Students' use of a rubric for research theses

Frans J. Prins, Renske de Kleijn & Jan van Tartwijk

Students' use of a rubric for research theses

Frans J. Prins[a]*, Renske de Kleijn[b] and Jan van Tartwijk[b]

[a]*Faculty of Social and Behavioural Sciences, Department of Education, Utrecht University, Utrecht, The Netherlands;* [b]*Faculty of Social and Behavioural Sciences, Educational Development and Training, Utrecht University, Utrecht, The Netherlands*

A rubric for research theses was developed, based on the manual of the American Psychological Association, to be used as an assessment tool for teachers and students. The aim was to make students aware of what is expected, get familiar with criteria, and interpret teacher and peer feedback. In two studies, it was examined whether students use and value these functions. In the first study, a rubric was provided to 105 Educational Sciences students working on their bachelor's thesis. Questionnaire data indicated that students did value the rubric for the intended functions, although rubric use was not related to ability. In a panel interview, teachers stated that the number of proficiency levels should be increased to be able to distinguish between good and excellent students adequately, and that a criterion concerning student's role during supervision should be added. Therefore, in the second study, 11 teachers were interviewed about their motives to give high grades and about the supervision process. This lead to an extra criterion concerning student's role during supervision and an additional proficiency level to assess excellent performance. It is argued that an adequate course organisation is conditional for the rubric's effectiveness.

Keywords: rubrics; teacher feedback; peer feedback; bachelor's thesis

Introduction

Nearly all undergraduate and graduate students in the Netherlands conclude their studies by conducting a research project and writing a thesis. Writing a research thesis is a complex task, as: (a) students are given a lot of leeway to determine the focus of their research themselves, (b) the duration of the project is often at least half a year, which is longer than general coursework, (c) students are supervised individually or in groups rather than in a classroom setting and (d) the goal of the task is twofold, given that students learn to conduct and report research, and are assessed on their research and writing skills. Quite often, students are confronted for the first time with all that they have learnt (theory, academic skills and research crafts), and have to combine and utilise these skills to complete their research project.

For students, conducting research at a university is a *learning task* in which they are guided by a teacher. It is common for teachers to devote a lot time providing feedback on drafts of students' research papers. It is questionable whether the feedback is received and understood well by all students. For instance, de Kleijn et al. (2013) found that feedback on master's theses is perceived to be not very goal-related. As Sadler (2010) argues, in order to be successful at conducting a learning

---

*Corresponding author. Email: f.j.prins@uu.nl

task, and to apply teacher feedback effectively, students should be aware of what is expected from them. Sadler labels that *task compliance*, which refers to 'the congruence between the type of response stipulated in the task specifications and the type of response actually submitted by a student' (Sadler 2010, 543). It is our experience that many students in higher education have trouble understanding what is exactly expected from them during their research project and writing their research thesis. Secondly, according to Sadler (2010), students should be able to make *quality judgments*, which often require holistic judgements in which multiple criteria are attended to simultaneously. Thirdly, students should have deep knowledge of and be able to use abstract *criteria* properly. Most criteria in writing assessment do not have sharp boundaries and, thus, these abstract criteria may be problematic for students. In our opinion, a well-designed rubric for research theses may be part of the solution for the troubles students encounter with task compliance, quality and criteria. Consequently, we felt the urge to design a rubric for research theses that let students know what is expected of them as well as let teachers provide effective and efficient feedback.

In this paper, we present two studies that are aimed at designing a rubric for providing feedback, as well as for assessing the quality of a research thesis in the domain of the social sciences. Before we present the studies, we will address several possible functions of a rubric and the design principles we adopted when designing the first prototype of our rubric.

Functions of rubrics

A commonly accepted definition of a rubric is 'a document that articulates the expectations for an assignment by listing the criteria, or what counts, and describing levels of quality from excellent to poor' (Andrade and Du 2005, 1). Sadler (2009a) described a rubric as a cross-tabulation of criteria against standards. Criteria are arranged so that there is one row for each criterion, and each cell in that row describes the characteristics of a particular level or standard for that criterion. An assessor nominates the cell that best characterises the quality of each student work on each criterion (Sadler 2009a, 163). Several studies have investigated the summative use of rubrics and thus their validity, reliability and educational consequences (for an overview see Jonsson and Svingby 2007; Brookhart and Chen 2015). In the present study, the *formative* use of rubrics (for students) will be addressed. In other words, in this study, rubrics were used as a tool to guide and support learning rather than only to assess learning (e.g. Reddy and Andrade 2010).

Rubrics may serve different functions (e.g. Wolcott and Legg 1998). For instance, rubrics can be used as an *assessment tool for teachers* to assess students' products. Rubrics can also be used to make students *aware of what is expected of them* (e.g. Jonsson 2014), or to let students familiarise themselves with the criteria for a good research article. Moreover, rubrics can be used to *interpret the feedback* that a student receives from a teacher, and can be applied as an *assessment tool for students* to review the products of their peer in a peer assessment setting or self-assessment (e.g. Panadero and Romero 2014). A recent review by Panadero and Jonsson (2013) on the formative use of rubrics revealed that there are several ways for this to mediate improved student performance. Rubrics may increase transparency, reduce anxiety, aid the feedback process, improve students' self-efficacy and support student self-regulation. In addition, rubrics can be used to map curricular goals and design assessments (e.g. Tractenberg, Uman, and McCarter 2010).

In our study, we designed a rubric for research papers in order to make students more aware of the assessment criteria, use these criteria for peer review, and understand teacher and student feedback. The organisation of the course in which the rubric will be used allows for all these functions of the rubric. As O'Donovan, Price, and Rust (2001) point out, a rubric will be of limited practical use if presented on its own, without the benefit of explanation, exemplars and the opportunity of discussion.

Designing a rubric for the assessment of research papers

There are many types and formats of rubrics. Some contain several pages of specific criteria, whereas other rubrics fit on one page. Some rubrics are *holistic*, where scoring is based on the premise that the whole is worth more than the sum of its parts (Wolcott and Legg 1998, 71), or where a single score is assigned to a product based on the overall impression (Weigle 2002). Other rubrics are *analytical*, where writing products are rated on several aspects of writing or criteria, rather than given a single score, and thus providing more detailed information about the student's performance (Weigle 2002). Recent reviews by Reddy and Andrade (2010) and by Panadero and Jonsson (2013) have taken this into account. Reddy and Andrade (2010, 445) stated that a large majority of the studies reviewed did not describe the process of development of rubrics to establish their quality. Panadero and Jonsson (2013, 142), in turn, argued that there seems to be no studies of the effects of different rubric designs (such as holistic vs. analytical, few vs. several proficiency levels, etc.). Thus, the available reviews did not provide clear cut guidelines or principles for designing our rubric. Taking this into account, in this study we outline the design decisions that are to be made when designing a rubric, the different arguments that might play a role and the specific decisions that we made.

The first decision that has to be made while designing a rubric is whether to adopt a *bottom-up approach*, where teachers and students are asked to design criteria and quality standards, or a *top-down approach*, where expert knowledge and theory concerning writing and conducting research would be the major input for the design of criteria and quality standards. A *bottom-up* approach may have the advantage of creating ownership and support among teachers and students to actually facilitate the use of the rubric once it is designed. On the other hand, it can have the danger of resulting in a context-dependent idiosyncratic assessment instrument. Vice versa, a *top-down* approach has the advantage of being more generally applicable and possibly being more valid as it is based on theory. The disadvantage could be that the intended users do not accept the rubric. Our aim was that our rubric for research papers would be useful for all social sciences, and not only for our education department. Therefore, as Timmerman et al. (2011) did when they developed their rubric for science writing based on literature, we decided to adopt a *top-down* approach. We used the APA publication manual (American Psychology Association, 2009) to create the content of the rubric. The APA manual contains a description of the structure of a manuscript as well as reporting standards for research in the behavioural and social sciences. Experts from multiple specialisations have contributed to the realisation of the APA manual. Manuscript elements such as an introduction, method, results and a discussion section, and what they should contain, are accurately and concisely described in the APA manual. Another advantage of using the APA manual

for the content of the rubric for research papers is that students are able to consult the APA manual when they need extra information about the reporting standards.

Next, it has to be decided whether the rubric is *assessor-oriented* and/or *user-oriented* (Weigle 2002). An assessor-oriented scoring rubric is intended to guide the rating process and thus to support the teachers (e.g. Timmerman et al. 2011), whereas the user-oriented rubric is designed with a focus on providing useful information to support students with interpreting feedback and assessment. Since our main purpose of the rubric was to make students more aware of the assessment criteria, use these criteria for peer review, and understand teacher and student feedback, we chose a user-oriented rubric. This decision is reflected in study 1, where we explicitly asked students to indicate whether the content of the rubric was understandable and/or whether parts of the rubric needed clarification.

The third design decision concerns whether a *holistic* or *analytic* rubric is designed. Many writing specialist prefer analytical scoring rubrics because they are considered to provide more detailed information about student performance (Weigle 2002). Biggs and Tang (2011) and Sadler (2009a, 2009b), on the other hand, advocate holistic judgements in higher education. In holistic grading, Sadler (2009a) describes how the assessor progressively builds up a complex mental response to a student's work, making a qualitative judgement on its overall quality. The assessor may provide a rationale in a summary or as running comments on various features of the work. The descriptions in the cells of a holistic rubric are intended as indicative rather than definitive (Sadler 2009a). Biggs and Tang (2011) argue that analytic assessment gives students feedback on how well they are doing on each important aspect of the writing product or other task, but the *value* of the writing product is how well it makes the case or addresses the question *as a whole*. In our opinion, the quality of the different manuscript elements (e.g. introduction, method, results and discussion) should be judged as a whole. It may well be the case that when looking at a particular element of a research paper analytically, it can be judged as sufficient, whereas this element may essentially miss the point (e.g. in a method section describing the wrong instruments in a very accurate way, or in a discussion section describing a theory very accurately and concisely without relating it to and answering the research question).

Finally, a design decision has to be made with respect to the number of levels and how to go from rubric scores to a final grade. Part of this decision is determined by the range of performances that can be reasonably expected from of the population of students the rubric has been designed for (Weigle 2002). Our assessment instrument should at least make pass/fail decisions. Hence, there should be two levels. Moreover, we also wanted to distinguish between sufficient and good, which resulted in creating three levels. Creating the level descriptions was not an easy task. The starting point was the APA manual, from which we extracted the standards of the highest level (level 3). For the descriptors of level 2, we retained from level 3 what, in our opinion, is absolutely necessary for students on a bachelor or master level in the social sciences to score sufficiently (the 'must haves'); or, in other words, to pass. We decided to avoid adjectives such as 'sufficient' and 'good' because these already indicated the levels, and are thus as descriptors and do not contribute any added value. Instead, we chose adjectives such as *logical*, *accurate* and *complete*. For instance, for level 2, a summary of literature does not have to be perfect or complete, but most of the necessary information should be present. For level 3, the summary should be complete. Similarly, for level 2, with the exception of a few, most of the issues in the literature summary should be connected to the

research questions. For level 3, all important concepts described in the literature summary should be covered in the research questions in a logical manner.

To go from rubric scores to a grade, we adopted a global (holistic) system. In the Netherlands, there is a grading system on a 10-points scale, in which 6 is just passing and 10 is outstanding. We decided that all scores at level 1 would translate into a grade of 4 or lower, all scores at level 2 would translate into a grade between 6 and 7.5, and all scores at level 3 would translate into a grade of 8 or higher. When scores vary, for instance when students for some criteria score at level 1 and for other criteria at level 2, this would lead to a grade between 4 and 6.

The first prototype

Based on the sixth edition of the APA manual, a two-page rubric for research papers was developed by the three authors of this paper (see Table 1). The rubric comprised six criteria and three proficiency levels: 'insufficient' (level 1), 'must have' (level 2), and 'nice to have' (level 3). Level 1, more or less, represents the absence of the must-haves of level 2 in students' products. Since it is important to indicate in a rubric what students must show in their product to match a particular level (Jonsson 2014), instead of what is lacking, we decided that *displaying* level 1 has no added value for letting students know what is expected of them. In comparison, Timmerman et al. (2011) chose to display the lowest level as 'not addressed', which does not really provide extra information for the students and assessors. The rubric can be used to assess the quality of a research plan (for which the criteria concerning results and discussion are not applicable) and of a research paper.

We added a *feedback and assessment form* to the rubric in order to make it easy for teachers to provide additional feedback (i.e. providing the running comments, see Sadler 2009a) and to briefly substantiate their scores. For instance, in the feedback and assessment form, teachers can explain to students who scored at level 1 what was lacking, what was insufficient and how this can be repaired. After all, our rubric for research papers has an explicit formative function.

Study 1: Evaluating the first prototype

The main purpose of the first study was to examine how students used the first version of the designed rubric for research papers. We were particularly interested in students' awareness of the assessment criteria used in the rubric, their use of these criteria for peer review, and their understanding of teacher and student feedback.

The following research questions were addressed:

(1) How do students use the rubric?
(2) How do students perceive feedback that is provided based on the rubric?
(3) Do students with different ability levels value and use the rubric differently?
(4) Do students who use the rubric more achieve better?

Method

Participants

Participants of the first study were 200 students of two cohorts of educational sciences who enrolled in a bachelor's thesis project at a Dutch University, and 16

Table 1. Rubric for research theses – version 1.

S. no.	Criterion	Level 2 – must have	Level 3 – nice to have
1.	Manuscript structure[a]	All elements are connected and organised	All elements are **logically** connected **and keypoints within sections are** organised
		Research questions, hypotheses, research design, results, inferences and evaluations are related	Research questions, hypotheses, research design, results, inferences and evaluations are related **and form a consistent and concise argumentation**
2.	Introduction (APA manual 2.05)	The scientific/societal problem the study addresses is introduced, and it is made clear why this deserves new research	The scientific/societal problem the study addresses is introduced, it is made clear why this deserves new research **and that it makes an important contribution to the field**
		A summary of literature pertinent to the problem is provided	A **complete** summary of **the** literature pertinent to the problem is provided
		The research question is connected to (or the hypothesis is derived from) the problem description and the literature review	The research question is **logically** connected to (or the hypothesis is derived from) the problem description and the literature review, **and gives a precise and accurate description of what the researcher wants to find out**
3.	Methods[b] (APA manual 2.06)	The variables are operationally defined, information is provided about the sample and the sampling procedures, about the methods for data gathering, and about the design of the study	**The information provided permits experienced investigators to replicate the study:** the variables are **conceptually and** operationally defined, information is provided about the sample and the sampling procedures, about the methods for data gathering **and their reliability and validity**, and about the design of the study
4.	Results[c] (APA manual 2.07)	Information is provided about the collected data, the analyses and their results. This information enables the reader to evaluate the discussion that is to follow	The information that is provided about the collected data, the analyses and their results is **accurate and complete (e.g. missing data, significance, effect sizes)**. This information enables the reader to evaluate the discussion that is to follow
			If interventions or experimental manipulations were used, evidence is provided on whether they were delivered as intended and all important adverse events and/or side effects are detailed
5.	Discussion[d] (APA manual 2.08)	The results are used to answer the research question and are interpreted in relation to the work of others	The results are used to answer the research question **or underpin a statement of (non) support of hypotheses**
			The results are interpreted in relation to the work of others **and**

Table 1.   (*Continued*).

S. no.	Criterion	Level 2 – must have	Level 3 – nice to have
			are taking the limitations or weaknesses of the study into account
		Based on the interpretation, the implications and importance of the findings are discussed	Based on the interpretation, the implications and importance of the findings are discussed **and related to the scientific/societal problem the study addresses. Unresolved or newly arisen problems are described**
6.	Organisation and writing style (APA manual ch3)	The manuscript has an organised structure	The manuscript has a **sound** organised structure **with concise headings and paragraphs, continuity in words, concepts and thematic development**
		Language is precise	Language is precise **and scientific** (see 3.07 APA manual). **Transitional words maintain the flow of thought** (see 3.05 APA manual)
		Cited references in text and in a reference list are all according to the APA manual sixth edition	Cited references in text and in a reference list, tables and figures are all according to the APA manual sixth edition

[a]For information concerning manuscript structure, see APA manual sixth edition, section 2.05 (where the relation between hypotheses and research design is emphasised), section 2.07 (where relation between results, analyses, and the discourse that is to follow is emphasised) and section 2.08 (where the relation between hypotheses, results, theoretical and practical consequences is addressed).
[b]For a research plan, a specific paragraph concerning the *analyses* is required.
[c]For a research plan, a Results section is not applicable.
[d]For a research plan, a Discussion section is not applicable.

teachers who supervised research projects of bachelor and master students, and used the rubric for formative and summative assessments.

Instruments

The following data were collected:

Student's perception of the function of the rubric. Student's perception of the rubric was assessed with a questionnaire. Questions could be answered on a five-point scale, varying from 'totally disagree' to 'totally agree'. Five questions concerned assessment criteria (e.g. for me, the concepts in the rubric were understandable), three questions about working on the research plan (e.g. the rubric has helped me with assessing the quality of our research plan), five questions about working on the first draft of the research paper (e.g. the rubric helped me with determining points of improvement for our first draft of the research paper), eight questions about giving and receiving peer reviews (e.g. during providing my peer feedback, the rubric has

helped to determine the strong points of the first draft of the research paper I had to review) and five questions about the function of the rubric while working on their thesis (e.g. the rubric has helped me with understanding the feedback of my fellow students from the feedback form). For each category, students also had the opportunity to add remarks via open-ended questions about the use of the APA manual, and about benefits and barriers of the rubric.

Feedback perceptions. Using the scales of Strijbos, Narciss, and Dünnebier (2010), feedback perceptions were measured with a student questionnaire with 12 items for student and 12 items for teacher feedback. The questionnaire comprised four scales of three items each: fairness, usefulness, acceptance and willingness. Items could be answered on a five-point scale, varying from 'totally disagree' to 'totally agree'. An example item is 'I am satisfied with the feedback that we received'.

Students' ability level. For students who provided their consent, ability level was mapped by calculating the mean of the grades for five previous courses during their bachelor in Educational Sciences (on methodology, educational psychology, educational design, assessment and educational change).

Teacher's opinions. In a panel interview with teachers who supervised several research projects of bachelor and master students, and used the rubric for formative and summative assessments, we discussed the benefits and obstacles of the rubric. We asked them to reflect on: (1) their experience with using the rubric, (2) elements that were missing, (3) the levels of the rubric, (4) advantages, (5) what could help them using the rubric more effectively and (6) what needed to be addressed in a manual for teachers and students.

Procedure

The student research projects were conducted in groups of three students and lasted 20 weeks. The course was organised in phases. In the first phase, students worked on a research plan in which they composed a literature summary, formulated research questions and set up a methodology for their study. In an interactive session, students presented their research plan orally to their fellow students and received peer feedback. In the second phase, student groups gathered, analysed and interpreted data, and wrote a draft of their research thesis. At the end of the second phase, each student provided written peer feedback (a peer review) on the first draft of the research thesis of a fellow student group. Each student group received three peer reviews and teacher feedback. The teacher distributed the first drafts of the research theses among the students. The peer review, as well as the teacher feedback on the first draft of the research thesis, was based on the rubric. In the third phase, student groups used the peer feedback and teacher feedback to write their final draft of the research thesis. At the end of phase three, the student groups presented their research project at a student conference. Students were graded on three documents: their research plan (group work), a peer review (individual) and a final research thesis (group work). The research plan and the final draft of the research thesis were evaluated by two teacher assessors using the rubric.

The rubric was provided at the beginning of the course when students started preparing their research plan, in order to familiarise them with the standards and

criteria of a research paper. The coordinator of the course clarified in a plenary session when and how the rubric can be used. Students as well as teachers were explicitly asked to indicate whether the content of the rubric was understandable, and/or whether parts of the rubric needed clarification. Minor textual changes were applied to exclude ambiguity. The way the course was organised, and the way the rubric was used during several phases of the course, is in line with a social constructivist approach to assessment (Rust, O'Donovan, and Price 2005): i.e. students were actively engaged with assessment criteria and feedback, as providers and receivers. Moreover, the course provided opportunities to discuss and suggest improvements to the rubric, and to compare exemplars (their own work, that of a fellow student group, as well as three peer reviews and teacher feedback). All this may increase the practical use of the rubric (see O'Donovan, Price, and Rust 2001).

Questionnaires were administered at the last plenary assembly of the course. Data were analysed using both quantitative (reliability, factor analysis, correlation and regression analysis) and qualitative analyses (grounded theory).

Results

Use of the rubric: four functions

Factor analysis was conducted on the items concerning the student's perception of the function of the rubric of cohort 1, using maximum likelihood estimation, an Oblimin rotation and a scree plot for determining the number of factors, as suggested by Costello and Osborne (2005). The scree plot indicated four or two factors. We considered the four factor solution to be most informative, with eigenvalues ranging from 1.31 to 5.35. The first factor was interpreted as *exploring the criteria* (start phase), the second factor was labelled *judging, underpinning and advising on peer work*, the third was labelled *understanding (peer and teacher) feedback on own work* and the fourth was labelled *understanding and applying the criteria*. Scale scores were computed by averaging the items that loaded highest on a factor with a minimum factor loading of 0.40 (Stevens 1992). Means, standard deviations and reliabilities in both cohorts are shown in Table 2. The four factors are interpreted to reflect four functions that the rubric could have for students. The mean scores for the four factors indicate that students in general used the rubric for all of these four functions. Students used the rubric most for the function of understanding and applying the criteria, and least for the function of understanding (peer and teacher) feedback on their own work.

Feedback perceptions

A factor analysis was conducted with the same procedure for questionnaire items concerning students' perceptions of teacher and peer feedback. The scree plot indicated five or three factors, and based on the interpretation and its relation to the original scales of Strijbos, Narciss, and Dünnebier (2010), we opted for the five factor solution with eigenvalues ranging from 1.30 to 5.98. Again scale scores were computed by averaging the items that loaded highest on a factor, with a minimum factor loading of 0.40. In Table 2, the mean scores for the feedback perception scales are presented. As can be seen, in general, students perceived the feedback as fair and useful and did not reject it. Interestingly, it appears that students perceived the feedback of their teacher more positively than the feedback from their peers.

Table 2.   Descriptive statistics for rubric-related variables ($N = 193$) and achievement variables ($N = 143$).

Measure	N items	M	SD	α Cohort 1	α Cohort 2
Rubric function					
Exploring the criteria (start phase)	6	3.63	0.68	0.86	0.81
Judging, underpinning, advising on peer work	5	3.66	0.77	0.82	0.86
Understanding (peer and teacher) feedback	3	3.38	0.94	0.77	0.82
Understanding and applying the criteria	6	3.90	0.65	0.84	0.82
Feedback perception					
Fairness and usefulness peer feedback	6	3.68	0.69	0.87	0.90
Willingness to use the peer feedback	3	4.13	0.78	0.90	0.89
Fairness and usefulness teacher feedback	6	4.08	0.89	0.92	0.96
Willingness to use the teacher feedback	4	4.51	0.79	0.82	0.83
Rejection of the feedback	4	2.01	0.78	0.70	0.70
Academic ability	5	7.41	0.59	0.71	0.68
Research plan	1	7.33	0.58	–	–
Peer review	1	7.93	0.72	–	–
Final thesis	1	7.75	0.66	–	–

Relation between rubric use, feedback perception, ability and achievement

Correlational analyses were performed to explore relations between student use of the rubric, feedback perceptions and achievement (see Table 3). The results show that students who used the rubric for providing feedback on peer work and for understanding received feedback also perceived their feedback more positively in terms of fairness and usefulness, and willingness to use. With respect to academic ability, no relations were found with rubric use. This indicates that the rubric is not systematically used more or differently by high- or low-ability students. With respect to student achievement it was found that students, after receiving lower grades for their research plan, started using the rubric more for understanding and applying the criteria. Additionally, students who rejected their feedback more also received lower grades for their own peer review. Lastly, students who found their peer feedback fair and useful also received higher grades for their final thesis.

Teacher panel

The teacher panel revealed six issues that should be addressed for the teachers to fully accept the rubric as a feedback and assessment instrument. First, the teachers urged for an extra proficiency level to distinguish excellent from good performance. Second, they missed that the abstract was not mentioned in the rubric. Third, in their opinion, reporting on reliabilities is a 'must have' and should be moved from level 3 to level 2. Fourth, during grading, all teachers take the supervision process and the role of the students during that process into account. Thus, they suggest that the role of the process needs to be integrated in the rubric. Fifth, teachers need a clear model to transform the rubric scores into a grade. Finally, teachers would benefit from the inclusion of pre-structured feedback in the method section to save time when providing running comments.

Table 3. Correlations between students' reported function of the rubric, feedback perceptions and academic ability.

S. no.	Measure	1	2	3	4	5	6	7	8	9
Rubric function										
1.	Exploring the criteria (start phase)	—								
2.	Judging, underpinning, and advising on peer work	0.277^*	—							
3.	Understanding (peer and teacher) feedback	0.419^*	0.197^*	—						
4.	Understanding and applying the criteria	0.524^*	0.603^*	0.419^*	—					
Feedback perception										
5.	Fairness and usefulness peer feedback	0.123	0.176^*	0.199^*	0.109	—				
6.	Willingness to use the peer feedback	0.093	0.250^*	0.163^*	0.258^*	0.579^*	—			
7.	Fairness and usefulness teacher feedback	0.078	0.036	0.185^*	0.125	−0.095	−0.036	—		
8.	Willingness to use the teacher feedback	0.074	0.156^*	0.178^*	0.223^*	−0.056	0.044	0.743^*	—	
9.	Rejection of the feedback	0.002	0.068	−0.009	−0.003	-0.166^*	−0.132	-0.186^*	-0.238^*	—
	Academic ability	−0.106	0.117	0.037	0.011	0.075	0.117	−0.023	0.034	−0.042
	Research plan	−0.017	−0.060	0.0352	$-\mathbf{0.176}^*$	0.154	0.126	0.043	−0.036	−0.063
	Peer review	0.072	−0.001	−0.020	0.020	0.044	0.017	0.156	0.108	$-\mathbf{0.204}^*$
	Final thesis	−0.012	0.107	0.077	0.023	$\mathbf{0.172}^*$	0.150	0.134	0.058	−0.167

$^*p < .05$.

Discussion study 1

First, the results of this study indicated that our two-page rubric, based on guidelines of the APA manual, fulfilled four functions for students. We base this conclusion on the factor structure we found that resembled the four functions of the rubric, and on the relatively high mean score on these four scales. The factor 'exploring the criteria' can be interpreted as an important condition for Sadler's (2010) *task compliance* to occur, as this indicates that students actually submit something that matches the task specifications. Therefore, students might need to explore the criteria in the first place. The factor 'judging, underpinning and providing advice on the work of peers' can be interpreted in terms of Sadler's (2010) *quality judgements,* where he argues that students should be able to make quality judgements in order to perform well. The factor 'understand (peer and teacher) feedback on own work' is assumed to contribute to students' quality judgement, as feedback on their own work can be seen as an example of a quality judgement. The factor 'understanding and applying the criteria' is interpreted in terms of Sadler's (2010) notion of students needing to be able to have deep knowledge of, and be able to use, abstract criteria. We thus see that the functions that the rubric can fulfil in the learning process of the student can be interpreted in terms of task compliance, quality judgments and using criteria. Relative high mean scores on these scales indicated that students perceived the rubric as useful when considering these four functions.

Second, it was found that the peer and teacher feedback provided with the rubric was perceived as fair and useful. Furthermore, students indicated a willingness to use the feedback. We interpret these findings as confirming that the rubric supports the feedback process between peers, as well as between the teacher and the students. Interestingly, the feedback of the teacher is found to be perceived more positively than the peer feedback. Students might expect that the teacher understands and applies the criteria better than peers do, which might explain the difference. An alternative explanation may be that, given that teachers are ultimately responsible for grading student work, students will always assign greater value to the feedback of the teacher compared to peer feedback.

Third, correlations in Table 3 suggest that students who score a low grade for their research plan were more likely to apply the rubric's criteria a bit later on during the course of their research project. Interestingly, ability was not related to student's use of the rubric. This suggests that it is mainly the task performance that triggers the use of the rubric, not whether students are strong or weak. These findings are in contrast to those of Price and Rust (1999), who found that the rubric was mainly used by the most motivated students. However, when looked at in a different light, it might be the case that students are motivated to use the rubric when they realise they might actually need it to improve their performance. When students used the rubric more often for judging, underpinning and advising on peer work, they considered the peer feedback as fair and useful, and were more willing to use the peer feedback. Thus, it seems that using the rubric makes the feedback that is received with the rubric more effective, which is in line with the findings of Price and Rust (1999).

Fourth, the results of the panel discussion indicated that the rubric needs additional features for the teachers to be enthusiastic about using it. Two issues can be easily addressed: including the abstract in the rubric and moving reliabilities from level 3 to level 2. Two other issues are taken up in a follow-up study, the second

study of this paper: the development of an extra proficiency level to distinguish excellent performance from good performance, and the development of a process rubric. The importance of the supervision process for students has been demonstrated in several studies (de Kleijn et al. 2012), but this seems to be the first study to outline that, when grading student's theses, the process also plays a role for teachers. Finally, in a third follow-up study, which will be planned and conducted in the near future, the score-grade transformation scale will be developed, and possibilities for using pre-structured feedback will be explored.

Study 2: Augmenting the rubric

Study 2 was aimed at augmenting the first prototype of the rubric for research reports, with an extra level for excellent performance and an extra criterion for student's role during the supervision process. First, with respect to the extra level for excellent performance, the APA did not provide enough ground for development, nor did we find useful starting points in literature about scientific writing or excellence in higher education. For determining the excellence level, we argue that it would be possible that excellence is shown in just one specific aspect of the research paper, for instance, in the introduction, where students could add new pioneering insights to the literature; or in the method as well as the results sections, where students could show extraordinary methodological skills. We, therefore, decided to not describe a next proficiency level but to make a list of points of excellence, related to the rubric criteria. Given the many possibilities for points of excellence, the list is not exhaustive. Actually, these points of excellence are comparable to the so-called latent criteria of Sadler (1989), who described them as 'those in the background, triggered or activated as occasion demands by some (existential) property of the work that deviates from expectation' (134). In other words, they pop up when student work triggers it. A holistic rubric leaves room for latent criteria. In order to create this list of points of excellence, we conducted interviews with experienced social science teachers who recently assigned a high grade to one or more students for their research thesis. This way, we were not only able to detect the actual points of excellence for content-related as well as process-related criteria, but also the teacher's reasons why they considered these particular points as excellent.

Second, with respect to the process criterion, we used a combination of *top-down* and *bottom-up* approach, and checked whether both approaches would converge to a single conclusion. For the *bottom-up* approach, we conducted interviews, and during the data collection and analyses, we came across the work of Hadwin, Wozney, and Pontin (2005). Therefore, we used their theory concerning the student's role during the educational conversations using a *top-down* approach. In their study, Hadwin et al. examined the transition of self-regulatory control from teacher to graduate student during so-called instructional conferences. Their theory about *self-regulatory ownership* distinguished amongst three levels: (1) *teacher-direct regulation*, in which teachers have the lead by demonstrating and instructing, (2) *co-regulation*, in which the teacher guides or prompts students to do the regulating themselves, or students request or prompt teachers to show them how to self-regulate, (3) *student-direct regulation*, in which students actively control and reflect on self-regulatory processes and products. These levels have clear differences. Moreover, the last level, where students show self-regulatory ownership, can be considered as desirable for students in higher education.

Method

Participants

Participants of study 2 were 12 experienced social science teachers who in 2012–2013 assigned a high grade (i.e. 8.5 or higher on a 10-point scale) to one or more bachelor and/or master students for their research paper. Teachers not from educational science did not use the rubric for their grading.

Instruments

Individual semi-structured interviews were conducted, based on the following questions: (1) Why did you grade this particular research paper with an 8.5 or higher? Be as specific as possible. (2) Did you grade one of your students a 6 or lower? If so, why? Be as specific as possible. (3) When do you give a student a failing grade? (4) How do you determine your grade? (5) Is the supervision process included in the grade you give to your students? If so, how?

Procedure

Coordinators of the social sciences curricula were asked which teachers in their department had assigned a high grade (8.5 or higher) to a student for their research paper in the past year. These teachers were invited for a 20-min interview. The interviews were recorded and transcribed.

Analyses

With respect to the points of excellence, two researchers analysed the interviews and made a list. They discussed their individual lists until consensus was reached and based on that a final list was made. With respect to the role of the research process, the interviews were analysed in relation to the three levels of Hadwin, Wozney, and Pontin (2005): teacher-direct regulation, co-regulation and student-direct regulation.

Results

The list with points of excellence was categorised in four main categories: theoretical insights of the study, carrying out of the study, writing style, and self-regulatory ownership. The examples in the list in Table 4 are by no means exhaustive and should be considered illustrative. Moreover, many teachers stated that an indication of a point of excellence is when students go beyond what is expected from them, outlined in the course objectives, and when they deliver a thesis that, depending on the results, could be publishable in a peer-reviewed journal.

With respect to the process criterion, the interviews not only provided examples for the level of self-regulatory ownership of Hadwin, Wozney, and Pontin (2005), but also for the level of teacher regulation and co-regulation. These examples are presented in Table 5.

Discussion study 2

This study revealed that excellence in thesis projects can be described as exceeding the course objectives or curricular goals. This is still quite general, as we know from

Table 4.   Points of excellence.

The theoretical insights of the study
Original approach in theoretical framework
Own conceptualisation of variables
Clear cut theoretical framework and research questions
New (recent) literature added to an existing theoretical framework
Outstanding practical relevance
The carrying out of the study
Analyses conducted that were not in the curriculum
Without help of the supervisor conducted complex analyses
Large sample size
Combination and integration of multiple methodologies
Substantial pilot study
Applying an adequate design in complex circumstances
Applying a complex design with complex research questions
Invested largely in raising the quality of the instruments
Writing style
The research paper is in well-written English instead of in the mother language
Self-regulatory ownership
Worked very independently
Introduced own ideas that exceed the standard curriculum
Setting challenging goals and meeting them
Proactive attitude during supervision meetings e.g. by setting the agenda
Communication with supervisor as equals
Self-proposed solutions rather than asking the supervisor open ended questions
Constructively critical towards supervisor feedback
Outstanding willingness to learn

previous studies that the goals of thesis projects are by no means clear cut and understood in the same way by students and teachers (de Kleijn et al. 2013). Therefore, based on the interviews, we created a list of examples of ways in which students can exceed these objectives or goals.

A second contribution of this study was that, based on a combination of a *top-down* approach using the work of Hadwin, Wozney, and Pontin (2005) and *bottom-up* approach using interviews with experienced thesis supervisors, we propose a rubric for judging the research process and the interaction between teacher and student.

This leads to version 2 of the rubric for research papers (see Table 6). The next step is to evaluate the use of version 2, with the extra proficiency level and the extra criterion about the supervision process. Future studies could address whether these process levels challenge students to self-regulate during supervision meetings with the teacher.

General discussion

The aim of our studies was to design a rubric for a research paper that is mainly user-oriented, that is, to make students more aware of the assessment criteria, use these criteria for peer review, and understand teacher and student feedback. The two studies resulted in a promising second version of the rubric, with four levels (from insufficient to excellence) that is ready to be evaluated.

Since there are many types and formats of rubrics, we made specific design decisions in order to create a rubric that serve the purposes that we had in mind. Unfortunately, the majority of the studies on the effects of rubrics lack a description

Table 5.   Levels of self-regulation.

Teacher regulation
The student meets her/his commitments and obligations
The student responds to the supervisor's instruction, explanations and demonstrations
The student responds to and processes the supervisor's feedback
Co-regulation
The responsibility for the research project is shared between student and supervisor
The student asks for help, confirmation, guidance, feedback and information
When asking the supervisor for help, the student provides relevant and concrete questions

of the development process of rubrics (Reddy and Andrade 2010, 445). We argue that, in order to bring research on the effectiveness of rubrics a step further, researchers should explicitly describe the aims of a rubric and connect these with their design decisions. That way, effects of rubrics with different aims and designs (such as holistic vs. analytical, few vs. several proficiency levels, etc.) can be studied systematically and clear cut guidelines or principles for designing rubrics can be provided. Furthermore, a strength of our second study is that we combined a *top-down* and *bottom-up* approach for augmenting the first version of the rubric. Empirical results concerning teacher perspectives on student's role during supervision converged with Hadwin, Wozney, and Pontin (2005) theory, which provides a strong theoretical and empirical basis for the different proficiency levels for the supervision process. In addition, with our way of creating a proficiency level for excellence we introduced a new approach for designing proficiency levels: not a specific description of the top level but a list of optional points of excellence, leaving room for latent criteria (Sadler 1989). Results showed that students, of both low and high ability, indicated that they used our rubric for the different purposes we designed it for, especially when their grade for their research plan was rather low.

During the design process, we noticed that just designing a useful and a positively perceived assessment instrument is not enough, and that the organisation of the course (i.e. providing the opportunity to use the rubric during the learning process) is a precondition for the effectiveness of the rubric. As Reddy and Andrade (2010) summarise it:

> The implication seems to be that simply handing out a rubric cannot be expected to have an impact on student work: students must be taught to actively use a rubric for self- and peer assessment and revision in order to reap its benefits. (445)

The courses in which the rubric was used contained moments of teacher feedback on the first draft of the research plan and the first draft of the research paper. In addition, *peer feedback* on the first draft of the research paper, based on the rubric, was incorporated into the course. What was still lacking was a manual for teachers and students, in which the rationale for the rubric is explained, and in which it is described how students can benefit from it. Still, that may not be enough for some students to use the rubric effectively. For students who submitted a research thesis of insufficient quality (level 1 for some criteria of the rubric), it may be of great value to discuss not only their performance with the teacher but also the assessment instrument and its use for self-assessment and peer assessment. Either the rubric did not provide sufficient information for these students about what was expected of them, or the rubric was not used. We should therefore aim at exchanging feedback as a dialogical process in which active engagement is played out (Nicol 2010), and in which students can ask for clarification of the assessment criteria.

Table 6.   Rubric for research theses – version 2.

S. no.	Criterion	Level 2 – sufficient	Level 3 – good	Level 4 – points of excellence
1.	Introduction (APA manual 2.05)	The scientific/societal problem the study addresses is introduced, and it is made clear why this deserves new research A summary of literature pertinent to the problem is provided The research question is connected to (or the hypothesis is derived from) the problem description and the literature review	The scientific/societal problem the study addresses is introduced, it is made clear why this deserves new research **and that it makes an important contribution to the field** A **complete** summary of **the** literature pertinent to the problem is provided The research question is **logically** connected to (or the hypothesis is derived from) the problem description and the literature review, **and gives a precise and accurate description of what the researcher wants to find out**	• Original approach in theoretical framework • Own conceptualisation of variables • Clear cut theoretical framework and research questions
2.	Methods[a] (APA manual 2.06)	Information is provided about the sample and the sampling procedures, about the methods for data gathering and their reliability and validity, and about the design of the study The variables are operationally defined	Information is provided about the sample and the sampling procedures, about the methods for data gathering and their reliability and validity, and about the design of the study The variables are **conceptually and** operationally defined **The information provided permits experienced investigators to replicate the study**	• Analyses conducted that were not in the curriculum • Without help of the supervisor conducted complex analyses • Large sample size • Combination and integration of multiple methodologies • Substantial pilot study • Applying an adequate design in complex circumstances
3.	Results[b] (APA manual 2.07)	Information is provided about the collected data, the analyses and their results. This information enables the reader to evaluate the discussion that is to follow	The information that is provided about the collected data, the analyses and their results is **accurate and complete (e.g. missing data, significance, effect sizes)**. This information enables the reader to evaluate the discussion that is to follow	• Applying a complex design with complex research questions • Invested largely in raising the quality of the instruments

(Continued)

Table 6. (Continued).

S. no.	Criterion	Level 2 – sufficient	Level 3 – good	Level 4 – points of excellence
			If interventions or experimental manipulations were used, evidence is provided on whether they were delivered as intended and all important adverse events and/or side effects are detailed	
4.	Discussion[c] (APA manual 2.08)	The results are used to answer the research question	**The results are used to answer the research question or underpin a statement of (non)support of hypotheses**	• New (recent) literature added to an existing theoretical framework
		The results are interpreted in relation to the work of others	The results are interpreted in relation to the work of others **and are taking the limitations or weaknesses of the study into account. Alternative explanations of the results are addressed**	• Outstanding practical relevance.
		Based on the interpretation, the implications and importance of the findings are discussed	Based on the interpretation, the implications and importance of the findings are discussed **and related to the scientific/societal problem the study addresses**	
			Unresolved or newly arisen problems are described	

(Continued)

Table 6. *(Continued).*

S. no.	Criterion	Level 2 – sufficient	Level 3 – good	Level 4 – points of excellence
5.	Organisation, manuscript structure[d], and writing style (APA manual 2.04, ch3)	The abstract is accurate, coherent and readable The manuscript has an organised structure All manuscript elements are connected and organised, i.e. research questions, hypotheses, research design, results, inferences and evaluations are related Language is precise and correct Cited references in text and in a reference list are all according to the APA manual 6th edition	The abstract is accurate, coherent, readable and **concise** The manuscript has a **sound** organised structure **with concise headings and paragraphs, continuity in words, concepts and thematic development** All manuscript elements are **logically** connected and **keypoints within sections are** organised, i.e., research questions, hypotheses, research design, results, inferences and evaluations are related **and form a consistent and concise argumentation** Language is precise, correct, **and scientific** (see 3.07 APA manual). **Transitional words maintain the flow of thought** (see 3.05 APA manual) Cited references in text and in a reference list, **tables, and figures** are all according to the APA manual 6th edition	• The manuscript is written in English
6.	Process[e]	Teacher regulation: • The student meets her/his commitments and obligations (e.g. meets deadlines, is polite, and	Co-regulation: • The responsibility for the research project is shared between student and supervisor	Self-regulatory ownership: • Worked very independently • Introduced own ideas that exceed the standard curriculum • Setting challenging goals and

(Continued)

Table 6. (Continued).

S. no.	Criterion	Level 2 – sufficient	Level 3 – good	Level 4 – points of excellence
		prepare meetings with the supervisor) • The student responds to the supervisor's instruction, explanations, and demonstrations • The student responds to and processes the supervisor's feedback	• The student asks for help, confirmation, guidance, feedback, and information • When asking the supervisor for help, the student provides relevant and concrete questions	• meeting them • Proactive attitude during supervision meetings e.g. by setting the agenda • Communication with supervisor as equals • Self-proposed solutions rather than asking the supervisor open ended questions • Constructively critical towards supervisor feedback • Outstanding willingness to learn

[a]For a research plan, a specific paragraph concerning the *analyses* is required.
[b]For a research plan, a Results section is not applicable.
[c]For a research plan, a Discussion section is not applicable.
[d]For information concerning Manuscript structure, see APA manual 6th edition, section 2.05 (where the relation between hypotheses and research design is emphasised), section 2.07 (where relation between results, analyses and the discourse that is to follow is emphasised), and section 2.08 (where the relation between hypotheses, results, theoretical and practical consequences is addressed).
[e]The process levels are based on the theory of self-regulatory ownership of Hadwin et al. (2005).

As Sadler (1989), (2014) argues, in order to be able as a student to use an assessment instrument effectively, and have knowledge concerning criteria and standards, it may be helpful to look at examples and to look over the shoulder of the expert assessor when (s)he is using the instrument to provide feedback or assess the quality of the product. After that, students should get hands-on experience with the assessment instrument when they use it for peer assessment. For some students, it may be necessary to make this *transfer of assessment knowledge* (O'donovan, Price, and Rust 2004) an explicit part of the course and of the learning process. It could also be valuable to combine rubrics and metacognitive activities during learning, such as self-assessment lessons, as Panadero (Panadero and Jonsson 2013; Panadero and Romero 2014) suggests.

Some issues are not resolved and some needs were still not granted. There is the question about how we should transform the scores for the criteria into a final grade. Should we allow for more differentiated level scores, that is, allow scores such as 1.5 or 2.5 instead of just indicating the level, and thus only allow level 1, 2 or 3? Furthermore, how much higher should your grade be when you have some of the points of excellence in your research paper? We do not favour a specific calculating system to go from level scores to final grade, as that may lead to a too analytic perspective on the use of the rubric. Whatever the choice, it is extremely important to describe our choices accurately in a manual for students and teachers. Finally, there is the wish for pre-structured feedback for teacher's convenience. That would be a next design step, after a positive evaluation of the rubric as it is now.

This rubric has been implemented in one specific situation. Using the dimensions of Price et al. (2011) for describing the type of assessment, our situation could be described as being: (a) simple, as it concerns only one assessment, (b) a combination of both formative and summative, and therefore focused on both learning and measurement, (c) rather individual for teachers as they use the rubric individually, and community-based for students as they work in groups and also provide feedback to other groups, (d) focused on effectiveness rather than efficiency, since considerable time is invested in students learning to understand the criteria through the peer feedback and (e) aimed at quality enhancement rather than assurance, as the main focus was on supporting students to understand the criteria and feedback they receive. Future studies could address whether rubrics, such as these, serve the same functions and are valued equally when they are implemented in ways other than was done in this study.

Disclosure statement

Notes on contributors

Frans J. Prins is an associate professor at the Department of Education and at the Educational Development and Training group of the Faculty of Social and Behavioural Sciences, Utrecht University, the Netherlands. His research interests include the effective use of feedback, fostering learner's motivation and self-regulation in primary, secondary and higher education.

Renske de Kleijn currently works as an assistant professor and educational consultant at the Educational Development and Training group of the Faculty of Social and Behavioural Sciences at Utrecht University in the Netherlands. Her research interests include the motivational effects of assessment and feedback, data literacy of teachers and blended learning.

Jan van Tartwijk is a full professor of Education at the Department of Education of the Faculty of Social and Behavioural Sciences, Utrecht University. He is the chair of the university's Graduate School of Teaching and he also chairs the faculty's Education Development and Training group. His research interests include teacher–student communication in the classroom, assessment and learning in the workplace, and the relation of assessment with variables, such as motivation and creativity.

References

American Psychological Association. 2009. *Publication Manual of the American Psychological Association*. 6th ed. Washington, DC.

Andrade, H., and Y. Du. 2005. "Student Perspectives on Rubric-Referenced Assessment." *Practical Assessment, Research & Evaluation*, 10 (3): 1–11. Retrieved from: http://PAREonline.net/getvn.asp?v=10&n=3

Biggs, J., and C. Tang. 2011. *Teaching for Quality Learning at University*. Maidenhead: Open University Press.

Brookhart, S. M., and F. Chen. 2015. "The Quality and Effectiveness of Descriptive Rubrics." *Educational Review* 67 (3): 343–368. doi:10.1080/00131911.2014.929565.

Costello, A. B., and J. Osborne. 2005. "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from your Analysis." *Practical Assessment Research & Evaluation* 10 (7): 1–9.

Hadwin, A. F., L. Wozney, and O. Pontin. 2005. "Scaffolding the Appropriation of Self-Regulatory Activity: A Socio-cultural Analysis of Changes in Teacher–Student Discourse about a Graduate Research Portfolio." *Instructional Science* 33: 413–450. doi:10.1007/s11251-005-1274-7.

Jonsson, A. 2014. "Rubrics as a Way of Providing Transparency in Assessment." *Assessment & Evaluation in Higher Education* 39 (7): 840–852. doi:10.1080/02602938.2013.875117.

Jonsson, A., and G. Svingby. 2007. "The Use of Scoring Rubrics: Reliability, Validity and Educational Consequences." *Educational Research Review* 2: 130–144.

de Kleijn, R. A. M., M. Tim Mainhard, P. C. Meijer, M. Brekelmans, and A. Pilot. 2013. "Master's Thesis Projects: Student Perceptions of Supervisor Feedback." *Assessment and Evaluation in Higher Education* 38: 1012–1026. doi:10.1080/02602938.2013.777690.

de Kleijn, R. A. M., M. Tim Mainhard, P. Meijer, A. Pilot, and M. Brekelmans. 2012. "Master's Thesis Supervision: Relations between Perceptions of the Supervisor–Student Relationship, Final Grade, Perceived Supervisor Contribution to Learning and Student Satisfaction." *Studies in Higher Education* 37 (8): 925–939.

de Kleijn, R. A. M., P. C. Meijer, M. Brekelmans, and A. Pilot. 2013. "Curricular Goals and Personal Goals in Master's Thesis Projects: Dutch Student-Supervisor Dyads." *International Journal of Higher Education* 2 (1): 1–11. doi:10.5430/ijhe.v2n1p1.

Nicol, D. 2010. "From Monologue to Dialogue: Improving Written Feedback Processes in Mass Higher Education." *Assessment and Evaluation in Higher Education* 35: 501–517.

O'Donovan, B., M. Price, and C. Rust. 2001. "The Student Experience of Criterion-Referenced Assessment (Through the Introduction of a Common Criteria Assessment Grid)." *Innovations in Education and Teaching International* 38 (1): 74–85.

O'Donovan, B., M. Price, and C. Rust. 2004. "Know What I Mean? Enhancing Student Understanding of Assessment Standards and Criteria." *Teaching in Higher Education* 9 (3): 325–335.

Panadero, E., and A. Jonsson. 2013. "The Use of Scoring Rubrics for Formative Assessment Purposes Revisited: A Review." *Educational Research Review* 9: 129–144.

Panadero, E., and M. Romero. 2014. "To Rubric or Not to Rubric? The Effects of Self-assessment on Self-regulation, Performance and Self-efficacy." *Assessment in Education: Principles, Policy & Practice* 21: 133–148. doi:10.1080/0969594X.2013.877872.

Price, M., J. Carroll, B. O'Donovan, and C. Rust. 2011. "If I was Going there I Wouldn't Start from Here: A Critical Commentary on Current Assessment Practice." *Assessment & Evaluation in Higher Education* 36 (4): 479–492.

Price, M., and C. Rust. 1999. "The Experience of Introducing a Common Criteria Assessment Grid Across an Academic Department." *Quality in Higher Education* 5 (2): 133–144.

Reddy, Y. M., and H. Andrade. 2010. "A Review of Rubric use in Higher Education." *Assessment and Evaluation in Higher Education* 35 (4): 435–448.

Rust, C., B. O'Donovan, and M. Price. 2005. "A Social Constructivist Assessment Process Model: How the Research Literature Shows us this could be Best Practice." *Assessment & Evaluation in Higher Education* 30 (3): 231–240.

Sadler, D. R. 1989. "Formative Assessment and the Design of Instructional Systems." *Instructional Science* 18: 119–144.

Sadler, D. R. 2009a. "Indeterminacy in the Use of Preset Criteria for Assessment and Grading." *Assessment and Evaluation in Higher Education* 34 (2): 159–179. doi:10.1080/02602930801956059.

Sadler, D. R. 2009b. "Transforming Holistic Assessment and Grading into a Vehicle for Complex Learning." In *Assessment, Learning and Judgement in Higher Education*, edited by G. Joughin, 45–63. New York: Springer Science+Business Media B.V. doi:10.1007/978-1-4020-8905-3_4.

Sadler, D. R. 2010. "Beyond Feedback: Developing Student Capability in Complex Appraisal." *Assessment and Evaluation in Higher Education* 35 (5): 535–550.

Sadler, D. R. 2014. "The Futility of Attempting to Codify Academic Achievement Standards." *Higher Education* 67: 273–288. doi:10.1007/s10734-013-9649-1.

Stevens, J. 1992. *Applied Multivariate Statistics for the Social Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.

Strijbos, J. W., S. Narciss, and K. Dünnebier. 2010. "Peer Feedback Content and Sender's Competence Level in Academic Writing Revision Tasks: Are they Critical for Feedback Perceptions and Efficiency?" *Learning and Instruction* 20: 291–303.

Timmerman, B., D. C. Strickland, R. L. Johnson, and J. R. Payne. 2011. "Development of a "Universal" Rubric for Assessing Undergraduates' Scientific Reasoning Skills Using Scientific Writing." *Assessment & Evaluation In Higher Education* 36 (5): 509–547. doi:10.1080/02602930903540991.

Tractenberg, R. E., J. G. Uman, and J. McCarter. 2010. "A Mastery Rubric: Guiding Curriculum Design, Admissions and Development of Course Objectives." *Assessment and Evaluation in Higher Education* 35 (1): 17–35.

Weigle, S. C. 2002. *Assessing Writing*. Cambridge: Cambridge University Press.

Wolcott, W., and Legg, S. M. 1998. *An Overview of Writing Assessment*. Urbana, IL: National Council of Teachers of English.