

Explicit Assessment Rubrics and L2 Motivation

An experimental study investigating the effects of implementing explicit assessment rubrics in a Swedish EFL context



LUND
UNIVERSITY

Author: Anthon Brunzell

Supervisor: Annika Lindskog

Term: Autumn 2020

Course: ÄEND14 English IV

Individual Paper (15 hp)

English Teacher Education

Abstract

With regards to increased research on the formative use of assessment rubrics and its effects on performance, this study highlights the need for more research on rubrics' effects on different kinds of motivation. Therefore, the study aims to investigate possible effects of implementing explicit assessment rubrics on L2 motivation in an EFL classroom. The study is quasi-experimental, which means that it uses a pre-post experimental method with two groups of students, one experimental group ($N = 15$) and one control group ($N = 16$). For testing, questionnaires were used and for the instructional treatment an explicit assessment rubric was constructed adapted to the context and recommendations from previous research. Data was collected during class with students' consent before and after the experiment. For the data analysis, a Wilcoxon Signed Ranks test was conducted together with a Mann Whitney U-test. The data analysis of the study showed that the experimental group increased Amotivation after the treatment ($P < 0.10$) as well as decreased their score of Extrinsic motivation ($P < 0.10$). The major conclusion of the essay is that that no matter how much positive or negative criteria compliance explicit assessment rubrics seems to promote, they do not in all cases prompt positive L2 Motivation effects. In addition, the study discusses the possible standardization of research specializing in assessment rubrics' effects on L2 motivation.

Keywords: formative assessment, rubrics, L2 motivation, self-determination theory, instrumentalism

Table of Contents

Introduction.....	4
Background.....	7
Self-Determination Theory	12
The model of use in present paper.....	13
Criteria compliance and Motivation.....	15
Assessment Rubrics.....	16
Skolverket and Assessment Rubrics.....	17
Materials and Method.....	19
Participants and Context.....	20
Procedure.....	21
Measuring Motivation.....	21
Construction of Assessment Rubrics.....	22
The Instructional Treatment.....	25
Statistical Analysis.....	27
Reliability and Validity.....	27
Ethical Considerations.....	29
Results.....	29
Amotivation.....	30
Self-Determined Motivation	35
Discussion.....	43
Autonomy, relatedness, and competence	44
Conclusion	45
References.....	47
Appendix A.....	50
Appendix B.....	58

Introduction

Formative assessment practises or in-process evaluations of student performances have been welcomed with optimism in the Swedish educational sector (Vetenskapsrådet, 2015). The Swedish Research Council's report of formative assessment research in a Swedish and international context concludes that while formative assessment practises are highly valued in Swedish schools, there are some uncertainties in their implementation. The authors furthermore claim that the implementation of formative assessment practises result in 'pseudo-formative' practises with teachers adopting a ritual approach toward assessment (p. 4). However, the authors also claim that the research conducted is insufficient. This study aims to test one example of formative assessment practise, explicit assessment rubrics, in a Swedish English as a foreign language context (henceforth EFL). The term 'assessment rubric' is in this study defined as a scoring guide designed to evaluate the quality of students' work in the form of written compositions, oral presentations or science projects (Popham, 1997).

In this study, I will argue that researchers have failed to assess the consequences that assessment rubrics may have on student motivation. Comprehensive research from the last two decades indicates that use of assessment rubrics typically improves student performance. However, it has failed to take into consideration effects on student autonomy and motivation (e.g. Torrance 2005; Brookhart & Chen, 2015). Rubrics used in accordance with EFL-teaching and L2 motivation has not been put through thorough research either. This essay can be seen as an attempt to atone for this in order to broaden the field researching assessment rubrics.

It is vital that teachers have competence not only in performative effects of assessment rubrics but also the motivational effects. Whatever students are motivated by may steer them from self-

autonomy to dependence, and as the Swedish curriculum puts it, ‘The task of the school is to encourage all students to discover their own uniqueness as individuals and thereby actively participate in the life of society by giving of their best in responsible freedom’ (Skolverket, 2019, p. 5). Swedish schools and teachers are therefore required to provide education that motivates through student autonomy. Critique has earlier been put forward by Torrance (2007) that assessment rubrics may lead to criteria compliance and *instrumentalism*, when students solely focus on achieving the specified criteria instead of all the other variable aspects that need to be fulfilled in school, such as linguistic competence and non-cognitive abilities. This would suggest that the motivational processes involved involve minimal amounts of autonomy and more of what Ryan and Deci (1985) would call *extrinsic motivation* or motivation by some external regulator, in this case a criterion for a grade. Therefore, in order to consider autonomy and eventual instrumentalism, the *self-determination theory* by the aforementioned authors is used. The theory assumes autonomy as one of the staples of motivation (or self-determination) and is therefore a theory that considers socio-contextual factors (Dörnyei & Ushioda, 2011). This study may therefore contribute to the field in that it investigates the degree to which explicit assessment rubrics encourages or discourages criteria compliance. Since instrumentalism leads to less student autonomy, a decrease in student motivation would be expected according to the theory.

An experimental approach will be used in order to find possible effects of explicit assessment rubrics on students’ self-determination. This means that one of two comparable groups will be receiving a treatment involving rubrics adapted to the context and tested before and after the experiment to measure motivation. The specific aim is to investigate possible effects of implementing explicit assessment rubrics on L2 motivation in a Swedish EFL context. Two research questions will be answered by the means of a data analysis with the null hypothesis that there is no

difference between the two groups and the alternative hypothesis that there is a difference between the groups in the parameters amotivation and self-determination.

- Is there a relation between using explicit assessment rubrics and amotivation in a group of EFL students at a Swedish secondary school?
- Is there a relation between using explicit assessment rubrics and self-determination in a group of EFL students at a Swedish secondary school?

The experiment itself involved two classes of a secondary school in the south of Sweden. Two teachers with the same experience worked with the same subject area for five weeks when the instructional treatment was implemented in one of the classrooms. Rubrics were used in both classrooms. However, while the experimental group had rubrics adapted to their needs and training in using them, there was a minimal presentation of rubrics directly from the Swedish curriculum without any guidance in the control group. In this way, the variable explicit assessment rubrics could be considered the factor determining eventual differences between the groups.

The study begins with a literature review that outlines the development of some overlapping research fields where the present study belongs. Theoretical concepts and assessment rubrics are then reviewed in a more detailed regard. These are used within the materials and method section. Statistics from the statistical analysis will be presented in the result section and discussed in the discussion section. Finally, a conclusion is found together with future lines for further research.

Background

There are many varying research topics that are touched upon in this essay and this literature review aims to outline them all. First in the following section a more general account of formative assessment and the recent history of assessment rubrics will be made. This part aims to guide the reader in where the research field comes from and where there are considerable gaps that need to be filled. In the second part, the theory that may help fill some of these gaps is outlined. In this part major concepts are defined and how the theory will be implemented is clarified. In the third part, assessment rubrics will be defined, and discussion follow on how these could be implemented in a Swedish context.

The concept formative assessment first attained popularity in the late 90s when proposals were made to transform assessment *of* learning into assessment *for* learning in order to make assessment support rather than discourage learning (Black & William, 1998; Torrance & Pryor, 1998). Since then, progress has been made in the research field of formative assessment. One practise that is designed to be used as a part of formative assessment frameworks and has gained traction the last decade is the usage of explicit assessment rubrics.

Explicit assessment rubrics have been regarded as plausible practises primarily in conjunction with general formative assessment practises. Because if these are missing assessment rubrics are generally considered inefficient (e.g. Jönsson & Panadero, 2013). This is because most studies that report a positive effect size assume that the rubrics are used formatively, i.e. during learning processes. Rubrics are therefore also intended to be used formatively in this text.

Explicit criterion-referenced assessment or assessment rubrics have been investigated along with the general interest for formative assessment during the last 20 years. However, it has

predominantly been seen as a good example of formative practise since Panadero and Jönsson's literature review (2013). The aim of their review was to examine if and how assessment rubrics had an effect on student performance. The method involved was a narrative content analysis that would identify some potential effects of using rubrics formatively. 21 articles in total were analysed. Results indicated that increased transparency, reduced student anxiety, an aided feedback process and improved student motivation were some of the major reasons behind increased performance by students when subjected to assessment rubrics. Factors that moderated the effects of using rubrics formatively may have been the use of meta-cognitive activities and comprehensive and long-term interventions (Panadero & Jönsson, 2013). This study remains a popular source when it comes to validating the use of assessment rubrics as it is one of the most cited with 320 citations.

However, Panadero and Jönsson's discoveries may be nuanced by a newer review by Brookhart and Chen (2015) encompassing 63 studies published from 2005 through October 2013. This review aimed not just to investigate performance but also self-regulation, according to Zimmerman (1986) a way of planning, organizing and self-controlling during a learning process. Overall, the studies examined showed an improvement in performance when rubrics were used. However, when it comes to self-regulation and motivation, results were mixed. Brookhart and Chen (2015) also state that the research field is very new in that it is out of its 'intimacy but not yet mature' and that the increasing number of studies are promising but not enough to establish any positive effects of using rubrics on student performance (p. 361). With this new light upon the field of assessment rubrics, depicting it as lacking in research, one may deduce that the field needs further motivational research. In the present study, contributions to the field may be made as we investigate one of the major critiques against explicit assessment rubrics.

The transparency of criteria that explicit assessment rubrics lead to may promote instrumentalism. Torrance (2007) discusses how new formative assessment methods in general may move students further from autonomy and closer to criteria compliance. While the goal of formative assessment always has been to transform assessment *of* learning into assessment *for* learning, Torrance warns that assessment rubrics may instead result in assessment *as* learning. In his study, a total of 237 learners and 95 assessors from post compulsory educational sectors such as higher education and work-based learning projects were interviewed. Results showed that formative feedback and explicit learning objectives led to learning ‘achievements’ attaining a more narrow and instrumental meaning. Learning was not any longer about learning, but rather about attaining expected or necessary grades for a reward such as specific permits and papers for work.

While this critique has gathered support, Jönsson and Panadero (2020) have since reviewed the criticisms and stated that there is a shortage of empirical evidence in these examples. In their review from 2020 they aimed to investigate (a) the main areas of criticism, (b) to what extent the claims are supported by empirical evidence and (c) to what extent they are not supported by empirical evidence. Studies were picked not from an extensive search of literature but out of the authors own reference library as well as out of three research experts’ recommendations. In total 27 studies were investigated, and 97 excerpts were picked out categorized as criticisms of assessment rubrics. The authors then used a content analysis in order to categorize the different criticisms. The results broadly rejected many of the criticisms as the 97 excerpts many times used anecdotal evidence. However empirical evidence is acknowledged in some cases, as in Torrance’s critique. Some aspects are also not accounted for.

In the aforementioned Swedish Research Council’s (2015) report of formative assessment research, the Swedish research was judged as scant and the implementations of formative practises

in Swedish schools as questionable. While formative assessment practises were found to be highly valued in Swedish schools, practises were not always implemented as prevalent research recommended but instead ended up as ‘pseudo-formative’ (Vetenskapsrådet, 2015, p.4). Teachers generally developed a ritual process when working with formative assessment and treated it as instrumental in attaining a goal rather than learning. The report offers a comprehensive review of the Swedish approach to formative assessment. It also indicates that studies investigating lower levels are few and that scientific basis for implementations at lower levels, i.e., elementary, and secondary school, is lacking. The report concludingly supports the case brought up by Torrance (2007) that assessment rubrics may lead to less focus on learning and more focus on criteria compliance. This may very well be relevant in research concerning assessment rubrics’ effect on student motivation, of which there generally has been very little research. In the following I discuss some examples that have similarities to present paper.

In 2018 Balan and Jönsson aimed to investigate student motivation and performance with increased explicitness of assessment criteria. The study tested students from four Swedish schools on performance as well as self-efficacy, i.e. one’s belief in one’s ability, before and after the study. The amount of explicitness was different in each school. The results showed that while performance increased in every school, the increase of explicitness did not correlate with the increase in performance. The schools with the biggest improvements were also the schools with the largest number of low-performing students. There were generally no changes in motivational measurements. The authors also state that while formative assessment also was a novelty in this study, the overall effects of formative assessment might have overshadowed the effectiveness of the explicitness. A number of shortcomings are listed at the end of the study, one of them being that samples should have more similar conditions, another one that students should be older and

a third that students should already be used to working with formative assessment-practises.

Balan and Jönsson's (2018) study is different from the current essay in that it measures self-efficacy and self-regulation (see *Previous Research*, p. 4) abiding to the SRL theory by Zimmerman (1986). In my study, a different theory will be addressed.

One of the few existing studies focusing on rubrics in English as a second language or foreign language was conducted by De Silva (2014). Specifically, De Silva investigated assessment rubrics' effect on ESL students writing and speech performance. Research questions included how exposure to these rubrics affected students' performance as well as if their understanding of the rubrics affected their self-assessment. A pre-post experimental study was conducted with a sample of 70 secondary school students in Sri Lanka. The students were divided into two classes. Results showed a significantly higher performance in both tasks for the class that received rubrics with explanations. This study gives further support for explicitness when working with assessment rubrics; however, it does not say anything about student motivation. Hence, this further highlights the need for more studies concerning the effect of assessment rubrics on ESL and EFL students' motivation.

To conclude, new popular research topics may not always prompt trustworthy conclusions. Assessment rubrics have over the last decade gained popularity among researchers because of the many positive side effects that increased transparency in assessment may have, however in a research field that is undeveloped. Assessment rubrics have a need for further validity analyses, studies on other aspects than performance and consequences of implications in different contexts such as among EFL students. Particularly, this paper may examine the notion of 'instrumentalism' which the sources reviewed indicate may be common when implementing assessment

rubrics. The manner of doing this may be by investigating students' motivation and autonomy which will be discussed below.

Self-Determination Theory

Within the past two decades, socio-contextual factors have moved to the centre of analysis of motivation. However, contemporary research has not only focused on certain contextual aspects such as the instructional environment and the cultural environment but also on how motivation and context can be integrated so that the complexities in the prevalent situations can be accounted for. During the last decades, motivational theorists have increasingly stressed the importance of socio-contextual factors affecting motivation (Dörnyei & Ushioda, 2011). Therefore, the underlying theoretical framework in this study must have an account of contextual factors integrated. A theory that does integrate contextual factors is the Self-Determination Theory (SDT) which will be outlined here.

Motivational research during the 1970s focused on two different types of motivation: *Intrinsic motivation* (IM) as in behaving for its own sake in order to experience some innate satisfaction and *extrinsic motivation* (EM) as in performing a behaviour to achieve a separable end such as receiving an award or getting a specific grade. However, Ryan and Deci (1985) replaced the dichotomy between IM and EM with a more elaborate construct called *the self-determination continuum* where motivational forms can be placed from least self-determined to most self-determined (Ryan & Deci, 1985). On this continuum, different types of motivation may co-exist, such as the enjoyment of learning a language as well as the value of using a language at work. The concepts EM and IM still exist, with the addition of *Amotivation* which could be described as the complete absence of any kind of self-determination. It is also this concept that will be

measured when this study looks at L2 motivation. Another staple of SDT is that self-determination is dependent on three contextual factors, the sense of *autonomy*, i.e., controlling one's own behaviour, *competence*, i.e., having a sense of accomplishment, and *relatedness*, i.e., feeling a connection to other individuals such as fellow peers and the teacher.

The theory altogether has been deemed as one of the most plausible to take into account social contexts and is generally deemed as a strong research area when looking at L2 motivation (Dörnyei & Ushioda, 2011). The theory therefore provides appropriate models for describing possible effects of assessment rubrics and how they may affect the contextual factors.

The model of use in present paper

In this section definitions will be outlined for each of the constructs that this essay aims to measure when it comes to L2 motivation. The model that will be described is made from two adaptations of the self-determination continuum involving extrinsic motivation, intrinsic motivation and amotivation. The different subtypes that constitute the continuum will be described as well as the measuring tool that will be used to collect data.

In 1997, Vallerand outlined a hierarchical model of extrinsic and intrinsic motivation in order to systematize literature on the subject as well as to identify psychological mechanisms underlying motivational changes. The model has been used as a central component of the self-determination continuum ever since (Gagne & Deci, 2005). Amotivation, which has earlier been described as the complete lack of self-determination will be understood parallel to these concepts but not in relation to them. It will therefore be measured separately. With a range from least self-determined to most self-determined, he distinguishes in total three subtypes of EM:

- External Regulation. The learner is self-determined to learn because of something that is attainable externally. An example of this subtype would be learning English in order to get a job which requires a certain proficiency in English.
- Introjected Regulation. The learner is self-determined to learn because of something that surroundings or cultural context values or requires. An example of this subtype would be learning English in order to be able to fit into society if English speaking is normative in the society.
- Identified Regulation. The learner is self-determined to learn because of something that may contribute to their identity. Learning English in order to develop a self-image that is reliant on a proficiency in English would be an example of this subtype.

Vallerands model will be used in this paper by adapting a measuring tool by Noels, Pelletier, Clément and Vallerand (2000) that is designed to measure the extrinsic subtypes in Vallerand's model as well as three intrinsic subtypes for IM outlined below:

- Knowledge. The learner is self-determined to learn because of the innate satisfaction in acquiring knowledge. Learning English because of the satisfaction in learning new words would be an example of this subtype.
- Accomplishment. The learner is self-determined to learn because of the innate satisfaction in accomplishing a task. Learning English because of the satisfaction in solving difficult grammar puzzles would be an example of this subtype.
- Stimulation. The learner is self-determined to learn because of the innate satisfaction in stimulating tasks. An example of this subtype would be learning English because of a good feel when pronouncing English words successfully.

In the study by Noels et. al. (2000), the authors aimed to assess the validity and reliability of an intrinsic and extrinsic motivation questionnaire. The study had a sample of 159 bilingual university students who were learning French as an L2. The questionnaire was administered in order to test amotivation and the different subtypes of motivation. Validity, reliability and intercorrelation analysis were then conducted. Self-determination subtypes were found correlated with student scores and therefore successfully measured. Their analysis determined the questionnaire as valid and reliable. This paper is particularly relevant for me as it presents this valid and reliable measuring tool of intrinsic and extrinsic motivation that I can use and adapt to my study.

Criteria compliance and Motivation

Extrinsic motivation may be closely linked to the concept of instrumentalism that Torrance (2005) refers to. Instrumentalism is the notion of behaviour being facilitated through any kind of external criteria, such as a passing grade (Torrance, 2005). In other words, students work not for themselves to be happy with their performance but because they simply need to attain a certain grade. In this case, the certain grade can be considered an *externally regulating factor* in the context of motivation.

Instrumentalism would then be an approach that primarily uses itself of extrinsic motivation since that side of the continuum is primarily regulated by external factors. The self-determination continuum can therefore be used in order to find whether Torrance's assumptions concerning explicit assessment rubrics are correct.

Assessment Rubrics

Assessment rubrics are documents outlining the assessment criteria that students are expected to fulfil. In relation to these criteria, their levels of quality are also described in the rubric (Reddy & Andrade, 2010). In a typical assessment rubric, there are two components: (1) evaluation criteria that list the content or the processes an assessor considers when looking at the quality of a student's work, (2) quality definitions that describes what skills a student must display in order to attain a level of achievement (Popham, 1997). Traditionally, rubrics have been used as an educator's tool when assessing student performance and grading their work. However, since the start of this century, researchers have been arguing for the use of assessment rubrics in a student-centred approach where students would use the rubrics as a part of formative assessment where they could be used for teaching as well as self-assessment (e.g. Reddy & Andrade, 2010).

One may also draw a line between rubrics used in holistic scoring and rubrics used in analytical scoring. In this study, holistic rubrics are defined as rubrics with broad quality definitions to one singular evaluation criteria. Analytical rubrics are defined as rubrics with quantitative quality definitions to several evaluation criteria. Holistic rubrics are typically used when assessing the broader qualities of student thinking, as well as higher-order skills in examination tasks which require no single correct answer, such as examination tasks testing categorization or argumentation. Analytical rubrics are on the other hand used when assessing more specific qualities of thinking and lower-order skills in tasks with singular correct answers (Jescovitch et. al., 2019). In the following section, the use and views of assessment rubrics in a Swedish EFL context will be examined.

Skolverket and Assessment Rubrics

For this study to be able to implement realistic rubrics in the experiment, these will need to adhere to what the Swedish National Agency of Education, or ‘Skolverket’, expresses about assessment rubrics. This is important as the rubrics used in this experiment needs to be generalizable to rubrics that could be constructed in other Swedish contexts.

In 2018 Skolverket released a material consisting of advice for teachers on how to set and follow up on grades equally. The curricular view of assessment rubrics is here clarified, mainly in one sentence where rubrics are described as ‘worth-while’ for students to assess their own knowledge (Skolverket, 2018, p. 29). Therefore, teachers using self-assessment in their instruction may according to Skolverket profit from using assessment rubrics.

While emphasizing the value of using assessment rubrics, Skolverket also advises caution when constructing them. One risk is according to Skolverket that a far too analytical or detailed rubric may regulate the teaching so that flexibility and adaptability is affected negatively. It may also decrease what the authors call “forward-looking feedback” as in feedback that is meant to point at what students should do differently in the future (Skolverket, 2018, p. 19). The usage of more holistic assessment rubrics is neither addressed nor discussed in the advice. While the material seems to offer teachers professional freedom when grading and embed recommendations on using assessment rubrics, it also advises teachers to be cautious when using them.

Material from Skolverket that gives more distinct advice on using formative assessment and rubrics is the supportive material from 2011. The material was made as a guidance for assessment practises and was meant to be used by schools in developing teacher competence. Several

chapters concern formative assessment and ways that the student may observe its own learning. This is very similar to the student-centred approach observed in the newer material.

The second to last chapter, 'Bedömningsituationer', is meant to give advice for teachers when working with both summative and formative assessment and is the only material from Skolverket where their view of using rubrics is made clear. First, the material advises the teacher to get a clear assumption of what their students are supposed to learn. Secondly, the material emphasizes that students should get explicit instruction of what skills they are expected to obtain and how. Thirdly, there are also suggestions using peer-assessment and self-assessment with the students. In order to make the goals concrete and distinct, the material finally advises on using assessment rubrics.

The material does not offer concrete advice to construct appropriate rubrics but merely gives an example from DiNO, a support for assessment designed for natural science subjects (Skolverket, 2014). The only statement about what a rubric consists of is that both qualitative and quantitative knowledge could be specified in a rubric. The material also suggests instructing students to self-assess with the help of a rubric in order to individualise work with rubrics. Just as in the older material from 2011, self-assessment is recommended as a way of working with assessment for learning.

Concludingly, we can say that while Skolverket through the years have increasingly suggested formative assessment practises, the teacher has been given ample freedom to implement these (cf. Hult, 2017). It is up to the teacher how to interpret what a 'good' assessment rubric consists of. Although a frequent suggestion in these materials is to put the student in the centre and work with self-assessment – this will thereby be the single most valuable value to adhere to in the

construction of assessment rubrics in this study. Explicit instruction is also recommended. However, Skolverket has expressed recommendations to be cautious when working with assessment rubrics which also tells of a discrepancy between Skolverket and the educational sector. For this study, rubrics will be designed with these considerations and the involved teachers' preferences in mind.

Materials and Method

In this section materials and methods used in the experiment will be outlined. Two English classes of 28 and 25 students respectively were designated as experimental group and control group. Both classes were in their first secondary school year (year 7 in *högstadiet*) and had two teachers that designed teaching units together and therefore had a similar pedagogical approach in both classes. An instructional treatment involving thorough use of explicit assessment rubrics was implemented in the experimental group for five weeks where the researcher would besides train students in using the rubrics administer motivational measurements before and after the intervention period.

In this manner, this study manipulates processes in a group of EFL learners and then compares the outcome with a comparable control group (Nunan, 1992, pp. 24-28). By the help of the final data analysis, the changes in Amotivation and Self-Determination subtypes will be deemed statistically significant or non-significant. If there are significant changes in any of the parameters, it will be possible to determine a possible relation between usage of explicit assessment rubrics and L2 motivation as the research questions requires. The instructional treatment that is administered in this case to the experimental group will therefore attempt to make assessment rubrics and learning objectives explicit. In this section I am to present details of the context and

procedure, how the materials (measurement tools and the explicit assessment rubrics) were constructed, what statistical analysis was used and considerations of reliability, validity, and ethicality.

Participants and Context

55 EFL students between the ages 13-14 in a secondary school in the south of Sweden were asked to participate in the study. The school was chosen as it looks typical for a Swedish school as it is close to the Swedish mean in average score (*meritvärde*), 224 compared to the mean 230 (Skolverket, 2019). Only students who were Swedish speakers and enrolled in two different classes of English were retained for the analysis to establish equity between the two groups. This was vital as it ensures validity for the experimental study (Nunan, 1992). Students who did not answer the questionnaire were left out either because of non-attendance or because the teacher deemed that it was inappropriate for these students to partake in the questionnaire, mainly because of difficulties understanding the Swedish language.

The teachers concerned were both teachers with experience of several levels in the Swedish school system and the usage of rubrics. However, the teachers normally used the rubrics provided by the curricular knowledge requirements (*Kunskapskrav*). They would make these explicit normally during the beginning of every teaching unit, but not throughout. The students who had their first year with these teachers were according to the teachers used to assessment rubrics. English classes were normally distributed on three occasions per week with 40 minutes scheduled on each class. While the experimental group was assigned a treatment with explicit assessment rubrics, the control group were assigned no treatment at all.

Procedure

One may divide the experimental approach of this study in three parts. (1) A pre-test and post-test were conducted in both groups measuring initial Amotivation as well as values across all subtypes of extrinsic and intrinsic motivation one week before the instructional treatment and one week after the treatment. (2) The instructional treatment involving explicit assessment rubrics adapted to the Swedish curriculum and the context was applied to the experimental group and the same instructional treatment without these was applied to the control group throughout a subject area over five weeks. (3) Student scores in the questionnaires were input into the statistical tool SPSS (Version 26) and examined using tests of normal distribution, the Mann Whitney U Test, the Wilcoxon Signed Ranks Test as well as descriptive statistics with the two instructional treatments being independent variables and the motivational scores as well as the SDT parameters dependent variables. All steps will be described here in greater detail.

Measuring Motivation

The pre-test as well as post-test were made using a questionnaire aiming to measure Amotivation as well as inclination toward the different subtypes under EM and IM. The items used in the questionnaire were adapted from the Language Learning Orientations Scale by Noels et. al. presented above (2000). Since the scale was originally made for testing intrinsic and extrinsic subtypes and tested as valid and reliable, it was found most appropriate for a study measuring all subtypes. However, it needed to be translated into Swedish for the student's English proficiency not getting in the way of them giving truthful answers on their motivation.

There are primarily two sections in the questionnaire, one measuring AM and another measuring the subtypes of EM and IM. For each questionnaire item 6-point scales were used for students to

rate the extent to which the reason in the item corresponded to their personal reasoning from ‘Stämmer inte alls väl’ to ‘Stämmer mycket väl’. To measure the amount of Amotivation that students feel toward studying English, five items were designed in order to see how students comprehend the purpose of studying English (E.g. ‘Jag förstår inte varför jag pluggar engelska’) and mostly adapted from the items designed by Noels. The mean result of these five items would then be considered the AM score for each student, 1.0 would be considered the least possible while 6.0 would be the highest possible AM score. In the second section measuring EM and IM subtypes, every subtype had three items each that also were adapted from the items used by Noels. The mean score of these three were then calculated to constitute the score of the subtype. Scores were put into databases up for analysis.

Construction of Assessment Rubrics

How the assessment rubrics used in the experiment were designed is described here. Throughout their construction, efforts were made to align the contents of the rubrics with the Swedish curriculum and the involved teachers’ preferences. Since the rubrics required an understanding from the students, the rubrics were written in Swedish. However, in this section formulations from the rubric will be translated to English.

First of all, the two constructs that were chosen to be assessed in the speaking tasks by the teachers were adopted from Lgr 11 (Skolverket, 2019). (1) The ability to speak as in a presentation (*mundlig framställning*) and (2) the ability to interact in English (*mundlig interaktion*). These curricular specifics are going to form the two broader dimensions in the assessment rubric. This study will also follow the recommendations on using analytical assessment rubrics since the skills concerned are neither analytical nor assessed summatively. Four evaluation criteria for

each dimension were defined by the teachers in this particular speaking task. These can be found in Table 1 in the rightmost row.

Table 1. *The two dimensions as well as the evaluation criteria*

Knowledge Requirements	Evaluation Criteria
Muntlig framställning	Understandable pronunciation Correct wording Coherent sentencing Reliance on a manuscript
Muntlig interaktion	Understandable sentencing Correct wording The use of linguistic strategies Understandable pronunciation

However, these evaluation criteria were not stated explicitly in the rubrics for the students. Both teachers preferred that the knowledge requirements and the evaluation criteria were concretized by setting specific objectives instead of directly transferring the formulations that exist within the curriculum to the rubrics. Furthermore, since Skolverket has shown encouragements for the use of student-centred rubrics, the researcher made them readable by the students themselves. This

meant that quality definitions from the curriculum such as ‘The student articulates coherently’ was made into formulations such as ‘Most sentences you say are connected’. The result of this concretization when it comes to the basic objectives is displayed in the rightmost row of Table 2. Some sentences use more general wording than they could have on demand from the teachers who insisted that they wanted students to reach for not just specific objectives but also more general objectives, ‘speaking freely’ is an example of this.

Table 2. *Evaluation Criterion concretized.*

Knowledge Requirements	Evaluation Criteria concretized
Muntlig Framställning	<p>You speak understandably, and easily.</p> <p>You use the correct term several times when speaking English.</p> <p>Many sentences you say are connected.</p> <p>You are partly speaking freely in the presentation.</p>
Muntlig Interaktion	<p>You make yourself understandable in conversations with other people.</p>

	<p>You use the correct term several times when speaking English.</p> <p>When you do not know the word, you are sometimes able to explain it in other words.</p> <p>You speak understandably, and easily.</p>
--	--

A system of progression was then added as two more rows were added for the Swedish passing grades from left to right. The quality definitions were developed according to the progression detailed by the knowledge requirements in the curriculum. Thereby, quality definitions were written on three levels, E-level, C-level, and A-level, from lowest pass to highest. For increased student-centring and possibilities for the students to self-assess, checkboxes were added to every quality definition for students to fill in. Colours were added and the end result can be seen in the Appendix.

Concludingly, the assessment rubric that is constructed within this study is in accordance with Skolverket as it is student-centred and in accordance with the teachers' preferences.

The Instructional Treatment

The five weeks where the experiment would unfold would consist of a teaching unit designed entirely by the two participating teachers and the experimental treatment would run as long as the

unit itself. The whole purpose of the unit was for the students to prepare for two oral examinations, an oral presentation, and an oral discussion. The beginning of the unit consisted of research on the topic of life in a particular global destination such as Madagascar, Hong Kong, Tasmania, Jamaica, and more. Students in groups of three or four were then supposed to come up with a narrative story for the destination. They were all assigned a character that needed a story for one of the global destinations. They made presentations on their computers and prepared their speeches. The oral discussions were taped and not supervised for the teacher's convenience. In the end of the project students filmed themselves doing the presentation. The contents of the teaching unit were not important for this study per se, but a coherent and common task was needed for the experimental treatment and as the study did not mean to interfere in the teachers planning, a speaking task was appropriate for both teachers. How assessment rubrics were made explicit during this unit will be explained in further detail in this next section.

The experimental treatment aimed to use student-centred and explicit assessment rubrics. Explicitness of the rubrics was achieved through two primary ways: (1) Regular sessions with the researcher every third class where the rubric was presented, and students were reminded of the objectives in each column. At the beginning of class, the researcher encouraged students to look at their own assessment rubrics while the same rubric was also projected onto the board. All the objectives that were written on the rubrics were further explained as to why their fulfilment were of importance in their speaking tasks. (2) Opportunities for students to ask questions about the rubric were presented when the researcher was present in the classroom, one or two times per week. This was a major part of the constant guidance that the researcher tried to give during the classes as has been proven necessary when working with rubrics (De Silva, 2014). Students were

also presented opportunities to fill out the checkboxes on the rubrics. However, this was not obligatory or supervised.

Statistical Analysis

The datasets containing Amotivation scores and SDT subtype scores were input into SPSS (version 26). First some tests of frequency were run, wherein the groups' mean score on all the tests and the normal distribution were calculated – this to establish the homogeneity of the individuals' scores. This is important since parametric tests such as a paired *t*-test or *ANOVA* assumes that there is homogeneity between data in which situations these types of tests would be run (Laerd Statistics, n.d.).

If data was not found to be homogeneous, a non-parametric approach was chosen instead with a *Mann Whitney U* test first to check for significant differences *between* the two groups. Then, a *Wilcoxon Signed Ranks* test was conducted to find significant differences *within* the groups before the experiment, and after. To clarify, the first test deals with significant discrepancies between the groups, while the second test deals with significant discrepancies between the pre-test and the post-test results.

Reliability and Validity

The experimental method is designed to be as reliable and valid as possible. A control group is used where circumstances such as L1s, gender distribution and socioeconomic backgrounds are as close to the experimental group as possible to check after a cause-and-effect relationship when applying a treatment to the experimental group. With a cause-and-effect relationship internal validity is established. What may be of bigger importance is the external validity, or if results can

be applied outside of the context of the study (Nunan, 1992). Efforts have been made to choose a school that looks typical for a Swedish schoolboy looking at the average score. These background statistics have been specified in the *Participants and Context* section. However, there are still variations in background context, the biggest one may be that the school is rural. *Rural school* will in this study be defined by a school 25-45 minutes from an urban area with a population over 3000. In comparisons with similar studies this difference in context must be taken in account since one must consider that the societal factors in this classroom may be different from the ones prevalent in an urban school. Given that it was not possible to randomly assign students to the experimental and control groups, it was important that these two classes were evenly matched when it comes to background abilities, organisation, and teaching methods (Nunan, 1992, p. 41). However, there is still little possibility of generalizing these results without a much bigger and randomized sample.

When it comes to internal validity and reliability, the questionnaire by Noels has been subjected to internal reliability and validity analysis and deemed as sufficient in both areas (2000). Although, the items have been translated and efforts have been made to make the translations as correct as possible. Since the original questionnaire was meant for university students, some questions were also adapted to fit EFL learners in a Swedish secondary school. In the adaptation, it was important that the questions would convey the same meaning and thereby match the English equivalent. When the post-test was designed, most questions were articulated differently or kept completely but with the original meaning still prevailing. The result of the adaptation can be seen in the Appendix.

Concludingly, the overall validity and reliability of this study is reliant on equivalent groups and that the internal validity and reliability of the questionnaire by Noels (2000) is retained even after

the translation. In all cases however, caution should be exercised when generalizing the results of this study overall. This study could be considered a pilot study, where research techniques and material can be replicated with larger and more generalizable samples.

Ethical Considerations

Students were informed that participation was voluntary and that their responses would remain confidential and anonymous. The questionnaires were filled out during class time without a time limit. Negations against participation was collected by the means of an online consent form directed at parents. The reason as to why negations were collected and affirmations were not was to direct responsibility from the teachers, as not to weigh them down with additional labour. In general, efforts were made not to affect the teachers' work all too much by delegating work relating to the research to the researcher. When questions were asked about the study or contents of the study, the researcher would answer these. Contents of the experiment would be explained by the researcher and the teacher would only be required to do the work that they normally do.

Results

The aim of this essay is to investigate possible effects of implementing explicit assessment rubrics on L2 motivation in a Swedish EFL context. Specifically, the study seeks to determine if explicit assessment rubrics affects the experimental group's Amotivation or Self-Determination. First the data analysis in Amotivation will be presented and then data analysis of the self-determination subtypes will be presented. A disclaimer to the reader is that the statistical tests that are conducted, a Mann Whitney U-test and a Wilcoxon Signed Ranks test as a part of the calculations transform data into ranks. Therefore, scores will sometimes be provided as 'ranks' instead

of their original score. What could be said from the results will be discussed in the Discussion section.

Amotivation

An initial holistic overview of the statistics shown indicates that there has been an increase in Amotivation in both groups. The mean score of the students in the control group reports an increase from 1.7 Amotivation to a score of 1.9 while the experimental group has increased from 1.3 to 1.5. What will be tested is if there is a probability (P) that there is a significant difference between the scores and between the groups. The confidence interval used is 90 % which means a significant difference would be $P < 0.10$.

Table 3. *Descriptive Statistics*

		Descriptive Statistics				
Group		N	Mean	Std. Deviation	Minimum	Maximum
Control	amot1	16	1.662	.7219	1.0	3.4
	amot2	16	1.875	1.1636	1.0	5.4
Exp	amot1	15	1.307	.5063	1.0	2.8
	amot2	15	1.493	.6923	1.0	3.6

First of all, the Kolmogorov-Smirnov as well as Shapiro-Wilk tests were used in order to calculate normal distribution. As the tests showed no normal distribution in the data as indicated by the significance ($P < 0.10$) and data was ordinal, it was found appropriate to use a non-parametric test.

Table 4. *Tests of Normality*

		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
Group	Statistic	Df	Sig.	Statistic	df	Sig.	
amot1	control	.179	16	.179	.861	16	.020
	exp	.272	15	.004	.681	15	.000
amot2	control	.226	16	.028	.751	16	.001
	exp	.238	15	.022	.729	15	.001

a. Lilliefors Significance Correction

Secondly, the Mann Whitney U test was conducted in order to find any differences between the groups. This is done by running the test on every individual's average Amotivation score on the post-test, the variable here is called *amot2*. The test showed that there was a significant difference in Amotivation according to Table 5 ($P < 0.10$). As Table 6 can tell us, the experimental group's Amotivation has increased.

Table 5. *Test Statistics for Amotivation*

amot2

Test Statistics ^a	
Mann-Whitney U	161.000
Wilcoxon W	351.000
Z	-1.670
Asymp. Sig. (2-tailed)	.095

a. Grouping Variable: group

Thirdly, the Wilcoxon Signed Ranks test was conducted to find significant differences within the groups between the pre-test and the post-test. Table 6 shows the increases in value per each group between the two tests. *Positive Ranks* shows how many individuals increased their Amotivation and *Negative Ranks* shows how many individuals decreased their Amotivation. From this table, we see how the control group had an approximate equal distribution of Positive, Negative, and Tied ranks while the experimental group had substantially more individuals improving or tying their Amotivation from the first test before the treatment with only three getting a lower Amotivation.

Table 6. *Improvements between tests*

Ranks

Group			N	Mean Rank	Sum of Ranks
control	amot2 - amot1	Negative Ranks	6 ^a	7.42	44.50
		Positive Ranks	8 ^b	7.56	60.50
		Ties	2 ^c		
		Total	16		
exp	amot2 - amot1	Negative Ranks	3 ^a	3.00	9.00
		Positive Ranks	7 ^b	6.57	46.00
		Ties	5 ^c		
		Total	15		

a. amot2 < amot1

b. amot2 > amot1

c. amot2 = amot1

When it comes to the statistical significance, there was no statistically significant change in the Amotivation expressed by the control group, while there was a statistically significant change in Amotivation for the experimental group ($P < 0.10$) as we can see in Table 7.

Table 7. *Test Statistics for Amotivation*

Test Statistics^a

group		amot2 - amot1
control	Z	-.505 ^b
	Asymp. Sig. (2-tailed)	.613
exp	Z	-1.913 ^b
	Asymp. Sig. (2-tailed)	.056

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

Altogether this result tells us that the control group has not acquired any overall change in Amotivation apart from some increased distribution of high and low scores which results in a higher standard deviation. Since the groups were unrelated to each other, this must be considered a result of some other random factor that we could not account for. The result also tells us of an experimental group that has overall increased their Amotivation scores with statistical significance after the experiment. Since both tests show a significant difference between groups as well as within groups this strongly indicates that the instructional treatment has affected the experimental group's Amotivation.

Self-Determined Motivation

What can be said about the Self-Determination of the individuals in this study is that most changes in the subtypes are marginal in both groups. In Table 8, the mean scores of individuals are displayed with the minimum and maximum score.

Table 8. *Descriptive Statistics*

		Descriptive Statistics				
Groups		N	Minimum	Maximum	Mean	Std. Deviation
exp	ex.reg1	21	1.00	6.00	3.3771	1.23473
	ex.reg2	19	1.33	5.33	3.1905	1.12413
	int.reg1	21	1.00	5.33	2.9010	1.27024
	int.reg2	19	1.33	5.33	2.9795	1.18345
	id.reg1	21	1.66	6.00	4.5357	1.27981
	id.reg2	19	2.00	6.00	4.2426	1.41856
	know1	21	1.33	5.66	3.2190	1.31296
	know2	19	1.33	5.00	3.2605	1.07427
	accomp1	21	1.33	6.00	3.2667	1.41728
	accomp2	19	1.00	5.00	2.8216	1.15175

	stim1	21	1.00	5.00	3.1076	1.32538
	stim2	19	1.00	5.33	2.4347	1.20161
	Ex.Mot1	21	1.33	5.55	3.6046	1.08830
	Ex.Mot2	19	2.22	5.11	3.4709	.99496
	Int.Mot1	21	1.33	5.33	3.1978	1.19486
	Int.Mot2	19	1.33	5.11	2.8389	.97512
	Valid N (listwise)	15				
control	ex.reg1	29	2.33	6.00	3.7872	1.05083
	ex.reg2	24	1.33	5.33	3.9417	1.13220
	int.reg1	29	1.00	6.00	2.8600	1.21304
	int.reg2	24	1.33	5.33	3.3092	1.22815
	id.reg1	29	1.00	6.00	4.1755	1.46787
	id.reg2	24	1.66	6.00	4.3300	1.34782
	know1	29	1.33	6.00	3.2617	1.39521
	know2	24	1.00	6.00	3.2187	1.67584
	accomp1	29	1.00	5.66	3.1000	1.37779
	accomp2	24	1.00	5.66	2.8713	1.43349
	stim1	29	1.00	6.00	3.3879	1.49096

stim2	24	1.00	6.00	2.9421	1.61735
Ex.Mot1	29	1.61	5.67	3.6076	1.00845
Ex.Mot2	16	2.78	5.33	4.0319	.87292
Int.Mot1	29	1.22	5.55	3.2499	1.22154
Int.Mot2	15	1.22	5.00	2.8184	1.19368
Valid N (listwise)	15				

Although, there are few discernible differences in the data, one subtype stands out when conducting the between-groups Mann Whitney U-test. In Table 9, the difference between the groups' score of External Regulation in the post-test (*ex.reg2*) is deemed significant ($P < 0.10$). This means that the two groups have gone opposite ways when it comes to External Regulation. According to Table 8, the experimental group has decreased its mean External Regulation score after the treatment while the control group has increased it. In Table 10 this is visualised by how the experimental group has a smaller mean rank than the control group.

Table 9. *Test Statistics for Self-Determination Subtypes*

Test Statistics^a						
	ex.reg2	int.reg2	id.reg2	know2	accomp2	stim2
Mann-Whitney U	137.500	192.000	223.500	221.500	222.000	197.000
Wilcoxon W	327.500	382.000	413.500	521.500	412.000	387.000
Z	-2.224	-.883	-.111	-.160	-.148	-.762
Asymp. Sig. (2-tailed)	.026	.377	.912	.873	.883	.446

a. Grouping Variable: Groups

Table 10. *Ranks in the experimental group and in the control group.*

Ranks

	Groups	N	Mean Rank	Sum of Ranks
ex.reg2	exp	19	17.24	327.50
	control	24	25.77	618.50
	Total	43		
int.reg2	exp	19	20.11	382.00
	control	24	23.50	564.00
	Total	43		
id.reg2	exp	19	21.76	413.50
	control	24	22.19	532.50
	Total	43		
know2	exp	19	22.34	424.50
	control	24	21.73	521.50
	Total	43		

accomp2	exp	19	21.68	412.00
	control	24	22.25	534.00
	Total	43		
stim2	exp	19	20.37	387.00
	control	24	23.29	559.00
	Total	43		

Just like the way amotivation was analysed, the Wilcoxon Signed Ranks test will now follow where the changes in subtype scores before the treatment and after the treatment are analysed. This time, Table 11 is sufficient to show that only one subtype has a significant change ($P < 0.10$) in Stimulation. From Table 8 we can discern that both groups decreased their Stimulation score. However, this change can be tracked in both groups and so must be affected by some other factor that both groups experienced. Neither External Regulation is found to have a significant difference after the treatment for the experimental group, however this does not negate the results of the Mann Whitney U-test that clarified a clear difference between the two groups.

Table 11. *Test Statistics for Subtypes*

		Test Statistics^a					
Groups		ex.reg2 - ex.reg1	int.reg2 - int.reg1	id.reg2 - id.reg1	know2 - know1	accomp2 - ac- comp1	stim2 - stim1
exp	Z	-1.368 ^b	-.341 ^b	-1.133 ^b	-.379 ^b	-1.574 ^b	-1.815 ^b
	Asymp. Sig. (2-tailed)	.171	.733	.257	.705	.116	.070
control	Z	-.366 ^c	-1.446 ^c	-.314 ^c	-.751 ^c	-1.336 ^b	-1.724 ^b
	Asymp. Sig. (2-tailed)	.714	.148	.754	.453	.182	.085

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

c. Based on negative ranks.

There could also be a point in looking at the self-determination continuum overall. What this means is instead of analysing measures of specific subtypes, an analysis could be made on several subtypes. An average score was calculated based on the three subtype scores on each side of the continuum, a score of Extrinsic motivation and a score of Intrinsic motivation. When running a Mann Whitney U-test to compare how the groups scored on the post-tests, a significant difference was found between the groups' Extrinsic motivation ($P < 0.10$), with the variable being *Ex.Mot2*. This can be seen in Table 12.

Table 12. *Test Statistics for Extrinsic and Intrinsic motivation*

Test Statistics^a				
	Ex.Mot1	Ex.Mot2	Int.Mot1	Int.Mot2
Mann-Whitney U	296.000	97.500	298.000	136.500
Wilcoxon W	731.000	287.500	529.000	256.500
Z	-.167	-1.806	-.128	-.208
Asymp. Sig. (2-tailed)	.867	.071	.898	.835
Exact Sig. [2*(1-tailed Sig.)]		.071 ^b		.837 ^b

a. Grouping Variable: Groups

b. Not corrected for ties.

The groups were similar in most regards when it comes to their score on separate self-determination subtypes on the self-determination continuum. However, the score of External Regulation was found to be significantly less for the experimental group, meaning that the instructional treatment led to individuals being less motivated by external goals such as getting a job in the future. The possible reasons behind this discrepancy need a broader review of possible determining factors.

Discussion

In this section results from the data analysis will be discussed accordingly with the two research questions and possible explanations. The results when it comes to amotivation and the self-determination continuum will be discussed first and then follow by possible explanations to the results according to self-determination theory and previous research.

Between-groups and within-groups testing show that the instructional treatment has led to a significant increase in amotivation. This means that the experimental group with 90 % confidence ($P < 0.10$) has increased its amotivation between the two tests as well as increased it in comparison to the control group. An increase in the parameter amotivation means that the experimental group has become less self-determined to study English after the experiment.

The statistical analysis displayed a relation between the self-determination continuum and the instructional treatment. The experimental group's scores of extrinsic regulation and extrinsic motivation were significantly lower than the control group's corresponding scores. Since there was no detected significant change between the two tests according to the Wilcoxon Signed Ranks test, this must be a result of the control group achieving a higher External Regulation without the explicit assessment rubrics and the experimental group achieving a lower score with the explicit assessment rubrics. This means that the experimental group was less self-determined to study English by external criteria such as grades. Why the treatment with explicit assessment rubrics might have given these effects will be discussed in the following section.

Autonomy, relatedness, and competence

Since the three major staples deciding self-determination are autonomy, relatedness, and competence, it is most logical to assume that the decrease in motivation detected is a result of a lack in one of these three. Previously, there was a discussion on whether explicit objectives may lead to a sense of ‘instrumentalism’, less student autonomy and thereby less motivation (cf. p. 7). This notion is based on the underlying critique that Torrance (2007) brought up. However, what was also stated earlier on was that instrumentalism would most likely involve externally regulating factors. Assessment rubrics would then lead to an increase in extrinsic motivation and not the decrease that happened in this case. Of course, this essay’s initial assessment of instrumentalism might also be wrong since instrumentalism has not been scientifically tested in correlation to self-determination (Jönsson & Panadero, 2020). Otherwise, a new point could be made about what kind of motivation explicit assessment rubrics encourage.

This point being that explicit assessment rubrics may not encourage students to learn in any way more than they would in the case of this control group. One of the major responses that Jönsson and Panadero (2020) provided to the ‘instrumentalism’ critique was that controversy might naturally erupt in the question as it relates to ‘whether we value this compliance with expectations as a desirable outcome of education or not, or whether intended learning should be seen as superficial and conformance per se’ (p. 10). While this is a very valid point, these results indicate the kind of compliance that would be preferable in order for students to achieve objectives is actually decreasing with the use of explicit assessment rubrics. If explicit assessment rubrics could be considered an efficient way of prompting criteria compliance that leads to achievements, then External Regulation would likely rise or remain the same with this intervention. This result thereby questions the notion brought up by Jönsson and Panadero.

The results also stand in contrast with the results found by Balan and Jönsson (2018). While the study by the latter found an increased self-regulation among students, this study finds less self-determination. This might however be explained by some difference in theory as self-regulation is a different concept. Self-determination theory was chosen as it is a theory that takes in account socio-contextual factors which it could be argued self-regulation does not (Dörnyei & Ushioda, 2011). Brookhart and Chen (2015) pointed out that assessment rubrics and their effects on motivation have received mixed scientific results and that the scientific endeavours have been generally undeveloped. Therefore, research on assessment rubrics' relation with motivation need to be standardized with theories that take in account social contexts.

Conclusion

The aim of this essay was to investigate possible effects of implementing explicit assessment rubrics on L2 motivation in a Swedish EFL context. To do this the essay needed to resolve two questions: 'Is there a relation between using explicit assessment rubrics and L2 motivation in a group of EFL students at a Swedish secondary school?' and 'Is there a relation between using explicit assessment rubrics and Self-Determination in a group of EFL students at a Swedish secondary school?' (p. 2). There are reasons to believe the two initial research questions have been answered. The first relation that was found with the usage of explicit assessment rubrics was that motivation to learn English as a secondary language decreased. The second was that students seemed to be less externally motivated to learn English as a secondary language. The major conclusion that can be drawn from the results of this essay is that no matter how much or little 'instrumentalism' explicit assessment rubrics seems to promote in a Swedish EFL context, they do not in all cases prompt positive L2 Motivation effects. What practising teachers can learn from this study is that they, depending on the kinds of motivation that they want to facilitate, need to

take in account that explicit assessment rubrics do not necessarily act as efficient external regulators of motivation. The effects do not differ from the usage of assessment rubrics taken directly from the curriculum.

In this study, I have argued that while the research field investigating explicit assessment rubrics' effects on student performance is well developed, it is still lacking when it comes to its width. There is still a need for more research on motivation, lower levels of education and EFL classrooms etc. This study has contributed to the field by applying research to these areas. It is one of the first studies to consider socio-contextual factors when measuring motivation in the field of assessment rubrics. Hopefully, it will contribute to standardizing research on possible instrumental teaching practises such as assessment rubrics. The study is therefore not only relevant to this research field but perhaps also to the broader field developing formative assessment practises. Although, for this standardization to happen, theoretical frameworks that take into account socio-contextual factors like the one in this study need to be used.

The results of this study seem to add further content to the research on assessment rubrics' effects on motivation, providing to the mixed results that the research is characterized by. Therefore, it is preferable that research on assessment rubrics receive a more substantial entry into the research field on L2 motivation. The field has already established several elaborate models on measuring motivation like the self-determination theory displayed here, the self-regulation theory and the L2 motivational self-system et cetera (cf. Dörnyei & Ushioda, 2011). Some of the broader conclusions that can be drawn for future research are (1) research on assessment rubrics examining specific motivational concepts need to be standardized, (2) The concept of 'instrumentalism' that has reoccurred throughout this essay need to be examined further.

References

- Black, P. & Wiliam, D. (1998) Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.
- Brookhart, M.S. & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, 67(3), 343-368.
- De Silva, R. (2014). Rubrics for Assessment: Their Effects on ESL Students' Authentic Task Performance. 4th CELC Symposium Proceedings, National University of Singapore.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum.
- Dörnyei, Z. and Ushioda, E. (2011) *Teaching and Researching: Motivation*. (2nd Ed.) Pearson, Harlow.
- Gagne, M., & Deci, E. L. (2005). Self-Determination Theory and Work Motivation. *Journal of Organizational Behavior*, 26(4), 331–362. <https://doi.org/10.1002/job.322>
- Hult, F. (2017). More than a lingua franca : Functions of English in a globalized educational language policy. *Language, Culture and Curriculum*, 30(3), 265-282.
- Jescovitch, L., Scott, E., Cerchiara, J., Doherty, J., Wenderoth, M., Merrill, J., Urban-Lurain, M. & Haudek, K. (2019). Deconstruction of Holistic Rubrics into Analytic Rubrics for Large-Scale Assessments of Students' Reasoning of Complex Science Concepts. *Practical Assessment, Research & Evaluation*, 24(7), 1-12.

Mann-Whitney U test using SPSS Statistics. (n.d.) Laerd Statistics. Accessed December 11, 2020 from <https://statistics.laerd.com/spss-tutorials/mann-whitney-u-test-using-spss-statistics.php>

Noels, K.A., Pelletier, L.G., Clément, R. and Vallerand, R.J. (2000) Why are you learning a second language? Motivational Orientations and Self-Determination Theory. *Language Learning* 50, 57-85.

Nunan, D. (1992). *Research methods in language learning*. Cambridge, Cambridge University Press.

Pandero, E. & Jönsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9(0), 129-144.

doi:<http://dx.doi.org/10.1016/j.edurev.2013.01.002>

Popham, W.J. (1997). What's wrong – and what's right – with rubrics. *Educational Leadership* 55(2), 72–5.

Reddy, Y.M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435–448.

<http://dx.doi.org/10.1080/02602930902862859>.

Sadler, D.R. (2009) Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34:2, 159-179, DOI: [10.1080/02602930801956059](https://doi.org/10.1080/02602930801956059)

Skolverket. (2011). *Kunskapsbedömning i skolan - praxis, begrepp, problem och möjligheter*.

Skolverket. (2014). Bedömningsstöd i biologi i grundskolan. Accessed September 24, 2020

from <https://www.skolverket.se/undervisning/grundskolan/bedomning-i-grundskolan/bedomningsstod-i-amnen-i-grundskolan/bedomningsstod-biologi-grundskolan>

Skolverket. (2018). *Allmänna råd om betyg och betygsättning*.

Skolverket. (2019) *Läroplan för grundskolan samt för förskoleklassen och fritidshemmet* (Rev. Uppl.).

Skolverket. (2019). SALSA - Skolenheters resultat av slutbetygen i årskurs 9 med hänsyn till elevsammansättningen. Accessed November 20, 2020 from <https://www.siris.skolverket.se/siris/f?p=101:165:0::NO:RIR:>

Torrance, H. & Pryor, J. (1998) *Investigating formative assessment: teaching, learning and assessment in the classroom*. Buckingham, Open University Press.

Torrance, H., Colley, H., Garratt, D., Jarvis, J., Piper, H., Ecclestone, K. & James, D. (2005) *The impact of different modes of assessment on achievement and progress in the learning and skills sector*. London, LSDA for the LSRC.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education*, 3(14), 281-294.

Vallerand, R.J. (1997) Toward a hierarchical model of intrinsic and extrinsic motivation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 271-360). San Diego, CA: Academic Press.

Vetenskapsrådet. (2015) *Formativ bedömning på 2000-talet - en översikt av svensk och internationell forskning*. (Hirsch, Å. & Lindberg, V. Eds.)

Zimmerman, Barry. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology*, 11. 307-313. 10.1016/0361-476X(86)90027-5.

Appendix A

The questionnaire will be written here, first follows a section which measures Amotivation.

Att bry sig om Engelska

Kryssa i hur mycket du tycker påståendena stämmer.

Jag förstår inte varför jag pluggar engelska.

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag tycker att jag slösar bort tiden när jag pluggar engelska.

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Det finns ingen idé med att plugga engelska.

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag bryr mig inte om att plugga engelska.

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag förstår inte varför jag ska jobba på engelskalektionerna

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Section break. Next follows a section that measures the three subtypes, External Regulation, In-
trojected Regulation and Identified Regulation.

Jag pluggar engelska för att...

Jag pluggar engelska för att det är förväntat att jag ska plugga engelska

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att få ett jobb i framtiden

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att få en hög lön i framtiden

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar till engelska för att visa att jag kan bidra till samhället genom språk

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att jag skulle skämmas om jag inte kunde prata med folk från
andra länder

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att om jag bara hade kunnat ett språk hade jag skämts

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att jag skulle vilja vara en person som kan prata engelska bra

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att jag skulle vilja vara en person med självförtroende att prata med engelsktalande människor

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att jag skulle vilja vara en person som pratar flera språk

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Section break. At this point, intrinsic motivational subtypes are measured. Knowledge, Accomplishment and Stimulation.

Jag pluggar engelska för att jag tycker om att lära mig om engelsk kultur

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att jag tycker om känslan när jag lär mig vad nya ord betyder

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att det är roligt att lära mig om engelsktalande människor

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för hur det känns när jag förvånar mig själv på engelskalektionerna

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för hur skönt det känns när jag klarar av något svårt på lektionen

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för en skön känsla som uppstår när jag gör något svårt på en engelskalektion

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att det känns mycket skönt att lyssna på engelska när det talaras

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att det känns mycket skönt att prata engelska

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att jag skulle vilja vara en person som kan komma i kontakt med andra kulturer

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

The next questionnaire was purposefully made to measure the same parameter as in the first questionnaire however with differing questions.

Att bry sig om Engelska 2

Kryssa i hur mycket du tycker påståendena stämmer.

Det finns ingen anledning för mig att plugga engelska

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

När jag gör engelskauppgifter känns det som att jag slösar bort tiden

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag bryr mig inte om att plugga engelska

1	2	3	4	5	6
Stämmer inte alls väl				Stämmer mycket väl	

Jag bryr mig inte om att plugga engelska.

1	2	3	4	5	6
Stämmer inte alls väl				Stämmer mycket väl	

Det finns ingen anledning till att jobba på engelskalektionerna

1	2	3	4	5	6
Stämmer inte alls väl				Stämmer mycket väl	

Section break. Next follows a section that measures the three subtypes, External Regulation, In-
trojected Regulation and Identified Regulation.

Jag pluggar engelska för att...

Jag pluggar engelska för att det är förväntat av mig

1	2	3	4	5	6
Stämmer inte alls väl				Stämmer mycket väl	

Jag pluggar engelska för att jag behöver det för att få jobb

1	2	3	4	5	6
Stämmer inte alls väl				Stämmer mycket väl	

Jag pluggar engelska för att kunna tjäna mycket pengar i framtiden

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar till engelska för att man som svensk medborgare ska kunna språk

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att jag skulle skämmas om jag inte kunde prata med folk från andra länder

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att om jag bara hade kunnat ett språk hade jag inte kunnat bidra till samhället

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att jag skulle vilja vara en person som kan prata engelska bra

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att jag skulle vilja vara en person med självförtroende att prata ned engelsktalande människor

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att jag skulle vilja vara en person som kan komma i kontakt med andra kulturer

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Section break. At this point, intrinsic motivational subtypes are measured. Knowledge, Accomplishment and Stimulation.

Jag pluggar engelska för att jag tycker om att lära mig om engelsktalande länder

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att jag tycker om att lära mig nya ord

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att det är roligt att lära mig om nya språk

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för hur det känns när jag förvånar mig själv på engelskalektionerna

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för hur skönt det känns när jag klarar av något svårt på en engelskalektion

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att det är skönt att lyssna på engelska

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att jag njuter av att prata engelska

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Jag pluggar engelska för att det är skönt att läsa engelska texter

1 2 3 4 5 6

Stämmer inte alls väl

Stämmer mycket väl

Appendix B

	✓	E	✓	C	✓	A
--	---	---	---	---	---	---

Muntlig framställning	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<p>Du pratar begriplig och enkel engelska.</p> <p>Du använder rätt ord flera gånger när du pratar engelska.</p> <p>Många meningar som du säger hänger ihop.</p> <p>Du pratar delvis fritt i presentationen.</p>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<p>Du har få hinder när du pratar engelska.</p> <p>Du använder rätt ord de flesta gånger när du pratar engelska.</p> <p>De flesta meningar du säger hänger ihop.</p> <p>Du pratar mestadels fritt i presentationen.</p> <p>.</p>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<p>Du pratar avancerad engelska utan hinder.</p> <p>Du använder rätt ord när du pratar engelska.</p> <p>Du talar sammanhängande.</p> <p>Du pratar fritt i presentationen.</p>
	✓	E	✓	C	✓	A
Muntlig interaktion	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<p>Du kan göra dig förstådd i samtal med andra.</p> <p>Du använder rätt ord flera gånger när du pratar engelska.</p> <p>När du inte kan ord kan du ibland förklara ordet på ett annat sätt.</p> <p>Du använder ett begripligt uttal när du talar engelska med andra.</p>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<p>Du har få hinder när du pratar engelska med andra.</p> <p>Du använder rätt ord de flesta gånger när du pratar engelska.</p> <p>När du inte kan ord kan du förklara ordet på ett annat sätt.</p> <p>Du använder ett tydligt uttal när du talar engelska med andra.</p>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<p>Du pratar engelska utan hinder med andra.</p> <p>Du använder rätt ord när du pratar engelska.</p> <p>När du inte kan ord kan du förklara ordet på ett annat sätt.</p> <p>Du använder ett tydligt uttal när du talar engelska med andra.</p>

