *If students accept the teachers' expectations and behavior toward them then they will be more likely to act in ways that confirm the teacher's initial expectations.* (p. 96)

This means that teachers' expectations can be motivating for students that are already, generally speaking, performing rather well. However, weaker students about whom teachers have rather negative expectations can experience a further reduction of motivation. In this respect, teacher expectations toward students share properties with grades, which are characterized by Dönryei (2001) concerning their motivational impact as that they "tend to aggravate social inequality as the strong get stronger and the weak get weaker" (p. 131).

One may also see positive aspects in this, as the possibility to influence students to act according to their teachers' expectations may also be seen as a chance for motivating them. Also, there is evidence that most beliefs teachers hold about students are more or less accurate (Wigfield & Harold, 1992, p. 96). It is likely that these two factors combined can have a great impact on student motivation. Wigfield and Harold (1992) also list evidence that teachers' expectations tend to be "negative […] rather than positive" (p. 96). From this, we can deduce that when assessing students, teachers should try to avoid communicating their (negative) expectations towards students as well as possible (of course, generally, they should always try to do that).

The way teachers prepare students for grading is also significant for student motivation, as Ames (1992) observes:

> *Because evaluation is one of the most salient features of the classroom, students' motivation to learn can be easily undermined by how evaluation occurs.* (p. 340)

In order to reduce these effects as far as possible, the conditions of the assessment should be made as clear as possible. This can be done by providing "sufficient advance warning and information" about the assessment and especially "clear specifications of criteria that will be used for marking" (Dörnyei, 2001, p. 94). To help the students feel less anxious about grades, the teacher can explain the whole procedure used to them, for example by considering the case of a fictional sample student. Dörnyei (2001, p. 132) suggests to also give *models* of what *exemplary performance* is to give students an idea of what achievements lead to a grading at the top end of the scale. However, this example should not come from one of the students, or the other students should at least not be aware of this, in order not to promote social comparison among the students. This way, it can at least be assumed that if the students are still afraid, it is because of the circumstance that they are being assessed and not because of the method that is used. Also, the fact that the conditions of the assessment are known to everybody from the beginning can help to take away the 'two camps' feeling because the students are not thinking of the teacher as the person who can give poor grades, but of someone who applies the criteria that are known to and understood by all.

The way teachers disclose grades to students is one of the ways in which they communicate their expectations and prejudices to the students, and an especially problematic one. There are several ways in which teachers often make grades public, such as "public pronouncement of grades", "displays of selected papers and achievements" and "wall charts detailing students' achievement and ability grouping" (Dörnyei, 2001, p. 92). Knowing that one's grades are made public increases the pressure because the other members of one's learning group will get to know them. Ames (1992) argues that

> *[…] social comparison may be among the most potent factors contributing to a negative motivation pattern.* (p. 328)

There is probably no way that teachers can keep students from finding out about each others' grades in a short time and may see it as a ranking of their performance. Still, when teachers make grades or levels of achievement public, they give this ranking an additional official, institutional weight. This way, it "can easily contribute to a

hierarchy of perceived ability which translates into motivational inequities" (Ames, 1992, p. 342). As a consequence, teachers should avoid this.

The next motivational aspect of assessment that will be discussed is self-assessment. As Oscarson tells us, it is "central to language learning" (Oscarson, 1989, p. 6) as it helps the students in forming an idea about their own ability level; furthermore, it creates an atmosphere of trust between the teacher and the students because she trusts them to "be honest in evaluating their own work" (Dörnyei, 2001, p. 133). There are several reasons why a method of assessing in-class participation should also involve student self-assessment. Kirk (2003) notes:

> *Durch [Schülerselbstbewertung ist …] die Wahrscheinlichkeit hoch […], dass auch die Lernleistung positiv beeinflusst wird.* (p. 62)

The *Lehrplan für die Sekundarstufe II* uses the notion of self-assessment when it argues as follows:

> *Eine Konzeption von Unterricht, die die Eigenverantwortlichkeit der Lernenden fördert, muss diese kontinuierlich an der Überprüfung ihres Lernstandes und Lernerfolges beteiligen und dabei ihre Fähigkeit zur selbstständigen Bewertung eigenen Lernens (self-assessment) entwickeln.* (Ministerium, 1999, p. 105)

Although the *Lehrplan* highlights the importance of self-assessment, it does not anchor it as an obligatory element of the grading process (in general, how exactly it should be integrated into school teaching remains unclear). However, the positive effects of self-evaluation require that it plays an actual role in the grading process. This can be done by giving the students "self-evaluation tools" (Dörnyei, 2001, p. 133) on the basis of which they can assess their own performance, for example an appropriate questionnaire. Other ways, however much more time-intensive, would be student reports on the work they did over certain periods of time or study diaries (*Lerntagebücher*). (Kirk, 2003, p. 99) The final step of the grading process should then "be the product of *two-way negotiation*" (Dörnyei, 2001, p. 133) that can for example have the form of an interview. Finally, there should be an opportunity for the students to evaluate the teacher as well (Dörnyei, 2001, p. 133), as this can help to establish a relationship of mutual trust when it comes to fair grading.

Teacher Motivation

It should also be kept in mind that a method for assessing in-class participation should not only be motivating for students, but for teachers as well. They should feel

comfortable using it. For this to be the case, it should be based on an academic approach to testing, so that the classical research paradigms of objectivity, reliability and validity are fulfilled (they will be discussed in more detail in the next section). The teachers should also have the feeling that they are giving the students fair treatment. This means, on the one hand, that the procedure used assesses student performance in a scientifically correct way, but on the other hand that teachers should have the feeling that they are treating their students as humans and not as objects that are only distinguished by certain characteristics. This also means that there still has to be some flexibility to it, so that it only aids teachers in giving grades and does not restrict them. Finally, and maybe most importantly, the procedure should be practical. As teachers usually have a very high workload, and it is often hardly possible for them to complete all tasks given to them diligently, a procedure of grading oral participation should be as little extra work as possible. The importance of this requirement must not be underestimated: If a procedure is too complicated or too time-consuming, many teachers will simply not use it, and continue to use the unsystematic procedures described in the introduction. Of course, these different criteria already imply that a compromise has to be found between the scientifically ideal process and the practically most manageable one.

II.3 Scientific Requirements

Scientific criteria are of overwhelming importance when it comes to measuring and assessing behavior and performance. As Young and He (1998) argue,

> *[These] criteria – [objectivity], reliability and construct validity – are familiar to test designers because they are essential qualities if tests are to be used as ways of measuring learners' abilities.* (p. 1)

The reason for this is that only when a method fulfills scientific criteria can it be assumed that there are no significant measurement mistakes (of course, there can never be absolute surety about this). The three criteria that Young and He mention, objectivity, reliability and validity, have been established as the principles of scientific testing and assessment design in a vast majority of the research literature. Objectivity can be achieved by the exclusion of all subjective influences from the assessment (Kirk, 2003, p. 40). Reliability means that the measuring is reproducible and therefore constant and exact (Jürgens, 2005, p. 76). Finally, validity means that the assessment measures exactly what it is supposed to measure, i.e., that all that is to be part of the assessment is assessed and that no other factors alter the result

(Jürgens, 2005, p. 77). Each of those criteria puts high demands on a method that aims at assessing in-class participation, as will be examined in the following. Additionally, more systematic issues about grading and assessment will be addressed.

II.3.1 Objectivity

Issues with observation

During the first step of the evaluation procedure, the criterion of objectivity is especially important. That is because this step involuntarily involves acts of observation; in our case, this is most likely the teacher observing the students' performance. Ingenkamp and Lissmann's (2008) definition of objectivity highlights the possible issues that can occur:

> *Eine Messung ist dann objektiv, wenn intersubjektive Einflüsse der Untersucher möglichst ausgeschaltet werden können.* (p. 52)

While this definition is slightly tautological (as it is set against the opposition between objectivity and subjectivity), it reveals the main problem during the act of observation: subjective perceptions of the one observing.

> *Die Beobachtung unterliegt den gleichen individuellen Mechanismen wie jeder andere Wahrnehmungsprozess.* (Jürgens, 2005, p. 90)

It is known today that human perception is indeed subject to many different factors that have a very significant influence on how information is taken in. This is not only due to the limited capacity of our senses and the fact that much perception takes place unconsciously. Our pre-knowledge about a situation influences our perception just as our perception adds to our knowledge, or, as Atteslander (1975) puts it more dramatically,

> *Wir glauben nur, was wir sehen – leider sehen wir nur, was wir glauben wollen.* (p. 138).

Of course, teachers at work are as much subject to these limitations as any other people are. In the classroom, this can for example mean that a teacher does not notice that a student is trying to participate in class because the teacher expects the student not to do it. As a consequence, the teacher's expectation is confirmed, which can then even reinforce the expectation. A scientific framework is necessary in order to cope with these kinds of self-fulfilling prophecies. According to Jürgens (2005), this can be done by systematizing observation as far as possible:

> *Die unbedingte Einbindung des Beobachtungsgeschehens in einen theoretischen Kontext und die systematische Planung der Beobachtung sollen einen möglichst hohen Grad an Kontrollierbarkeit ermöglichen, als Vorraussetzung der Wiederholbarkeit und Objektivität der Beobachtung.* (p. 94)

Observation can be classified in the continuum between *naïve* and *systematic observation*. The difference between those two extremes consists in whether the observation is done with a certain intention or objective in mind or not (Ingenkamp & Lissmann, 2008, p. 74). A related distinction is that between *systematic* and *unsystematic observation*. Systematic observation involves a clearly defined purpose and a high degree of control on the side of the teacher, while unsystematic observation takes place in the form of intuitive perceptions (Jürgens, 2005, p. 98).

Observation in school, as all observation, can be classified somewhere between those extremes. As systematic observation, in both definitions, is the more desirable form (Jürgens, 2005, p. 98), Jürgens suggests a number of approaches that can be applied when observing students that should change observation into that direction. Unsystematic or naïve perception is not generally wrong, but should serve as an initiation into systematic perception. This means that special occurrences in students' behavior, even if perceived by a teacher unconsciously and recollected later, can serve as possible subjects of further investigation (for example, aggressive behavior during a group work phase) (Jürgens, 2005, p. 98). After that occurrence, the teacher should observe if the behavior takes place regularly in comparable situations. Of course, due to the limits of human perception teachers cannot pay attention to a large number of students at the same time; the procedure outlined above does not really involve actively trying to monitor the students all the time, but rather to keep observations that have been made in mind and remember them when something similar happens. In this way, behavior that is typical for a certain student can be identified.

In order to be able to avoid judging the quality of students' contributions on the basis of subjective and maybe even subconscious perception only, Jürgens (2005, p. 110) suggests that the teacher adopts a reference of distinguished levels of complexity and achievement by which students' contributions are measured. As an example that he considers widely applicable, he gives the distinction between reproduction of information, reorganization, transfer and problem-solving thinking, with the first representing the lowest and the last representing the highest level of achievement

(Jürgens, 2005, p. 110). Evaluating students' contributions by the level of achievement they can be assigned to helps to avoid an overwhelmingly subjective judgment by the teacher. In the context of this thesis, this system is only intended to serve as an example for frameworks, and not as a necessary distinction to make, other systems that better fit the individual context (but have been fixed beforehand and are applied systematically) are just as possible. Jürgens (2005, p. 102) also argues that concrete frameworks and solutions should always be left to the individual teachers or schools.

While teachers tend to give differentiated evaluations of written work, providing feedback on aspects such as content and use of language, there is a tendency to sum up and combine the individual factors of in-class participation and treat them as one factor (Kirk 2003, p. 15). This bears a great danger of generalization. For example, if a student's oral contributions in class are very infrequent and the student often disturbs the teaching, the tendency of many teachers is to also rank the quality of the student's oral contribution in class as worse than they are (after all, they may be very good) (Jürgens, 2005, p. 86). Consequently, teachers should treat the different aspects of student performance that they grade as in-class participation separately during the assessment process. This does, of course, still mean that there can be one final grade for in-class participation (there has to be, anyway).

Oral assessment has more characteristics that can compromise the objectivity of a procedure. Rapp (1979, p. 34) explains that in every situation of direct communication between assessor and assessed (and those are among the roles that teachers and students play in class), there are emotional exchanges between them that mainly take place at a subconscious level. For example, it is common for people to have expectations about other people's behavior. In the case of teachers, this means that it is not unlikely that when in class, they have expectations about the behavior of their students, especially about the quality and quantity of their performance in class. Research has shown that behavior that matches the expectations of an observer is very likely to be noticed more strongly, while behavior that does not match them may even be blanked out completely (Jürgens, 2005, p. 84). Applied to the situation of teachers observing students' behavior in class, this can lead to the situation that the performance of students of whom the teacher believes that they are good is perceived as even better, and the performance of those believed to be rather weak as even weaker.

Concerning the quality of student answers, a typical feature of classroom conversations has to be brought into consideration: Very often, when one student cannot answer a question, teachers tend to pass the question on to other students. Also, they often give additional hints on the answer when passing on the question to another student. The student who can finally give the answer makes an especially positive impression on the teacher, while the others do not (Jürgens, 2005, p. 86). What has to be kept in mind, however, is that the student who finally gives the answer has more time and more information (hints) than the others, and therefore, the circumstances of the answer are not comparable. One realizes that in this situation, not only the criterion of objectivity is violated, but also the one of validity, as in this case, the individual students' skills that are measured are not the same. While the first student who answers faces the task of presenting knowledge of the subject of the question, for the other students the task becomes more and more to decode the teacher's hints successfully.

Issues with recording information

Coming to the second stage, recording information, it is important to note that a kind of recording sheet for taking down teacher observations is a necessary requirement of systematic observation (Jürgens, 2005, p. 100). This sheet can only fulfill its function if it reflects the observation principles of the teacher; that is, a teacher has to determine what she wants to observe and how she wants to record it, and use a sheet that reflects those considerations (Jürgens, 2005, p. 100). Jürgens suggests a very detailed sheet that contains a table with a large number of student characteristics; the teacher records if and how far these characteristics occur with the individual student. Ingenkamp and Lissmann (2008, pp. 89-91) promote the usage of recording sheets that have different student characteristics that can be rated on a scale. Jürgens (2005, p. 107) also admits that for the practical requirements of everyday teaching practice, a more simple recording structure can be more adequate. Rather than taking general notes for every lesson only, he propagates also recording information about students in cases when the teacher observes something special about a student.

Another significant point deals with the temporal connection between stages one and two. In general it can be said that the longer the period of time between those two stages, that is the amount of time that passes after a teacher notes a student's

behavior in class and before she records it in some way, the higher the risk that what is recorded differs from what was taken in. Jürgens (2005) expresses this by saying:

> *Liegen zwischen der beobachteten Situation bzw. dem Beobachteten Verhaltensausschnitt […] längere zeitliche Distanzen, kann es zu beträchtlichen Verzerrungen bzw. Akzentuierungen oder Nivellierungen kommen.* (p. 85)

The requirement that the time in between has to be short is less problematic in the case of oral exams, which are usually limited to one or several, few points in time. In the case of grading oral participation in class, it means that teachers have to invest considerable amounts of time into using a proper procedure. Jürgens (2005, p. 86) suggests that one week between the first two stages is already too long; it would be optimal if this could be carried out after every lesson or at least every teaching day.

II.3.2 Reliability

Grading standards

A general problem that all grading has to cope with is that of which standard is to be applied when judging the students' achievements. It is obvious that standards are necessary; teachers cannot grade students' performance without any frame of reference that tells them what can be expected. However, if students' achievements are compared to unreasonable standards, this calls into question the accuracy of the procedure. Therefore, this issue relates to the criterion of reliability. The established term for it in the German research literature is *Bezugsnormproblematik* (Jürgens, 2005, p. 45; Ingenkamp & Lissmann, 2008, p. 63). Commonly, three principal standards of grading that can be applied are distinguished: The first one, the *intraindividual standard* (German term: *intraindividueller Maßstab*), takes into consideration the development of the individual learner. This makes it possible to create an individual learner profile and allows for differentiated support and assessment, as the problems and advances of learners can be compared over time. (Jürgens, 2005, p. 45). Secondly, the *interindividual standard* (*interindividueller Maßstab*) uses the achievements of the learning group that an individual is part of as a reference. In the case of assessment in schools, this would most likely be the *Klassenverband* or the year, or even bigger groups of students (Jürgens, 2005, p. 45). Thirdly, there is the *criterion-based standard* (*kriteriumsorientierter Maßstab*). This standard originated from the scientific movement of *criterion-referenced instruction* (Mager, 1975, p. 5). It uses a fixed learning target as a measure, and thus allows for a

greater control of how far curricular aims have been accomplished (Jürgens, 2005, p. 46).

Each of those criteria, considered on their own, have their weaknesses. The first two fail to take into consideration as a factor the learning targets set by the curriculum. The third one, on the other hand, is not able to consider the progress that individual learners make, because it fails to take into account that the skill levels inside classes always vary at the beginning of a school year. One learner can achieve worse results than another one, and still have made more learning progress than the other. The second criterion has been the subject of very heavy criticism and is, at this time, not considered acceptable as a basis for grading from a scientific point of view (Jürgens, 2005, p. 47). The first and third criteria have weaknesses that can be balanced by the respective other one. The weakness of the intraindividual criterion can be dealt with by also taking into account the learning targets that the learners are supposed to meet. The shortcomings of the third, learning-target based approach can be balanced by also considering individual factors. Jürgens argues that a combination of both factors with a heavy emphasis on the third one has proven most reasonable. (Jürgens, 2005, pp. 52-53). One has to be aware, however, that this would be treading virgin soil, as the commonly used standards are based on the intraindividual and interindividual criteria (Ingenkamp & Lissmann, 2008, p. 63).

Moving back from a purely theoretical perspective to our specific situation of a method that assesses oral participation, how can such standards be characterized here? When weighing learners' oral participation in order to express them in the form of grades, it is most important for the teachers to have a clear idea of what the learners are supposed to be able to do, both in terms of (oral) language skills and in terms of knowledge. These expectations should be based on the curriculum. As a second factor, however, they should also bear in mind the individual progress of the learner. As this is supposed to be the lesser of the two factors, it could be treated as a bonus that learners get on their grade if they made much progress. In some cases, albeit only in extreme ones, it should also be possible to reduce grades if teachers have the feeling that some students put very little effort into their achievements. This idea is connected to the notion of the so-called *pädagogische Zensur*, a grade that is supposed to motivate learners and honor their efforts, which Jürgens considers an advantage of the intraindividual criterion (Jürgens, 2005, p. 47).

II.3.2 Validity

Subject of grading

Ingenkamp and Lissmann (2008) sum up the meaning of the criterion of validity in the following way:

> *Die Gültigkeit oder Validität eines Verfahrens sagt aus, ob tatsächlich das gemessen wird was man messen will, und nicht irgendetwas anderes.* (p. 57)

Accordingly, an important question relating to the criterion of validity is what exactly should be subject to grading. In order to approach this very important topic, we will first look at a list of "Beurteilungsgegenstände für 'mündliche Zensuren'" (Kirk, 2003, p. 27) that Kirk collected from various scientific sources:

> *Frequenz der Unterrichts, Qualität spontaner Unterrichtsäußerungen, Qualität der Antworten auf Lehrerfragen (bei bzw. ohne Meldung), Leistungen in unterrichtlichen Miniprüfungen, Leistungen in ‚Nicht-Klassenarbeiten' wie schriftlichen Lernkontrollen, Kurztests, Vorführungen vor der Klasse, Qualität der Hausaufgaben, Vergessen von Arbeitsmaterialien, […], Mitarbeit bei Projekten, Vorlesen und Vortragen von Texten, mündliches Referieren, […] die allgemeine (qualitative und quantitative) Beteiligung am Unterrichtsgespräch, der Nachweis der Erledigung mündlicher Hausaufgaben, […] Hörverständnis, Verwendung der Fachsprache, […] Bereitschaft zur Teamarbeit, Einfühlungsvermögen entwickeln, planen und organisieren, Verantwortung übernehmen, Selbstkritik.* (p. 27)

These lists seem like a more detailed breakdown of the subjects of grading that are mentioned in the legal specifications examined above. In the face of these many possible factors (and more could be found), one of the main challenges for teachers is to decide which ones to include in the assessment and how to weight them. Indeed, Kirk (2003) notes that:

> *[…Es besteht] eine große Unsicherheit der Lehrer/innen darüber, welche Schüleraktivitäten bei der mündlichen Leistungsbeurteilung zu berücksichtigen sind.* (p. 15)

This decision cannot be taken over completely by a method of assessing in-class participation as this would deprive the teachers of their latitude and their right to have individual preferences. Only a rough classification can be made. However, it has to be kept in mind that the students are faced with a very dissatisfying situation as well if it is possible that they have to face a new concept of what aspects of their behavior are to be graded every time they get a new teacher (Kirk, 2003, p. 16). When the students perceive that what is graded depends on teacher preferences this

can also, for them, undermine the idea that grading is done in a reasonable and just way (if they ever had such an idea).

A first approach is to try to sort these different factors of grading in some way. Most generally, they can be divided into cognitive and non-cognitive ones. The term *cognitive features* (*Kognitive Merkmale*) denotes how well students are able to understand tasks and problem-solving strategies and cope with the content that is presented to them in general. *Non-cognitive features* (*Nichtkognitive Merkmale*) designate the individual ability to work independently and effectively and to interact with others in a productive way (Jürgens, 2005, p. 88).

Let us first have a more detailed look at cognitive features. An important distinction that has to be made in this area is that of quantity and quality. On the one hand, the quantity of student contributions and participation should play a role, as it reflects that a student puts effort into the lesson (Jürgens, 2005, p. 112). On the other hand, if quantity plays too great a role, this implies, for the students, that the only purpose of participation is to show activity in order to get a good grade, rather than to understand and contribute participation of high quality (Jürgens, 2005, p. 112). For Jürgens, this argument outweighs the first, so that he suggests that quality should be more important than quantity, and that the latter should only be treated as a relatively small bonus or subtraction from the student's grade. While the exact degree to which quality and quantity influence the final grade may be subject to debate and should probably best be left to the individual teacher, it becomes clear from the above considerations that the teachers should at least be constant in how they deal with them.

Many foreign language teachers may want to distinguish between the technical knowledge of their students in a subject and their language competency. In general, it holds true that language competency and technical competency are connected whenever students participate in class verbally (and in writing as well) (Kirk, 2003, p. 27). In the case of a foreign language class, this becomes especially important, as the intellectual content of students' utterances is, of course, limited by their command of the foreign language.

It is extremely hard to give any rule of thumb of how non-cognitive features should be graded. One could call into question whether they should be part of the grade for in-class participation at all when they are already graded as *Kopfnoten*, because this

would mean that they would be taken into account twice when determining students' grades.

Opportunity to contribute

Research suggests that how well individual students can participate in class depends on the form of teaching that is chosen (Kirk, 2003, p. 54). We have seen that there is a large number of factors that can possibly be included in the assessment of in-class participation. Therefore, it is a central requirement of validity that teachers provide the students with differentiated opportunities to participate. Kirk (2003) says:

> *Übertragen auf den Bereich des Mündlichen lässt sich schulstufenübergreifend fordern, dass aufgrund der Heterogenität der Gesprächsfähigkeit und der Vielfalt der Leistungsanforderungen differenzierte Lernsituationen geschaffen werden müssen.* (p. 54)

This problem is especially big in classes which are still mainly taught using *teacher-centered instruction* (*Frontalunterricht*).

II.4 Summary of Key Issues

The ideal method of assessing in-class participation that has been discussed should, from a theoretical point of view, fulfill the following criteria:

1. Institutional Requirements

- It fulfills the legal obligation to assess in-class participation
- It allows for a differentiated assessment of student performance through grades
- It weighs in-class participation and written achievements equally
- It allows teachers to place different emphasis when teaching in the *Sekundarstufe I* or *II*
- It gives teachers methodological support with the whole procedure
- It also allows for assessing personal and social characteristics of students
- It makes grading criteria and grades transparent to the students

2.1 Motivational Requirements (Concerning Students)

- It avoids creating the idea that students have failed when they get poor grades, e.g. through chances to improve their grades
- It helps dealing with students' (foreign) language anxiety

- It allows teachers to control demotivating aspects in their behavior when dealing with grades
- It preserves the positive motivational effects of grades
- It allows teachers to prepare students for the grading
- It is transparent to both teachers and students
- It allows the teacher to avoid the “two camps” effect
- It allows the teachers to present grades to students privately and avoid embarrassment of the students in any way
- It gives the students an opportunity to reflect on their performance and, ideally, to join in a discussion in which they can influence (to a certain extent) their final grade
- It helps avoid the ‘vicious circle’ of weaker students losing further motivation because of bad grades

2.2 Motivational Requirements (Concerning Teachers)

- It can be carried out easily by teachers
- It does not take up much time
- It should be possible to integrate the method into a typical teaching schedule
- Its practicability should be convincing for teachers
- Its scientific foundation should be convincing for teachers
- It should allow teachers enough latitude to make adjustments for individual cases

3. Scientific Requirements

- It helps the teachers with making the movement from naïve or unsystematic to systematic observation
- It supports the teachers in finding focus points for observation that allow them to observe systematically (as a tendency)
- It allows teachers to define for themselves the elements they want to assess as in-class participation (providing the method, not the subject of grading)
- It provides teachers with a framework of rating through which they can distinguish between different achievement levels of student performance

- It supports teachers in keeping apart different variables of student behavior
- It structures teachers' take-in of student behavior in such a way so that it minimizes subjective perception
- It provides the teachers with a means of recording their observations
- It organizes the assessment procedure so that the time between observation and recording is not too long
- It allows the teachers to choose the right standard to compare students' achievements to
- It leaves teachers enough latitude, for example for giving pedagogical grades
- It is applicable in a number of different teaching environments, so that the teachers can give students different opportunities to participate

Concurrences and Areas of conflict

Several requirements that occur in different sections of the list are very similar. For example, the demand that assessment criteria should be transparent to the students is at the same time the result of motivational considerations and part of education legislature (maybe the education authorities applied the same motivational concepts?). In the same way, it is necessary for teachers to control their attitude towards the students not only because the students can perceive it and it can then have a negative motivational effect, but also because it compromises the teachers' own ability of objective observation. It is necessary to consider different dimensions of student behavior and determine exactly what the subject of grading is in order to maintain two of the central research paradigms: objectivity and validity. The desire to be able to adjust the method of grading to special cases can be justified both on the basis of motivational reflections and of scientific procedure.

In accordance with the design of this investigation, the list that has just been presented is of a purely theoretical nature. But before practical considerations are applied to it, it has to be noted that it is, in itself, of a paradoxical nature, as there are requirements in it that are hard to fulfill more completely without at the same time fulfilling others to a lesser extent. It has been explained why, from a scientific point of view, it is necessary that the recording of the students behavior that a teacher has observed and considered relevant for the students assessment is carried out in regular intervals that should occur very frequently. However, investing this amount of time and planning clashes with the requirement that a method of assessing in-class

participation should be practical, easy to use and possible to be integrated into a typical teaching schedule. In the same way, the scientific demands on the method already determine many aspects of the assessment process, but the necessity of the teacher having pedagogic latitude and being able to customize the process has been put forward as well. These conflicting ideas have to be reconciled when it comes to developing a method and putting it into practice by finding the right balance between both extremes.

Another qualification of the list has to be made: While it states properties that a method of assessing in-class participation should feature and allow the teachers to do, there are points that cannot be put into practice by using a certain method only. For example, controlling their attitude towards students is something that teachers have to be involved with personally; it cannot be done only by using the right method. Teachers still have to control their emotional attitudes towards their students, an important and difficult task which a method cannot relieve them of. In general, a method can only provide a framework that supports teachers with carrying out the assessment.

3 – The Requirements and the Real World – The Study

III The Requirements and the Real World – The Study

III.1 Study Design

Purpose

In the first section of this thesis, criteria for a method of assessing in-class participation have been examined and compiled. With a theoretical list of properties for such a method established, the logical next step is to examine how they relate to the actual teaching practice. This means that, going back to Anthony's differentiation between approach, method and technique explained in the introduction, the next step is to see how the techniques that can be derived from the framework for a method that has been developed relate to actual teaching practice (within our scope, North Rhine-Westphalian *Gymnasien*). The study conducted in the context of this thesis examined what methods teachers use for evaluating in-class participation and how they fulfill the criteria that have been discussed.

Research Methodology

The most basic choice of research methodology is that between a qualitative and quantitative approach. While qualitative research uses very open methods to investigate in areas of which little is known, quantitative research is based on research hypotheses that have been developed out of assumptions about the area that are then tested and verified or falsified. Dörnyei (2003) gives a more exact definition of quantitative research:

> *After all, the essential characteristic of quantitative research is that it employs categories, viewpoints, and models that have been precisely defined by the researcher in advance, and numerical or directly quantifiable data are collected to determine the relationship between these categories [...].* (p. 14)

Carefully studying this definition suggested that a quantitative approach was the most appropriate in this context. The study was supposed to find out about teachers' opinions about the criteria developed in the theoretical section of this paper and their teaching practices. Thus, there were already a number of categories and viewpoints provided, and the main task of the study was to collect data that gives us information about these. The quantifiability of the data collected then made it relatively easy to relate it to the theoretical considerations made above. For the same reason, and based on the distinction of quantitative and qualitative data made above, it would have been hard to relate the very open data collected in qualitative research to the categories

and viewpoints in the theory section in a systematic fashion. As a consequence, a quantitative approach was chosen.

A questionnaire was selected as the instrument of data collection for the following reason. According to Dörnyei (2003), the main types of data a questionnaire can yield are "*factual*, *behavioral* and *attitudinal*" (p. 8). These are exactly the different types of data that the survey was interested in, namely the attitudes of teachers towards the criteria for the assessment of in-class participation, the behavioral pattern by which they carry out the assessment, and the demographic background against which to set this information. Furthermore, questionnaires as a research tool also have considerable practical advantages. Firstly, there is their extreme efficiency. They take a relatively long time to create, as a number of different factors have to be considered when structuring and phrasing the questions. Then, however, they can be reproduced and given to any number of participants. Thus, even large amounts of data can be collected quite easily. A carefully designed and structured questionnaire also controls the way in which the replies can be given, ensuring thereby that the responses are of a common form and therefore comparable, so that they can be processed together. Therefore, as a final step of the survey process, statistical analysis renders clear statements that can be related to the questions that guided the design of the questionnaire (Dörnyei, 2003, pp. 9-10). Other tools of quantitative research, such as interviews and observations designed to yield quantifiable data, cannot be used as easily and efficiently in the context of this survey. Finally, as the survey was to deal with teachers' attitudes and behavioral patterns, a questionnaire that is directly addressed to them is the most easily administrable and most straightforward way of reaching a sufficiently great number of teachers. As filling in a questionnaire is one of the simplest and quickest ways in which one can take part in a scientific survey, this choice also maximizes the number of those who actually participate by reducing the likelihood that someone does not take part because of a high workload or time problems.

Implementation

The basic options for the implementation of the study were either to design a paper-based questionnaire or an electronic one. There are two principal ways of carrying out the distribution of the questionnaire, either to select and invite the participants or to make the questionnaire available to the target group and hope for them to take part. These two strategies for the implementation of the study left four principal

strategies out of which the one that was most advantageous for the individual situation had to be selected. The study described here used an electronic questionnaire that was made available online for members of the target group without prior selection. This decision was based on the following considerations: Firstly, an electronic approach made the handling of the data significantly more easy and fail-safe. The questionnaire in its electronic form could not be altered by mistake (which could happen to printed questionnaires, e.g. through bad copying or damage to the paper in any stage of the survey). The electronic format also ensured that the participants responded to the questionnaire in exactly the way that was intended in its design, as they could only select replies that have been inserted into the questionnaire form. This procedure also minimized the potential for errors: The answers to each question were directly saved to a database as a numerical value and could later be imported into any statistical software. No counting out of the results and coding had to be done, which is a substantial source for errors (Dörnyei, 2003, p. 101).

While the advantages of choosing an electronic solution over a paper-based one were mainly administrative, the main reason for making it publicly available was to increase the number of participants. Selecting participants and inviting them to take part in a controlled environment would have meant that they would have had to be contacted first, and, if they had agreed to take part, invited to meet at an assigned date and place where the questionnaire could be filled in. The approach of making it publicly available was not only more convenient to the researcher, as he does not have to arrange sessions where the participants can take the questionnaire, but for them as well, as they could participate wherever and whenever they wanted to. This was likely to lower the threshold potential participators had for taking part. Of course, this approach also had its disadvantages: There was no personal invitation, which again lowered the likeliness that people would partake in the survey. Those who decide to participate could also reconsider this at any time, and as this was probably of low priority to them, there was a certain likeliness that other activities or lack of interest would dissuade them from taking part. There was also the problem of respondent self-selection. This occurs because there may be special characteristics in the people who participate that make them do that (Dörnyei, 2003, pp. 75-76); for example, in our case, teachers who were very aware of the issues about grading may have been more likely to participate than those who were not, resulting in a bias in the data collected. However, this is a problem that can never be avoided completely,

as “Self-selection is inevitable to some extent because few questionnaire surveys can be made compulsory” (Dörnyei, 2003, p. 75), and this was definitely also true for this questionnaire.

An obvious form of implementing a combination of an electronic and open questionnaire is to make it available online via the internet. There are a number of online survey services that allow researchers to put their questionnaires online and make them available via a URL. This has the considerable advantage that the URL can be passed on via electronic communication, such as E-Mail and newsgroups. For this survey, the service *oFb – der onlineFragebogen* was selected (Leiner (2009); see appendix A). Firstly, it was especially developed for scientific research and therefore allows complete customization of the questionnaire to give the researcher as much latitude as possible. Secondly, it is publicly available without restrictions for academic purposes. Thirdly, the service is able to export the results of the survey into formats that can be loaded directly into standard statistical software, such as SPSS.

Questionnaire design

For designing the questionnaire, the five steps drawn up by Dörnyei (2003) were pursued:

- *Deciding on the general features of the questionnaire, such as the length, the format, and the main parts.*
- *Writing effective items/questions and drawing up an item pool*
- *Selecting and sequencing the items*
- *Writing appropriate instructions and examples*
- *Piloting the questionnaire and conducing item analysis* (p. 16)

Additionally, a pilot study with a simple questionnaire was conducted before these steps to gather information for the composition of the main questionnaire.

Pilot Study

To prepare for the designing of the questionnaire, a pre-test was carried out with ten teachers as participants. This questionnaire contained five rather general questions about in-class assessment (see appendix B) that were designed to give a rough impression of what results might be expected so that the questions in the final questionnaire could be set as to render clearer results. The title of this pilot questionnaire was “Questions on ‘Mündliche Mitarbeit’”, because, on the one hand, the term *mündliche Mitarbeit* has a more closely defined meaning than *Sonstige Leistungen* and, on the other hand, is more familiar than *in-class participation*. This

way, it was easier for the participants to answer the questions without much background information about what exactly was meant.

Question one asked the participants whether they were satisfied with their system of assessing in-class participation. Question two concerned the amount of time they were investing into that work, firstly on each teaching day and secondly in addition to that when giving quarterly grades. Question three asked them whether they were sometimes worried that they were treating students unjustly, and if so, then why. Question four was intended to find out if teachers would be interested in learning about a different method. In the fifth question, the participants were asked to rate the importance of a number of features in a method of assessing in-class participation.

The test results can be found in appendix C to this thesis. The answers to questions one, three and four suggest that English teachers were mostly satisfied with their method of assessing in-class participation, but were also sometimes worried about that they might be not assessing their students in a fair way and interested in learning about alternative ways of assessing it. This at least partially paradox attitude suggested that in addition to the special requirements from the theory sections, a section in the main questionnaire should ask about the general attitude teachers have towards methods of assessing in-class participation and whether they have been introduced to or tried to learn about alternatives. Question two asked about the amount of work teachers invested into in-class assessment each week and at the end of each quarter term. The picture drawn by the results showed that the amount of time invested was between 0 and 2 hours in the first and between 0 and 3 hours in the second case. This information was not sufficient to draw a picture of when and how long teachers carry out the assessment work, as it did not capture the work invested in other intervals, for example weekly or monthly. After all, every teacher might have an individual time schedule for taking notes on in-class participation. As a consequence, more detailed questions about this were drawn up for the study questionnaire. The fifth question was aimed at highlighting what features of a method of in-class assessment teachers find particularly important. All the features given were in the majority ranked as important or very important. If this tendency can also be found in the main study, this attitude should be checked and put into perspective by not only asking about attitudes but also about behavioral patterns that reflect in how far teachers actually put this attitude into practice.

General features of the questionnaire

A crucial question that constitutes a great problem in the composition of the questionnaire is that of its length. As a big number of different aspects are addressed in the theory section, an intuitive suggestion is to write one question for each of these aspects. When this was done for a first draft, the basic question catalogue contained nearly 60 questions, the full questionnaire with repetitions for reliability checks, which will be discussed below, contained 100 questions (see appendix D). Tests showed that someone who did not know the questionnaire needed more than 45 minutes to answer it, which was clearly too long. Therefore, instead, two to four items were selected from each section of the theory section.

Dörnyei (2003, p. 18) suggests as a guideline that a questionnaire should not consist of more than 4-6 pages and should not take longer than 30 minutes to complete. The final version of the questionnaire consisted of 6 pages, including one instructions page and, according to tests, took 20-25 minutes to complete. In terms of length, a compromise had been found between the desire to include large amounts of content and the need for brevity.

The issue of anonymity was not very problematic in the context of this survey due to the online approach that was chosen. The demographic details in the questionnaire include gender, age, the number of years that the participant has worked as a teacher and how much teaching the participant does in the *Sekundarstufe I* and *II*. Once the survey was taken, the results were saved to a database together with only a unique serial identification number that distinguished the different cases. Because the survey was carried out via the internet and (of course) the IP addresses of the participants were not recorded, there was no way to find out about the identity of a participant. Following a suggestion by Dörnyei (2003, p. 23), a statement was added to the instructions at the beginning of the questionnaire assuring the participants of the anonymity of the survey.

Question writing and item pool / Selecting and sequencing the questions

As a first step in the question writing process, research hypotheses were formulated based on the requirements compiled in the theory section (Bryman & Cramer, 1990, p. 4). As has been pointed out above, the study was on the one hand interested in whether teachers agree with a certain requirement and on the other hand whether

they follow this requirement in their teaching practice. Therefore, there were two hypotheses for each section, namely the following:

> Teachers are aware of the institutional requirements of assessing in-class participation.
>
> Teachers fulfill the institutional requirements of assessing in-class participation in their teaching practice.
>
> Teachers are aware of the requirements of assessing in-class participation relating to student motivation.
>
> Teachers fulfill the requirements of assessing in-class participation relating to student motivation in their teaching practice.
>
> Teachers are aware of the requirements of assessing in-class participation relating to teacher motivation.
>
> Teachers fulfill the requirements of assessing in-class participation relating to teacher motivation in their teaching practice.
>
> Teachers are aware of the scientific requirements of assessing in-class participation relating to teacher motivation.
>
> Teachers fulfill the scientific requirements of assessing in-class participation relating to teacher motivation in their teaching practice.

Based on these hypotheses, a list of questions covering all hypotheses was compiled, containing concrete issues from the respective part of the theory section. An ordinal scale was chosen for the majority of the questions. This was done because the questions were supposed to query teachers about their attitudes and beliefs about certain statements (which were based on the requirements from the theory section). These queries could be implemented as interval scales because those require a metric relationship between the answers, that is, answer 4 must be measurably twice as high as answer 2 (in a measure connected to the question), which is not possible when asking about attitudes and beliefs (Argyrous, 2000, p. 12). Modal scales do not allow the researcher to create answers that can be ranked in a hierarchy at all. They are appropriate for some questions (such as asking for the gender of the participants). For the questions about attitudes and beliefs, however, ordinal scales, which allow for a ranking of the answers but do not require a metrical relation among them, are the most useful ones. Specifically, the subtype of Likert scales was used (Dörnyei, 2003, pp. 36-37). These questions have answers that have a range of answers that reaches from one extreme answer to a question to the other. For example, a question with a Likert scale would be “Do you think student motivation is important”, with the answers ranging from 1 (strongly disagree) to 5 (strongly agree). Different

answer ranges are possible. The first distinction that can be made between them is between an even or odd number of answering possibilities. Dörnyei (2003) explains:

> *Some researchers prefer using an even number of response options because of the concern that certain respondents might use the middle category ('neither agree nor disagree,' 'not sure,' or 'neutral') to avoid making a real choice, that is, to take the easy way out.* (p. 37)

While this is probably true, a point can be made for odd numbers as well. If there is no middle item, then participants are not able to express that they have a neutral opinion towards a certain question. They are forced to choose either one or the other tendency without having a tendency of opinion. The consequence is that tendencies are created literally out of thin air. The researcher considered this issue more problematic than the point Dörnyei makes about odd-numbered scales and therefore, odd-numbered ordinal scales were used. Next, the number of answers had to be determined. Three answers were too little to allow for differentiated answers, as this left only one answer to express agreement, one to express disagreement and a neutral one. Seven answers, on the other hand, could offer too many possibilities of differentiation, as participants would then have to choose between three different levels of agreement and disagreement. As a consequence, 5-step scales were used. The labels for the individual answer were designed after the model of semantic differential scales. That means that only two labels were given for the two extreme values of the scale, leaving the other three open. This defined a continuum between the two extremes (such as "strongly disagree" and "strongly agree"). This had the advantage that the wording of the in-between answer did not confuse or irritate participants, which could otherwise happen if they found the labels inappropriate (Dörnyei, 2003, p. 39). In addition, modal questions were used in situations where the answers did not have a semantic ordering that allowed for the usage of an ordinal scale.

The first list of questions contained 57 items (see appendix D), which was, according to all external information collected for the questionnaire design section of this paper, a much too voluminous quantity. Additionally, repetitions of questions in different versions appeared to be necessary as a means of checking the reliability of the measurement. If the reliability had had to be checked for each question, it would have made a number of more than 100 questions inevitable, resulting in a clearly unacceptable length.

In order to adopt a system that leads to an acceptable amount of questions, advice offered by the *Statistik-Beratungs-Centrum StatBeCe*, a free service offered to researchers at Bielefeld University, was taken. It consisted of firstly reducing the number of research hypotheses to one from each subsection from the first part of the paper and then selecting two to three variables/questions to cover each hypothesis. This goes along with the advice Dörnyei (2003) gives:

> *Because of the fallibility of single items, there is a general consensus among survey specialists that more than one item is needed to address each identified content area, all aimed at the same target but drawing upon slightly different aspects of it.* (p. 43)

A hypotheses tableau was used for the assignment of variables and the planning of the questionnaire. This is a table that features one row for each hypothesis and one column for each variable and marks in the cells that show to which hypothesis a question can be related. The hypothesis tableau can be found in appendix E. In the realization as a questionnaire, each variable was represented by a question. The questions were taken from the already existing item pool of 57 questions described above, and selected firstly because they were typical representatives of the category of requirements they belong to and secondly because they appeared especially interesting. The list of 27 questions can be found in appendix F. The final amount of questions actually appearing in the questionnaire was 35, however, because some questions could not be realized as one but had to be split up into multiple items (for example, the question about the amount of time teachers spend on carrying out the assessment had to be represented as six questions to allow teachers to give different answers for the different time intervals “each lesson”, “each school day”, “each school week”, “each school month”, “in addition to that each quarter term” and “in addition to that each half term”; see appendix G).

Quality management

Clearly, the three scientific quality criteria that have been worked with in the theory section, objectivity, reliability and validity, also had to be fulfilled by the study. They will not be defined again because this has already been done above. The first one, objectivity, is mainly concerned with the design and wording of the questions. In order to measure what the questionnaire was supposed to measure and not to create a bias in the answers, the questions had to be put in a clear, unambiguous, neutral and, especially important, non-suggestive way (Dörnyei, 2003, pp. 52-56). To ensure this,

the questions used were checked and revised by several people with academic backgrounds in different fields who gave feedback on whether they understood the questions and found any of the answers suggestive.

Instead of a repetition-based method that requires questions to be asked several times in different ways, an internal consistency-approach that can cover both aspects was chosen to ensure reliability and validity (Dörnyei, 2003, p. 110). This is a method that measures the consistency among different questions that measure the same field to determine the degree of reliability and validity. Dörnyei (2003) describes its importance this way:

> *Internal consistency is generally seen as the psychometric prerequisite for any scientific survey measurement.* (p. 111)

Internal consistency is measured by the *Cronbach Alpha coefficient* (Dörnyei, 2003, p. 112). For example, to check the degree of internal consistency, one could use items 19 to 24 of the questionnaire (the questions relating to the time teachers invest into the assessment of in-class participation) and calculate the *Cronbach Alpha coefficient* to see if the amounts of time given as answers to the individual questions are possible when compared to the amounts of time given as answers to the other questions. The results of the *Cronbach Alpha coefficient* procedure will be included in the section on the study results.

In the end, the following question configuration was chosen: Four questions asked about the institutional context of in-class assessment, based on issues from the respective theory section. Four questions inquired about student motivation, two about teacher motivation. Six questions queried about the scientific requirements of in-class assessment. Five aimed at delivering information about the amount of time teachers spend on the assessment. Another six dealt with attitudes about and experiences with in-class assessment in general. Finally, the last four questions surveyed the demographic background of the participants.

In the following the structuring of the questionnaire and the individual sections will be discussed. The decision was made to group the questions, following the advice that in order to avoid frustrating the participants, one has to "ensure that the respondents' overall impression is that the structure is well-organized and orderly" (Dörnyei, 2003, p. 60). In agreement with another suggestion of Dörnyei's, namley that "questions that deal with the same *topic* should be grouped together" (Dörnyei,

2003, p. 60), this most obvious grouping was chosen. When one applies the design principle that not too many questions should be presented to the participant at one time, it seems adequate to more or less reserve one page in the questionnaire for one topic. This division lead to a total of seven pages, with the first being an introduction and instructions page, the second being dedicated to the institutional background of in-class assessment, the third one dealing with aspects of student and teacher motivation, the fourth one relating to the scientific standards of in-class participation, the fifth one asking various questions about the amount of time teachers spend on the assessment, the sixth one asking some general questions and surveying the demographic background of the teachers, and the seventh and last one being a "thank you" note.

Writing appropriate instructions and selecting the layout

A separate introduction section and a message of thanks at the end of the questionnaire are very important, as they

> *play an important role in determining the respondents' feelings toward the questionnaire and in specifying how they should go about answering the items.* (Dörnyei, 2003, p. 26)

The introduction section briefly explained the purpose and design of the study and also the target audience, so that potential participants who were not teachers of English at a Gymnasium were dissuaded from participating. After that, the introductory page contained the statement that all the information collected would be handled anonymously and some information about the length of the questionnaire and the time it took to fill it in. Finally, there was a note of appreciation for participating that was taken up again in the short message on the last page (see appendix G).

The layout of the questionnaire was chosen as one of several possible templates provided by *oFb – der onlineFragebogen*. It could not be customized. This was not necessary, however, because the template had been professionally developed by the provider of the online questionnaire service and satisfied the common layout standards for academic research (named, for example, in Dörnyei (2003, pp. 19-20)).

Report on the Survey

The questionnaire was composed as described above in the second half of 2008. A final test period for the questionnaire was held in November 2008. After some more

changes based on the feedback gained from test participants (see above), the survey went online in December 2008. Using the database of schools in North Rhine-Westphalia provided by the *Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen*, E-Mail addresses of a number of schools were obtained. They were contacted and asked to pass on the request of participation to the English teachers working there. Against the researcher's expectations and despite of precautions that had been taken, the number of participants increased only very slowly. By the end of January 2009, only six people had taken part. As the required number of participants is much higher (see discussion on participant numbers in section 2.2), the researcher decided to try to reach potential participators in non-electronic ways as well.

A new survey period was set from March to May 2009. 35 schools in North Rhine-Westphalia were selected at random and their postal addresses obtained from the database of the *Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen*. A letter was sent to each of these schools, containing a covering letter to the principals that introduced the survey and asked for their support (see appendix H). Also included were twenty invitations to take part in the survey that were requested to be handed out to the English teachers at that school. This measure showed some effect, so that the number of participators started to increase. In the survey period from March to May 2009, the number of people who took the survey reached 60.

III.2 Results

III.2.1 Quality of Data Collected

A first consideration concerning the quality of the data collected is whether the sample (the number of participants) is big enough. Not all participants have completed the questionnaire. Those who have not cannot be considered for analysis because there would be a bias in the data if a single case were to be considered for the analysis of some items but not for other items. The number of complete sets of answers is 60. Dörnyei (2003) suggests that

> *From a purely statistical point of view, a basic requirement is that the sample should have a* normal distribution*, and a rule of thumb to achieve this […] is that the sample should include 30 or more people.[…]*
>
> *Because in L2 studies meaningful correlations […] have often been as low as 0.30 and 0.40, a good rule of thumb is that we need around 50 participants to*

> *make sure that these coefficients are significant and we do not lose potentially important results* (p. 74)

It is suggested that, according to the first criterion, at least 30 participants are necessary; the second one requires at least around 50. Therefore, the number of 60 seems to be sufficient.

As has been described in the section on study design, an internal consistency-approach is used to ensure the reliability and construct validity of the results. According to this procedure, the *Cronbach Alpha coefficient* should now be calculated to test them. The following table shows the results of the test:

Questions No.	**Cronbach Alpha Coefficient**
1.1-1.5	0.903
4.1-4.3	0.804
10-11	0.556
19-24	0.842
32-33	0.961

The items selected for the test are those that feature either multi-scale questions (such as questions 1.1-1.5, where multiples scales ask about the familiarity with several legal documents) or items between which a direct relationship can be assumed (such as questions 19-24, where the answer to one question is connected to the other answers because the different questions describe the amount of time spent on one activity over different time spans). The *Cronbach Alpha coefficient* should be larger than 0.7 to reflect construct validity. This is the case for all the *Cronbach Alpha coefficients* that have been calculated except for the one for questions 10 and 11. However, applying the test to only two items is a borderline case in which deviation from the 0.7 guideline are possible (Dörnyei, 2003, p. 112). Therefore, and as all the other *Cronbach Alpha coefficients* are a lot higher than 0.7, the reliability and validity test can be considered a success.

A distinction that is crucial to the interpretation of the study results has to be made: That between descriptive and inferential statistics. Dörnyei (2003) describes the two types in the following way:

> *Descriptive statistics are used to summarize sets of numerical data in order to conserve time and space. It is obvious that providing the* mean *and […] the* standard deviation *of the results, we have achieved a well-rounded description of the scores that would satisfy most purposes. […] In order to venture any generalization concerning the wider population and not just the particular sample, […] we need to employ* inferential statistical procedures. *[…] One important feature of statistical significance is that it is the function*

> *of not only the magnitude of the result but also of the* size of the sample *investigated. (pp. 114-116)*

An investigation such as the one that has been carried out here cannot claim significance in the sense of inferential statistical procedures because of the size of the sample. 60 participants are enough for a descriptive statistical analysis, as has been determined earlier in this section. They do not, however, suffice to achieve statistical significance when the wider population we would draw conclusions on, the English teachers in North Rhine-Westphalia, is a high multiple of the size of the sample (Dörnyei, 2003, p. 116). Therefore, the approach of descriptive statistics is used here.

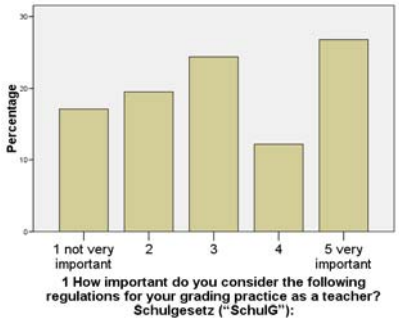
III.2.2 Method of Analysis

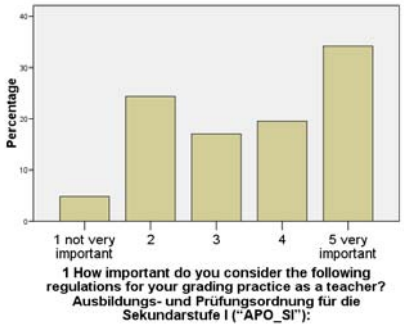
The analysis of the data collected builds on standard statistical operations based on the distinction between descriptive and inferential statistics made above. A univariate analysis is carried out for each individual variable. Measures of central tendency (in most cases the median, as the arithmetic mean cannot be applied to ordinal scales), ranges of dispersion and distributional analysis are used to describe these results. Bivariate analysis in the form of contingency tables will be used in cases where the relationship between two variables is of interest and the data allows this kind of investigation. This is firstly because a complete bivariate analysis would contain 595 individual comparisons, which would exceed the space available for this thesis. Also, the number of participants of 60 is too low to allow for a systematic bivariate analysis. If this number is split up into smaller groups because two variables are analyzed at the same time, the resulting groups can easily be so small that they have single-digit sizes, and such groups then do not fulfill the minimum of the required sample size as described in section 2.1. For the same reason, multivariate analysis with more than two variables will not be carried out at all. A brief introduction to the statistical terms used for the analysis of the results can be found in appendix L.

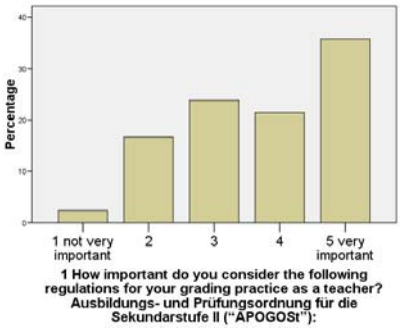
III.2.3 Description of the results

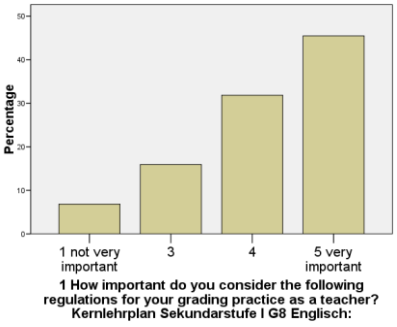
A presentation of the raw data can be found in appendix J. Appendix K contains a summary of the raw data generated by the software used for the analysis (SPSS 15) that gives detailed figures explaining the number of cases considered and the distribution of the answers as absolute numbers and as percentages. The question numbers refer to those in the questionnaire (an overview can be found in appendix F).

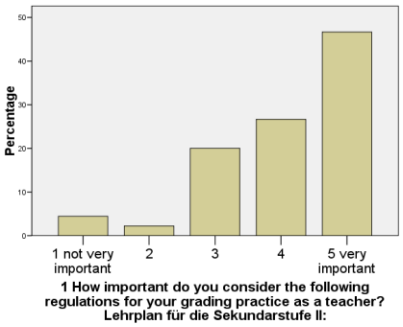
Questions 1-4





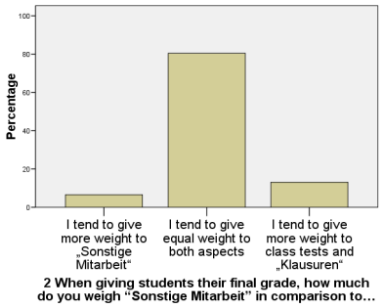




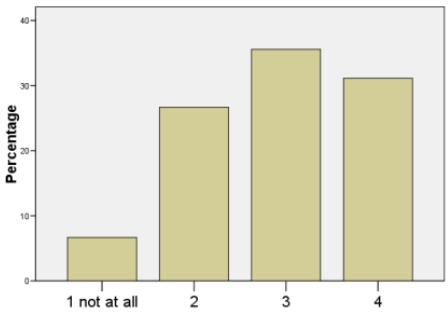


Question one asked teachers about how important they consider different legal documents concerning education for their grading practice. The importance was to be rated on a scale from 1 (not very important) to 5 (very important). The first item, *Schulgesetz*, was rated with a median of 3, which is the lowest median of all items of this question, and a low quartile of 2. 61% of the answers were assigned to answer 3 or below. The answers to the second item, *Ausbildungs- und Prüfungsordnung für die Sekundarstufe I*, are more balanced, with a median of 4 and 46.3% of the answer being 3 or below, however, still with a low quartile of 2. The *Ausbildungs- und Prüfungsordnung für die Sekundarstufe II* has a median of 4 and a low quartile of 3, with 42.9% of the results being 3 or below, thereby still being rated a bit more important than the *Ausbildungs- und Prüfungsordnu ng für die Sekundarstufe I*. The *Kernlehrplan Sekundarstufe I* has a median of 4 and also a low quartile of 4, with only 22.7% of the answers being the middle answer or below. The item *Lehrplan für die Sekundarstufe II* has the same median and answer 3 and below make up 26.7% of the results. The low quartile is 3, rating it as slightly less important than *Kernlehrplan Sekundarstufe I*. Although the items were not ordered intentionally, there is a movement from a normal distribution to a negative skew, indicating that the further down the list an item is found, the more importantly it is rated.

Question two asked the participants how much they graded *Sonstige Mitarbeit* in comparison to written exams. There were three possible answers representi ng either that *Sonstige Mitarbeit* was considered more important, both were given equal weight, or that written exams were more important. By far the greatest number of participants replied that they gave equal weight to both factors (80.4%). 13% said they weighed written exams more, 6.5% said they weighed *Sonstige Mitarbeit* more.

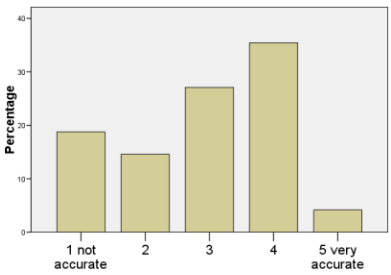


2 When giving students their final grade, how much do you weigh "Sonstige Mitarbeit" in comparison to...

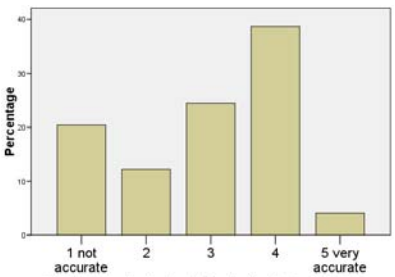


3 In how far do the aspects of „Sonstige Leistungen“ that you grade differ between Sekundarstufe I and Sekundarstufe II?

Question three asked whether the participating teachers graded different aspects as *Sonstige Mitarbeit* in *Sekundarstufe I* and in *Sekundarstufe II*, with the answer ranging from 1 (not at all) to 5 (very much). Except for the fact that no one chose answer number 5, we have a nearly normal distribution with 3 being the answer that was selected most often (35.6%), but 4 (31.1%) and 2 (26.7%) being selected nearly as often and the extreme answers hardly selected.

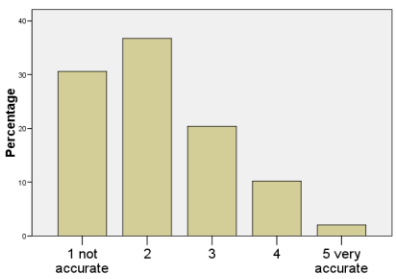


4 In how far do the individual „Kopfnoten“ give an accurate account of the personal and social performance of students? Zuverlässigkeit/Sorgfalt:



4 In how far do the individual „Kopfnoten“ give an accurate account of the personal and social performance of students? Leistungsbereitschaft:

Question four asked about the teachers' opinion on whether the three *Kopfnoten* give a reasonable account of students' performance, the range of possible answers starting at 1 (not accurate) and ending at 5 (very accurate). The answers to the first two *Kopfnoten*, *Leistungsbereitschaft* and *Zuverlässigkeit/Sorgfalt*, are rather similar. The median is 3, about 60% of the answers selected 3 or below (57.1% for the first and 60.4% for the second item), the low quartile is 2 and the high quartile is 4. There is a negative skew that, however, only spreads from options 1 to 4, leaving out option 5. All in all, the agreement seems to be average. This is different,

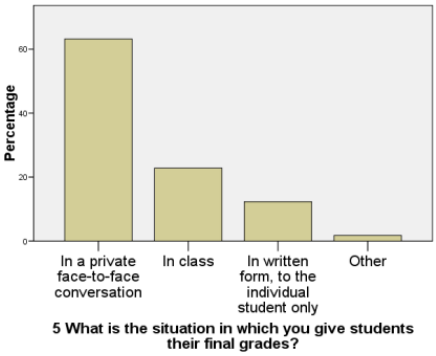


4 In how far do the individual „Kopfnoten“ give an accurate account of the personal and social performance of students? Sozialverhalten:

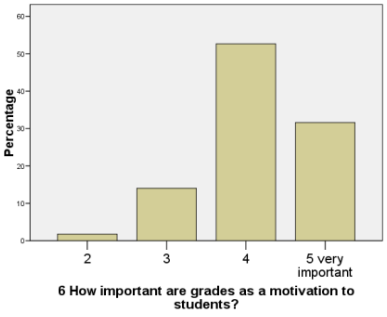
however, for the third item, *Sozialverhalten*. Answers 1 to 3 make up nearly 90 % of the answers, forming a strong positive skew that is reflected in a median of 2 and a low quartile of 1 and a high quartile of 3. This seems to mark a strong disagreement.

In general, the section consisting of questions 1 to 4 is the one which the smallest numbers of participants answered, with the number of partakers who answered this question being in most cases between 40 and 45 out of 60.

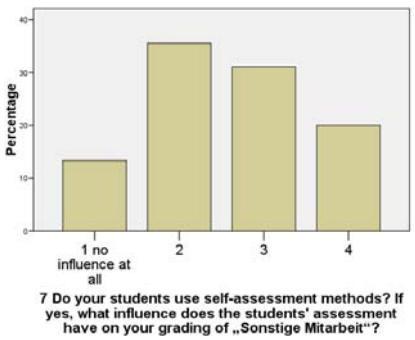
Questions 5-11



Question 5 is a nominally-scaled question that asks about the situation in which teacher give students their final grades. The answer "In a private face-to-face conversation" was by far the one that was most often selected (63.2% of the answers), "In class" accounted for 22.8% of the answers, "In written form, to the individual student only" for 12.3%. 1.8% said they preferred other ways. The answer "In written form, publicly" was not selected at all.

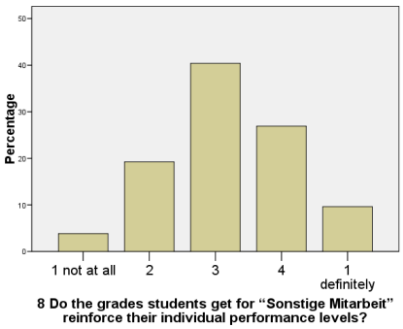


Question 6 was designed to find out how important teachers consider grades as a motivation for students, offering answers on a ordinal scale that begins with 1 (not important at all) and ends with 5 (very important). The median of the answers the participants gave is 4, with the low quartile being 4 as well and the high quartile being 5. This makes a negative skew that, however, has 4 and not 5 as its center with the highest amount of selections (52.6%).



Question 7 was directed towards teachers who let their students use self-assessment tools and asks them how big an influence the students' self-assessment has on their grades. The ordinal scale has answers that range from 1 (no influence at all) to 5 (a lot of influence). The median is 3, the low quartile 2 and the high quartile 3. This means that the answers the participants gave have the form of a near normal distribution with a positive skew. The cumulated percentages confirm this, as 80% of the results are

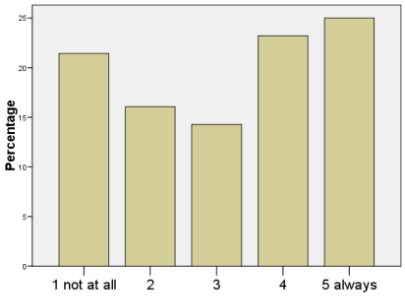
located at 3 or below. The number of answers was rather low in this case, with 45 out of 60 participants answering.



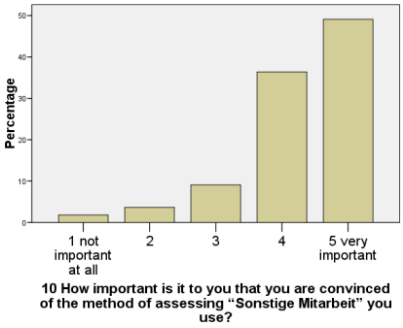
8 Do the grades students get for "Sonstige Mitarbeit" reinforce their individual performance levels?

Question 8 is an ordinally-scaled question about the opinion of the teachers: Do the grades students get for *Sonstige Mitarbeit* reinforce their individual performance levels? The answers range from 1 (not at all) to 5 (definitely). The median is 3, the low quartile is 3 as well and the high quartile is 4. The answers are in a normal distribution with a very slight negative skew.

Question 9 asks participants whether they give their students a say in what grades they get for *Sonstige Mitarbeit*. Again, the scale is ordinal with 5 possible answers starting with 1 (not at all) and ending with 5 (always). The median is 3, the low quartile is 2 and the high quartile is 4.75. There is an inverted normal distribution that represents a tendency towards extremes. A slightly negative skew highlights that the tendency is more to the positive extreme (4 and 5) than to the negative one (1 and 2).



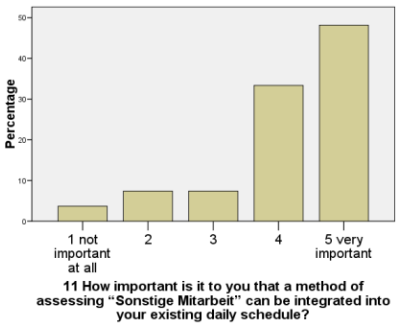
9 Do you give students a say in what grade they get for "Sonstige Mitarbeit"?



10 How important is it to you that you are convinced of the method of assessing "Sonstige Mitarbeit" you use?

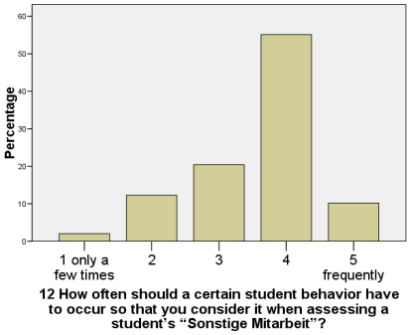
Question 10 is the first question to deal with teacher motivation. It asks teachers to rate how important it is for them that they are convinced of the method of assessing *Sonstige Mitarbeit* they use on a scale ranging from 1 (not important at all) to 5 (very important). There is a strong negative skew that is reflected in the measures of central tendency and dispersion: The median is 4, the low quartile is 4 as well, the high quartile is 5. Only 14% of the answers replied with 3 or below. This result indicates that there is strong agreement that being convinced of the method they use is very important for the teachers.

Question 11 deals with teacher motivation, too. It asks how important it is that a method of assessing *Sonstige Mitarbeit* can be integrated into an existing daily schedule. There is also a strong negative skew with a median of 4, a low quartile of 4 and a high quartile of 5, with 18.5% percent rating it 3 or below. This suggests that

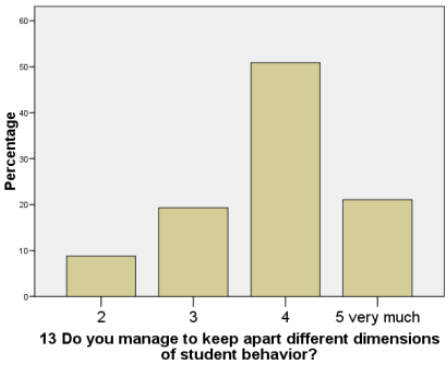


the agreement to the subject of the question being important is very strong, even though slightly less strong than with question 10.

Questions 12-18



Question 12 begins the set of questions that deal with the three scientific criteria of objectivity, reliability and validity. It asks for the participants' opinion on how often a certain kind of student behavior should have to occur before it is considered for grading *Sonstige Mitarbeit*. A five-point ordinal scale asks participants to select answers beginning with 1 (only a few times) and 5 (frequently). The results show a normal distribution with a strong negative skew, with is reflected by a median of 4, a low quartile of 3 and a high quartile of 4. For 55.1% of the answers, 4 was chosen, which is more than all other answers together. This suggests that most participants believe that a certain kind of behavior should not have to occur frequently, but at a level just below that.



Question 13 asks the participants whether they manage to keep apart the different dimensions of student behavior, with the answers ranging from 1 (not at all) to 5 (very much). Again, 50.9% of the participants chose answer 4, which is more agreement than was given to all other answers. Similarly to question 12, there is a normal distribution with a negative skew, with a median of 4, a low quartile of 3 and a high quartile of 4. No participant answered that they did not manage to keep different dimensions of student behavior apart, and most believed that they managed to do that just short of "very much".

Question 14 asks teachers to select which aspects of student behavior they graded as part of *Sonstige Mitarbeit*. The list of answers is scaled nominally and consists of 12

items of which any number can be selected. These are the selections sorted by frequency of selection (as total numbers out of 60):

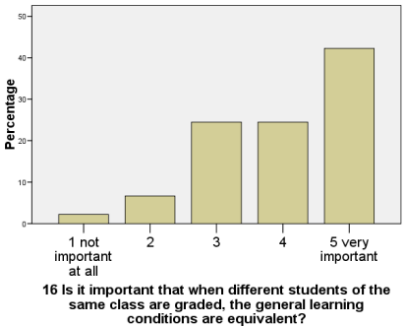
Quality of oral participation	57
Frequency of oral participation	56
Presentations	52
Quality of homework	51
Doing homework	49
Tests	48
Creativity	36
Orderly and complete workbooks	29
Regular presence in class	23
Social behavior in class	16
Protocols	15
Others	11

In question 15, participants who chose the item “Others” at Question 14 could specify this in a text field. The extra items were added by the participants, all were mentioned only once unless otherwise noted:[1]

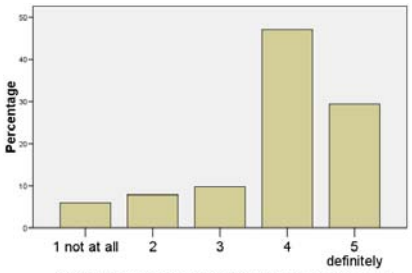
Participation in projects and group assignments (named 2 times)
Extra credits
Contributing material for the lessons
Arranging expert visits
Pronunciation and intonation
Readiness to elaborate on complex tasks
Style of working in silence
Ability and willingness to focus on the task
Dealing with other students' contributions (taking things into consideration, developing an argument etc.)
Communication skills, attitude
Portfolio, reading log, posters done on a special topic, a folder done on a special topic
2-minute talks and other regular contributions within the weekly programme
Communicative aspects: e.g. use of classroom phrases; avoiding/speaking English during phases of group work

[1] To the entertainment of the readers, one of the replies to question 15 should be mentioned although it has been removed from the result list. One of the participants answered this question with the comment: “The political idea of Sonstige Mitarbeit suffers from severe systematic mistakes.”

In question 16, the teachers are asked to rate how important it is that when different students of the same class are graded, the general learning conditions are equivalent on a scale from 1 (not important at all) to 5 (very important). 15 out of 60 teachers did not answer this question. The median is 4, the low quartile is 3 and the high quartile is 5. There is a strong negative skew, and 5 is the answer that was selected most often with 42.2%, followed by 3 and 4 with 24.4% each. This can be interpreted as the teachers seeing the statement in the question as correct, but disagree on how important it really is, although more than 42% see it as very important.



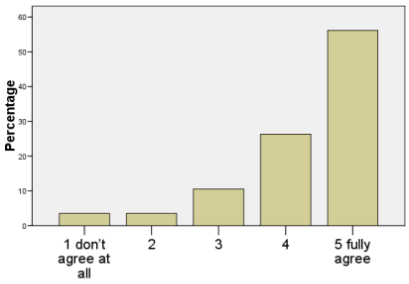
16 Is it important that when different students of the same class are graded, the general learning conditions are equivalent?



17 Does the fact that the different dimensions of students' oral language competency are connected have significance for your assessment of "Sonstige Mitarbeit"?

Question 17 asks the participants: "Does the fact that the different dimensions of students' oral language competency are connected have significance for your assessment of *Sonstige Mitarbeit*?" The answering scale is ordinal with answers ranging from 1 (not at all) to 5 (definitely). The median is 4, the quartile is 4 and the high quartile is 5. There is a distinct negative skew, but the answer most frequently selected is 4 (47.1%). 5 is selected relatively frequently as well (29.5%), and the answers 1 to 3 only make up 23.5% of the answers. The agreement with the statement in the question is very strong, but answer 4 is selected more frequently than the answer of maximal agreement, 5.

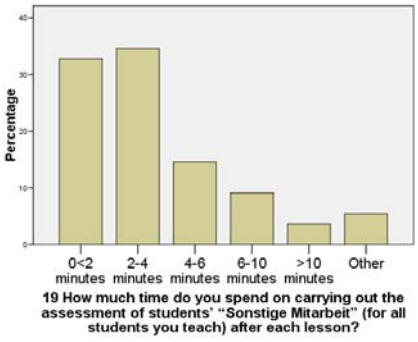
In question 18, teachers were asked to say how much they agree with the statement: "It is essential to fair grading that the teaching methods used in class must be varied." The spectrum of the answers goes from 1 (don't agree at all) to 5 (fully agree). The median is 5, the low quartile is 4 and the high quartile is 5. There is a very strong negative skew with 56.1% of the replies choosing 5. This shows that there is very strong agreement with the statement in question.



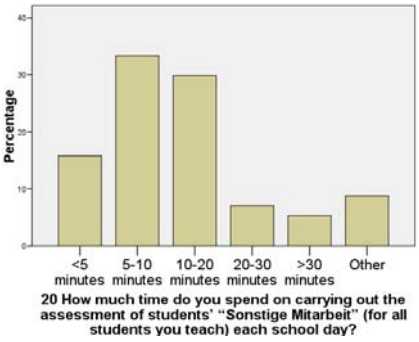
18 Do you agree with the following statement: „It is essential to fair grading that the teaching methods used in class must be varied?"

Questions 19-24

In questions 19-24, the participants were asked how much time they spend on carrying out the assessment of students' *Sonstige Mitarbeit* on the basis of given

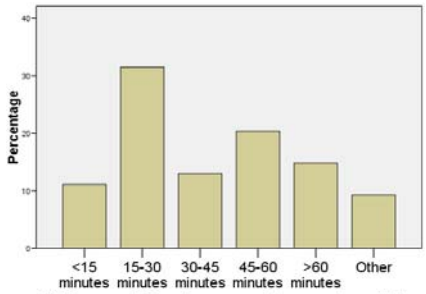


periods of time that they could choose from. The periods of time will be referred to by numbers, starting from the shortest, which was in every case the first answer, to the longest period, which was the last one and therefore is assigned the highest number. First, the participants were asked to say how much time they spend on assessing *Sonstige Mitarbeit* after each lesson, for all students they teach. The median is 2, the low quartile is 1 and the high quartile is 3. This is reflected in a very strong positive skew, indicating that most teachers spend very little time under these conditions. 67.3 percent of the participants spend 4 or less minutes. 5.5% of the participants chose the answer "Other", and, as all periods of time are covered by the different answers, this is likely to mean that they spend no time at all on assessment after each lesson.

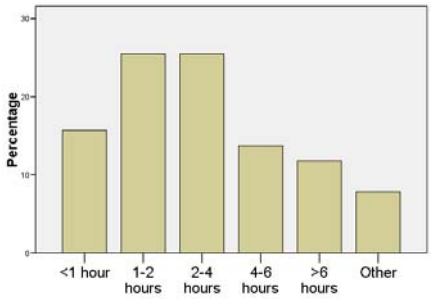


Question 20 asks teachers how much time they spend on the assessment each school day. The median value is 3, the low quartile value is 2, and the high quartile value is 3 as well. The results are distributed in something remotely like a normal distribution with a distinct positive skew, so that the maximal number of answers goes to answer 2 with 33.3%, followed by answer 3 with 28.3%. Only 15 % chose answer 1, however. It can be summarized that a vast majority of the teachers spend between 5 and 20 minutes on the assessment of *Sonstige Mitarbeit* each school day. Again, there are 8.8% who chose "Other", probably indicating that they spend no time at all that way.

Question 21 asks the same question as the two questions before, this time about a whole school week. The results are in a complex distribution that has two adjacent normal distributions, one with the maximum at answer 2 and one with the maximum at answer 4. This is also shown by that the low quartile is 2 and the high quartile 4.25; the median of 3 is not a good indicator for the results this time, as answer three is the one that has been third least often selected out of seven. As a consequence of these results, we can deduct that there are two different tendencies – one centered around spending 15-30 minutes on the assessment per school week, and one centered



21 How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) each school week?

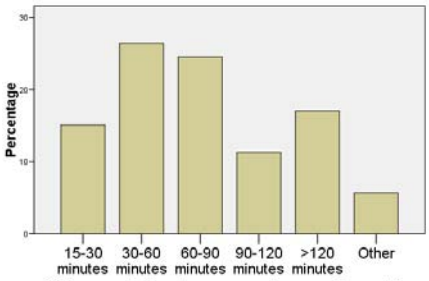


22 How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) each school month?

around 45-60 minutes. 9.3% seem to spend no time on the assessment at all in that period.

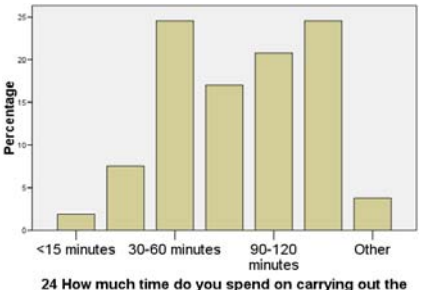
Question 22 asks about the time spent on the assessment of *Sonstige Mitarbeit* in a whole school month. There is a normal distribution with a slight positive skew; the median has the value 3, with 2 being the low quartile and 4 being the high one. As a result, we can see a tendency centering around spending between 1 and 4 hours per month (answers 2 and 3 combined; both represent 25.5% of the results). 7.8% seem to be spending no time at all during that period.

Question 23 is intended to find out how much time teachers spend on the assessment each quarter in addition to the time they already spent in regularly during the term (as extra time for determining the grades). Again, the results are in a complex distribution with two tendencies. One is centered around the maximum answer 3 (with 26.4% of the answers selecting 3) and is by far the stronger one. The other has its maximum in answer 6 (17%). This is reflected by the measures of central tendency and dispersion. The low quartile is 3, that is at the center of the first tendency, and the high quartile 5, that is at the center of the second tendency. The median value of 4 highlights how much stronger the first tendency is. 5.7% of the participants seem not to be spending any time on the assessment each quarter.



23 How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) each quarter (on determining Quartalsnoten only , not considering the time you ...

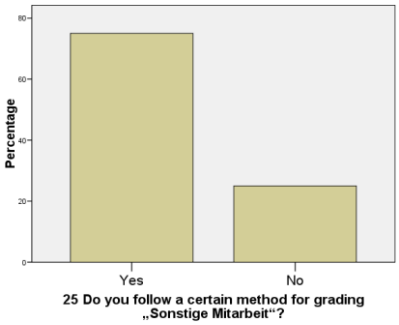
Question 24 asks about the time additionally spent at the end of teach half term. Once again, we have a complex distribution with two tendencies that, this time, seem nearly equally strong. One has its maximum in answer 3 (24.5%) and one, in answer 6 (24.5%). The low quartile of 3 and the high quartile of 6 reflect this distribution. In this case,



24 How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) each half term (on determining the overall grade, not considering the time you spe...

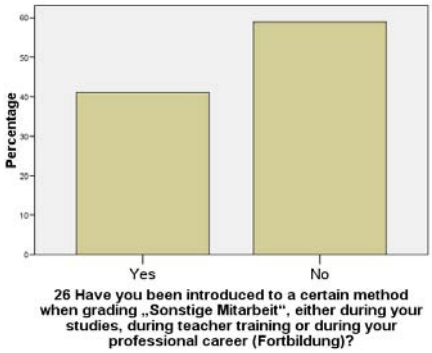
the median value of 4 is not a reasonable measuring instrument, as 4 is only the fourth most selected answer out of 7. As a consequence, we can distinguish one tendency centered around spending 30-60 minutes at the end of each half term and another centered around spending more than 120 minutes. 3.8% seem to be spending no time on assessment at all at the end of each half term.

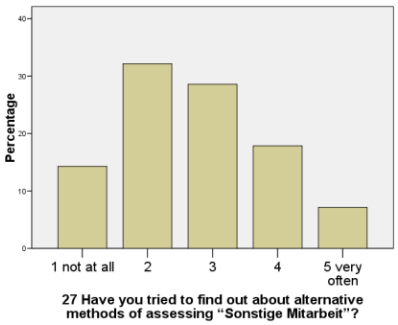
Questions 25-30



Question 25 is the first one of the general questions on methods of assessing in-class participation. It asks teachers whether they follow a certain method for the assessment. The answering possibilities are scaled nominally and only give the choices "Yes" and "No". 75% of the participants answered "Yes", 25% answered "No", which renders a clear majority.

Question 26 asks teachers whether they have been introduced to a certain method of grading *Sonstige Mitarbeit* and also leaves them the answers "Yes" and "No". 58.9% answered "No", 41.1% answered "Yes", which is a distinct tendency towards "No".





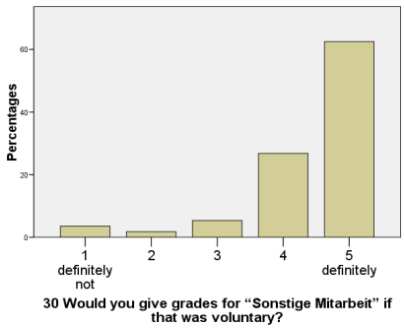
Question 27 asks the participants whether they have tried to find out about alternative methods. The answers range on an ordinal scale from 1 (not at all) to 5 (very often). The median is 3, the low quartile is 2 and the high quartile is 3.75. There is a positive skew, which is reflected in that 2 is the most frequent answer with 32.1%, and 75% of the replies chose answer 3 or below. This suggests the tendency that the participants have rather not tried to find out about alternative methods.

Question 28 asks the participants how satisfied they are with their method of assessing *Sonstige Mitarbeit*. Again, there is a 5-point ordinal scale

with answers starting with 1 (not satisfied at all) and ending with 5 (very satisfied). The results show a negative skew, with a median value of 4, a low quartile of 3 and a high quartile of 4. Four is by far the most frequent answer (40%). There is a tendency for the participants to say that they are satisfied with their method, but relatively few selected the maximum degree of satisfaction, represented by answer 5 (16.1%).

Question 29 is a nominally scaled list question again. It asks teachers to give reasons why they are not satisfied with their method if this is the case. Nine possible answers were given, one of them being the item "Others". These are the results sorted by frequency of selection:

Others	58
Too complicated	56
Not in perfect accordance with the legal regulations	56
Not motivating to the students	54
Too time-consuming	51
Unsystematic	50
Too irregular	45
Not objective enough	44
Not reliable enough	42



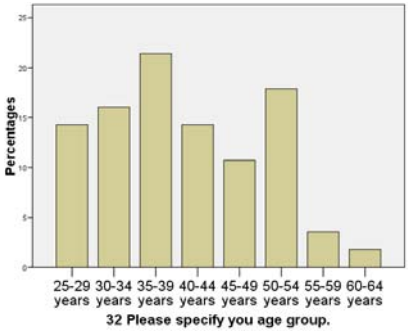
In question 30, the participants are asked to say whether they would give grades for *Sonstige Mitarbeit* whether that was voluntary or not. The answers are arranged in an ordinal scale and range from 1 (definitely not) to 5 (definitely). There is a very strong negative skew, with a median value of 5, a low quartile of 4 and a high quartile of 5. 62.5% chose answer 5, answers 4 and 5 together make up 89.3% of the replies. It can be seen that there is a very strong tendency to give a positive answer to this question. When examining the graph, it may strike one that, even though agreement with the answers seems to lower consistently from answer 5 to 2, answer 1 has been selected almost twice as often (3.3%) as answer 2 (1.7%). However, in absolute numbers, this means that two people chose answer 1 compared to one who chose 2, so the significance of this difference can be doubted.

Questions 31-34

The last four questions deal with the demographic background of the participants. Question 31 asks about their gender. 60.7% are female, 39.3% are male.

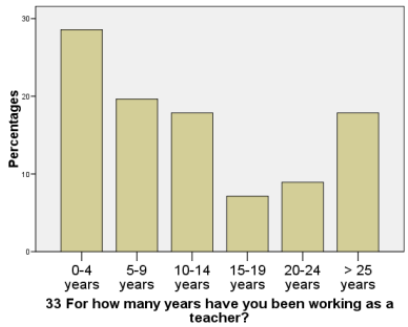
Question 32 asks about the age group of the participants. The answers are organized in ordinal scales and organized as five-year periods that the participants can select. These are the results:

25-29 years	14.3%
30-34 years	16.1%
35-39 years	21.4%
40-44 years	14.3%
45-49 years	10.7%
50-54 years	17.9%
55-59 years	3.6%
60-64 years	1.8%
>70 years	0%



Question 33 asks for many years the participants have been working as teachers. The setup of the question is similar to that of question 32. The results are:

0-4 years	28.6%
5-9 years	19.6%
10-14 years	17.9%
15-19 years	7.1%
20-24 years	8.9%
>25 years	17.9%



Finally, the last question asks the participants in which sections of the *Gymnasium*, *Sekundarstufe I* or *II*, they do most of their teaching, again using a nominal scale. These are the results:

More in *Sekundarstufe I* than in *Sekundarstufe II*	19.6%
To the same degree in *Sekundarstufe I* and *II*	41.1%
More in *Sekundarstufe II* than in *Sekundarstufe I*	35.7%
Other	3.6%

Next, selected aspects of the bivariate analysis will be discussed (for the contingency tables used as a basis for this section see appendix M). As has already been explained, only selected aspects will be discussed; firstly because a full analysis cannot be accomplished within this paper; secondly because the data material, especially the number of participants, only allows a well-founded analysis in special cases when at least one dimension of the contingency table contains no more than two or three items.

Firstly, the variable gender is set against other variables in order to highlight differences between the two sexes. Question two, asking about how teachers weigh *Sonstige Mitarbeit* in comparison to written exams, is the first question where a

significant difference can be detected. While 94.1% of the male participants weigh both aspects equally, only 75% of the women do. 5.9% of the males give more weight on exams opposed to 17.9% of the females. 7.1% of the women give more weight to *Sonstige Mitarbeit*. Women seem to be more ready to deviate from giving both aspects equal weight.

Applying the gender distinction to question 5 (What is the situation in which you give students their final grades?), we see that women tend towards giving them in a private face-to-face conversation (76.5%), while men only do that to a lesser extent (45.5%) and rather use different ways, such as giving them in class (3.,8%) or in written form to the individual student (18.2%).

Women seem to follow a certain method towards grading *Sonstige Mitarbeit* less often than men (70.6% vs. 81.8%, question 25). This can be set against question 28, as men seem to be a lot more satisfied with their method of assessing *Sonstige Mitarbeit* than women. 77.3% of the men answered this question with 4 or 5, while only 47.1% of the women gave such answers. Instead, the answer women gave most often was the average of 3 (38.2%).

Another observation can be made when one uses the age group variable to differentiate the results. In order not to split the groups of participants up into too many subgroups that cannot be analyzed in a scientific manner anymore, the variable age group has been recoded into a one that differentiates between age group 25-49 years on the one hand and 50->70 years on the other hand. When using this variable to differentiate question 26, one learns that 44.2% of the members of the younger age group have been introduced to a method of assessing *Sonstige Mitarbeit* during their professional training, while this is only true for 30.8% of the older age group.

On the basis of question 34 that asked participants in which section of the *Gymnasium* they do most of their teaching, questions 25 and 26 can be differentiated. Set against question 25 (Do you follow a certain method for grading *Sonstige Mitarbeit*), this variable reveals that 87% of those who teach to the same degree in *Sekundarstufe I* and *II* answer with “Yes”, but only 70% of those who teach more in the *Sekundarstufe II* and only 54.5% of those who teach more in the *Sekundarstufe I*. For question 26 (Have you been introduced to a method of grading *Sonstige Mitarbeit* during your professional training), we learn that 43.5% of those who teach to the same degree in *Sekundarstufe I* and *II* answer with “Yes”, and so do 45% of

those who teach more in the *Sekundarstufe II*, but only 18.2% of those who teach more in the *Sekundarstufe I*.

One last observation can be made: Setting question 25 (Do you follow a certain method for grading *Sonstige Mitarbeit*) against question 28 (How satisfied are you with your method) shows that those who do not follow a certain method as a tendency answer question 28 with answer 3 (50%), while only 28.5% chose answers 4 and 5. Out of those who do, 69% chose answers 4 and 5. This means that satisfaction with the own way of assessing *Sonstige Mitarbeit* is significantly higher among those who say that they employ a certain method. Another possible interpretation is that teachers who do not follow a certain method are not very concerned with how or even how well they assess in-class participation and therefore answer question 28 with the neutral answer, i.e. 3.

III.2.4 Interpretation of the results

A distinction can be made concerning the various items of question 1. The varying legal regulations are rated as more important for the own teaching practice as they become more concrete: The *Schulgesetz* is rated least important of the five items, while the *Lehrpläne* for *Sekundarstufe I* and *II* for the subject the participants are teaching are rated most important of all. This goes along with the observation that the number of participants who did not answer the items in question one at all decreases in the same way in which the importance is rated. The second question shows that, all in all, teachers consider the legal regulation that in-class participation and written exams are to be weighed equally for the final grade, binding for them. As one might expect, if teachers deviate from this rule it is most often in the direction of weighing written exams more; women are more likely to deviate from that rule than men. Question 3 shows no clear tendency on whether teachers grade different aspects as in-class participation in *Sekundarstufe I* and *II*. Question 4 differentiates between three *Kopfnoten*. While for the two first ones, *Leistungsbereitschaft* and *Zuverlässigkeit/Sorgfalt*, the rating is average with, on one hand, a large number of answer 1, 2 and 3, four as the highest rating suggests that there are also a number of teachers who rather agree to the usefulness of *Kopfnoten*. This is completely different for the third *Kopfnote*, *Sozialverhalten*. The answers are overwhelmingly negative, marking a very strong disagreement with this *Kopfnote*. One of the possible interpretations for this is that it is an answer that the teachers give on the question what should be or what can be assessed as student behavior in class. While a number

of teachers believe that willingness to perform, accuracy and reliability are features that can be assessed, a vast majority thinks that this is not the case for social behavior.

The degree to which teachers go along with the institutional requirements can be classified as average. How binding they find the documents of educational legislature seems to be vastly dependent on how concrete and exclusively relevant for their subject they are. For the two sample pieces of educational legislature that were included in the questionnaire, the result is that by and large, teachers abide by it. They have a critical but also differentiated attitude towards the *Kopfnoten*, which also represent educational legislature.

The results of question 5 show that most teachers give students their final grades in a private face-to-face conversation, with in class being the second most frequent, but nearly only a third as often as the first. The third most frequent way is to do it in written form, to the individual student only. The first and third way are in agreement with the motivational requirements that have been stipulated in this paper, as they minimize the social-comparison-aspect of the grade-giving. About a fifth of the teachers, those who use the practice of giving grades in class, are not in agreement with this motivational requirement. Women seem to consider this more than men, as about 85% of them choose the ways that are unproblematic from a motivational point of view, while only a little less than 65% of the men do this and the other men choose ways that are more likely to cause the problems described in the theory section.

This can be set into contrast to the answers to question 6, as they show that, in general, there is a very strong agreement among the teachers questioned that grades and assessment is a strong motivational factor. Question 7 asks about one aspect in which motivational considerations can be applied. Of the teachers in the survey, three thirds use self-assessment tools, which shows that their importance is recognized. However, only a rather small number of the teachers make the next step and actually consider the results of the self-assessment when it comes to determining the grade. Question 8 stands in contrast to the results of question 6, which shows that nearly all teachers consider grades an important motivational factor; in question 8, only an average number of teachers express the believe that grades reinforce students' individual performance levels in class. One possible reason for this could

be that some teachers lack the background knowledge on motivational issues. One also gets a very diverse picture out of the replies to question 9, as there seems to be a wide disagreement about whether students should have a say in what grade they get. The inverted normal distribution of the results shows a tendency towards either one extreme or the other: Either students have no say at all, or a rather strong one, where the second tendency seems to be the stronger one. This is in agreement with the motivational principle of two-way negotiation.

In general, teachers recognize that motivation is a central issue for grading and for a method for assessing in-class participation. They also try to put motivational ideas into practice in several ways. However, there are results that suggest that some teachers lack the background knowledge on the motivational issues of grading, which both affects their beliefs about motivational conditions and their motivational behavior in class. This could be an aspect of teacher training that could be improved if it really turned out to be a problem.

Questions 10 and 11 show that for the teachers, their own motivation when it comes to assessing in-class participation is very important as well. For a vast majority, it is very important that they are convinced of the method they use (question 10), although this does not reveal what makes up that conviction for the individual teacher. Similarly, teachers also find it very important that their method can be integrated into a fixed daily schedule. As has been suspected in the theory section, it turns out that teacher motivation (and the factors connected with it, such as a scientific design that is appealing to them) actually is a very important aspect of a method of assessing in-class participation. No method can be successfully implemented without agreement and a certain amount of motivation on the side of the teacher, and it has been confirmed that the teachers have expectations that the method should fulfill.

Question 12 is the first of four questions that deal with the scientific requirement of objectivity as it relates to the requirement of systematic observation. It is also related to reliability, as it deals with the preconditions for reliable, that is, reproducible measurement. The result that more than half of the teachers who participated in the survey believe that a certain kind of behavior should occur rather frequently (4 on a scale of 5) is a first indicator that teachers have a good understanding of objectivity issues and reliability. Question 13 shows that, in general, the teachers think that they

are able to differentiate between the various dimensions of student behavior; there is, at the same time, an awareness of that this can hardly ever be achieved in full, as only relatively few teachers chose the highest answer; a little more than a quarter of the teachers seem to have problems with keeping the different dimensions apart.

Questions 14 and 15 are the first to deal with the issue of validity, but are connected to reliability as well. It shows that what teachers consider when assessing in-class participation is very varied. While there are factors that nearly everyone considers, such as quantity and frequency of oral participation, there are also a number of factors that only three thirds of the teachers, or half or one fourth of them, or only individual teachers consider. As has been discussed in the theory section on validity, this is a significant problem because validity can only be established on the grounds that the subject of grading is clearly defined. The strong variation in the list from question 14 does not mean that one teacher grades different aspects at different times. But is means that for students, different aspects of their behavior are graded at different times, varying from subject to subject or from teacher to teacher within a subject when there is a change.

Question 16 deals with reliability. The first noteworthy fact about the results is that a relatively large number of people did not answer it. This could mean that the question has not been put in a concrete enough manner, but could also mean that some teachers have never considered that issue. Of the teachers who answered, nearly all rated something between average and high importance, with an especially high number who thought it very important. This suggests that there is an awareness of reliability issues. It does not, however, say anything about how well they manage to put this awareness into practice, but, in any case, being aware of a problem is a reasonable foundation for dealing with it.

Question 17 is concerned with objectivity. The outcome shows that the participants are aware of that the different dimensions of students' oral language competency are connected and that they must consider this fact in their assessment of in-class participation. This means that they know that the quality of what a student says is limited by other aspects of language competency, such as vocabulary, ect.

Question 18 is the last one dealing with validity. The teachers are clearly aware of that teaching methods have to be varied in order to assess students in a fair manner. Together with question 12, this draws a diverse general picture concerning how well

teachers deal with validity issues. Question 18 is an example of high teacher awareness, question 12 is an example of rather mixed awareness, although this may be related to the teachers wish for latitude in assessing their students as well.

Questions 19-24 deal with the time used for the assessment of students' in-class participation and is thus also related to the requirement of objectivity. The amount of time spent after each lesson is rather small. The biggest single group spends 2-4 minutes on it, but there are also a large number who spend even less time or do not take notes on students' achievements after each lesson at all. This group is a lot smaller when considering whole school days, and more than sixty percent spend between 5 and 20 minutes assessing in-class participation each day. The tendencies become more varied when it comes to school weeks, although the results for school months suggest a tendency towards spending between one and four hours. The amounts of extra time spent on determining the quarter and half-term grades again are very varied, although they show two different tendencies. As one does not really know what exactly the way the different teachers assess in-class participation is, it is especially difficult to make any statement about the longer periods of time concerning whether the amount of time spent is too large or too small. One can, however, make two observations. Firstly, as a general tendency, the amount spent after each lesson is very short, or even no time at all. This would be the ideal period of time for recording observations but, as has been argued before, is hard to integrate into a typical teaching schedule at school and also too time-consuming. The same tendency also shows that most teachers invest a relatively constant amount of time each day, which, as has been argued in the respective theory section, is also acceptable. Furthermore, as the very high *Cronbach Alpha coefficient* suggests, the consistency among the answers to questions 19-24 that one individual participant gave is considerable. This indicates that the recording of observations and the further steps in the grading process are done consciously and in a systematic way.

From what can be gathered from questions 12, 13, 17 and 19-24, teachers seem to be good at dealing with the requirement of objectivity. There is a definite awareness of the issues connected with it. Similarly to what has been said on reliability, awareness is no insurance that they will be able to act accordingly, but questions 19-24 also suggest that, based on the information they gave during the survey, they also manage that rather well.

The next group of questions deals with general questions about assessing in-class participation. Question 25 reveals that 75% of teachers follow a certain method when assessing it. It is, of course, a positive observation that a majority as great as three quarters follows a certain method. One the other hand, because of the big need for there being a method at all that has been explained in the theory section, it would be desirable to have even higher numbers here. It would be utopian to demand a result of 100% of teachers following a certain method, but in this case the maxim has to be "as many as possible".

The result of question 26 that nearly 60% of the teachers have not been introduced to a method of assessing in-class participation during any part of their professional training reveals one of the problems immanent to the topic. Of course, it is connected to the problem that there is no received set of methods with a theoretical foundation from which one could choose, so that there is little that could be presented to the new teachers. But this, of course, results in teachers having had no proper introduction to the specific issues with assessing in-class participation and having not even been told why it is necessary to use a certain consistent and well-developed method at all. Without this understanding, it is not surprising that a number as high as 25% of the participants still think that it is sufficient to assess in-class participation without a certain method. However, the number of participants who have not been introduced to a method is considerably smaller among younger teachers than among older ones, although it is still clearly higher than 50%. Also, participants teaching to a greater extent in the *Sekundarstufe II* have been introduced to a method twice as often as those who teach to a greater extent in the *Sekundarstufe I*. To examine this difference, one should aim to find out whether there is a possible difference in the training the two groups of teachers get.

This result might also bear on the figures for question 27, where three quarters of the participants answered with 1, 2 or 3 to the question whether they have tried to find out about alternative methods of assessing in-class participation themselves. Several possible reasons can be identified. Firstly, it could be that people think their own method is sufficient (see question 28); secondly, as has already been argued, without an introduction to the specific problems with in-class participation, one might not realize why a certain method or a method at all is necessary; thirdly, it might also be that participants thought that there were no proper methods available, as there is no prominent literature that focuses on the issue of in-class participation.

The results of question 28 are very surprising. More than 50% of the participants answered the question of how satisfied they are with their method of assessing in-class participation with 4 or 5, with the biggest group of repliers having selected 4. Women seem to be more skeptical of their method, as they gave more average ratings, while, of the men, more than 75% said that they were rather satisfied or very satisfied. This number is surprising in the light of the fact that nearly 60% of the participants have not been introduced to a method of grading in-class participation and that only a certain percentage follow one at all. However, this satisfaction fits in well with the results for question 27, which show that most teachers who participated have also not tried to find out about other methods.

The findings of question 29 stand in contrast to those of question 28. In the last question, a majority of participants said that they were rather satisfied or very satisfied with their method of assessing in-class participation. The instructions for question 29 were “If you are not satisfied with your way of grading *Sonstige Mitarbeit*, what are the problems?”. The high satisfaction expressed in answer to question 28 suggests that the number of answers to question 29 should be rather low. But on the contrary, six of the nine items were selected 50 or more times by a total of 60 participants. The other three items were still selected 42 or more times. This suggests that, even though the total satisfaction with the method the teachers are using is high, they also see a lot of problems. A possible interpretation for this is that teachers know that their method of assessing in-class participation is faulty, but they believe that there is no better one. Surprisingly, the item that was chosen most often was “Others”. That means that even though the selectable items were chosen in order to cover very prominent problems, the teacher still see a number of other problems that exceed the range of the question. This question did not give participants the opportunity to add more items in an extra text field, so that it is not possible to find out what exactly the other problems are. The rest of the items reflect issues that have been dealt with in the theoretical section of this paper. Teachers also see them as weaknesses of the methods they use, which shows that their rather high degree of satisfaction must be the results of some kind of weighting that made them decide that the advantages of their methods are greater than the known weaknesses.

The above results have shown that the assessment of in-class participation is not a trivial task and that teachers are aware of that. Nevertheless, question 30 (Would teachers gives grades for in-class participation if that was voluntary?) shows that

they would still give grades for in-class participation if the decision whether to do that or not were up to them. Especially because of the many complications about assessing in-class participation, teachers seem to believe that grading it is so important that it is worth finding a way to deal with those complications.

4 – Synthesis and Conclusion

IV - Synthesis and Conclusion

The final task of this thesis is to find a synthesis between findings of the theoretical section and the results of the study that was carried out. As the study has been designed on the basis of the theoretical section, the most interesting question is how the study results can be compared to the theoretical requirements.

The criteria for a method of assessing in-class participation that were called institutional vary in how much they are taken up and fulfilled by the teachers. This is not unproblematic, as they do not, such as the other criteria, represent scientific considerations on what would be ideal, but real and binding legal regulations. However, only a small percentage of teachers actually violate those regulations; most just consider those that are not very concrete as less important for their grading practice. There is also variation in how they deal with motivational problems. Some of the findings suggest that many teachers have a good understanding of the motivational issue of grading. However, the study has also shown that some teachers lack the necessary knowledge to implement motivational behavior in their teaching practice. A significant result that should not be forgotten is that it is quite important for the teachers to find the method they are using motivating and satisfying for themselves. The aspects of the scientific requirements that have been examined also shed a very good light on the teachers: Most were highly aware of the problems with objectivity and reliability, as there were always tendencies which indicated that the majority successfully implemented the criteria drawn up. This is especially significant as objectivity is probably the most important of the three criteria, as it can easily be compromised but is most essential for the whole grading process. Only the third criterion, validity, has produced more mixed results. It is very important that what is graded is very clearly defined, both at the intraindividual and the interpersonal level.

It has to be kept in mind that a set of requirements such as the one that has been compiled in the first section of this paper has, to the author’s knowledge, never been created before (in such a specific way for the area of assessing in-class participation); it can be assumed that when the participants filled in the questionnaire form they did not have a complete theoretical framework in mind, but rather answered out of individual beliefs and based on their own teaching practice. Therefore, the fact that they agreed with and fulfilled most of the requirements drawn up (with the exception

of some motivational and validity issues) is nothing short of astonishing. This is because of the fact that the majority of teachers have never been introduced to methodology during their professional training and also never tried to find out about it themselves. It turns out that, based on the information they gave away in the survey, they made many correct motivational decisions based on either intuition, experience, or methodological knowledge they have from other areas of teaching. This is a considerable achievement that should be held in high esteem. It is also a vague indicator of consistency between the theoretical section of this paper and the beliefs of the teachers examined, suggesting a likeliness that, to a certain degree, either both are correct or false.

However, this result is not an argument for the general superfluity of a method for assessing in-class participation. On the contrary, it even shows that a method would be extremely helpful, as the strong correspondence between the theoretical criteria and the findings on the participants' teaching practice means that a method would fit in and would probably be well received. Also, even though the participating teachers fulfilled the theoretical criteria in a surprisingly precise way, these were only tendencies. When between 50% and 90% of the teachers questioned agree with a certain principle for a method of assessing in-class participation, this still means that there are a number of teachers who answered differently and who should at least be offered information on why their choice might not be ideal from a theoretical point of view. Also, a well-founded method should not only tell people to act according to principle but also communicate at least a small degree of theoretical understanding for it (for example, a teacher should not only be told that it is a bad idea to put up a list with final grades in class but also understand why). Finally, there are some areas in which the attitudes and professional behavior of teachers are not completely in accordance with the theoretical ideas, mainly student motivation and validity.

Again, it should be made explicit that this paper did not intent to criticize or excoriate teachers in any way. The criteria to which their beliefs about teaching and teaching practices have been compared were only drawn up in this paper and, though the author believes them to be reasonable, are only a theoretical construct and have no binding nature for teaching in any way.

Now, with the investigation closed and a synthesis drawn, there are a number of further projects that would forward the process of finally providing a methodology

for assessing in-class participation. Firstly, a more in-depth version of the survey that considers the results of the first one and the ways in which it can be improved could deepen the understanding of teachers' demands, beliefs and practices concerning the assessment of in-class participation. It would especially be helpful to find a setup in which a much higher number of participants can be achieved, so that the collected data can undergo systematic bivariate and multivariate analysis. Another reasonable undertaking would be to collect different ways in which teachers assess in-class participation in a series of qualitative case studies. One could then use a framework such as the one developed here to examine those ways and find elements or maybe even complete methods which satisfy all the theoretical criteria applied. A third area of investigation would be how the assessment of in-class participation is being dealt with during teacher training, at the university and during the *Referendariat* and later training. It has been shown that only a minority of the participants of the survey had been introduced to ways of assessing in-class participation, although the tendency was for this number to increase among younger teachers. It could then also be examined how teachers could be introduced to the detailed knowledge that goes along with carrying out a method of assessment, especially in the area in which the study suggested that there may be shortcomings, such as student motivation issues and validity of assessment.

The author hopes that his investigation of the circumstances, requirements and practices concerning in-class participation will satisfy the readers' ideas about academic standards and might indeed serve the underlying purpose of helping do develop the respective methodology, as it seems to be needed. Clearly, only a small portion of the work necessary to do this has been done; hopefully, this thesis can serve as a contribution.

5 – Works Cited

V - Works Cited

Books and articles from books:

Ames, C. (1992). Achievement Goals and the Classroom Motivational Climate. In D. Schunk (Ed.), *Student Perceptions in the Classroom* (pp. 327-348). Hillsdale, NJ: Erlbaum.

Argyrous, G. (2000). *Statistics for Social & Health Research*. London: SAGE Publications.

Atteslander, P. (1975). *Methoden der empirischen Sozialforschung*. Berlin: Schmidt.

Bryman, A., & Cramer, D. (1990). *Quantitative Data Analysis for Social Scientists*. London: Routledge.

Dörnyei, Z. (2001). *Motivational Strategies in the Language Classroom*. Cambridge: Cambridge Universtiy Press.

Dörnyei, Z. (2003). *Questionnaires in Second Language Research*. Boca Raton: Routledge.

Ingenkamp, K, & Lissmann, U. (2008). *Lehrbuch der pädagogischen Diagnostik* (6th ed.). Weinheim: Beltz.

Jürgens, E. (2005). *Leistung und Beurteilen in der Schule* (5th ed.). Sankt Augustin: Academia Verlag.

Kirk, S. (2003). *Beurteilung mündlicher Leistungen*. Bad Heilbrunn: Verlag Julius Klinkhardt.

Mager, R. (1975). *Preparing Instructional Objectives* (2nd ed.). Belmont, CA: Lake Publishing Co.

Pintrich, P., & Schrauben, B. (1992). Students' Motivational Beliefs and Their Cognitive Engagement in Classroom Academic Tasks. In D. Schunk (Ed.), *Student Perceptions in the Classroom* (pp. 149-184). Hillsdale, NJ: Erlbaum.

Rapp, G. (1979). Formen und Funktionen der Lernerfolgskontrolle. In Maria Böhmer (Ed.), *Lernerfolgskontrolle*. Königsstein/TS: Scriptor.

Wigfield, A., & Harold, R.D. (1992). Teacher Beliefs and Children's Achievement Self-Perceptions: A Developmental Perspective. In D. Schunk (Ed.), *Student Perceptions in the Classroom* (pp. 95-122). Hillsdale, NJ: Erlbaum.

Young, R., & He, A.W. (1998). *Talking and Testing*. Amsterdam: Benjamins.

Articles from journals:

Anthony, E. M. (1963). Approach, method and technique. *English Language Teaching*, *17*, 63-67.

Oscarson, M. (1989). Self assessment of language proficiency: rationale and applications. *Language Testing*, *6*, 1-13.

Sources for Educational Legislature: (The name of the Department of Education is given differently in the different sources; it refers to the same department, however.)

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (2008a). *Arbeitsverhalten und Sozialverhalten in Zeugnissen*. Retrieved from http://www.schulministerium.nrw.de/BP/ImFokus/Handreichung_AS/RS_HandreichungArbeitsundSozialverhalten.pdf

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (2007a). *Ausbildungs- und Prüfungsordnung Sekundarstufe I – APO-S I*. Retrieved from http://www.schulministerium.nrw.de/BP/Schulrecht/APOen/APO_SI.pdf

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (2007b). *Kernlehrplan für das Gymnasium – Sekundarstufe I in Nordrhein-Westfalen. Englisch*. Heft 3417. Frechen: Ritterbach Verlag GmbH.

Ministerium für Schule und Weiterbildung, Wissenschaft und Forschung des Landes Nordrhein-Westfalen. (1999). *Richtlinien und Lehrpläne für die Sekundarstufe II – Gymnasium/Gesamtschule in Nordrhein-Westfalen. Englisch*. Heft 4704. Frechen: Ritterbach Verlag GmbH.

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (2009a). *Schulgesetz des Landes Nordrhein-Westfalen*. Retrieved from http://www.schulministerium.nrw.de/BP/Schulrecht/Gesetze/SchulG_Info/Schulgesetz.pdf

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (2008b). *Verordnung über den Bildungsgang und die Abiturprüfung in der gymnasialen Oberstufe (APO-GOSt)*. Retrieved from http://www.schulministerium.nrw.de/BP/Schulrecht/APOen/APOGOSt.pdf

Other internet sources:

Leiner, Dominik. (2009). *oFb - der onlineFragebogen → Fragebogen online erstellen, Befragung im Internet durchführen*. Retrieved from http://ofb.msd-media.de/index.php

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (2009b). *Schuldatenbank des Ministerium für Schule und Weiterbildung*. Retrieved from http://www.schulministerium.nrw.de/BP/_suche1/SchuleSuchen/index.html

Additional sources used as background for statistic analysis:

Georgii, H. (2007). *Einführung in die Wahrscheinlichkeitstheorie und Statistik* (3rd ed). Berlin: de Gruyter.

Kähler, W. M. (2008). *Statistische Datenanalyse.* (5th ed.). Wiesbaden: Vieweg.

Computer software used:

Noonan, J. (2006). SPSS für Windows (Version 15.0) [Computer software]. Chicago, IL, USA: SPSS Inc.

Somasegar, S. (2008). Microsoft Office Excel (Version 2007) [Computer software]. Redmond, WA, USA: Microsoft Corporation.

Somasegar, S. (2008). Microsoft Office Word (Version 2007) [Computer software]. Redmond, WA, USA: Microsoft Corporation.

Appendix A – oFb Homepage

Homepage of the web service *oFb – der onlineFragebogen* that contains information about the project and the services offered for academics.

Willkommen auf der oFb-Homepage!

oFb - der onlineFragebogen

Der onlineFragebogen bringt Ihre Online-Befragung schnell und zuverlässig ins Internet. → Details

Was kostet oFb?

Für wissenschaftliche Befragungen ist (und bleibt) der onlineFragebogen kostenlos.

Für Befragungen, die einem Unternehmen dienen, bitten wir freundlich um eine kleine Spende. Wenn Sie in einem privaten oder wissenschaftlichen Projekt mit oFb sehr zufrieden waren, freuen wir uns natürlich auch über eine Unterstützung für den onlineFragebogen. → Mithilfe & Spenden

Warum gerade oFb?

oFb wurde speziell für die wissenschaftliche Online-Befragung konzipiert und wird stetig für die tägliche Forschungspraxis weiterentwickelt. Die fast uneingeschränkte Flexibilität von oFb ist nach wie vor konkurrenzlos.

Aber Sie entscheiden: → Günstige Alternativen

Weitere Angebote

Ergänzend zum onlineFragebogen bieten wir professionelles Fragebogen-Review, die fachkundige Umsetzung von Fragebögen für das Internet, die Übernahme von Layouts in oFb sowie die Implementierung von Spezial-Lösungen im Fragebogen (z.B. Online-Auswertung, Reaktionszeitmessung, etc.). → Dienstleistungen

---

Was bietet oFb?

- Komfortables Erstellen eines Online-Fragebogens
- Äußerst flexibel für fast jede Herausforderung
- Grafische und numerische Skalen-Verankerung
- Einbindung von Grafiken, Bildern und HTML
- Randomisierung
- Datenübernahme nach SPSS (inkl. Labels)
- Rücklauf-Kontrolle in Echtzeit
- Zugriffskontrolle via Passwort, IP-Adresse oder durch personalisierte Links inkl. E-Mail-Versand
- oFb läuft direkt im Internet (keine Installation)
- Selbstverständlich werbefrei

Features für Anspruchsvolle und Profis

- Frei programmierbare Filter
- Frei programmierbares Layout
- Dynamische Inhalte
- Gleichverteilte Randomisierung (Urnen)
- Unterstützung von Panel-Befragungen
- Untersützung internationaler Schriftzeichen (Japanisch, Koreanisch, etc.)
- Valider HTML-Code

Sozialforschung im Internet → Links

- **Online Forschung**
  Umfangreiche Linkliste und einige Artikel rund um das Thema Online Forschung.
- **GIR-L**
  Die German Internet Research List ist eine Fach-Mailingliste für alle, die tiefer in das Thema Onlineforschung einsteigen wollen.
- **Deutsche Gesellschaft für Online-Forschung**
- **Web Survey Methodology Portal**
- Es gibt auch psychologische Experimente im Internet, z.B. im W-Lab, ein Psychologisches Labor im WWW, am Hanover College oder im Web-Labor für experimentelle Psychologie.

Login

Login:
Passwort: OK

→ Demo-Account
→ Fragebogen-Demo

Aktuell

22.08.2009
Doppelt so viele Teilnehmer - kostenlos
Das SoSci Panel, ein Pool von Interviewpartnern, ist im August auf über 20.000 Teilnehmer gewachsen. Wer eine qualitativ hochwertige Befragung ohne kommerziellen Hintergrund durchführt, kann mit dem SoSci Panel die eigene Stichprobe verdoppeln.
→ Infos und Bedingungen

Projektvorstellung

Im Feld seit 02.08.2009
persönliche Windenergie
Eine Befragung, die die Möglichkeiten einer individuellen und persönlichen Energiegewinnung durch Wind untersucht.

→ Mehr laufende Projekte

269 Projekte im Feld
57 Studien in Vorbereitung

Appendix B – Pre-Test for Study

The questionnaire that was used for carrying out the pre-test for the study.

Questions on Assessing "Mündliche Mitarbeit"

Please answer the following questions about your experiences with assessing your students oral participation during lessons:

1 Are you satisfied with your system of assessing a students' "Mündliche Mitarbeit"?
1) completely 2) mostly 3) partly 4) not at all

2 How many hours do you spend on assessing your students' "mündliche Mitarbeit" during the term each week?
1) 0-1 hours 2) 1-2 hours 3) 2-3 hours 4) 3 hours or more

How many hours do you spend on determining the students grades when you give the quater term grades?
1) 0-1 hours 2) 1-2 hours 3) 2-3 hours 4) 3 hours or more

3 Are you sometimes worried that you may not be assessing your students fairly?
1) never 2) sometimes 3) frequently

If not, what are the reasons?

- ○ It is not possible to pay enough attention to the individual students
- ○ The structure of school lessons makes it impossible
- ○ I do not have enough time to record details about every student
- ○ Other: _____

_____

4 Would you be interested in receiving information about a different method of assessing the "mündliche Mitarbeit" of your students?
1) very interested 2) interested 3) not interested

5 In your opinion, how important are the following features of a method of assessing the "mündliche Mitarbeit"?

transparent
1) very important 2) rather important 3) rather not important 4) not important

conforming to ministry regulations
1) very important 2) rather important 3) rather not important 4) not important

simple / easy to use
1) very important 2) rather important 3) rather not important 4) not important

being motivating for the teachers
1) very important 2) rather important 3) rather not important 4) not important

being motivating for the students
1) very important 2) rather important 3) rather not important 4) not important

fulfilling scientific criteria (objectivity, reliability)
1) very important 2) rather important 3) rather not important 4) not important

What other criteria do you find very important? _____ ____

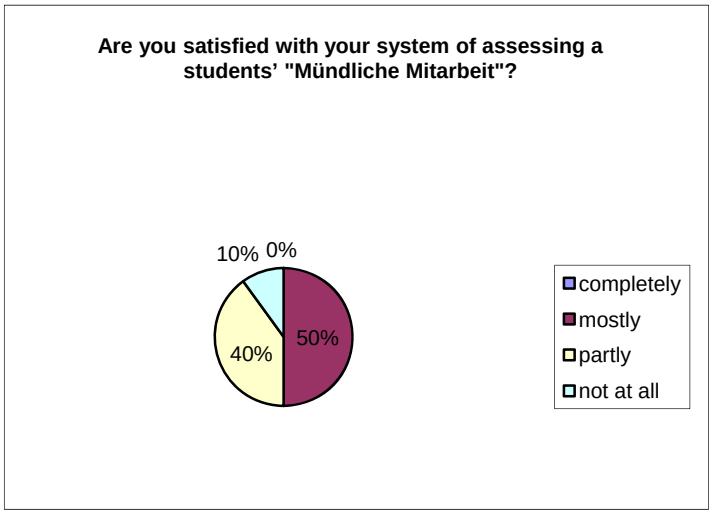
Thank you for your participation!

Appendix C – Results of Pre-Test

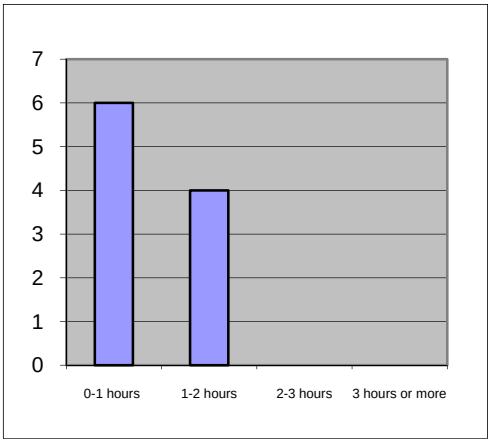
A visualization of the results of the Pre-Test.

Results of pre-test:

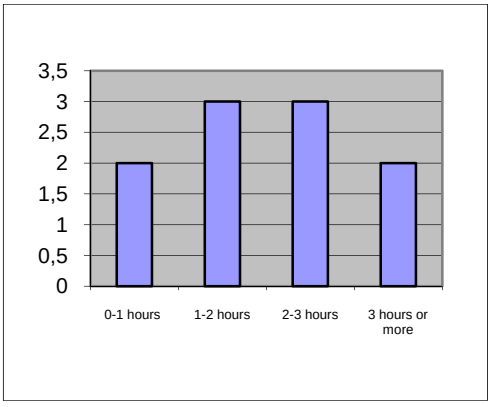
Are you satisfied with your system of assessing a students' "Mündliche Mitarbeit"?



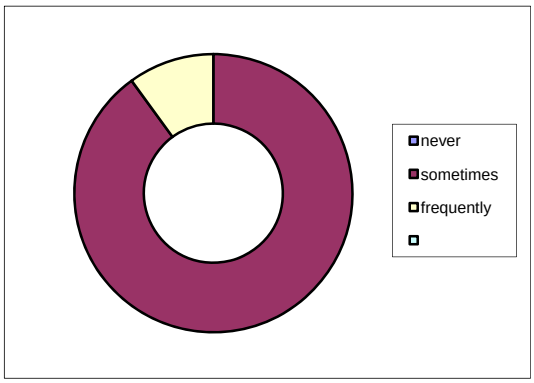
How many hours do you spend on assessing your students' "mündliche Mitarbeit" during the term each week?



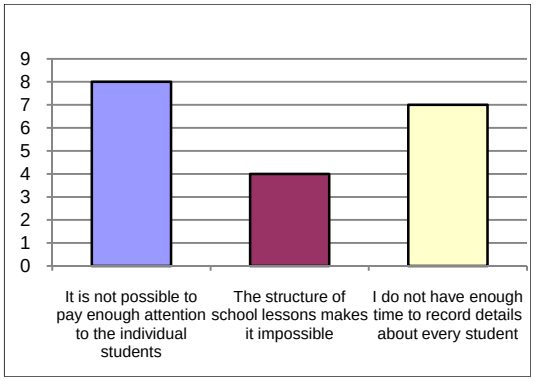
How many hours do you spend on determining the students' grades when you give the quarter term grades?



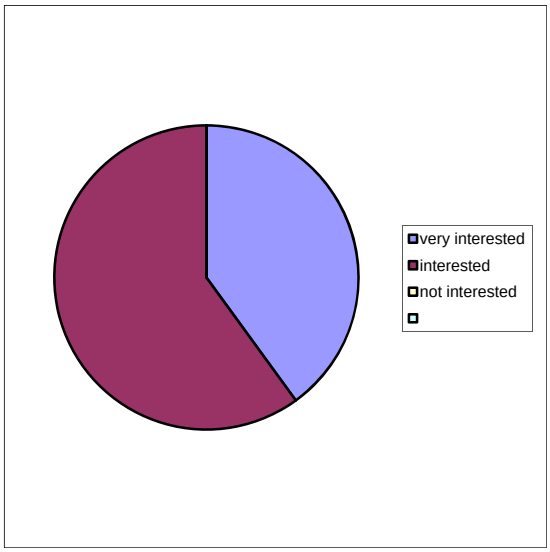
Are you sometimes worried that you may not be assessing your students fairly?



If not, what are the reasons?

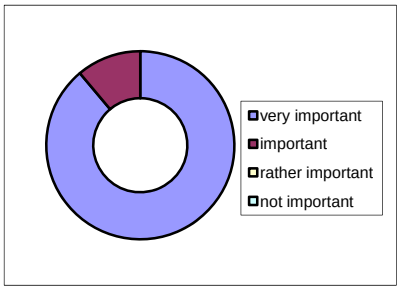


Would you be interested in receiving information about a different method of assessing the "mündliche Mitarbeit" of your students?

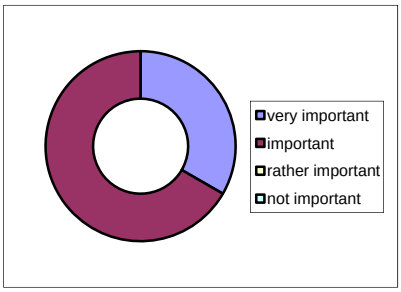


In your opinion, how important are the following features of a method of assessing the "mündliche Mitarbeit"?

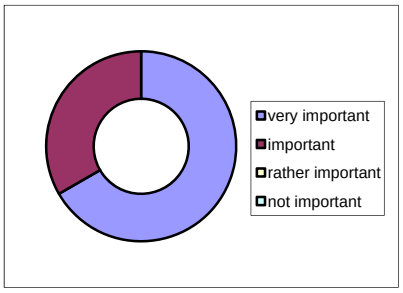
Transparent



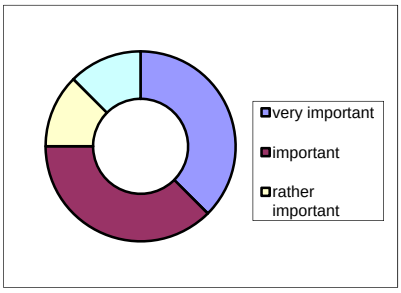
Conforming to ministry regulations



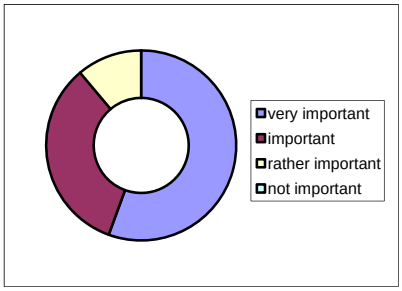
Simple / Easy to use



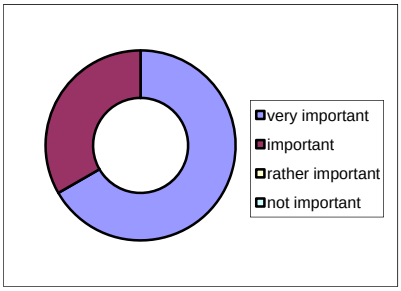
Being motivating for the teachers



Being motivating for the students



Fulfilling scientific criteria (objectivity, reliability)



Appendix D – Full Catalogue of Question Items

The full catalogue of questions that was created for the questionnaire and from which items were chosen for the final questionnaire.

Teachers on In-Class Assessment

Institutional Background:

1. Which regulations concerning the assessment of in-class participation in the North Rhine-Westphalian "Schulgesetze" do teachers know? Schulgesetz / Ausbildungs- und Prüfungsordnungen for the Sekundarstufe 1 / 2 / Lehrpläne for English for the Sekundarstufe 1 / 2)

2. How important do the teachers consider those regulations for their teaching practice?

3. How do teachers weigh the assessment area of "Sonstige Mitarbeit" (Mitarbeit im Unterricht, Hausaufgaben, Tests) in comparison to written tasks like exams when it comes to determining the final grade?

4. Do teachers put emphasis on different aspects of student participation in the Sekundarstufe 1 and 2?

5. In how far do the individual „Kopfnoten" give a reasonable account of the personal and social performance of students? Leistungsbereitschaft / Zuverlässigkeit/Sorgfalt / Selbstständigkeit / Verantwortungsbereitschaft / Konfliktverhalten / Kooperationsfähigkeit

Student Motivation:

6. How important for the teachers is it that standards for grading in-class participation are transparent to students?

7. Do teachers give students the opportunity to improve their grades for in-class participation if they are not satisfied with them? If yes, what do they look like?

8. How much time do teachers invest into making grading standards for in-class participation transparent to the students?

9. What are the circumstances in which teachers give students their final grades? In a private conversation / In class / In written form, to the individual student only / In written form, publicly

10. Do teachers give students a real say in what grades they get?

11. Do the teachers consider foreign language anxiety as a factor when grading in-class participation?

12. Do the teachers know possibilities of controlling how much of a burden students see in in-class participation grades? (open question)

Based on the teachers' experiences, …

13. How important are grades as a motivation to students?

14. Do students perceive grades for in-class participation to be a burden?

15. Do students perceive grades for in-class participation as a positive motivational factor?

16. Can the teachers' behavior influence whether students perceive grades for in-class participation as a burden

17. Can the teachers' behavior influence whether students perceive grades for in-class participation as motivation?

18. How important is it for the grading process that students practice self-assessment of their in-class participation grade?

19. Do the grades students get for in-class participation reinforce their individual performance levels (e.g. good students get good grades and are thus very motivated, bad students get bad grades and are poorly motivated)?

20. What do students believe that in-class participation in the foreign language classroom assesses? Choices: Language competency / Quality of the content of statements / Frequency of participation / Others

21. How much does foreign language anxiety influence students' in-class participation?

22. How big is the influence of the school situation on the in-class participation performance of students (awareness of social pressure)?

Teacher Motivation:

23. How important is it that teachers are convinced of the method of assessing in-class performance they use?

24. How important is it that a method of assessing in-class performance is simple to carry out?

25. How important is it that a method of assessing in-class performance can be carried out quickly?

26. How much latitude should a method of assessing in-class participation leave to the individual teacher (if more latitude means that the method will cause more work)?

27. How important is it to teachers that a method of assessing in-class participation can be integrated into their existing daily schedule?

Scientific Criteria:

28. Do teachers think that their method of grading in-class participation is based on a reasonable methodology?

29. Do teachers base their assessment of in-class participation on things they observe only a few times or only on things they notice very often?

30. Do teachers try to keep apart different dimensions of student behavior (like quantity / quality of contribution, concentration, cooperativeness)

31. Do teachers manage to keep apart different dimensions of student behavior (like quantity / quality of contribution, concentration, cooperativeness)

32. What sources did teacher use to acquire that methodology? Scientific texts / Practical teacher's handbooks / Other literature / Designed it by oneself

33. Do teachers think they miss a lot of student behavior in class?

34. Do teachers think that the grade students get for in-class participation would be very different if they had more time to observe?

35. How important is it that different teachers use the same grading standards when assessing in-class participation?

36. Which aspects of student behavior do teachers grade as part of in-class participation (excluding „Kopfnoten")? Choose one or more: Quality of oral participation / Frequency of oral participation / Social behavior in class / Regular presence in class / Doing homework / Quality of homework / Tests / Orderly and complete workbooks / Others

37. Which standards do teachers apply when assessing their students' performance?
Rate degree: Intraindividual (considering learning progress) / Interindividual (comparison to other students) / Criteria-based / Others

38. Are teachers using any of the following aids for assessing in-class participation?
List with grades for every lesson / List with grades that are taken regularly in other intervals / Observation sheet for individual students / List with symbolic remarks other than grades (e.g. +, 0, -) / Others

39. How much time do teachers spend on carrying out the assessment of students' in-class participation: Each lesson / Each schoolday / Each schoolweek / Each schoolmonth / Each quarter (on determining Quartalsnoten) / Each half term (on determining the overall grade)

40. How important is it that teachers try to be as objective as possible when assessing in-class participation?

41. Is the teacher's subjective perception when assessing student performance a problem?

42. How important is it that when students are graded, the general conditions have to be equivalent?

43. Do the teachers manage to keep general conditions in the same situation similar when grading students?

44. Is the so-called "self-fulfilling prophecy (the effect that teachers tend to pay more attention to student behavior that meets their expectations) a problem for teachers?

45. How important is it that one teacher grades performances of the same kind by the same standards at different points in time?

46. How important is it that a teacher tries to get several instances of a particular student's behaviour in a certain situation (e.g. when reading out a dialogue) before giving grades?

47. Is it a big problem if a lot of time passes between observing students and noting down the observation?

48. Is it important for assessing in-class participation that the different factors of students' oral statements are interrelated (e.g. language competency influences the quality of content)?

49. How important is it to give students exact and clear tasks and instructions when observing their in-class participation?

50. How important is variation of teaching method in order to give students different circumstances in which they can participate?

General:

51. How important is it to teachers to learn about alternative methods of assessing in-class participation?

52. Have teachers been trying to find out about alternative methods of grading in-class participation?

53. How satisfied are teachers with their method of assessing in-class participation?

54. Would teachers give grades for in-class participation if that was voluntary?

Demographics

55. Gender

56. Age

57. How many years has the participant worked as a teacher?

Appendix E – Hypotheses Tableau

The hypotheses tableau that was created to organize the study.

Hypotheses Tableau

	Question	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27
No.	**Hypothesis**																											
1.1	Teachers are aware of legal regulations concerning grading practice	X			X																							
1.2	Teachers comply with legal regulations concerning grading practice in their teaching		X	X																								
2.1	Teachers are aware of the nature of optimal grading conditions						X	X	X																			
2.2	Teachers control their teaching practice and behavior in class in order to provide optimal grading conditions					X				X																		
2.3	Teachers want a method for assessing in-class participation that is motivating for them as well										X	X																
3.1	Teachers consider objectivity issues in their grading practice												X	X		X		X										
3.2	Teachers consider reliability issues in their grading practice												X				X											
3.3	Teachers consider validity issues in their grading practice														X				X									
4.1	The teachers are committed to assessing in-class participation well																			X	X	X	X	X				
5.1	Demographics																								X	X	X	X

Appendix F – Table of Questions

An overview over the questions used in the final version of the questionnaire, encompassing its index number, the hypothesis a question relates to, the number it is assigned in the questionnaire and a short summary of the question.

Table of Questions

Question No.	Hypothesis No.	Questionnaire Item No.	Question (short version)
Q1	1.1	1	How important do you consider the following regulations for your teaching practice? Schulgesetz (“SchulG”) \ Ausbildungs- und Prüfungsordnung for the Sekundarstufe I (“APO_SI”) \ Ausbildungs- und Prüfungsordnung for the Sekundarstufe II (“APOGOSt”) \ Kernlehrplan Sekundarstufe I G8 Englisch \ Lehrplan für die Sekundarstufe II
Q2	1.2	2	In the legal regulations for public schools in NRW, the assessment area “Sonstige Mitarbeit” or “Sonstige Leistungen” refers to how students participate in class, how well and how regularly they do their homework and how well they perform at tests during the lessons. When giving students their final grade, how much do you weigh “Sonstige Mitarbeit” in comparison to written exams?
Q3	1.2	3	In how far do the aspects of student participation that you grade vary from Sekundarstufe I to Sekundarstufe II?
Q4	1.1	4	In how far do the individual „Kopfnoten“ give a reasonable account of the personal and social performance of students? Leistungsbereitschaft \ Zuverlässigkeit \ Sorgfalt \ Selbstständigkeit \ Verantwortungsbereitschaft \ Konfliktverhalten \ Kooperationsfähigkeit
Q5	2.2	5	What is the situation in which you give students their final grades? In a face to face-conversation \ In class \ In written form, to the individual student only \ In written form, publicly
Q6	2.1	6	How important are grades as a motivation to students?
Q7	2.1	7	Do your students use self-assessment methods? If yes, what influence does the students' assessment have on you grading of "Sonstige Mitarbeit"
Q8	2.1	8	Do the grades students get for “Sonstige Mitarbeit” reinforce their individual performance levels (e.g. good students get good grades and are thus very motivated, bad students get bad grades and are poorly motivated)?
Q9	2.2	9	Do you give students a real say in what grade they get for “Sonstige Mitarbeit”?
Q10	2.3	10	How important is it that you are convinced of the method of assessing “Sonstige Mitarbeit”-performance you use?
Q11	2.3	11	How important is it to you that a method of assessing “Sonstige Mitarbeit” can be integrated into your existing daily schedule?
Q12	3.1, 3.2	12	How often does a certain student behavior have to occur so that you consider it when assessing a student’s “Sonstige Mitarbeit”?
Q13	3.3	13	Do you manage to keep apart different dimensions of student behavior (such as quantity / quality of contribution, concentration, cooperativeness) when grading?

Question No.	Hypothesis No.	Questionnaire Item No.	Question (short version)
Q14	3.3	14-15	Which aspects of student behavior do you grade as part of “Sonstige Mitarbeit” (excluding „Kopfnoten“)? Choose one or more: Quality of oral participation \ Frequency of oral participation \ Social behavior in class \ Regular presence in class \ Doing homework \ Quality of homework \ Tests \ Orderly and complete workbooks \ Others
Q15	3.1	19-24	How much time do you spend on carrying out the assessment of students’ "Sonstige Mitarbeit": Each lesson / Each schoolday / Each schoolweek / Each schoolmonth / Each quarter (on determining Quartalsnoten) / Each half term (on determining the overall grade)
Q16	3.2	16	How important is it that when different students of the same class are graded, the general conditions (setting, preparation time, ect.) are equivalent?
Q17	3.1	17	Does the fact that the different dimensions of students’ oral language competency are connected (e.g. language competency influences the quality of content) have significance for you assessment of “Sonstige Mitarbeit”?
Q18	3.3	18	Do you agree with the following statement: "It is essential to fair grading that the teaching methods used in class must be varied"?
Q19	4.1	25	Do you follow a certain method for grading "Sonstige Mitarbeit"?
Q20	4.1	26	Have you been introduced to a certain method when grading "Sonstige Mitarbeit", either during your studies, during teacher training or during your professional career (Fortbildung)
Q21	4.1	27	Have you tried to find out about alternative methods of assessing “Sonstige Mitarbeit”?
Q22	4.1	28	How satisfied are you with your method of assessing “Sonstige Mitarbeit”?
Q23	4.1	30	Would you give grades for “Sonstige Mitarbeit” if that was voluntary?
Q24	5.1	31	Please specify your gender.
Q25	5.1	32	Please specify you age group.
Q26	5.1	33	For how many years have you been working as a teacher?
Q27	5.1	34	In which sections of the Gymnasium do you do most of your teaching?

Appendix G – Questionnaire

The final version of the questionnaire. It was published under the name *TOICA – Teachers on In-Class Assessment*.

Teachers' Views on In-Class Assessment

The purpose of this questionnaire is to examine the way teachers deal with the assessment of "Sonstige Mitarbeit". The term "Sonstige Mitarbeit refers to different aspects of how students participate in class. The aim of the questionnaire is to find out what strategies teachers adopt when grading "Sonstige Mitarbeit" and what they pay attention to when grading.
In order to ensure that the results are valid, the study will only focus on English teachers at a "Gymnasium" in North Rhine-Westphalia.

Of course, all information collected will be handled completely anonymously. The questionnaire contains 35 questions and takes about 20 minutes to fill in.

Many thanks for the time and effort you are investing by participating.

Next

1. How important do you consider the following regulations for your grading practice as a teacher?

	not very important			very important	cannot answer
Schulgesetz ("SchulG")	○	○	○	○ ○	○
Ausbildungs- und Prüfungsordnung für die Sekundarstufe I ("APO_SI")	○	○	○	○ ○	○
Ausbildungs- und Prüfungsordnung für die Sekundarstufe II ("APOGOSt")	○	○	○	○ ○	○
Kernlehrplan Sekundarstufe I G8 Englisch	○	○	○	○ ○	○
Lehrplan für die Sekundarstufe II	○	○	○	○ ○	○

2. In the legal regulations for public schools in NRW, the assessment area "Sonstige Mitarbeit" or "Sonstige Leistungen" refers to how students participate in class, how well and how regularly they do their homework and how well they perform at tests during the lessons. When giving students their final grade, how much do you weigh "Sonstige Mitarbeit" in comparison to written exams?

- ○ I tend to give more weight to „Sonstige Mitarbeit"
- ○ I tend to give equal weight to both aspects
- ○ I tend to give more weight to class tests and „Klausuren"
- ○ Other

3. In how far do the aspects of „Sonstige Leistungen" that you grade differ between Sekundarstufe I and Sekundarstufe II?

The curriculum for English in the Sekundarstufe I (G8) gives the following examples for „Sonstige Leistungen": "...die kontinuierliche Beobachtung der Leistungsentwicklung im Unterricht (verstehende Teilnahme am Unterrichtsgeschehen sowie kommunikatives Handeln und Sprachproduktion schriftlich wie vor allem mündlich), wobei individuelle Beiträge zum Unterrichtsgespräch sowie kooperative Leistungen im Rahmen von Team- und Gruppenarbeit zu beachten sind" (p. 48)

The curriculum for English in the Sekundarstufe II defines „Sonstige Mitarbeit" as follows: „Dazu gehören Beiträge zum Unterrichtsgespräch, die Leistungen in Hausaufgaben, Referaten, Protokollen, sonstigen Präsentationsleistungen, die Mitarbeit in Projekten und Arbeitsbeiträge". (p. 100)

Based on that distinction, in how far do the aspects of student participation that you grade differ between Sekundarstufe I and Sekundarstufe II?

							cannot answer
not at all	○	○	○	○	○	very much	○

4. In how far do the individual „Kopfnoten" give an accurate account of the personal and social performance of students?

	not accurate				very accurate
Leistungsbereitschaft	○	○	○	○	○
Zuverlässigkeit/Sorgfalt	○	○	○	○	○
Sozialverhalten	○	○	○	○	○

Next

5. What is the situation in which you give students their final grades?

- ○ In a private face-to-face conversation
- ○ In class
- ○ In written form, to the individual student only
- ○ In written form, publicly (e.g. putting up lists)
- ○ Other

6. How important are grades as a motivation to students?

							cannot answer
not important at all	○	○	○	○	○	very important	○

7. Do your students use self-assessment methods? If yes, what influence does the students' assessment have on your grading of „Sonstige Mitarbeit"?

							cannot answer
no influence at all	○	○	○	○	○	a lot of influence	○

8. Do the grades students get for "Sonstige Mitarbeit" reinforce their individual performance levels (e.g. good students get good grades and are thus very motivated, weak students get bad grades and are poorly motivated)?

							cannot answer
not at all	○	○	○	○	○	definitely	○

9. Do you give students a say in what grade they get for "Sonstige Mitarbeit"?

							cannot answer
not at all	○	○	○	○	○	always	○

10. How important is it to you that you are convinced of the method of assessing "Sonstige Mitarbeit" you use?

							cannot answer
not important at all	○	○	○	○	○	very important	○

11. How important is it to you that a method of assessing "Sonstige Mitarbeit" can be integrated into your existing daily schedule?

							cannot answer
not important at all	○	○	○	○	○	very important	○

Next

12. How often should a certain student behavior (e.g. helping other students as a positive aspect or disturbing class work as a negative aspect) have to occur so that you consider it when assessing a student's "Sonstige Mitarbeit"?

							cannot answer
only a few times	○	○	○	○	○	frequently	○

13. Do you manage to keep apart different dimensions of student behavior (such as quantity / quality of contribution, concentration, cooperativeness)?

							cannot answer
not at all	○	○	○	○	○	very much	○

14. Which aspects of student behavior do you grade as part of "Sonstige Mitarbeit" (excluding „Kopfnoten")? Choose one or more:

- ☐ Quality of oral participation
- ☐ Frequency of oral participation
- ☐ Social behavior in class
- ☐ Regular presence in class
- ☐ Doing homework
- ☐ Quality of homework
- ☐ Tests
- ☐ Orderly and complete workbooks
- ☐ Presentations
- ☐ Protocols
- ☐ Creativity
- ☐ Others

15. If you chose "Others" under the last question, please explain briefly what you mean.

16. Is it important that when different students of the same class are graded, the general learning conditions (such as setting, preparation time, ect. in which the assessment takes place) are equivalent?

							cannot answer
not important at all	○	○	○	○	○	very important	○

17. Does the fact that the different dimensions of students' oral language competency are connected (e.g. language competency influences the quality of content) have significance for your assessment of "Sonstige Mitarbeit"?

							cannot answer
not at all	○	○	○	○	○	definitely	○

18. Do you agree with the following statement: „It is essential to fair grading that the teaching methods used in class must be varied?"

							cannot answer
don't agree at all	○	○	○	○	○	fully agree	○

Next

19. How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) after each lesson?

- ◯ 0<2 minutes
- ◯ 2-4 minutes
- ◯ 4-6 minutes
- ◯ 6-10 minutes
- ◯ >10 minutes
- ◯ Other

20. How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) each school day?

- ◯ <5 minutes
- ◯ 5-10 minutes
- ◯ 10-20 minutes
- ◯ 20-30 minutes
- ◯ >30 minutes
- ◯ Other

21. How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) each school week?

- ◯ <15 minutes
- ◯ 15-30 minutes
- ◯ 30-45 minutes
- ◯ 45-60 minutes
- ◯ >60 minutes
- ◯ Other

22. How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) each school month?

- ◯ <1 hour
- ◯ 1-2 hours
- ◯ 2-4 hours
- ◯ 4-6 hours
- ◯ >6 hours
- ◯ Other

23. How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) each quarter (on determining Quartalsnoten only , not considering the time you spend on that regularly during the term)

- ◯ <15 minutes
- ◯ 15-30 minutes
- ◯ 30-60 minutes
- ◯ 60-90 minutes
- ◯ 90-120 minutes
- ◯ >120 minutes
- ◯ Other

24. How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) each half term (on determining the overall grade, not considering the time you spend on that regularly during the term)

- ◯ <15 minutes
- ◯ 15-30 minutes
- ◯ 30-60 minutes
- ◯ 60-90 minutes
- ◯ 90-120 minutes
- ◯ >120 minutes
- ◯ Other

Next

25. Do you follow a certain method for grading „Sonstige Mitarbeit“?

Yes ○ ○ No

26. Have you been introduced to a certain method when grading „Sonstige Mitarbeit“, either during your studies, during teacher training or during your professional career (Fortbildung)?

Yes ○ ○ No

27. Have you tried to find out about alternative methods of assessing “Sonstige Mitarbeit”?

							cannot answer
not at all	○	○	○	○	○	very often	○

28. How satisfied are you with your method of assessing “Sonstige Mitarbeit”?

							cannot answer
not satisfied at all	○	○	○	○	○	very satisfied	○

29. If you are not satisfied with your way of grading „Sonstige Mitarbeit“, what are the problems? You can choose multiple items.

- ☐ Unsystematic
- ☐ Not reliable enough
- ☐ Too complicated
- ☐ Too time-consuming
- ☐ Too irregular
- ☐ Not objective enough
- ☐ Not motivating to the students
- ☐ Not in perfect accordance with the legal regulations
- ☐ Other

30. Would you give grades for “Sonstige Mitarbeit” if that was voluntary?

definitely not ○ ○ ○ ○ ○ definitely

31. Please specify your gender.

- ○ female
- ○ male

32. Please specify you age group.

- ○ 25-29 years
- ○ 30-34 years
- ○ 35-39 years
- ○ 40-44 years
- ○ 45-49 years
- ○ 50-54 years
- ○ 55-59 years
- ○ 60-64 years
- ○ 65-69 years
- ○ >70 years

33. For how many years have you been working as a teacher?

- ○ 0-4 years
- ○ 5-9 years
- ○ 10-14 years
- ○ 15-19 years
- ○ 20-24 years
- ○ > 25 years

34. In which sections of the Gymnasium do you do most of your teaching?

- ○ More in Sekundarstufe I than in Sekundarstufe II
- ○ To the same degree in Sekundarstufe I and II
- ○ More in Sekundarstufe II than in Sekundarstufe I
- ○ Other

35. Are there any remarks on this questionnaire that you would like to make?

36. If you want to receive the results of the study as soon as they have been interpreted, please enter a valid e-mail address here. The e-mail address you enter cannot be linked to the answers you gave in the questionnaire in any way.

Next

Thank you for your participation!

Appendix H – Letter to Principals

A letter that was sent to the principals of a number of schools in North Rhine-Westphalia asking them to support the study by asking the English teachers at their school to participate.

JENS PETER FISCHER
Hauptstrasse 32
33790 Halle

Jens Peter Fischer, Hauptstrasse 32, 33790 Halle

Xxxxxxxx-Xxxx-Gymnasium
Xxxxxx Xxx 00
33333 Xxx Xxxxxxxx

Halle, 18. Februar 2009

An die Schulleitung

Bitte um Unterstützung bei Online-Umfrage für Masterarbeit

Sehr geehrte Damen und Herren,

ich bin Student an der Universität Bielefeld mit dem Studienziel Lehramt an Gymnasien und Gesamtschulen in den Fächern Englisch und Mathematik. Zurzeit schreibe ich meine Masterarbeit über Gütekriterien für die Bewertung der Sonstigen Mitarbeit im Unterricht. Im Rahmen dieser Arbeit führe ich eine Studie unter Englischlehrer/innen durch, mit der die Einstellungen der Lehrer/innen zu den einzelnen Gütekriterien erforscht werden sollen.

Um die Arbeit auf eine wissenschaftlich solide Grundlage zu stellen, benötige ich in etwa 100 Lehrer/innen, die an einer Umfrage teilnehmen. Die Teilnahme kann von einem beliebigen Ort aus über das Internet durchgeführt werden und nimmt ca. 10 Minuten in Anspruch.

Ich würde Sie gerne bitten, die beiliegenden Zettel mit Einladungen zur Teilnahme an der Studie den Englischlehrer/innen an Ihrer Schule zukommen zu lassen. Ihre Unterstützung würde mir sehr viel weiter helfen, da die erforderlichen 100 Teilnehmer erfahrungsgemäß nicht leicht zu finden sind. Ich bedanke mich schon im Voraus für Ihre Hilfe.

Mit freundlichen Grüßen,

Jens Peter Fischer

Appendix I – Flyer for English Teachers

A flyer that was passed on to English teachers via mail and internet communication, asking them to participate in the study.

English Teachers needed for Scientific Survey

I am a student of English at Bielefeld University studying to become a teacher at a Gymnasium. As part of my master's thesis, I am conducting an online survey about teachers' attitudes towards grading "Sonstige Mitarbeit".

It is rather difficult to find enough teachers that participate. If you are an English teacher at a Gymnasium, it would be a great help if you took part in the survey. It can be done online, from any place convenient to you. Simply type in the following URL and answer the questions:

http://ofb.msd-media.de/toica

The survey is completely anonymous and will only take about 10 minutes to complete.

Many thanks for the time and effort you are investing by participating.

Kind regards,

Jens Peter Fischer

Appendix J – Presentation of Raw Data

A table containing the raw data that was collected through the questionnaire.

	ID	SchulG	APOSI	APOSII	KernLPSI	LehrplanSII	DistSMKI	DiffSISII	KNLeist	KNZuv	KNSoz	SitFinGrad
1	76	1	5	5	5	4	2	4	4	4	2	3
2	77	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99
3	79	1	2	3	3	3	2	4	3	3	3	2
4	80	4	4	4	5	5	2	3	3	3	3	1
5	81	5	5	5	5	5	2	2	4	4	2	1
6	82	5	5	5	5	5	2	3	2	2	2	2
7	83	5	5	5	5	5	2	2	4	4	1	2
8	84	5	5	5	5	5	2	2	3	3	3	1
9	86	3	3	3	5	5	2	1	4	3	3	1
10	87	5	5	5	5	5	2	1	1	1	1	1
11	88	-1	-1	-1	4	4	2	2	2	2	1	1
12	89	-99	-99	-99	4	4	-99	3	4	4	1	2
13	90	2	3	3	5	5	2	2	4	4	3	1
14	91	5	5	5	5	5	3	-1	3	4	4	3
15	92	2	2	3	3	3	3	3	4	4	2	1
16	93	1	4	4	4	3	2	4	3	4	3	2
17	94	-1	2	3	5	5	2	3	4	3	2	1
18	95	2	4	4	3	3	3	4	4	4	2	1
19	96	2	4	4	4	4	2	3	1	1	1	1
20	97	3	4	4	4	4	1	-1	1	1	1	2
21	98	5	5	5	5	5	1	3	1	1	1	1
22	99	3	5	5	5	5	2	3	4	4	2	1
23	100	4	4	4	4	4	2	3	5	5	2	1
24	110	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	3
25	115	5	5	5	5	5	2	3	3	3	3	2
26	116	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	3
27	121	2	2	2	4	2	2	3	4	-99	2	1
28	123	2	2	2	1	1	2	4	4	4	4	3
29	125	3	3	4	4	4	2	4	3	3	3	3
30	127	3	2	2	-1	3	1	2	3	2	4	3

	GradesMoti	ImpSelfAss	IndivPerfLe	StudGrade	ConvMetho	IntrMethod	AssessOfte	DiffDimApa	WhatSL_Q	WhatSL_F	WhatSL_S	WhatSL_R
1	4	2	3	.	4	5	4	4	2	2	1	1
2	-99	-99	-99	-99	-99	-99	-99	-99	1	1	1	1
3	4	2	4	5	5	5	4	4	2	2	1	1
4	5	-1	3	1	5	5	3	4	2	2	1	1
5	2	2	3	2	1	-1	-1	4	2	2	1	2
6	4	2	2	4	5	2	5	5	2	2	2	2
7	3	3	4	3	5	4	2	5	2	2	2	1
8	4	2	2	1	5	5	2	5	2	2	1	1
9	4	3	4	3	4	5	3	3	2	2	1	1
10	5	1	3	4	5	-1	4	3	2	2	1	1
11	4	-1	-1	1	4	4	2	2	2	2	2	1
12	5	3	3	3	4	5	-1	3	2	2	1	1
13	4	4	3	5	4	5	4	4	2	2	2	1
14	4	-1	4	3	5	5	3	3	2	2	1	1
15	4	-1	3	2	4	4	3	4	2	2	1	2
16	4	2	4	1	5	4	5	2	2	2	1	1
17	4	3	3	4	5	4	4	5	2	2	1	2
18	4	1	3	2	3	4	4	2	2	2	1	1
19	4	3	5	2	4	2	4	2	2	2	1	1
20	3	3	3	5	3	4	3	3	2	2	2	2
21	5	-1	3	4	5	5	4	4	2	2	1	1
22	4	4	3	5	5	4	4	4	2	2	2	2
23	5	3	2	4	4	5	5	4	2	2	1	1
24	5	-1	3	1	4	4	-1	5	2	1	1	1
25	5	1	1	5	5	5	5	5	2	2	1	1
26	5	4	5	5	3	-1	-1	5	2	2	1	2
27	3	3	4	1	2	3	4	5	2	2	2	2
28	5	2	4	1	5	5	4	4	2	2	1	2
29	4	4	3	3	4	5	3	4	2	2	1	2
30	3	2	5	2	5	3	3	4	2	2	2	1

	WhatSL_D	WhatSL_Q	WhatSL_T	WhatSL_O	WhatSL_Pr	WhatSL_Pr	WhatSL_C	WhatSL_O	GenCondE	SFPProble	VaryMeth	TFA01
1	2	2	2	2	2	1	2	1	4	4	3	2
2	1	1	1	1	1	1	1	1	-99	-99	-99	-99
3	2	2	2	1	2	1	2	1	5	5	4	2
4	2	2	2	1	2	1	1	1	5	4	5	3
5	2	2	2	1	2	1	2	1	5	4	5	5
6	2	2	1	1	2	2	1	1	3	1	4	1
7	2	2	1	1	2	1	2	2	-1	4	5	4
8	2	2	2	2	2	1	1	1	5	1	2	2
9	2	2	2	2	2	1	2	1	3	4	4	2
10	2	2	2	2	1	1	1	1	5	4	5	1
11	2	2	2	2	2	1	1	1	2	-1	3	2
12	2	2	2	2	2	2	2	1	-1	5	5	-99
13	2	2	2	2	2	2	2	1	3	4	2	2
14	2	1	2	1	2	2	1	1	-1	-1	3	1
15	2	2	2	2	2	1	2	1	4	4	5	4
16	2	2	2	2	2	2	1	1	-1	3	5	1
17	2	2	2	2	2	2	2	1	5	5	5	1
18	2	2	2	2	2	1	1	1	5	2	4	2
19	2	2	2	1	2	1	1	1	4	4	5	1
20	2	2	2	2	2	2	2	2	-1	4	5	2
21	2	2	2	2	2	2	2	1	5	5	5	3
22	2	2	2	1	2	1	1	1	5	4	5	4
23	2	2	2	1	2	1	2	1	1	4	5	3
24	1	1	1	1	1	1	2	2	5	5	4	1
25	2	2	2	2	2	1	2	1	5	2	5	1
26	2	2	2	2	2	1	2	1	3	-1	5	3
27	2	2	2	2	1	1	1	1	3	3	3	1
28	2	2	2	2	2	2	2	1	4	5	5	3
29	2	2	2	2	2	2	1	1	3	4	4	2
30	2	2	1	1	2	1	2	2	5	4	5	3

	TFA02	TFA03	TFA04	TFA05	TFA06	Method	IntroToMet	TriedAltMet	SatisMetho	GradesIfVol	MC_UN	MC_NRE
1	2	2	2	4	4	1	1	3	3	5	2	2
2	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	1	1
3	2	3	3	4	6	1	1	4	3	5	1	2
4	3	4	5	3	3	1	2	2	4	1	1	1
5	5	5	5	6	6	1	2	5	4	5	1	1
6	2	2	2	5	6	1	1	3	5	5	1	1
7	3	2	-99	2	3	1	1	4	5	5	1	1
8	3	4	5	6	6	1	1	4	5	5	1	1
9	2	2	4	4	4	1	2	3	4	5	1	2
10	2	3	3	2	3	1	2	1	4	4	1	1
11	3	3	-99	4	-99	1	1	2	2	3	1	2
12	6	6	-99	5	5	1	1	2	3	3	1	1
13	1	1	1	3	4	1	2	2	3	5	1	2
14	1	1	1	3	3	1	2	3	4	4	2	1
15	3	3	3	2	2	1	2	1	3	4	1	2
16	1	1	1	2	3	1	2	2	4	5	1	2
17	3	5	5	6	6	1	1	4	4	4	1	1
18	1	1	1	5	5	1	1	2	2	4	1	2
19	2	2	2	2	3	2	2	4	2	3	2	1
20	4	-99	-99	-99	-99	-99	-99	-99	-99	-99	1	1
21	3	4	4	4	5	2	2	2	3	5	1	2
22	3	4	3	4	4	1	1	2	4	5	1	1
23	3	5	3	6	6	2	2	3	3	5	2	2
24	1	2	2	7	1	1	2	5	4	4	1	1
25	6	6	6	3	3	1	2	1	3	4	1	1
26	5	-99	3	7	7	1	1	2	4	5	1	1
27	1	1	1	3	3	2	2	3	3	2	2	2
28	3	4	3	3	3	1	1	5	4	4	1	1
29	3	4	4	4	4	1	1	4	4	4	1	1
30	4	5	4	6	6	2	2	3	3	5	2	1

	MC_TC	MC_TTC	MC_TI	MC_NOE	MC_NMTS	MC_NPCL	MC_OT	Gender	Age	YearsTeac	SekIoII
1	2	2	2	1	1	1	1	2	2	2	3
2	1	1	1	1	1	1	1	-99	-99	-99	-99
3	1	1	1	2	2	1	1	1	5	4	2
4	1	2	2	1	1	1	1	2	3	1	1
5	2	2	1	1	1	1	1	1	3	3	2
6	1	1	1	1	1	1	1	2	2	2	3
7	1	1	1	1	1	1	1	2	7	6	3
8	1	1	1	1	1	1	1	1	3	3	3
9	1	1	1	2	1	1	1	2	1	1	2
10	1	1	1	1	1	1	1	2	3	3	1
11	1	1	1	2	1	1	1	1	1	1	2
12	1	1	1	1	1	1	1	2	4	4	2
13	1	1	1	2	1	1	1	1	1	1	2
14	2	1	2	1	1	2	1	1	2	1	3
15	1	1	1	2	1	1	1	1	3	2	1
16	1	1	2	2	1	1	1	2	6	6	2
17	1	1	1	1	1	1	1	2	4	3	3
18	1	1	2	2	1	1	1	1	1	1	1
19	1	1	2	1	1	1	1	1	2	2	1
20	1	1	1	1	1	1	1	-99	-99	-99	-99
21	1	1	1	1	1	1	1	1	1	1	1
22	1	1	1	1	1	1	1	1	1	1	2
23	1	2	1	1	1	1	1	1	5	5	3
24	1	1	1	1	2	1	1	2	6	5	2
25	1	1	1	2	2	1	1	1	6	6	3
26	1	1	1	1	1	1	1	2	5	4	3
27	1	1	2	2	1	1	1	1	2	1	1
28	1	1	1	1	1	1	1	2	4	3	2
29	1	1	1	1	1	1	1	1	2	2	2
30	1	2	2	1	1	1	1	1	3	2	3

	ID	SchulG	APOSI	APOSII	KernLPSI	LehrplanSII	DistSMKI	DiffSISII	KNLeist	KNZuv	KNSoz	SitFinGrad
31	132	-1	-1	-1	5	5	2	4	2	2	3	1
32	134	3	-99	3	-99	-99	3	-1	4	4	2	1
33	137	3	4	4	5	5	2	4	3	3	1	1
34	140	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	1
35	144	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	1
36	155	1	2	2	4	3	2	3	1	1	1	2
37	156	4	5	5	5	4	2	4	1	1	1	5
38	159	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	1
39	160	1	3	3	1	1	2	1	1	1	1	2
40	162	3	3	3	3	3	2	3	1	1	1	2
41	163	3	2	2	5	5	2	3	2	2	2	1
42	167	-1	3	5	3	5	2	4	2	2	2	1
43	168	5	5	5	5	5	2	2	5	4	1	1
44	169	1	1	2	1	3	3	4	1	3	2	2
45	172	5	5	5	4	5	2	2	4	4	4	2
46	174	4	5	5	5	5	2	4	3	3	2	1
47	175	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	1
48	176	-99	-99	-99	-99	-99	-99	2	4	4	5	1
49	177	1	1	1	4	4	2	2	4	5	1	1
50	178	-99	-99	-99	-99	-99	-99	3	3	3	2	2
51	182	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	1
52	184	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	1
53	187	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99
54	200	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99
55	202	3	2	2	4	4	2	2	3	3	3	1
56	203	5	2	3	3	5	2	2	4	4	2	1
57	204	2	3	3	4	4	2	4	4	4	2	1
58	209	2	4	4	3	3	2	4	2	2	2	1
59	210	-1	-1	-1	-1	-1	3	-1	1	1	1	1
60	229	4	-1	-1	4	4	2	3	4	3	4	1

	GradesMoti	ImpSelfAss	IndivPerfLe	StudGrade	ConvMetho	IntrMethod	AssessOfte	DiffDimApa	WhatSL_Q	WhatSL_F	WhatSL_S	WhatSL_R
31	5	-99	-1	4	5	5	4	2	2	2	2	1
32	4	4	4	4	5	5	4	4	2	2	1	2
33	3	2	-1	2	4	4	2	3	2	2	1	2
34	4	4	2	5	5	5	4	4	2	2	2	2
35	5	2	2	1	4	4	-1	4	2	2	1	2
36	4	1	3	1	3	4	3	4	2	2	1	1
37	3	2	-1	5	5	5	-1	3	2	2	1	2
38	4	1	5	5	5	5	4	3	2	2	2	2
39	5	2	3	2	4	1	3	3	2	2	1	1
40	4	4	1	5	5	5	2	5	2	2	2	1
41	4	3	2	4	5	4	4	4	2	2	1	1
42	4	4	3	4	4	4	5	3	2	2	2	2
43	4	-1	3	1	5	5	4	4	2	2	1	2
44	3	-1	4	2	5	4	4	5	2	2	1	1
45	4	4	4	5	4	3	3	4	2	2	1	1
46	5	3	4	5	5	5	4	4	2	2	2	2
47	4	3	-1	3	4	4	4	4	2	2	1	2
48	5	-1	3	4	3	3	4	4	2	2	1	1
49	4	3	4	4	5	2	4	5	2	2	1	2
50	5	-1	2	3	-1	5	4	4	2	2	1	1
51	5	2	4	2	5	5	4	3	2	2	2	1
52	4	2	2	1	4	5	4	4	2	2	1	2
53	-99	-99	-99	-99	-99	-99	-99	-99	1	1	1	1
54	-99	-99	-99	-99	-99	-99	-99	-99	1	1	1	1
55	4	2	3	3	4	2	2	4	2	2	2	2
56	3	3	2	5	-1	5	-1	4	2	2	1	1
57	4	2	3	1	4	5	-1	4	2	2	1	1
58	4	-1	4	4	4	4	4	5	2	2	1	1
59	5	1	5	4	2	1	1	4	2	2	1	1
60	5	3	2	5	5	4	4	4	2	2	1	1

	WhatSL_D	WhatSL_Q	WhatSL_T	WhatSL_O	WhatSL_Pr	WhatSL_Pr	WhatSL_C	WhatSL_O	GenCondE	SFPProble	VaryMeth	TFA01
31	2	1	2	1	2	1	1	1	-1	4	5	2
32	2	2	2	1	1	1	1	1	-1	3	4	2
33	2	2	2	2	2	1	1	1	-1	-1	3	1
34	1	2	1	1	1	1	2	1	4	4	5	3
35	2	2	2	2	2	2	2	2	4	5	5	6
36	1	2	2	1	2	1	2	1	2	3	3	2
37	2	2	2	1	2	2	1	1	-1	4	5	5
38	2	1	2	1	2	1	2	1	5	2	5	1
39	2	2	2	1	2	1	1	1	-1	1	4	1
40	1	2	1	2	2	1	2	2	3	5	5	2
41	2	2	2	2	2	1	2	1	5	4	5	4
42	2	2	2	1	2	1	2	1	4	5	4	2
43	2	2	2	1	2	1	2	1	4	4	5	2
44	2	2	2	2	2	2	2	1	3	5	5	1
45	1	2	1	2	2	1	2	1	3	4	4	2
46	2	2	2	2	2	2	2	2	-1	5	5	3
47	2	2	2	1	2	1	1	1	4	4	1	1
48	2	1	2	2	2	1	1	1	5	4	5	1
49	2	2	1	2	2	1	2	2	5	3	5	1
50	2	2	2	1	2	1	2	1	5	-1	4	4
51	2	2	2	1	2	1	2	1	5	-1	5	6
52	1	2	2	1	2	1	2	1	2	2	4	6
53	1	1	1	1	1	1	1	1	-99	-99	-99	-99
54	1	1	1	1	1	1	1	1	-99	-99	-99	-99
55	2	2	1	1	2	1	2	1	3	4	4	1
56	2	2	2	2	2	2	2	1	4	5	5	2
57	2	2	2	1	2	1	1	2	3	4	4	2
58	2	2	2	1	2	1	2	1	5	5	4	-99
59	1	1	2	1	2	1	2	2	4	5	1	1
60	1	2	2	2	2	1	1	2	-1	5	5	2

	TFA02	TFA03	TFA04	TFA05	TFA06	Method	IntroToMet	TriedAltMet	SatisMetho	GradesIfVol	MC_UN	MC_NRE
31	3	4	4	5	5	1	1	2	4	5	1	1
32	3	4	3	5	7	1	1	4	5	4	1	1
33	2	3	-99	-99	-99	2	2	3	4	5	1	1
34	4	5	4	4	4	1	1	4	4	4	1	1
35	2	5	5	7	5	1	2	5	3	5	1	1
36	3	3	3	4	5	1	2	3	3	5	1	2
37	5	5	5	6	6	1	2	3	5	5	1	1
38	2	4	3	3	6	2	2	2	1	5	2	2
39	2	2	2	4	5	2	2	1	4	5	1	1
40	6	6	6	6	6	1	1	4	4	5	1	1
41	4	4	3	2	2	1	2	3	4	5	1	1
42	1	2	1	2	2	1	2	1	3	5	2	1
43	2	3	3	3	5	1	2	2	5	4	1	1
44	1	1	1	6	6	2	1	1	3	5	2	1
45	2	2	3	3	3	2	2	3	4	5	1	1
46	3	4	2	3	5	2	2	3	3	5	1	2
47	1	2	2	4	5	2	2	3	3	4	1	2
48	2	2	2	-99	-99	1	2	3	4	5	1	2
49	2	2	2	4	4	1	1	1	5	5	1	1
50	3	-99	-99	-99	4	1	2	2	2	5	1	2
51	6	6	6	6	6	2	1	2	2	5	1	1
52	2	2	1	3	3	1	1	2	4	4	1	1
53	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	1	1
54	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	1	1
55	2	2	2	3	3	2	2	2	5	4	1	1
56	3	5	4	5	6	1	1	4	4	5	1	1
57	2	2	2	3	4	1	1	2	5	5	1	1
58	2	2	2	2	2	1	2	1	4	5	1	1
59	6	6	6	4	5	1	2	2	1	1	2	2
60	2	2	2	3	3	1	2	3	4	5	1	1

	MC_TC	MC_TTC	MC_TI	MC_NOE	MC_NMTS	MC_NPCL	MC_OT	Gender	Age	YearsTeac	SekIoII
31	1	1	1	1	1	1	1	1	1	1	1
32	1	1	1	1	1	1	1	1	6	6	4
33	1	1	2	1	1	1	1	1	2	1	1
34	1	1	1	1	1	1	1	1	3	3	3
35	1	2	2	2	1	1	1	1	3	1	3
36	1	1	1	1	2	2	1	2	6	6	2
37	1	1	1	1	1	1	1	2	4	3	2
38	1	2	2	2	1	2	1	2	4	2	3
39	1	1	1	1	1	1	1	2	7	6	2
40	1	1	1	1	1	1	2	1	4	4	4
41	1	1	1	1	1	1	1	1	4	2	2
42	1	1	1	2	1	1	1	1	6	6	3
43	1	1	1	1	1	1	1	1	3	2	2
44	1	1	1	1	1	1	2	1	6	5	3
45	1	1	1	1	1	1	1	2	6	5	3
46	1	2	1	2	1	1	1	1	3	1	1
47	1	1	2	2	1	1	1	1	2	1	2
48	1	1	1	1	2	1	1	1	6	6	1
49	1	1	1	1	1	1	1	1	5	3	3
50	1	1	2	1	2	1	1	1	6	6	2
51	1	1	2	2	1	1	1	1	2	1	2
52	1	1	1	1	1	1	1	2	8	6	2
53	1	1	1	1	1	1	1	-99	-99	-99	-99
54	1	1	1	1	1	1	1	-99	-99	-99	-99
55	1	1	1	1	1	1	1	2	5	3	3
56	1	1	1	1	1	1	1	2	3	2	2
57	1	1	1	1	1	1	1	1	3	2	2
58	1	1	1	1	1	1	1	2	4	3	3
59	2	2	2	2	1	2	1	2	5	5	2
60	1	1	1	1	1	1	1	1	1	1	3

Appendix K – Summary of Raw Data

A summary of the data collected during the survey generated by the software *SPSS 15*.

Summary of Raw Data

1 How important do you consider the following regulations for your grading practice as a teacher? Schulgesetz ("SchulG"):

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	cannot answer	5	8,3	10,9	10,9
	1 not very important	7	11,7	15,2	26,1
	2	8	13,3	17,4	43,5
	3	10	16,7	21,7	65,2
	4	5	8,3	10,9	76,1
	5 very important	11	18,3	23,9	100,0
	Total	46	76,7	100,0	
Missing	-99	14	23,3		
Total		60	100,0		

1 How important do you consider the following regulations for your grading practice as a teacher? Ausbildungs- und Prüfungsordnung für die Sekundarstufe I ("APO_SI"):

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	cannot answer	4	6,7	8,9	8,9
	1 not very important	2	3,3	4,4	13,3
	2	10	16,7	22,2	35,6
	3	7	11,7	15,6	51,1
	4	8	13,3	17,8	68,9
	5 very important	14	23,3	31,1	100,0
	Total	45	75,0	100,0	
Missing	-99	15	25,0		
Total		60	100,0		

1 How important do you consider the following regulations for your grading practice as a teacher? Ausbildungs- und Prüfungsordnung für die Sekundarstufe II ("APOGOSt"):

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	cannot answer	4	6,7	8,7	8,7
	1 not very important	1	1,7	2,2	10,9
	2	7	11,7	15,2	26,1
	3	10	16,7	21,7	47,8
	4	9	15,0	19,6	67,4
	5 very important	15	25,0	32,6	100,0
	Total	46	76,7	100,0	
Missing	-99	14	23,3		
Total		60	100,0		

1 How important do you consider the following regulations for your grading practice as a teacher? Kernlehrplan Sekundarstufe I G8 Englisch:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	cannot answer	2	3,3	4,3	4,3
	1 not very important	3	5,0	6,5	10,9
	3	7	11,7	15,2	26,1
	4	14	23,3	30,4	56,5
	5 very important	20	33,3	43,5	100,0
	Total	46	76,7	100,0	
Missing	-99	14	23,3		
Total		60	100,0		

1 How important do you consider the following regulations for your grading practice as a teacher? Lehrplan für die Sekundarstufe II:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	cannot answer	1	1,7	2,2	2,2
	1 not very important	2	3,3	4,3	6,5
	2	1	1,7	2,2	8,7
	3	9	15,0	19,6	28,3
	4	12	20,0	26,1	54,3
	5 very important	21	35,0	45,7	100,0
	Total	46	76,7	100,0	
Missing	-99	14	23,3		
Total		60	100,0		

2 When giving students their final grade, how much do you weigh "Sonstige Mitarbeit" in comparison to written exams?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	I tend to give more weight to „Sonstige Mitarbeit"	3	5,0	6,5	6,5
	I tend to give equal weight to both aspects	37	61,7	80,4	87,0
	I tend to give more weight to class tests and „Klausuren"	6	10,0	13,0	100,0
	Total	46	76,7	100,0	
Missing	-99	14	23,3		
Total		60	100,0		

3 In how far do the aspects of „Sonstige Leistungen“ that you grade differ between Sekundarstufe I and Sekundarstufe II?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	cannot answer	4	6,7	8,2	8,2
	1 not at all	3	5,0	6,1	14,3
	2	12	20,0	24,5	38,8
	3	16	26,7	32,7	71,4
	4	14	23,3	28,6	100,0
	Total	49	81,7	100,0	
Missing	-99	11	18,3		
Total		60	100,0		

4 In how far do the individual „Kopfnoten“ give an accurate account of the personal and social performance of students? Leistungsbereitschaft:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1 not accurate	10	16,7	20,4	20,4
	2	6	10,0	12,2	32,7
	3	12	20,0	24,5	57,1
	4	19	31,7	38,8	95,9
	5 very accurate	2	3,3	4,1	100,0
	Total	49	81,7	100,0	
Missing	-99	11	18,3		
Total		60	100,0		

4 In how far do the individual „Kopfnoten“ give an accurate account of the personal and social performance of students? Zuverlässigkeit/Sorgfalt:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1 not accurate	9	15,0	18,8	18,8
	2	7	11,7	14,6	33,3
	3	13	21,7	27,1	60,4
	4	17	28,3	35,4	95,8
	5 very accurate	2	3,3	4,2	100,0
	Total	48	80,0	100,0	
Missing	-99	12	20,0		
Total		60	100,0		

4 In how far do the individual „Kopfnoten“ give an accurate account of the personal and social performance of students? Sozialverhalten:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1 not accurate	15	25,0	30,6	30,6
	2	18	30,0	36,7	67,3
	3	10	16,7	20,4	87,8
	4	5	8,3	10,2	98,0
	5 very accurate	1	1,7	2,0	100,0
	Total	49	81,7	100,0	
Missing	-99	11	18,3		
Total		60	100,0		

5 What is the situation in which you give students their final grades?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	In a private face-to-face conversation	36	60,0	63,2	63,2
	In class	13	21,7	22,8	86,0
	In written form, to the individual student only	7	11,7	12,3	98,2
	Other	1	1,7	1,8	100,0
	Total	57	95,0	100,0	
Missing	-99	3	5,0		
Total		60	100,0		

(„In written form, publicly“ was not selected by any of the participants)

6 How important are grades as a motivation to students?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	2	1	1,7	1,8	1,8
	3	8	13,3	14,0	15,8
	4	30	50,0	52,6	68,4
	5 very important	18	30,0	31,6	100,0
	Total	57	95,0	100,0	
Missing	-99	3	5,0		
Total		60	100,0		

(“1 – not important at all” was not selected by any of the participants)

7 Do your students use self-assessment methods? If yes, what influence does the students' assessment have on your grading of „Sonstige Mitarbeit“?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	cannot answer	11	18,3	19,6	19,6
	1 no influence at all	6	10,0	10,7	30,4
	2	16	26,7	28,6	58,9
	3	14	23,3	25,0	83,9
	4	9	15,0	16,1	100,0
	Total	56	93,3	100,0	
Missing	-99	4	6,7		
Total		60	100,0		

(“5 – a lot of influence” was not selected by any of the participants)

8 Do the grades students get for “Sonstige Mitarbeit” reinforce their individual performance levels?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	cannot answer	5	8,3	8,8	8,8
	1 not at all	2	3,3	3,5	12,3
	2	10	16,7	17,5	29,8
	3	21	35,0	36,8	66,7
	4	14	23,3	24,6	91,2
	5 definitely	5	8,3	8,8	100,0
	Total	57	95,0	100,0	
Missing	-99	3	5,0		
Total		60	100,0		

9 Do you give students a say in what grade they get for “Sonstige Mitarbeit”?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1 not at all	12	20,0	21,4	21,4
	2	9	15,0	16,1	37,5
	3	8	13,3	14,3	51,8
	4	13	21,7	23,2	75,0
	5 always	14	23,3	25,0	100,0
	Total	56	93,3	100,0	
Missing	-99	3	5,0		
	System	1	1,7		
	Total	4	6,7		
Total		60	100,0		

10 How important is it to you that you are convinced of the method of assessing "Sonstige Mitarbeit" you use?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	cannot answer	2	3,3	3,5	3,5
	1 not important at all	1	1,7	1,8	5,3
	2	2	3,3	3,5	8,8
	3	5	8,3	8,8	17,5
	4	20	33,3	35,1	52,6
	5 very important	27	45,0	47,4	100,0
	Total	57	95,0	100,0	
Missing	-99	3	5,0		
Total		60	100,0		

11 How important is it to you that a method of assessing "Sonstige Mitarbeit" can be integrated into your existing daily schedule?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	cannot answer	3	5,0	5,3	5,3
	1 not important at all	2	3,3	3,5	8,8
	2	4	6,7	7,0	15,8
	3	4	6,7	7,0	22,8
	4	18	30,0	31,6	54,4
	5 very important	26	43,3	45,6	100,0
	Total	57	95,0	100,0	
Missing	-99	3	5,0		
Total		60	100,0		

12 How often should a certain student behavior have to occur so that you consider it when assessing a student's "Sonstige Mitarbeit"?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	cannot answer	8	13,3	14,0	14,0
	1 only a few times	1	1,7	1,8	15,8
	2	6	10,0	10,5	26,3
	3	10	16,7	17,5	43,9
	4	27	45,0	47,4	91,2
	5 frequently	5	8,3	8,8	100,0
	Total	57	95,0	100,0	
Missing	-99	3	5,0		
Total		60	100,0		

13 Do you manage to keep apart different dimensions of student behavior?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	2	5	8,3	8,8	8,8
	3	11	18,3	19,3	28,1
	4	29	48,3	50,9	78,9
	5 very much	12	20,0	21,1	100,0
	Total	57	95,0	100,0	
Missing	-99	3	5,0		
Total		60	100,0		

("1 – not at all" was not selected by any of the participants)

14 Which aspects of student behavior do you grade as part of "Sonstige Mitarbeit"? Quality of oral participation:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	2	57	95,0	100,0	100,0
Missing	1	3	5,0		
Total		60	100,0		

14 Which aspects of student behavior do you grade as part of "Sonstige Mitarbeit"? Frequency of oral participation:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	2	56	93,3	100,0	100,0
Missing	1	4	6,7		
Total		60	100,0		

14 Which aspects of student behavior do you grade as part of "Sonstige Mitarbeit"? Social behavior in class:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	2	16	26,7	100,0	100,0
Missing	1	44	73,3		
Total		60	100,0		

14 Which aspects of student behavior do you grade as part of "Sonstige Mitarbeit"? Regular presence in class:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	2	23	38,3	100,0	100,0
Missing	1	37	61,7		
Total		60	100,0		

14 Which aspects of student behavior do you grade as part of "Sonstige Mitarbeit"? Doing homework:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	2	49	81,7	100,0	100,0
Missing	1	11	18,3		
Total		60	100,0		

14 Which aspects of student behavior do you grade as part of "Sonstige Mitarbeit"? Quality of homework:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	2	51	85,0	100,0	100,0
Missing	1	9	15,0		
Total		60	100,0		

14 Which aspects of student behavior do you grade as part of "Sonstige Mitarbeit"? Tests:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	2	48	80,0	100,0	100,0
Missing	1	12	20,0		
Total		60	100,0		

14 Which aspects of student behavior do you grade as part of "Sonstige Mitarbeit"? Orderly and complete workbooks:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	2	29	48,3	100,0	100,0
Missing	1	31	51,7		
Total		60	100,0		

14 Which aspects of student behavior do you grade as part of "Sonstige Mitarbeit"? Presentations:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	2	52	86,7	100,0	100,0
Missing	1	8	13,3		
Total		60	100,0		

14 Which aspects of student behavior do you grade as part of "Sonstige Mitarbeit"? Protocols:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	2	15	25,0	100,0	100,0
Missing	1	45	75,0		
Total		60	100,0		

14 Which aspects of student behavior do you grade as part of "Sonstige Mitarbeit"? Creativity:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	2	36	60,0	100,0	100,0
Missing	1	24	40,0		
Total		60	100,0		

14 Which aspects of student behavior do you grade as part of "Sonstige Mitarbeit"? Others:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	2	11	18,3	100,0	100,0
Missing	1	49	81,7		
Total		60	100,0		

16 Is it important that when different students of the same class are graded, the general learning conditions are equivalent?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	cannot answer	12	20,0	21,1	21,1
	1 not important at all	1	1,7	1,8	22,8
	2	3	5,0	5,3	28,1
	3	11	18,3	19,3	47,4
	4	11	18,3	19,3	66,7
	5 very important	19	31,7	33,3	100,0
	Total	57	95,0	100,0	
Missing	-99	3	5,0		
Total		60	100,0		

17 Does the fact that the different dimensions of students' oral language competency are connected have significance for your assessment of "Sonstige Mitarbeit"?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	cannot answer	6	10,0	10,5	10,5
	1 not at all	3	5,0	5,3	15,8
	2	4	6,7	7,0	22,8
	3	5	8,3	8,8	31,6
	4	24	40,0	42,1	73,7
	5 definitely	15	25,0	26,3	100,0
	Total	57	95,0	100,0	
Missing	-99	3	5,0		
Total		60	100,0		

18 Do you agree with the following statement: „It is essential to fair grading that the teaching methods used in class must be varied?"

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1 don't agree at all	2	3,3	3,5	3,5
	2	2	3,3	3,5	7,0
	3	6	10,0	10,5	17,5
	4	15	25,0	26,3	43,9
	5 fully agree	32	53,3	56,1	100,0
	Total	57	95,0	100,0	
Missing	-99	3	5,0		
Total		60	100,0		

19 How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) after each lesson?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	0<2 minutes	18	30,0	32,7	32,7
	2-4 minutes	19	31,7	34,5	67,3
	4-6 minutes	8	13,3	14,5	81,8
	6-10 minutes	5	8,3	9,1	90,9
	>10 minutes	2	3,3	3,6	94,5
	Other	3	5,0	5,5	100,0
	Total	55	91,7	100,0	
Missing	-99	5	8,3		
Total		60	100,0		

20 How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) each school day?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	<5 minutes	9	15,0	15,8	15,8
	5-10 minutes	19	31,7	33,3	49,1
	10-20 minutes	17	28,3	29,8	78,9
	20-30 minutes	4	6,7	7,0	86,0
	>30 minutes	3	5,0	5,3	91,2
	Other	5	8,3	8,8	100,0
	Total	57	95,0	100,0	
Missing	-99	3	5,0		
Total		60	100,0		

21 How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) each school week?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	<15 minutes	6	10,0	11,1	11,1
	15-30 minutes	17	28,3	31,5	42,6
	30-45 minutes	7	11,7	13,0	55,6
	45-60 minutes	11	18,3	20,4	75,9
	>60 minutes	8	13,3	14,8	90,7
	Other	5	8,3	9,3	100,0
	Total	54	90,0	100,0	
Missing	-99	6	10,0		
Total		60	100,0		

22 How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) each school month?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	<1 hour	8	13,3	15,7	15,7
	1-2 hours	13	21,7	25,5	41,2
	2-4 hours	13	21,7	25,5	66,7
	4-6 hours	7	11,7	13,7	80,4
	>6 hours	6	10,0	11,8	92,2
	Other	4	6,7	7,8	100,0
	Total	51	85,0	100,0	
Missing	-99	9	15,0		
Total		60	100,0		

23 How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) each quarter (on determining Quartalsnoten only , not considering the time you spend on that regularly during the term)?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	15-30 minutes	8	13,3	15,1	15,1
	30-60 minutes	14	23,3	26,4	41,5
	60-90 minutes	13	21,7	24,5	66,0
	90-120 minutes	6	10,0	11,3	77,4
	>120 minutes	9	15,0	17,0	94,3
	Other	3	5,0	5,7	100,0
	Total	53	88,3	100,0	
Missing	-99	7	11,7		
Total		60	100,0		

24 How much time do you spend on carrying out the assessment of students' "Sonstige Mitarbeit" (for all students you teach) each half term (on determining the overall grade, not considering the time you spend on that regularly during the term)?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	<15 minutes	1	1,7	1,9	1,9
	15-30 minutes	4	6,7	7,5	9,4
	30-60 minutes	13	21,7	24,5	34,0
	60-90 minutes	9	15,0	17,0	50,9
	90-120 minutes	11	18,3	20,8	71,7
	>120 minutes	13	21,7	24,5	96,2
	Other	2	3,3	3,8	100,0
	Total	53	88,3	100,0	
Missing	-99	7	11,7		
Total		60	100,0		

25 Do you follow a certain method for grading „Sonstige Mitarbeit"?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	Yes	42	70,0	75,0	75,0
	No	14	23,3	25,0	100,0
	Total	56	93,3	100,0	
Missing	-99	4	6,7		
Total		60	100,0		

26 Have you been introduced to a certain method when grading „Sonstige Mitarbeit“, either during your studies, during teacher training or during your professional career (Fortbildung)?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	Yes	23	38,3	41,1	41,1
	No	33	55,0	58,9	100,0
	Total	56	93,3	100,0	
Missing	-99	4	6,7		
Total		60	100,0		

27 Have you tried to find out about alternative methods of assessing “Sonstige Mitarbeit”?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1 not at all	8	13,3	14,3	14,3
	2	18	30,0	32,1	46,4
	3	16	26,7	28,6	75,0
	4	10	16,7	17,9	92,9
	5 very often	4	6,7	7,1	100,0
	Total	56	93,3	100,0	
Missing	-99	4	6,7		
Total		60	100,0		

28 How satisfied are you with your method of assessing “Sonstige Mitarbeit”?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1 not satisfied at all	2	3,3	3,6	3,6
	2	5	8,3	8,9	12,5
	3	16	26,7	28,6	41,1
	4	24	40,0	42,9	83,9
	5 very satisfied	9	15,0	16,1	100,0
	Total	56	93,3	100,0	
Missing	-99	4	6,7		
Total		60	100,0		

29 If you are not satisfied with your way of grading „Sonstige Mitarbeit“, what are the problems?
Not reliable enough:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1	42	70,0	70,0	70,0
	2	18	30,0	30,0	100,0
	Total	60	100,0	100,0	

29 If you are not satisfied with your way of grading „Sonstige Mitarbeit“, what are the problems? Too complicated:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1	56	93,3	93,3	93,3
	2	4	6,7	6,7	100,0
	Total	60	100,0	100,0	

29 If you are not satisfied with your way of grading „Sonstige Mitarbeit“, what are the problems? Too time-consuming:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1	51	85,0	85,0	85,0
	2	9	15,0	15,0	100,0
	Total	60	100,0	100,0	

29 If you are not satisfied with your way of grading „Sonstige Mitarbeit“, what are the problems? Too irregular:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1	45	75,0	75,0	75,0
	2	15	25,0	25,0	100,0
	Total	60	100,0	100,0	

29 If you are not satisfied with your way of grading „Sonstige Mitarbeit“, what are the problems? Not objective enough:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1	44	73,3	73,3	73,3
	2	16	26,7	26,7	100,0
	Total	60	100,0	100,0	

29 If you are not satisfied with your way of grading „Sonstige Mitarbeit“, what are the problems? Not motivating to the students:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1	54	90,0	90,0	90,0
	2	6	10,0	10,0	100,0
	Total	60	100,0	100,0	

29 If you are not satisfied with your way of grading „Sonstige Mitarbeit“, what are the problems? Not in perfect accordance with the legal regulations:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1	56	93,3	93,3	93,3
	2	4	6,7	6,7	100,0
	Total	60	100,0	100,0	

29 If you are not satisfied with your way of grading „Sonstige Mitarbeit“, what are the problems? Other:

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1	58	96,7	96,7	96,7
	2	2	3,3	3,3	100,0
	Total	60	100,0	100,0	

30 Would you give grades for “Sonstige Mitarbeit” if that was voluntary?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	1 definitely not	2	3,3	3,6	3,6
	2	1	1,7	1,8	5,4
	3	3	5,0	5,4	10,7
	4	15	25,0	26,8	37,5
	5 definitely	35	58,3	62,5	100,0
	Total	56	93,3	100,0	
Missing	-99	4	6,7		
Total		60	100,0		

31 Please specify your gender.

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	female	34	56,7	60,7	60,7
	male	22	36,7	39,3	100,0
	Total	56	93,3	100,0	
Missing	-99	4	6,7		
Total		60	100,0		

32 Please specify you age group.

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	25-29 years	8	13,3	14,3	14,3
	30-34 years	9	15,0	16,1	30,4
	35-39 years	12	20,0	21,4	51,8
	40-44 years	8	13,3	14,3	66,1
	45-49 years	6	10,0	10,7	76,8
	50-54 years	10	16,7	17,9	94,6
	55-59 years	2	3,3	3,6	98,2
	60-64 years	1	1,7	1,8	100,0
	Total	56	93,3	100,0	
Missing	-99	4	6,7		
Total		60	100,0		

33 For how many years have you been working as a teacher?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	0-4 years	16	26,7	28,6	28,6
	5-9 years	11	18,3	19,6	48,2
	10-14 years	10	16,7	17,9	66,1
	15-19 years	4	6,7	7,1	73,2
	20-24 years	5	8,3	8,9	82,1
	> 25 years	10	16,7	17,9	100,0
	Total	56	93,3	100,0	
Missing	-99	4	6,7		
Total		60	100,0		

34 In which section of the Gymnasium do you do most of your teaching?

		Frequency	In Percent	Valid In Percent	Cumulated In Percent
Valid	More in Sekundarstufe I than in Sekundarstufe II	11	18,3	19,6	19,6
	To the same degree in Sekundarstufe I and II	23	38,3	41,1	60,7
	More in Sekundarstufe II than in Sekundarstufe I	20	33,3	35,7	96,4
	Other	2	3,3	3,6	100,0
	Total	56	93,3	100,0	
Missing	-99	4	6,7		
Total		60	100,0		

Appendix L – Explanation of Statistical Terms

An introduction to the statistical terminology used for the analysis of the survey data.

Explanation of statistical terms:

Measure of central tendency and measure of dispersion:

For items with ordinal scales such as they have been used in the questionnaire, the *median* is the appropriate measure of central tendency. It is defined as "the mid-point in a distribution of values" (Bryman & Cramer, 1990, p. 82):

> *Imagine that the values in a distribution are arrayed from low to high […]; in this example, the median is the middle value […]. When there is an even number of values, the average between the two middle values is taken.* (Bryman & Cramer, 1990, p. 82-83)

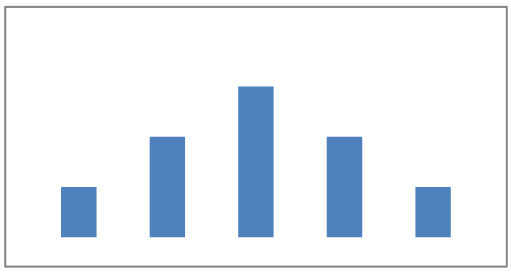
The respective measures of dispersion are the *high and low quartiles* or *1st and 3rd quartile* (Argyrous, 2000, p. 64; p. 75). They are defined in the following way:

> *If all the cases in a distribution are ranked from lowest to highest, the median is the value that divides the data in half.* (Argyrous, 2000, p. 67)
>
> *If the median divides the data in half, the low quartile is the value that divides the first half of the data, and therefore the value located at the end of the first quarter of the data; similarly, the high quartile is the value located at the end of the third quarter of the data. (own definition)*

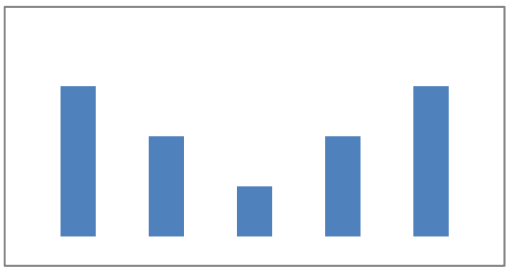
For items with modal scales such as they have been used in the questionnaire, the *mode* is the measure of central tendency, representing the single most frequent value (Argyrous, 2000, p. 65).

Distributions:

A diagram and the values associated with it have a normal distribution if the maximum of the individual values is taken on by the middle value or middle values of the range of results and the other values are in decreasing order when compared from the middle values to the outer values. An inverted normal distribution is the complement to a normal distribution, with the highest values at the outer limits of the scale and decreasing towards the middle. (Argyrous, 2000, pp. 112-131)

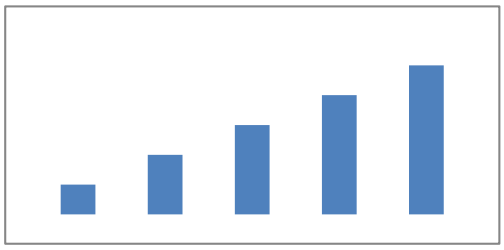


normal distribution

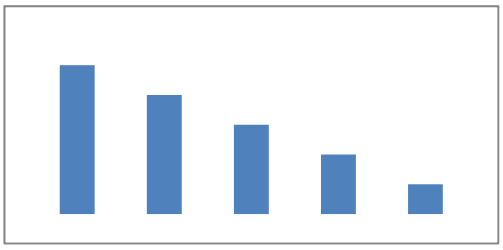


inverted normal distribution

A diagram and the values associated with it have a negative skew if the maximum is taken on by a value or values higher than the middle value. A diagram with a positive skew has the maximum taken on by a value that is lower than the middle. (Argyrous, 2000, p. 72)

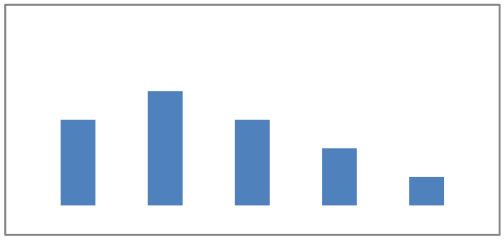


negative skew

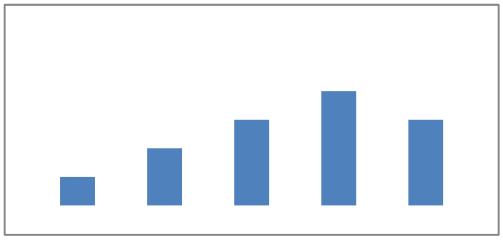


positive skew

A diagram and the values associated with it have a normal distribution with a positive / negative skew if the results follow the curve of a normal distribution but are shifted towards lower / higher values (Argyrous, 2000, p. 72).

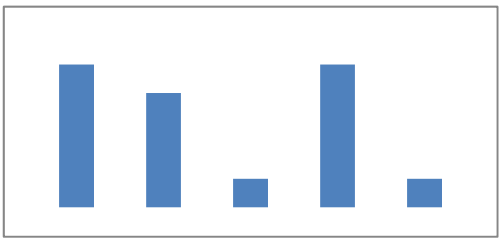


normal distribution with positive skew



normal distribution with negative skew

A diagram and the values associated with it have a complex distribution or none of the above distributions can be identified or if the results are a composite of several of the above distributions (Argyrous, 2000, pp. 112-131).



complex distribution

Appendix M – Contingency Tables

Contingency tables for selected questions generated by the software *SPSS 15*.

25 Do you follow a certain method for grading „Sonstige Mitarbeit“? * 28 How satisfied are you with your method of assessing “Sonstige Mitarbeit”? Kreuztabelle

			28 How satisfied are you with your method of assessing “Sonstige Mitarbeit”?					Gesamt
			1 not satisfied at all	2	3	4	5 very satisfied	1 not satisfied at all
25 Do you follow a certain method for grading „Sonstige Mitarbeit“?	Yes	Anzahl	1	3	9	21	8	42
		% von 25 Do you follow a certain method for grading „Sonstige Mitarbeit“?	2,4%	7,1%	21,4%	50,0%	19,0%	100,0%
	No	Anzahl	1	2	7	3	1	14
		% von 25 Do you follow a certain method for grading „Sonstige Mitarbeit“?	7,1%	14,3%	50,0%	21,4%	7,1%	100,0%
Gesamt		Anzahl	2	5	16	24	9	56
		% von 25 Do you follow a certain method for grading „Sonstige Mitarbeit“?	3,6%	8,9%	28,6%	42,9%	16,1%	100,0%

31 Please specify your gender. * 2 When giving students their final grade, how much do you weigh “Sonstige Mitarbeit” in comparison to written exams? Kreuztabelle

			2 When giving students their final grade, how much do you weigh “Sonstige Mitarbeit” in comparison to written exams?			Gesamt
			I tend to give more weight to „Sonstige Mitarbeit“	I tend to give equal weight to both aspects	I tend to give more weight to class tests and „Klausuren“	I tend to give more weight to „Sonstige Mitarbeit“
31 Please specify your gender.	female	Anzahl	2	21	5	28
		% von 31 Please specify your gender.	7,1%	75,0%	17,9%	100,0%
	male	Anzahl	0	16	1	17
		% von 31 Please specify your gender.	,0%	94,1%	5,9%	100,0%
Gesamt		Anzahl	2	37	6	45
		% von 31 Please specify your gender.	4,4%	82,2%	13,3%	100,0%

31 Please specify your gender. * 5 What is the situation in which you give students their final grades? Kreuztabelle

			5 What is the situation in which you give students their final grades?				Gesamt
			In a private face-to-face conversation	In class	In written form, to the individual student only	Other	In a private face-to-face conversation
31 Please specify your gender.	female	Anzahl	26	5	3	0	34
		% von 31 Please specify your gender.	76,5%	14,7%	8,8%	,0%	100,0%
	male	Anzahl	10	7	4	1	22
		% von 31 Please specify your gender.	45,5%	31,8%	18,2%	4,5%	100,0%
Gesamt		Anzahl	36	12	7	1	56
		% von 31 Please specify your gender.	64,3%	21,4%	12,5%	1,8%	100,0%

31 Please specify your gender. * 25 Do you follow a certain method for grading „Sonstige Mitarbeit“? Kreuztabelle

			25 Do you follow a certain method for grading „Sonstige Mitarbeit“?		Gesamt
			Yes	No	Yes
31 Please specify your gender.	female	Anzahl	24	10	34
		% von 31 Please specify your gender.	70,6%	29,4%	100,0%
	male	Anzahl	18	4	22
		% von 31 Please specify your gender.	81,8%	18,2%	100,0%
Gesamt		Anzahl	42	14	56
		% von 31 Please specify your gender.	75,0%	25,0%	100,0%

31 Please specify your gender. * 28 How satisfied are you with your method of assessing “Sonstige Mitarbeit”? Kreuztabelle

			28 How satisfied are you with your method of assessing “Sonstige Mitarbeit”?					Gesamt
			1 not satisfied at all	2	3	4	5 very satisfied	1 not satisfied at all
31 Please specify your gender.	female	Anzahl	0	5	13	11	5	34
		% von 31 Please specify your gender.	,0%	14,7%	38,2%	32,4%	14,7%	100,0%
	male	Anzahl	2	0	3	13	4	22
		% von 31 Please specify your gender.	9,1%	,0%	13,6%	59,1%	18,2%	100,0%
Gesamt		Anzahl	2	5	16	24	9	56
		% von 31 Please specify your gender.	3,6%	8,9%	28,6%	42,9%	16,1%	100,0%

32 Please specify your age group * 26 Have you been introduced to a certain method when grading „Sonstige Mitarbeit“, either during your studies, during teacher training or during your professional career (Fortbildung)? Kreuztabelle

			26 Have you been introduced to a certain method when grading „Sonstige Mitarbeit“, either during your studies, during teacher training or during your professional career (Fortbildung)?		Gesamt
			Yes	No	Yes
32 Please specify your age group	1,00	Anzahl	19	24	43
		% von 32 Please specify your age group	44,2%	55,8%	100,0%
	2,00	Anzahl	4	9	13
		% von 32 Please specify your age group	30,8%	69,2%	100,0%
Gesamt		Anzahl	23	33	56
		% von 32 Please specify your age group	41,1%	58,9%	100,0%

34 In which sections of the Gymnasium do you do most of your teaching? * 25 Do you follow a certain method for grading „Sonstige Mitarbeit“? Kreuztabelle

			25 Do you follow a certain method for grading „Sonstige Mitarbeit“?		Gesamt
			Yes	No	Yes
34 In which sections of the Gymnasium do you do most of your teaching?	More in Sekundarstufe I than in Sekundarstufe II	Anzahl	6	5	11
		% von 34 In which sections of the Gymnasium do you do most of your teaching?	54,5%	45,5%	100,0%
	To the same degree in Sekundarstufe I and II	Anzahl	20	3	23
		% von 34 In which sections of the Gymnasium do you do most of your teaching?	87,0%	13,0%	100,0%
	More in Sekundarstufe II than in Sekundarstufe I	Anzahl	14	6	20
		% von 34 In which sections of the Gymnasium do you do most of your teaching?	70,0%	30,0%	100,0%
	Other	Anzahl	2	0	2
		% von 34 In which sections of the Gymnasium do you do most of your teaching?	100,0%	,0%	100,0%
Gesamt		Anzahl	42	14	56
		% von 34 In which sections of the Gymnasium do you do most of your teaching?	75,0%	25,0%	100,0%

34 In which sections of the Gymnasium do you do most of your teaching? * 26 Have you been introduced to a certain method when grading „Sonstige Mitarbeit“, either during your studies, during teacher training or during your professional career (Fortbildung)? Kreuztabelle

			26 Have you been introduced to a certain method when grading „Sonstige Mitarbeit“, either during your studies, during teacher training or during your professional career (Fortbildung)?		Gesamt
			Yes	No	Yes
34 In which sections of the Gymnasium do you do most of your teaching?	More in Sekundarstufe I than in Sekundarstufe II	Anzahl	2	9	11
		% von 34 In which sections of the Gymnasium do you do most of your teaching?	18,2%	81,8%	100,0%
	To the same degree in Sekundarstufe I and II	Anzahl	10	13	23
		% von 34 In which sections of the Gymnasium do you do most of your teaching?	43,5%	56,5%	100,0%
	More in Sekundarstufe II than in Sekundarstufe I	Anzahl	9	11	20
		% von 34 In which sections of the Gymnasium do you do most of your teaching?	45,0%	55,0%	100,0%
	Other	Anzahl	2	0	2
		% von 34 In which sections of the Gymnasium do you do most of your teaching?	100,0%	,0%	100,0%
Gesamt		Anzahl	23	33	56
		% von 34 In which sections of the Gymnasium do you do most of your teaching?	41,1%	58,9%	100,0%

E r k l ä r u n g

Hiermit erkläre ich, dass ich die vorliegende Masterarbeit selbständig verfasst und gelieferte Datensätze, Zeichnungen, Skizzen und graphische Darstellungen selbständig erstellt habe. Ich habe keine anderen Quellen als die angegebenen benutzt und habe die Stellen der Arbeit, die anderen Werken entnommen sind -einschl. verwendeter Tabellen und Abbildungen -in jedem einzelnen Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht.

Bielefeld, den 28.09.2009 Jens Peter Fischer

(Unterschrift)