

INCREASING THE RELIABILITY OF SCORES OF CLASSROOM OBSERVATION INSTRUMENTS FOR PRE-SERVICE TEACHERS?

A.M. Broos

S1380117

a.m.broos@student.utwente.nl

EDUCATIONAL SCIENCE AND TECHNOLOGY
UNIVERSITY OF TWENTE

SUPERVISOR

M. Dobbelaer

m.j.dobbelaer@utwente.nl

K. Schildkamp

k.schildkamp@utwente.nl

EXTERNAL ORGANISATION

Christelijke Hogeschool Ede (CHE)

EXTERNAL SUPERVISOR

A. van Tilburg

amtilburg@che.nl

UNIVERSITY OF TWENTE.

TABLE OF CONTENT

Acknowledgement	3
Abstract	4
1. Introduction	5
1.1 Problem Statement	5
1.2 Theoretical Conceptual Framework	7
1.2.1 Student Assessment in Authentic Practice	7
1.2.2 Classroom Observation	8
1.2.3 Rubrics as Observational Instrument	11
1.2.4 Reliability and Validity	14
1.2.5 Theoretical Validity Framework	15
1.3 Research Question and Model	17
2. Research Design and Methods	18
2.1 Research Design	18
2.2 Participants	19
2.3 Instrumentation	21
2.4 Procedure	23
2.5 Data Analysis	24
3. Results	26
3.1 Focus Group 1 – CHE Teachers	26
3.2 Focus Group 2 – ICOs and Mentors	26
3.3 Argument based approach to the Validity of the Original- and Experimental Rubrics	28
3.4 Experiment	30
4. Conclusion and Discussion	36
4.1 Discussion	36
4.2 Limitations and Recommendations	37
4.3 Conclusion	38
Reference List	40
Appendix A: Rubric VM1.2 Extern	43
Appendix B: Informed Consent Form	45
Appendix C: Form Descriptive Characteristics Participants Focus Group	46
Appendix D: Script regarding the Focus Groups	47
Appendix E: Forms with Rubric Elements for the Focus-Group with the CHE- Teachers	48
Appendix F: Questions for the Focus Groups with ICOs	52
Appendix G: Data Focus Group 1, with the CHE-Teachers	53
Appendix H: Adjusted Rubric for the Second Focus Group of Phase 1	56
Appendix I: Script Regarding the Experiment	57
Appendix J: Original Rubric for VM1.2, Fragment 1	59
Appendix K: Original Rubric for VM1.2, Fragment 2	60
Appendix L: Experimental Rubric for VM1.2 Fragment 1	61
Appendix M: Experimental Rubric for VM1.2 Fragment 2	62
Appendix N: Form Descriptive Characteristics Participants Experiment	63
Appendix O: Transcription Focus Group with ICOs	64

ACKNOWLEDGEMENT

First of all I would like to thank the different supervisors that have helped me along the process of this research. As there are:

- M. van Geel – my first supervisor who helped me with my first draft with regard to holistic- and analytic assessment
- M. Dobbelaer – who kept patience when I was under pressure because of data-gathering deadlines and gave feedback on the report
- K. Schildkamp – who supervised me, encouraged me and also gave feedback on the report
- A. van Tilburg – who was my first point of contact for this study's institution. She made this study possible by making time for this study on the ICOs training day.

In addition I would like to thank:

- My girlfriend and fiancé, A. van den Brink – who always has my back and believed in me. She was understanding when I had to spend more time on this study than I had anticipated on.
- My parents – for encouraging me to always challenge myself
- All the CHE-teachers, ICOs and mentors - who gave input during the focus groups or participated in the experiment
- All of the colleagues – who I could spar with in the process

ABSTRACT

One of the challenges teacher educators face today is to assess preservice teachers in a reliable and valid manner. Since teacher quality influences student results (Hattie, 1996), reliable and valid assessment is important to ensure and maintain teacher quality. More and more evidence emerges suggesting that assessment by observation is more suitable than using traditional measures such as knowledge tests or portfolio's. However, assessment by observation comes with several issues, such as the lack of suitable standards, dealing with difference in interpretation of the observation instruments or grades, or different assessment focus of assessors resulting in unreliable results. This study researches rubrics as observation instruments for the assessment of preservice teachers. This results in the following research question: *How can the construct validity of a rubric be improved, and to what degree can this improvement increase the inter-rater reliability of the rubric's scores?* In this mixed method research, it is attempted to improve the construct validity of the rubric by organizing two focus groups (Phase 1), resulting in an experimental rubric. In the quantitative part (Phase 2), the Interrater Reliability (IRR) of the original rubric is investigated and compared to the IRR of the experimental rubric. Therefore, two lessons are being assessed by 70 participants in the control- and experimental group. The results show that, by use of the theoretical validity framework of Kane (2006), the construct validity of the experimental rubric has increased relative to the original rubric. However, the results also show that more adjustments could have been made in order to increase the construct validity. With regard to the IRR, calculations of the ICC, Fleiss' Kappa and an analyzation of dichotomous data with Fisher's Exact test show no increase of the IRR of the experimental rubric relative to the original rubric. Moreover, this study indicates that the IRR for the rubric is poor. Several recommendations are being made with regard to the assessment instrument, for example to compose an assessment manual with general scoring guidelines. Moreover, it is suggested to implement a reliability threshold in order to warrant reliability of preservice teacher assessment.

Keywords

Classroom Observation Instruments, Construct Validity, Inter-rater Reliability, Rubric Use, Pre-service Teachers

1. INTRODUCTION

This first chapter of this research report firstly contains the *problem statement*, where the problem is introduced and the goal and focus of the study is explained. Secondly, in the *theoretical and conceptual framework* the main concepts of this study are discussed, that is “assessment of pre-service teachers in authentic practice”, “Observation Instruments”, “rubrics as observational instrument”, “reliability and validity” and finally the “Theoretical validity framework” of Kane (2006). Thirdly, the *research question* is stated.

In this report, pre-service teachers are referred to as *students*, children attending the primary school are referred to as *pupils*, graduated primary school teachers are simply referred to as *teachers* and teachers of educational programs are referred to as *teacher educators*.

1.1 Problem Statement

One of the common challenges teacher educators face today is to assess students in a reliable and valid manner. More and more evidence emerges suggesting that *observation of students* is more suitable to assess teaching, instead of traditional, less authentic measures (Pecheone & Chung, 2006, p. 23) such as knowledge test, portfolio-assessment or pupil results. Ideally, scores are exact representatives of the assessed knowledge, skill, attitude and/or behaviour of the student, independent of assessment and assessor. However, observations comes with subjectivity, since the assessor needs to interpret what he sees. The challenge arises to keep assessment through observations reliable and valid.

Reliable and valid student assessment is not only challenging, it is also important in order to measure educational quality (Creemers, Kyriakides & Antoniou, 2012). In short, reliability represents the degree to which input repeatedly gives the same output, validity represent the degree to which a measurement measures what it is intended to measure (Vos, 2009). There is a numerous amount of convincing evidence that teacher quality influences student results, as is concluded in the literature study of Hatty (1996) therefore reliable student assessment is needed to ensure their quality. Moreover, student assessment serves a substantial gatekeeping function to determine which teachers are capable to start teaching. Finally, reliable assessment of students is needed for the accreditation of teacher educational programs, which is currently an international issue (Bryant, Maarouf, Burcham, & Greer, 2016).

Nevertheless, assessment through observation brings several validity and reliability issues. With regard to validity, first correct *standards* are needed to assess students (Darling-Hammond, 2006, p.124). However, there are many different views on what those correct standards should comprehend in order to assess whether a student is a good teacher (Crebbin, 1997) and the same goes for the definition of “good” teaching (Sato, 2014). The search for this definition started decades ago but its outcome is still evolving (Margolis & Doring, 2013). Therefore, different classroom observation instruments include different indicators of good teaching. For instance, the Framework for Teaching (FFT) includes indicators with regard to reflecting on teaching and communication with pupils’ parents, where Classroom Assessment Scoring System (CLASS) only focusses on the student’s contact with the pupil. The developers of CLASS are convinced that particularly the contact

between pupil and teacher determines the development of the child and therefore the quality of education (Stuhlman, Hamre, Downer, & Pianta, 2010). Not only different indicators can be used, assessors can also have a different definition or standards on what is "good" (Erdosy, 2003). At the moment there is no consensus on the correct standards for a "good" teacher. Secondly, these standards, or criteria, can be *interpreted differently by the assessor*, "resulting in different meanings for the same scores"(Jeong 2015, p. 2). Therefore, one observation can lead to divergent scores from different assessors. Thirdly, teachers can have a *different expectation* regarding the same grade (Jeong, 2015). One teacher may have higher expectations of an "A" than another. Fourthly, issues immerge on how to correctly observe *teachers' influence* on students' behavior and results. According to Darling-Hammond (2006, p.121) it is difficult to uncover if changes in the student's attitude or performance can be attributed to the student's behaviour or are due to other circumstances. And finally, even if specific criteria are given, the assessor's *focus* can be different, resulting in different scores (Jeong 2015).

With regard to reliability, four causes for a low measurement reliability can be named. With regard to *observation sampling*, organizational factors often limit the amount of observations that can be done by the assessor. However, one observation from one assessor is as half as reliable as three lessons that are observed by three assessors (Hill, 2012). Secondly, regarding the *observation length*, the assessor must have the time to observe long enough to assess reliably. For instance, with an observation length of 30 minutes, more student lessons need to be observed in order to maintain the same level of reliability as when the whole lesson is observed (Hill, 2012). Thirdly, a *between-occasion error* can occur when circumstances unfairly influence the assessment outcomes, therefore decreasing reliability (Vos, 2009). To illustrate, when the assessor feels good during the first observation but is ill during the second, circumstances can decrease a score unfairly therefore decreasing the reliability. At last, *between-rater errors* can cause a decrease in reliability of the scores, for instance when different characteristics of assessors result in different outcomes (Vos, 2009).

This study attempts to gain insight in the reliability and validity of assessment of students through classroom observation in order to ensure minimum quality standards for primary school teachers. The primary schools rely on reliable and valid assessment to guarantee that recently graduated teachers possess the basic skills and knowledge to start working in the classroom. This study especially focusses on the influence of construct validity and the Inter-Rater Reliability (IRR). The influence of improving the construct validity of an observational instrument is investigated as a way to increase the IRR of the scores of such an instrument. This study therefore investigates to what extent an experimental rubric with an increased construct validity can lead to an increase of the IRR for that rubric.

This study therefore can contribute in the search on how to increase the reliability by investigating the scoring inference process of transferring the sample of observations into the observed score, according to Kane's framework (Kane, as cited in Cook, Brydges, Ginsburg, & Hatala, 2015). This framework is further discussed in the theoretical framework.

1.2 Theoretical Conceptual Framework

In the theoretical and conceptual framework the main concepts of this study are discussed. First of all, an introduction on assessment is given. Secondly, classroom observation is introduced as a way to assess students. Thirdly, rubrics are discussed as an instrument to observe students in authentic practice. Fourthly, reliability and validity are explained in order to address the reliability and validity of scores that originate from rubrics. Finally, the theoretical validity framework of Kane (2006) is introduced in order to build an argumentation around a rubrics validity.

1.2.1 Student Assessment in Authentic Practice

Assessment can be defined as “a broad term for the evaluation of ‘something’ for some purpose” (Halpern, 2013, p. 361). Specified to the educational context assessment starts with gathering and using data, drawing conclusions and finally improving educational outcomes based on these conclusions (Halpern, 2013, p. 361). Specified to teachers, teacher performance assessment “specifies what teachers ought to know and be able to do in real teaching contexts” (Caughlan & Jiang, 2014, p. 378).

Assessment can serve different purposes. First of all, *formative assessment* is focused on intermediate learning (Dixson, 2016) and therefore an important tool to teach students how to teach (Caughlan & Jiang, 2014, p. 378). In this case the assessment results can for instance be used halfway through the learning process to give feedback on the process or to further clarify learning goals and teacher’s expectations (American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education [AERA, APA, & NCME], 1999). This is also referred to as *low-stakes assessment*. Secondly, *summative assessment* is focused on measuring progress in a learning sequence (Dixson, 2016). This is also referred to as *high-stakes assessment* since the assessment can be used to determine if an assessed learner is capable to continue a learning process or to practice a profession. Thirdly, assessment also has purpose to *improve teaching* itself, according to Bryant et al. (2016). This is also a form of formative assessment but now the teacher can learn by analyzing the assessment scores. For instance, analyzing a learners’ weaknesses can give new insights for subsequent lessons or parts of the curriculum. Finally, summative assessment of the teacher or educational program can occur when accrediting an educational programs (Pecheone & Chung, 2006, p.25).

The many factors that influence teacher and classroom behavior make the assessment of students a *complex* endeavor, as stated in the problem statement. Due to this complexity, teacher performance assessment is bound to include a certain *subjectivity*, which can be at the expense of the reliability of the assessment. Or the assessment ignores the complexity, increasing reliability but decreasing validity. In other words, an observation instrument can never fully comprehend the complexity of the educational context, but still needs to be specific and directive for high-stakes assessment in order to decrease subjectivity and increase reliability.

There are different ways to assess students on their teaching skills and behavior. For graduated teachers, *pupils’ results* can be used, reasoning that teacher quality is an important factor with regard to pupil achievements (Hattie, 1996). Nevertheless, students often give only a small

percentage of all lessons, therefore their lessons often have insufficient impact on pupil results. Moreover, student results can also be influenced by factors that are not related to their teacher, such as maturation or homeschooling. Secondly, *knowledge tests* can be used to assess a student's content knowledge or pedagogical content knowledge. Thirdly, *portfolios* can be used such as the Performance Assessment for California Teachers (PACT) (Darling-Hammond, 2006, p.130). Fourthly, *pupil surveys* on teacher quality can contribute to assess student's teaching quality (Pavlina, Zorica, Pongrac, 2011), however it is questionable whether high ratings from the students can be attributed to student's teaching qualities (Leventhal, Perry and Abrami, 1977). Finally, *observing authentic practice* of teaching is increasingly considered as a more valid measure than traditional assessment methods that are focused on knowledge (Darling-Hammond, 2006, p. 121). And since this assessment method is currently rather common, this study focusses on observation and its instruments.

1.2.2 Classroom Observation

Observation occurs by almost everyone on every moment of the day. Young children learn by observing and imitating adult behavior (Schmidt, Butler, Heinz, & Tomasello, 2016). Nevertheless, human observations are subjective and therefore can be misleading (Connor-Greene, 2007). The need for validation of observation presents itself. Therefore, there are different types of observational instruments, each with their own content and protocol. To illustrate, several current teacher observation instruments are discussed, selected by convenience sampling (Vos, 2009). Afterwards, recommendations are being summarized with regard to observation instruments.

Types of Observational Instruments for Teachers

There are many different observational instruments for different contexts, with different content and purposes. Observation takes place in multiple sciences, such as biology, sociology, medicine, marketing, psychology and education (Girard & Cohn, 2016). For instance with regard to education, observations can serve to predict student outcomes, assess teacher quality, improve student learning or give feedback (Praetorius & Charalambous, 2018). However, observations regarding other sciences can have different purposes, such as estimating patient health, group dynamics and responses or consumer behavior.

These different context require different types of observational instruments. Girard and Cohn (2016) distinguish categorical and dimensional measurements. Instruments with *categorical measurement* restrict the assessor to a limited amount of possible answers (e.g. slow, medium or fast), whereas *dimensional measurements* restrict the assessor to numerical values, or ratings (e.g. a scale of 1-10). Observations can also be temporally categorized in event-based or interval-based observations (Girard & Cohn, 2016). For *event-based* observation, a predetermined event is observed for the duration of that whole event. *Interval-based* observations contain a predetermined duration in which multiple different events might occur. Finally, different types of instruments may require different *types of assessors*, as there are pre-trained assessors, peers, experts or simply unrelated assessors (Girard & Cohn, 2016).

With regard to teacher assessment, the observation instruments are used to 'measure observable behaviors of teaching practice that are demonstrated in the classroom, such as student engagement, use of questioning techniques, classroom management, and accuracy of presentation of content' (Joe, Tocci, Holtzman & Williams, 2013). There are many *different types* of observation instruments. This can be as simple as an instrument to measure the frequency of certain behavior by *tallying* its occurrence. Other instruments are more complex, like CLASS which contains different dimensions that are divided in *indicators and behavioral markers* for different levels of teacher skills (Pianta, Hamre and Mintz, 2012). Class can be categorized as a *rubric*, which give a description of each possible score for each scoring item, of which an example can be seen in Appendix A. This type of observation instrument will be discussed later on. Moreover, alternations of these rubrics are being made, for instance a *continua model* is developed by Grainger and Weir (2016). This instrument also gives indicators for different standards for each scoring item. A pass standard is given as well as a credit standard, a distinction standard and a high distinction standard. The assessor first estimates which standard is reached, but since this is a continua model the assessor can then indicate how far the student is in their process of reaching the next standard. This gives a more nuanced score and more room for the assessor to implicate the complexity of the assessment practice.

Current Observational Instruments for Teachers

With regard to teacher assessment, several observational instruments that have already been developed are now discussed. To begin with the CLASS, the UTEACH Observation Protocol (UTOP) for Math and Science, the Mathematical Quality of Instruction (MQI) and the Protocol for Language Arts Teaching Observation (PLATO). Finally, the FFT is discussed, which not an observation instrument but a broadly used definition of good teaching in the United States.

First of all, CLASS is developed by Pianta, Hamre and Mintz (2012) and contains three domains (emotional support, classroom organization and instructional support). Each domain is subcategorized in respectively three, three and five dimensions. Student engagement is not part of the domains but it is observed as a detached section. An observation cycle takes 15 minutes in which all three domains are simultaneously observed, which is later transformed to a score. In general, at least 4 cycles are recommended during one year for one teacher; however, this can differ with regard to the purpose of the observation. Indicators, examples and elaborate descriptions are given for each possible score of each dimension.

Regarding subject specific observation instruments, the UTOP has four sections, each divided in six to eight items, which are each scored on Likert scale from "Not observed at all" to "Observed to a great extent" (UTeach Natural Sciences, 2014). In this scale, each score is provided with an elaborate description and examples that regard both frequency of the observed behavior and its quality define the outcome. Secondly, the *Mathematical Quality of Instruction (MQI)* is specified for mathematical education for K-12 students, and is built on 5 elements: "richness of the mathematics; errors and imprecision; working with students and mathematics; student participation in meaning-making and reasoning; and connections between classroom work and mathematics" (K-12 Educational Team, 2010a). Lessons are videotaped, divided into fragments of five to seven-and-a-half minutes and then

scored on all five MQI-elements. The overall lesson is also scored on the five elements. To give a final score for a teacher, three lessons are assessed as described before, by two assessors. Finally, the *Protocol for Language Arts Teaching Observation* (PLATO) is developed for English language arts and consists of 13 elements, such as purpose, intellectual challenge, modeling and guided practice (K-12 Educational Team, 2010b). The observation process consists of cycles of each 15 minutes in which all 13 elements are scored. Not all 13 elements have to be present, but multiple different non-consecutive observations are needed to obtain a reliable and complete picture of the teacher's instruction and lesson content.

The research based *Framework for Teaching* (FFT) by Danielson (2013) framework is built on four domains, namely planning and preparation, the classroom environment, instruction and professional responsibilities. Each domain contains five or six components, which each contain two to five elements, leading to a total of 76 elements. For instance, one elements is "response to student misbehavior" as an element of "Managing student Behavior", which is a component of the domain "The Classroom Environment". Each element can be scored in four levels, namely "unsatisfactory", "basic", "proficient" and "distinguished", and each level is illustrated by critical attributes and examples. Further guidelines on how to observe and score teachers are not implemented in the framework. Training is recommended in order to use the framework.

Recommendations regarding Teacher Observation

The different observation instruments provide several recommendations regarding observing teachers. According to the CLASS, reliability increases with the *number of observations*, since more observations leads to a higher reliability of the measurement (Pianta et al., 2012). Therefore it is recommended to determine a minimum number of observations in order to guarantee reliable assessment. The UTOP on the other hand, is designed only to observe one lesson. Scoring only takes place after each observation, with help of video takes or field notes (UTeach Natural Sciences, 2014).

Secondly, the assessors needs to *remain objective*, score new cycles independent from scores of previous cycles, giving single incidents not too much weight with regard to the general score, score a dimension independently from the scores of other dimensions and finally, high scores do not mean perfection (Pianta et al., 2012).

With regard to the FFT, *consensus on teaching quality* and especially consensus on its corresponding language are essential with regard to developing teaching quality (Danielson, 2013). This consensus can for instance be obtained by rater training or repeated use of the rubrics as discussed later on.

Notable is that all observation systems that are discussed earlier recommend to observe all domains at once. This is in line with Van Merriënboer and Tjiam (2013) who found that complex tasks, such as teaching a class, must be trained as a whole. Some training programs divide the complex task into parts and train those independently. The whole task is only trained at the end of the program. Van Merriënboer and Tjiam (2013) advocate a whole task practice approach where all different parts of the task are trained at once. With regard to teacher assessment, this pleads for whole task assessment instead of the assessment of only parts of the complex task.

Moreover, the assessors need to have *enough time* to assess all different items (Joe et al., 2013). And it needs to be clear what *differentiates the scoring options* in order to obtain interrater agreement (Womack, 2011). But also the *external factors* that can influence an observation's outcome need to be described on the instrument, such as the school, class, time, date or teacher (Pianta & Hamre, 2009; Stuhlman et al., 2010). Finally, it is recommended that an *assessment manual* is developed in which protocols are described (Evers, Lucassen, Meijer, & Sijtsma, 2010), such as what sample of lessons needs to be observed (Grossman, Cohen, & Brown, 2014), the observation period, when to score the instrument and how to act during observation. An overview of these recommendations is given in Table 1. Since this study focusses on rubrics as observational instrument, also recommendations will be given regarding rubric design and the use of rubrics. However, first definitions, advantages and disadvantages of rubrics as observational instruments are discussed.

1.2.3 Rubrics as Observational Instrument

A commonly used *definition* comes from Andrade (2000; Reddy & Andrade 2010), describing a rubric as "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". Jeong (2015, p. 1) adds to this definition that a rubric is also a representation of the teacher's assessment construct, it clarifies teachers' expectations.

The use of rubrics offers several *advantages*. Rubrics can *increase reliability* of assessments (Jonsson and Svingby, 2007), since rubrics give standards for each possible score. When those scoring criteria are formulated specifically and clearly, the rubric gives a more directive input to the assessor compared to more general assessment instruments with less specified scoring criteria. Moreover, Reynolds (2010) claims that teachers increasingly view rubrics as *more accurate and efficient* assessment tools. Rubrics also provide *feedback* to the assessed persons (Bryant p. 81) and facilitate *transparent assessment* (Bryant p. 81; Jonsson, 2014). The assessed persons are given insight to which specific level is reached on a certain criterion, and what progress is still needed to reach the next level. The use of rubric consequently can contribute to *self-assessment* (i.e. measuring personal learning progress), *self-regulation* (i.e. regulating personal learning process) and *motivation* for learning (Bryant, p. 81).

On the other hand, *disadvantages* can also be named with regard to rubrics as observation instrument. Garcia-Ros (2011) observes teachers' resistance to use rubrics for assessment and Bryant et al. (2016, p. 82) admit that the use of rubrics is not without flaws. First of all, *ambiguous goals* or *different views* on teacher quality can still hinder the effectiveness of rubrics (Jeong, 2015, p11; Bryant et al., 2016, p. 82). Secondly, students that are assessed with rubrics can be *too assessment-oriented*, only focusing on how to get a high score on the rubric (Zeichner, 2010), which can make students too submissive and therefore hinder self-regulation of their learning-process and professional development. Thirdly, it is *hard to capture all of the criteria* in a rubric (Jeong, 2015, p11). Finally, issues also arise with regard to validity and reliability, as several studies suggest that *increasing IRR is often at the expense of validity* (Jonsson & Svingby, 2007; Johnson, Penny, &

Gordon, 2000). Studies regarding rubrics often focus on reliability and fail to address validity (Jeong, 2015). While reliability is a condition for validity (Vos, 2009), it cannot be stated that a reliable rubric is a valuable observation instrument, since a reliable rubric must still be validated before statements can be made regarding its value or quality.

The rubric *design* and its use are determinative for its effectiveness, its ability to provide useful information for its purpose (e.g. formative- or summative feedback, low- or high stakes assessment). First of all effective rubrics are *formulated correctly* (Bryant et al., 2016, p. 80; Jeong, 2015, p. 3). Sentences are linguistically correct and do not include double denials, incorrect repetitions or other similar faults. Precise and correct formulation is required to avoid multi-interpretable scoring indicators. Secondly, effective rubrics are *appraised frequently* (Bryant et al., 2016). In other words, the rubric is not used as a static instrument but it is evaluated and adjusted regularly in search of improvement, leading to a higher effectiveness. Thirdly, *attributes must be consequently used across levels* (Jeong, 2015). Within one criterion, each description of a scoring indicator must use the same concepts and formulations. To illustrate, when scoring a teacher's use of voice, it is confusing if the first scoring level contains the concept 'dynamic', the second 'energetic' and the third speaks of 'powerful'. Since these concepts subtly differ in meaning, it can unnecessarily complicate the scoring process.

Regarding the use of rubrics, *repeated use* is needed to gain consensus and develop a shared meaning (Darling-Hammond, 2006). Moreover, *training for the assessors* is required, that is training on how to assess students on their teaching quality. Multiple studies show that training assessors can increase the reliability of the assessor's scores (Eppich et al., 2015; Feldman et al., 2012; Lovorn & Rezaei, 2011). Finally, rubrics must contain *specific descriptions* of the criterion, complemented with *examples*, since this can increase reliability, according to Thaler, Kazemi and Huscher (2009).

Recommendations regarding rubrics as observational instrument

An overview the recommendations from the discussed observation instruments and the recommendations regarding the rubric's design and use is given in Table 1.

Table 1

Recommendations regarding rubrics as observational instrument

Category	Requirement	Explanation
1. Design	1.a Whole task assessment	Assess all domains of a complex task at once (Van Merriënboer & Tjiam, 2013).
	1.b Include external factors	Describe teacher-unrelated factors that can influence the observation result, e.g. time and date, school, teacher (Pianta & Hamre, 2009; Stuhlman et al., 2010).
	1.c Correct formulation	Linguistically correct sentences to avoid multi-interpretable indicators. No double denials, incorrect repetitions or other similar faults (Bryant et al., 2016).
	1.d Appraise frequently	Evaluate and adjust in search for improvement (Bryant et al., 2016).

	1.e Use attributes consequently across levels	Use the same concepts and formulation within the indicators of one criterion (Jeong, 2015).
	1.f Specific descriptions	Describe specific observable behavior: "the student gives a differentiated arithmetic lesson with use of materials that..." (Thaler, Kazemi & Huscher, 2009).
	1.g Add examples	For instance, with regard to the student-pupil relation: "the student says things like: 'I do like your clothes' or 'Good to see you today'" (Thaler, Kazemi & Huscher, 2009).
	1.h Differentiate in scoring options	Make sure the distinction between the different scoring descriptions is clear (Womack, 2011).
2. Observation protocol	2.a Develop an assessment manual	Protocols can be described in this manual, e.g. timing and duration of observations, the sample of lessons that need to be observed (Evers, Lucassen, Meijer, & Sijtsma, 2010)
	2.b Determine a minimal number of observations	More observations will increase the reliability of the final score, but also depends on the assessment goal (Pianta et al., 2012).
	2.c Choose type of rater suitable for assessment objective	Pre-trained assessors, peers, experts or simply unrelated assessors (Girard & Cohn, 2016).
3. Use of the instrument	3.a Remain objective	Score independent from previous observations (Pianta et al., 2012). Give incidents not too much weight with regard to the general score (Pianta et al., 2012). Score dimensions independent from other dimensions (Pianta et al., 2012). High scores do not mean perfection (Pianta et al., 2012).
	3.b Gain consensus on teaching quality	Develop a similar concept of teaching quality in order to interpret the instrument similarly (Danielson, 2013)
4. Assessor	4.a Use repeatedly	Repeated use in different situations to develop a shared meaning (Darling-Hammond, 2006, p. 10).
	4.b Training on how to use the instrument	To discuss observation protocols, the instrument, or collectively assessing students' lessons (Eppich et al., 2015; Feldman et al., 2012; Lovorn & Rezaei, 2011).
5. Feasibility	5.a Determine whether interval- or event based observation is desirable	To observe the entire lesson, event based observation is desirable. However, in practice it might be more feasible to pre-determine observation-length (Girard & Cohn, 2016).
	5.b Organize time to observe all items	Many items require more time to observe than when only few items need to be assessed (Joe et al., 2013).

1.2.4 Reliability and Validity

Test scores can yield a certain level of reliability and validity, which are investigated in this study. Therefore the definitions and different types of both reliability and validity are given. First of all, *validity* is defined by the American Psychological Association (as cited in Vos, 2009, p. 76) as “the appropriateness, meaningfulness, and usefulness of the specific inferences’ made from the measures”. Kane (2006, p. 8) also emphasizes that usefulness by stating that ‘validity has to do with the proposed interpretations and uses of test scores’. According to him, validation of test scores has a direct link to the consequences that they have, where especially unintended negative consequences are most contentious. The need arises to examine a test validity before it is used. For instance, reliable but invalid measures cause an increase of data that the researcher attempted to measure but did not intended to measure, i.e. the measurements are reliable but the wrong content is being measured. Therefore the wrong consequences might be applied.

Validity can be specified into several components. First, there is the distinction between external- and internal validity, where the *external validity* refers to the extent results can be generalized to other contexts. *Internal validity* indicates whether “the program, service, or whatever was studied, in fact, produced the observed outcome” (Hernon & Schwartz, 2009, p. 73). Furthermore, there are different types of test validity, of which three are discussed. Firstly, *content validity*, that is to what degree the items that are used for a construct represents the intended construct (Hernon & Schwartz, 2009). For instance, if an arithmetic test only includes items on vocabulary, the test has a low content validity. Secondly, to investigate the *criterion validity* multiple test on the same construct are compared to see if the intended construct is covered (Hernon & Schwartz, 2009). Thirdly, whether the items that are used cover all components of the construct refers to *construct validity* (Vos, 2009). To illustrate, an arithmetic test on operations should not only cover addition, but also subtraction, fractions, etcetera.

With regard to *reliability*, a definition is given by the American Psychological Association (as cited in Vos, 2009, p.76), namely “the degree to which observed scores are ‘free from errors of measurement’”. Unreliable scores cause an increase of incorrect data (e.g. random results) so that the measurement results deviate from what the researcher attempted to measure. Because of this incorrect data, unreliable measurements are consequently invalid.

Reliability can also be measured in different ways. First of all by the *internal consistency*, i.e. the correlation between items on the same construct (Hernon & Schwartz, 2009) or within-test consistency (Vos, 2009). Secondly, by measuring the *test-retest reliability*, for instance by measuring the correlation between the results of multiple similar test (Vos, 2009). Thirdly, by the *parallel test reliability*, where the results of similar but different items are compared (Vos, 2009). Finally, by calculating the *IRR* where the scores of two or more assessors on the same subject are compared (Hallgren, 2012). The IRR qualifies the degree of agreement among assessors. When independent assessors score items identically the agreement is higher than when there is a large difference in the scores from multiple assessors.

As described in the problem statement, this study focusses on construct validity and the IRR. Since multiple students of this study's educational program indicate that the different scoring items and its constructs are interpreted differently among assessors, the construct validity is further examined in order to clarify constructs used in the rubric. This indication of different interpretation is also a motive for an investigation of the IRR, along with the fact that multiple students state that rubrics are used differently among assessors. For instance, some use half points where others do not. But also that some assessors use the rubric strict and analytic where others use it more holistic, more as a guideline that can also be ignored, leading to a divergent score. Therefore, a lack of inter-rater agreement is expected. It is investigated whether an increase in construct validity also leads to an increase in IRR. To increase the construct validity, clarifications and examples of constructs are added, along with general guidelines in order to clarify how the rubrics are intended to be used. To argue whether this leads to an increased validity, the theoretical validity framework of Kane (2006) is used.

1.2.5 Theoretical Validity Framework

Kane (as cited by Cook et al., 2015) developed a theoretical framework for validating an observation instrument's scores, as can be seen in Figure 1. This framework helps to validate the decisions that are made based on observations. An interpretative argument is constructed to justify the conclusions and decisions that are made based on the observed performances. This interpretive argument consists of four inferences. Firstly, the observation instrument transfers the sample of observations into a score. This *scoring inference* is for instance dependent on the scoring procedure and its standardization, e.g. training assessors on the use of the observation instrument and its protocols. Secondly, the *generalizability inference* is to which extent the observed lessons can be generalized to the target domain, which can for instance be all lessons given by the teacher in that year. The observed scores can only be partly generalized to the overall target domain and the teacher's performance. This can for instance depend on number of samples or the characteristics of the samples. Thirdly, the *extrapolation inference* concerns the question whether the sample of observations represent the real-world performance. To what degree do the observation results predict the performance in real-life? Finally, the *implication inference* discusses what consequences suit the observations and whether these consequences are useful for the stakeholders, such as the assessor, the learner, or the society.

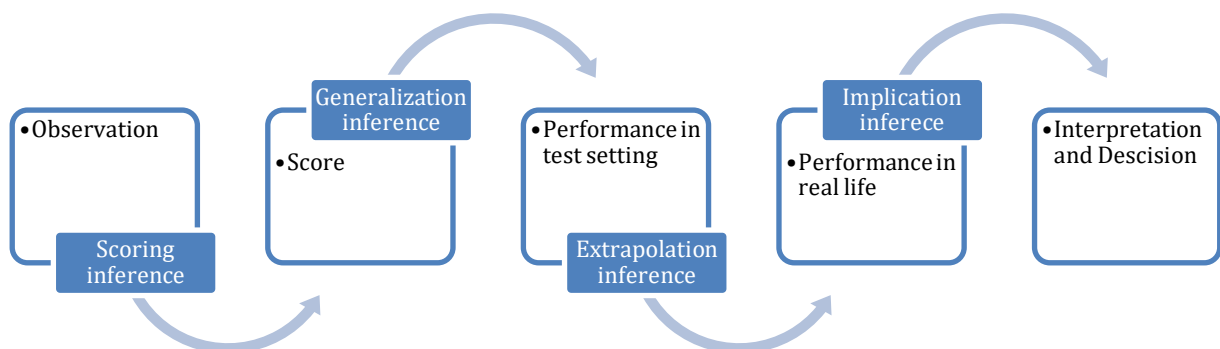


Figure 1. Overview of Kane's Validity Argument (Cook et al. 2015)

The approach of the interpretive argument depends on the stage that the test is in (Kane, 2006). When the test is in the *development stage*, the interpretive argument can be outlined before the test is developed. Possible sources of variance can be pre-determined and pilots can be done in order to be more consistent with the interpretive argument. Even the assumptions and inferences that underline the pre-determined interpretive argument can be adjusted when necessary. When the test is in *appraisal stage*, questionable assumptions are investigated, but also assumptions that can be checked easily in order to increase the confidence in the argument's claims. Also, hidden assumptions are searched in order to make them explicit since only then those assumptions can be evaluated. "An interpretive argument that has survived all reasonable challenges to its coherence and plausibility can be provisionally accepted, with the understanding that new evidence may undermine its credibility in the future." (Kane, 2006, p 26) Finally the interpretive argument can also be reviewed in hind-sight, during the *evaluation stage*. Insights that result from this evaluation can be used for the development of new test, when a different interpretative argument needs to be established.

The thoroughness of the evaluation of the interpretive argument depends on the different stakes that can be in play. For high-stakes assessment more extensive evidence is required with regard to the scoring inference than for low-stakes assessment (Kane, 2006). When an assessment instrument is simply used to give feedback on a student's learning curve, it is more likely that the teacher's implications based on the scores are harmless (Kane, 2006). When ambitious statements are being made based on the assessment (e.g. when a student fails this assessment, he or she is not fit to become a teacher), a robust and complete scoring argument is called for.

Scoring Inference

This study focusses on the scoring inference, the process of transferring observations into scores, based on a scoring rule as warrant for the scoring inference (Kane, 2006). No earlier research with regard to validity and reliability has been done with regard to the rubric used that are currently used at the institution of this study. Therefore this research started with the scoring inference as the first step for the validity argument. This scoring inference "*relies mainly on expert judgement about the appropriateness of the scoring rule and the thoroughness of quality control procedures.*" (Kane, 2006, p. 25) The inference is must be made plausible by providing evidence and arguments, and then it stands until is proven otherwise.

As described before, 'the warrant for the scoring inference is the scoring rule' (Kane, 2006, p.28). The scoring rule consists of the required guidelines in order to assign a score to the observed behavior. The backing of this warrant can consist of the outcomes of earlier studies, judgement of experts or empirical evidence. The warrant can for instance be the *substantiation of the scoring rules*, to assure that the scoring rule is administered as it is specifically intended to. This requires a clear distinction between scoring levels and clear rules on how to compute the final score. Another part of the warrant can be to make sure that the scoring rule is applied *accurately and consistently*. Accurately refers to the scoring being done as it is intended to. Consistently applying the scoring rule leads to reliable scoring despite of difference in circumstances. For instance, when the student is scored reliable despite the classes' pupil population, the assessor or the influence of time. The IRR

can be calculated as backing for this warrant (Kane, 2006). Accuracy and consistency can be warranted by using a reliability threshold to set conditions for assessors before they can use the instrument. Finally, it is also important to *address possible bias* in order to prevent systematic error that unintentionally influences the scores regardless of assessor or of the assessed student. For example, this may occur when an assessor only assesses one student for a longer period.

1.3 Research Question and Model

Based on the problem statement and the theoretical framework, it can be concluded that there is a need for reliable and valid assessment of students in educational programs. This study attempts to investigate how to improve the construct validity of rubrics as observation instrument. Secondly, it is studied whether this improvement in construct validity can lead to rubric scores with an increased IRR.

How can the construct validity of a rubric be improved, and to what degree can this improvement increase the inter-rater reliability of the rubric's scores?

Based on the theoretical framework a hypothesis was formulated, namely that more specific rubrics, enriched with clarifications and examples can lead to scores with a higher construct validity. In the first phase of this research, two focus groups were organized in order to develop an adjusted, 'experimental' rubric, enriched with clarifications and examples. In the second phase, assessors participate in an experiment where two fragments of student's lessons were assessed. The control group used the original rubric, whereas the experimental group used the experimental rubric. It was expected that the IRR of the scores resulting from this experimental rubric is higher than the IRR of the scores resulting from the original rubric. The methodology of this research is further explained in the next chapter.

2. RESEARCH DESIGN AND METHODS

2.1 Research Design

In this mixed method research, the construct validity of several constructs from an observational instrument was qualitatively investigated and improved (Phase 1). The instrument of this study was a rubric that was already in use in the educational program to assess lessons of pre-service teachers, see Appendix A. The increase of the construct validity was done by specifying and clarifying constructs based on the input of two focus groups, which was processed into a newly designed experimental instrument. To determine whether the construct validity has increased, the assumptions and evaluation criteria of the argument based approach of Kane (2006) are used, which are also described in the theoretical framework. In the quantitative part (Phase 2), the IRR of the original observation instrument was investigated and compared to the IRR of the experimental instrument. A posttest-only control group true experimental design was used as described by Vos (2009), where both groups assessed the same two students' lessons. However, the control group used the original rubric and the experimental group used the experimental rubric. The aim of this study was firstly to quantify the IRR of the rubric that is currently used, and secondly to uncover whether a rubric with a higher construct validity leads to a higher IRR. A model of this study's research design is presented in Figure 2.

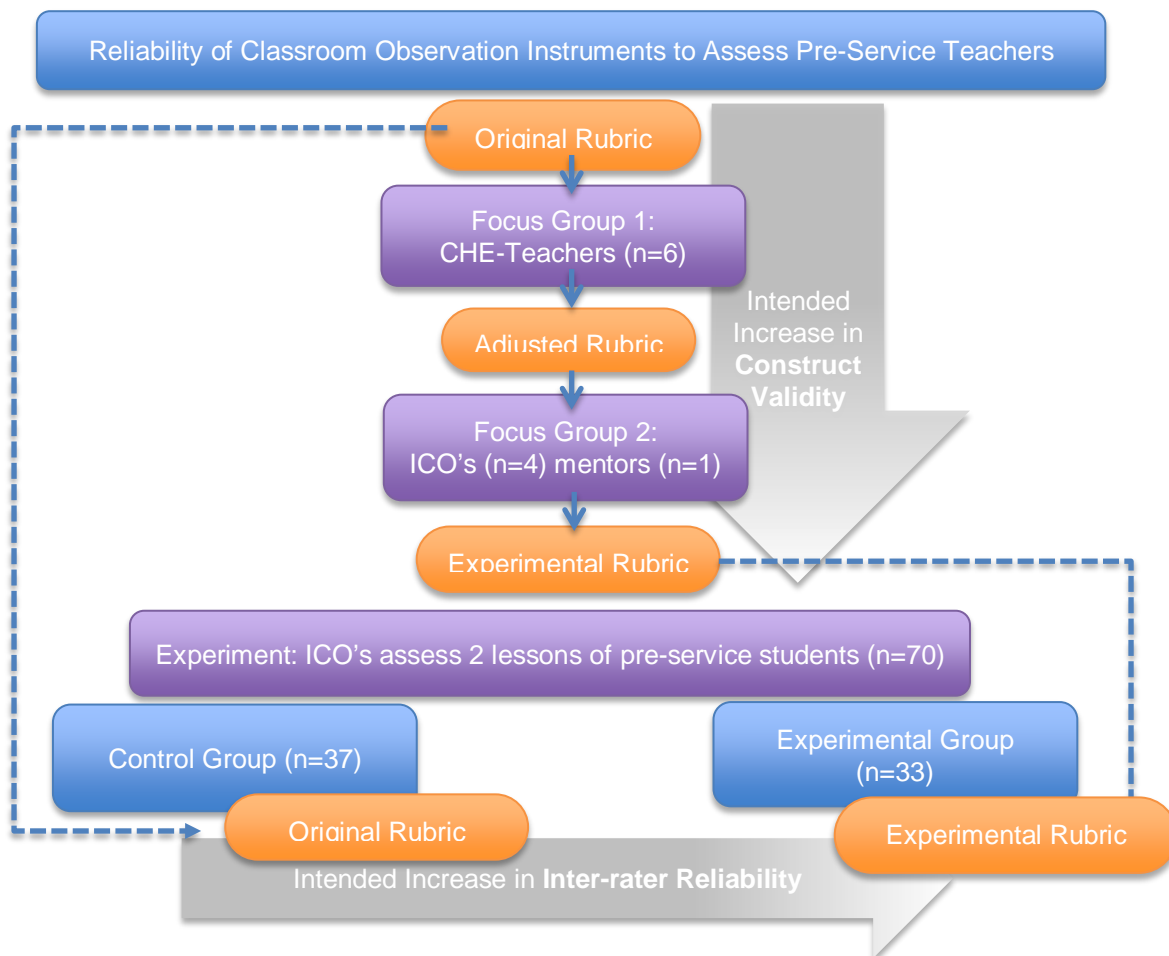


Figure 2: Model of Research Design

The institution for this research is a higher vocational education-institution providing several educations, e.g. communication, theology, journalism and education for primary school teacher. The institution is searching for ways to increase the validity and reliability of observational assessment of students. Currently rubrics are used to assess students in authentic practice. These rubrics are designed by teacher educators, and based on predefined outcomes, which are again based on the learning outcomes formulated by the Dutch government for teacher education for primary schools. In each quartile of each year a different rubric is used, with different scoring items. Each rubric has around 8 scoring items and for each criterion a 5 point scale from 0 to 4 is used that specifies for each score what is expected of the students to obtain that score (see Appendix A). Each scoring item again contains several constructs that are used to describe the criteria for each score. For instance for the scoring item "Prevalence" a construct can be "Act Suitable", measuring whether a student acts suitable with regard to unwanted behavior of the pupils. The scores of each item are summed up and converted into grades. For instance, when each scoring item is scored a 4, this results in a 10, the highest grade in the Dutch educational system. For these rubrics, the students need to receive a 6 or higher for the grade to be sufficient. The rubric of this study contains 4 scoring items, and at least 11 points need to be acquired in order to obtain a 6 as grade. This score can be obtained by any combination of the scoring items, for instance by scoring 4 points twice, 3 points once, and 0 points on the last scoring item.

The students are assessed three times, by at least two different assessors, both employed by the primary school of the student's internship. The first assessor supervises all the students of that primary school, the so called *ICO*. The second assessor is the teacher of the class where the student has his or her internship, the so called *mentor*. The average of those three grades must be sufficient or else the student must retake the course in the next quartile. Nevertheless, second time failure is almost always used to bar or delay a student. Hence, the rubrics for this research are used for summative, high-stakes assessment, but also formative, to give feedback to students on their performance. Since every scoring item is described in the rubric, the student can afterwards analyze the rubric to see which items are insufficient and what needs to be improved to gain higher scores. On top of that, the scores from the rubric can afterwards be further explained or elucidated by the *ICO*, however this is not obligatory or stimulated by the educational program. The student can always consult the mentor, who is also familiar with the rubric.

2.2 Participants

2.2.1 Participants of Phase 1

For the first focus-group in phase 1, all 52 Educational teachers from the CHE were invited. Seven CHE-teachers signed up for the meeting, of which eventually six were present. The six participants were asked to fill in a short questionnaire (see Appendix C) in order to collect descriptive information on sex, function, role with regard to the CHE and experience in assessing pre-service teachers. The group consisted of 4 males and 2 females. Furthermore, 5 participants were teachers and 1 was a teacher counsellor. All participants had at least 10 years of experience. Finally, 2 participants indicated to assess students 0 to 5 times per year as opposed to the other 4 participants, who indicated to give no assessments to students.

For the second focus-group in phase 1, all ICOs and mentors were approached that are in contact with this study's institution, and who work in Voorthuizen, Nijkerk, Nijkerkerveen, Barneveld and Kootwijkerbroek. These cities were selected for practical reasons, since this second focus-group took place in Voorthuizen. 7 ICO's and 1 mentors are willing to participate. However, 2 ICOs cancelled one day before the meeting and 1 ICO forgot the meeting, resulting in a group of 4 ICO's and 1 mentor. All 5 participants were female. Furthermore, two participants had less than 5 years of experience, one indicates 5-10 years and 2 indicate more than 10 years of experience. Finally, 1 participant did not assess students at the moment, 2 participants assessed 0 to 5 times per year, 1 participant assessed 5 to 10 times per year and 1 participant assessed 10 to 20 times a year. Table 2 gives an overview of these descriptive statistics for both the first and second focus-group.

2.2.2 Participants of Phase 2

The participants of the experiment in phase 2 were all ICOs that were present on an ICO-training on June 4, 2018. For this training, 181 ICOs were invited of which 77 were present. Assessors were randomly divided in both conditions by receiving a red or blue colored strip on entering the room, which was handed out alternately. Because of missing values, the results of 7 participants were inadmissible, therefore ignored in this study. This resulted in 37 participants in the control group and 33 participants in the experimental group. The sample consisted of 12 men (17%) and 52 women (74%). 6 participants did not fill in their gender (9%). Regarding to prior education, most participants had completed HAVO (N=41, 59%) and most participants were primary school teachers (N=41, 59%).

Table 2

Descriptive Statistics of the participants of the Focus Groups

Variable	Possible Outcomes	Focus Group 1 (n=6)	Focus Group 2 (n=5)	Total
Gender	Female	4	5	4
	Male	2	-	2
Role	Teacher	5	-	5
	Teacher Counselor	1	-	1
	ICO	-	4	4
	Mentor	-	1	1
Years of experience in this role	<5	-	2	2
	5-10	-	1	1
	>10	6	2	8
Amount of assessments given per year	0	4	1	5
	1-5	2	2	4
	5-10	-	1	1
	10-20	-	1	1
	>20	-	-	-

Table 3

Descriptive Statistics of the participants of the Experiment

Variable	Possible Outcomes	Control Group (n=37)	Experimental Group (n=33)	Total
Gender	Female	26	26	52
	Male	9	3	12
Prior Education	MAVO	3	4	7
Education	HAVO	20	21	41
	VWO	7	5	12
	Other	7	3	10
Function	Teacher	8	6	14
	Teacher/ICO	19	22	41
	IB/ICO	4	1	5
	Other	6	4	2
Assessment Training	No	10	7	17
Yes	27	26	53	
Rubric Experience	No	13	10	23
Yes	24	23	47	
ICO Experience	<5 Students	2	5	7
	5-10 Students	5	9	14
	10-20 Students	13	10	23
	>20 Students	17	9	26

Most participants have had any form of assessment training on rubric use (N=53, 76%) and most of them have had experience with this particular rubric (N=47, 67%). Finally with regard to their experience as an ICO, most participants indicated that they have trained more than 10 students. Table 3 gives an overview of these descriptive statistics for both the control- and experimental group.

Based on independent samples T-Tests with a significance level of $\alpha=.05$, there was no significant difference between the experimental- and the control group on the variables Gender ($p=.109$), Assessment Training ($p=.578$) and Rubric Experience ($p=.673$). Based on the Freeman-Halton extension of the Fisher exact probability test, there was also no significant difference between both groups on Prior Education ($p=0.614$), Function ($p=.522$) and ICO Experience ($p=.167$). It was concluded that both groups did not differ significantly on the measured characteristics as a result of randomly assigning participants to the control- or experimental group.

2.3 Instrumentation

For this study, a rubric was selected with the lowest grade average (7.3262) to avoid the ceiling effect (Vos, 2009). Moreover, multiple teacher educators already claimed that ICOs were having trouble interpreting this specific rubric. Since there was no evident difference in variance of assessors' scores

between rubrics, this had no influence in the selection process of the rubric. This rubric focusses on students' relation with the pupils, the way students stimulate the pupils feeling competent, how the students stimulate autonomy and finally the prevalence of the students in the classroom. Each of these four constructs has a 0-4 scale, therefore resulting in a total score between 0 and 16 points. These points are converted into grades, where 11 points is a 6, which is the minimal grade for the assessment to be sufficient.

2.3.1 Instrumentation of Phase 1

In phase two an experimental rubric is developed in an attempt to increase its construct validity. Two focus groups are organized, where the participants are asked to identify unclear constructs from the descriptions of the scoring items. The participants are also asked how these unclear descriptions can be clarified or specified, possibly by adding examples, as was recommended in the theoretical framework and Table 1. These clarifications can serve as backing for the warrant of the scoring rule of this study's rubric (Kane, 2006). For the first focus group in phase 1, the original rubric was used (see Appendix A). The participants were asked to sign a consent form (see Appendix B), and to fill in a short questionnaire on gender, function, role with regard to the CHE, role experience, student assessment experience (see Appendix C). This information was later on used to describe the participants. For both focus-groups a script was used, based on the script of Eliot and Associates (2005), as described in Appendix D. In this script, the participants were stimulated to share their honest, possibly deviant opinion, change opinion, and give new input. The participants were asked to clarify and/or exemplify constructs by writing comments on forms as can be seen in Appendix E. To illustrate, a construct that is discussed is "Undesired Behavior". Post-its were used to enable each participant to individually react on the comments of the other participants.

For the second focus group in phase 1, the adjusted rubric was used based on the outcomes of the first focus group (see Appendix H). Nine pre-determined questions were used based on five categories that describe the progressiveness of focus-group questioning (Krueger, 1994). The script that is being used (see Appendix D) can be complemented with the interview scheme resulting from these nine questions (see Appendix F). The first category is "Opening questions", such as "What are your experiences on assessing students?" Secondly, "Introductory questions", in this case "For which courses did you use rubrics to assess students?" Thirdly, "Transitional questions" as there is "What are the advantages of rubrics for assessing students in authentic practice?" These first three questions mainly used to introduce the subject at hand, and to activate the participants' prior knowledge. Fourthly the "Main questions", such as "What parts or concepts are still unclear to you?" These questions are used to uncover which parts of the rubric are unclear and which parts cause disagreement. Hereafter the participants are asked how these parts of the rubric can be reformulated in order to increase the construct validity. And finally "Ending questions", in this case "Is there something that has not yet been said, of which you are strongly convinced that it should be said, with regard to the use of rubrics on assessing students in authentic practice?" The participants were asked to sign a consent form (see Appendix B), and to fill in the same short questionnaire as the participants of the first focus group (see Appendix C).

2.3.2 Instrumentation of Phase 2

For the experiment, all participants received the same instruction to avoid intersession contamination (Vos, 2009). Therefore a script was read to minimize the influence of experimenter expectancy (Vos, 2009) (see Appendix H). In this script, the faithful subject method of Vos (2009) was applied and it was stressed that data is processed anonymously to avoid evaluation apprehension (Vos, 2009). The participants were asked to sign a consent form (see Appendix B), and to fill in a short questionnaire on gender, prior education, function, assessment training, rubric experience and ICO experience (see Appendix N).

For the control group of the experiment in phase 2, the original rubric was used as described in the introduction (see Appendix J, K). For the experimental group in phase 2, an experimental rubric was developed based on the input from the focus-groups (see Appendix L, M). It was intended that this adjusted experimental rubric has a higher construct validity and a higher IRR, as discussed in the results section. The development of this rubric is further described in the procedure.

2.4 Procedure

2.4.1 Procedure of Phase 1

For the first 20 minutes of the first focus-groups in phase 1, the teachers were paired. Each group receives two forms with each one element of a rubric's indicator (for instance, "*makes contact with the children*"). Each pair was asked to clarify and specify the selected indicator, as also explained on the forms in Appendix E. With six participants, three pairs were made, therefore the two remaining indicators were discussed by the two pairs that were done early. In the second 20 minutes, the participants were asked to review each other's forms individually by making comments on post-it's to indicate whether they agree or disagree and also explain their line of reasoning.

For the second focus group, the ICOs were questioned on the revised rubric from the previous focus group (see Appendix H), based on the pre-determined questions as can be seen in Appendix F). An assistant was instructed to record the meeting, take notes when speech was unclear and to refrain from the conversation. Both the researcher and assistant welcomed the participants.

2.4.2 Procedure of Phase 2

For the experiment in phase 2, the participants have alternately received a red or blue colored stripe on entering the room, indicating if they can take place in the left or right side of the room. The control group took place on the left side and the experimental group on the right side, with some rows of vacant tables in between.

The control group received a package of the informed consent-form (see Appendix B), two original rubrics, (one for each fragment, see Appendix J, K) and the form on descriptive characteristics of the participants (see Appendix N). The experimental group received a package of the informed content-form (see Appendix B), two experimental rubrics (one for each fragment, see Appendix L, M) and the form on descriptive characteristics of the participants (see Appendix N). For each package, all documents had an identical number that was related to one specific participant. First, the informed consent-form was filled in. Secondly, two fragments of 15 minutes of pre-service teachers' lessons were showed. The control group assessed both lessons with the original rubric, the experimental

group with the experimental rubric. Finally, the questionnaire was filled in for descriptive information (see Appendix N).

2.5 Data Analysis

2.5.1 Data Analysis of Phase 1

The data of the first focus group of phase 1 (see Appendix G) was used as input in an attempt to develop a clearer and specific rubric, thereby increasing the construct validity as discussed in the theoretical framework. The input from the forms was compared with the input from the post-its. If the new formulation was (nearly) undisputed, it was added to, or integrated in the rubric. If the new formulation was disputed in the post-its, the researcher considers if the new formulation was again adapted based on the input from the post-its. Finally, these changes were added to, or integrated in the rubric. This resulted in a revised, and expectedly clearer rubric.

The data of the second focus group of phase 1 was analyzed as recommended by Eliot and Associates (2005). The recordings were transcribed and analyzed by open- and axial coding in ATLAS.ti. Before the data was coded, additional notes from the assistant were added to the transcription and insignificant words were removed. Every sentence was labeled with a number corresponding to the participant. The coded data was analyzed to determine if further adjustments on the experimental rubric could be made in order to increase the construct validity. Examples of codes were “Undesirable behavior” that contains all fragments where the participants discussed how undesirable behavior of pupils can be clarified, or “Unforeseen circumstances” where the participants signaled that it was unclear how to assess students if no unforeseen circumstances occur.

2.5.2 Data Analysis of Phase 2

The data from the experiment of phase 2 first consisted of filled in assessment forms, which were transferred into SPSS. The data was checked for missing values or outliers (Bryant et al., 2016). Based on these results, the IRR is calculated as backing for the warrant of the scoring rule as discussed in the theoretical framework (Kane, 2006). Bresciani et al. (2009) state that the Intraclass Correlation Coefficient (ICC) can be used to calculate the IRR when multiple assessors score on an ordered scale. Hallgren (2012) adds that the ICC can also be computed when assessing with an interval or ratio scale. Therefore, to compare the IRR of both the control- and experimental group, the ICC was calculated for both groups.

Since all participants have assessed the same two fragments, the two-way random effects model was applied using a consistency definition (Hallgren, 2012). The single measures ICC was being used, since this study is about the IRR between individual assessors, and not as the IRR of the group as a whole (Hallgren, 2012). Therefore the ICC(2,1) is used, as described by Landers (2015). Since some of the data violated the normality assumption, Fleiss' kappa (Falotico & Quatto, 2014) was calculated in order to give indication of the agreement among multiple assessors in a similar way as Cohen's Kappa for two assessors, and the ICC does when the data is normally distributed. However, Fleiss' Kappa does not require normal distribution of the data. The kappa is calculated as proposed by Fleiss (1974):

$$K_{fleiss} = \frac{\bar{P} - \bar{P}e}{1 - \bar{P}e}$$

\bar{P} is calculated by:

$$\bar{P} = \frac{1}{r-1} \left(\frac{1}{nr} \sum_{i,j} r_{ij}^2 - 1 \right)$$

In this formula n is the number of subjects, r the number of assessors per subject and r_{ij} is the number of raters for the i th subject and the j th category.

$\bar{P}e$ is calculated by:

$$\bar{P}e = \sum_{j=1}^q p_j^2$$

In this formula q is the number of categories and p_j is calculated by:

$$p_j = \frac{1}{Nn} \sum_{Nn} n_{ij}$$

In this formula N is the number of items.

3. RESULTS

3.1 Focus Group 1 – CHE-Teachers

The first focus group in phase 1 is focussed on clarifying or exemplifying certain constructs of the rubric in order to increase its construct validity. Eight constructs are selected, such as “acts appropriate” or “unwanted behaviour”. The data from this focus group is processed into the first draft of an experimental rubric (see Appendix H).

On the basis of the comments on the eight chosen construct and the reactions on those comments (see Appendix G), clarifications or examples are formulated and added to the original rubric. No changes are made in the formulation of the scoring criteria of the rubric itself, only clarifications or examples are added. For instance, most participants agreed that unwanted behaviour was behaviour that was inconvenient with regard to the situation in the classroom. Moreover, each participant’s reaction included an addition, such as “unwanted behaviour is behaviour that can be hurtful to others”. Since all reactions included different reactions, the researcher choose a clarification of “Unwanted Behaviour” that covered most of the reactions, namely “Behaviour that does not suit the actual situation in the classroom. Behaviour that disrupts the class. Behaviour that is not agreed upon. EXAMPLE: talking when silence is required.” The clarifications or examples were added to the original rubric, as can be seen in Table 4, resulting in the first draft of the experimental rubric (Appendix H).

3.2 Focus Group 2 – ICOs and Mentors

The second focus group in phase 1 is focussed on receiving feedback on the first draft of the experimental rubric by ICOs or Mentors, that is the assessors of this rubric. The data from this focus group is used to adjust the experimental rubric to increase its construct validity further.

On the basis of the transcription and its coding (see Appendix O, the experimental rubric is adjusted. For each code, all corresponding quotations are reviewed. For instance, for the code “Unforeseen Circumstances” and its corresponding quotations (see Appendix O, nr. 197, 198, 199, 200) it is decided that a clarification is needed on what to do when there are no unforeseen circumstances. Therefore, the phrase “If there are no unforeseen circumstances, ask for the input of the children” is added. This is done for every code with its corresponding quotations.

The changes that are being made on the basis of the second focus group can be seen in Table 3. In addition to this table, another construct was clarified based on the input of the second focus group, namely: “...and takes into account the individual needs of the child”. This construct was not discussed during the first focus group, but during the second focus group it became clear that it needed clarification. This clarification was formulated as follows: “TRUE: make a child feel competent on what it can learn (zone of proximal development) False: when the teacher convinces a child of a competence, while the child has not acquired that competence yet or can’t learn it yet.” These changes resulted in the experimental rubric as can be seen in Appendix L and M.

Table 4

Results of the First- and Second Focus Group Regarding the Development of the Experimental Rubric

Scoring Item	Construct	Additions/Clarifications, based on the input from Focus Group 1	Changes based on the input from Focus Group 2
Relation	Student makes contact	TRUE: verbal/non-verbal, respond, ask questions about given answers, interaction. FALSE: Regularly ignoring children, focus lies only on their own lesson	-
	Conforms to child's perception on content- and relational level	Content level: which subject or lesson content is being chosen? Relational level: Does the student relate to age and personality?	"Personality" is changed into "Interest": "Does the student relate to age and interest?"
Competence	Give insight and trust in child's competence	Give insight: confirm children in what they are able to do, and let them name this themselves. Give trust: let them experience what they are able to do. Often refers to emotion.	"Perseverance" is added to "Often refers to emotion": "Often refers to emotion and perseverance"
	...and the actual (poss. unforeseen) situation in the classroom	TRUE: Join and go with ideas apart from own input and engaging children suitable to their competence FALSE: Hold all initiative to themselves	"and engaging children suitable to their competence" is replaced with "If there are no unforeseen circumstances, ask for the children's input."
Autonomy	The child is only requested to ask procedural questions	TRUE: Questions about how and what regarding the assignment, where the teacher answers all questions FALSE: Questions about how and what regarding the assignment where the children can have their own input	-
	...and stimulate reciprocal interaction between children	The teacher responds to suggested alternative of a pupil and gives space to discuss this with the teacher or other pupils. The child experiences it can have input.	Three changes result in the following description: " <u>In the context of autonomy</u> : The teacher responds to suggested alternative of a pupil (or asks for it) and gives space to discuss this with the teacher or other pupils. The child experiences it can have input. NOT: Allowing the children to collaborate."

Prevalence	Unwanted behaviour	Behaviour that does not suit the actual situation in the classroom. Behaviour that disrupts the class. Behavior that is not agreed upon. EXAMPLE: talking when silence is required.	-
	Acts suitable	If necessary, first ask for a clarification. Then clearly naming the problem of the behaviour. Not regarding the person, but the behaviour.	-

3.3 Argument based approach to the Validity of the Original- and Experimental Rubrics

This argument based approach to validation takes place during the appraisal stage (Kane, 2006). For this study, the test was already developed without explicitly formulating the interpretive argument of Kane (2006). The rubrics are both used for low- and high stakes applications. The rubrics are intended to give formative feedback, but are also used to bar students when multiple lessons are graded insufficiently. Therefore extensive appraisal of the interpretive argument is necessary (Kane, 2006). Based on the validity framework of Kane (2006) the backings for the warrant of the scoring rule are now discussed for both the original- and the experimental rubric, as discussed in the theoretical framework.

3.3.1 Warrant of the Scoring Rule of the Original Rubric

First of all it is discussed how the scoring rules are substantiated for the original rubric. The distinction of scales of each scoring item are indicated clearly. Differences between scales are emphasized by making the fonts bold. However, as a result of both focus groups it can be concluded that not all constructs used in the scoring items are clear to all of the participants. Sometimes participants have different concepts of constructs, or have trouble translating the construct to observable behavior in the classroom. In short, the distinctions may be clearly visible but it is questionable whether these distinctions are all understandable to the assessors. It is therefore also questionable whether all scoring items are appropriate for classroom interactions, since assessors have trouble translating constructs to observable classroom behavior. For the assessors it is clear how to compute the observed score for one lesson, and how to compute the final score from scores of multiple lessons since the latter is simply the average.

Secondly, the original rubric lacks measures to score accurate and consistent. Currently there is no reliability threshold for the ICOs with regard to this rubrics. Some ICOs are experienced while others just started assessing students. In all likelihood, this endangers the consistency among assessors. There is some obligatory training to assess students, although this consists of only a few moments per year. For instance during this training it is discussed how to use the observation instrument and how to interpret its content. It is questionable whether this is enough to achieve accuracy and consistency for the different rubrics, since every quartile another rubric is being used.

And finally with regard to assessment manuals, there is one available for each rubric. This includes information on the sample of observations required for that rubric, and more elaborate descriptions of the scoring items. Still, currently there is no manual that includes general scoring rules, such as 'take the text of this rubric as literally as possible', 'do not give half points' or 'do not change scores once you have calculated the final grade'.

Finally, possible bias by assessor is addressed by obliging that students need to be assessed by at least two different assessors. Mostly, one is the mentor and one is the ICO. When one of those two is incapable of assessing the students (e.g. because of sickness) the educational program arranges a teacher educator to visit the school and assess the student. It is also possible for the student to arrange assessment by a third assessor (e.g. a teacher from another class), although this is not obliged.

3.3.2 Warrant of the Scoring Rule of the Experimental Rubric

With regard to the substantiation of the scoring rules, the most eminent distinction of the experimental rubric from the original rubric are the clarifications and added examples with regard to several constructs, as can be seen in Table 4. Constructs that were unclear to the participants of both focus groups, were discussed and clarified by the participants themselves and the researcher. Therefore it is probable that these additions help to distinguish between scales and increase the appropriateness for classroom observation. When assessors receive more elaborate explanations and examples of constructs, it is probable that this increases the ability to observe these constructs. For this rubric it is also clear how to compute the observed score.

Secondly, there are more measures to score accurate and consistent for the experimental rubric compared to the original rubric. General scoring guidelines are added in the rubric since a general manual on scoring rules is absent. These guidelines first of all specify that the assessor must take the descriptions of each scoring criterion as *literally* as possible. In other words, the assessor is asked to leave any personal interpretation. Secondly, the assessor is asked to only *review content that is named in the scoring criteria*. E.g. when the rubric items only discuss pedagogical content, the assessor must not let the scores be influenced by the student's organization of the lesson. And finally, it is stressed that the calculation and reviewing of *the final score must not influence the result of individual scoring items*. When the assessment process leads to a lower final score than the assessor expected, he cannot change the scores of individual scoring items in order to increase the final score. It is intended that the addition of these general guidelines increase consistency of applying the scoring rules by different assessors. For this rubric there has not yet been any training of assessors, since the rubric was only recently developed for this study. In similar fashion with the original rubric, there is no reliability threshold for this rubric.

At last, the bias by assessor regarding the experimental rubric is addressed in the same way as the original rubric. At least two different assessors will need to have assessed the student in order to prevent assessor bias.

To conclude, the additions appear to clarify and specify the rubrics content and scoring rules. Therefore it is plausible that the construct validity of the experimental rubric has increased compared to the construct validity of the original rubric.

3.4 Experiment

3.4.1 General Analysis of Scores

During the experiment the control- and experimental group each assessed two lessons of students by watching fragments, resulting in 140 filled in rubrics. The percentages are calculated for each possible score on each scoring item for both the control- and experimental group as can be seen in Table 5 for fragment 1 and Table 6 for fragment 2.

The highest percentage for one score on a scoring item can be found in the prevalence scores of the experimental group with regard to fragment 2 (78.8%). Notable is that no participant of the control- or experimental group has given 0 points for prevalence for the first fragment. In addition, for fragment 2 only one participant has given a 4 score on one of the scoring items. The greatest difference between the control- and the experimental group can be found in the results of fragment 2, where 59.5% of the participants from the control group have given a 0 score for competence, compared to 33.3% of the participants from the experimental group.

Table 7 gives an overview of the average scores given for each item for the control- and experimental group. Notable is that the absolute differences between the average score of the control group and the average score of the experimental group are the largest for "Prevalence" of fragment 1 ($\Delta=.418$) and "Competence" of fragment 2 ($\Delta=-.314$). For the "Prevalence" scoring item, this difference even is significant ($p=.049$), which means that the control group has scored "Prevalence" higher ($M=2.054$) than the experimental group ($M=1.636$). Since only one scoring item is assessed significantly different by the control- and experimental group, it is assumed that combining the scoring items for the analysis will not lead to an overall significant difference.

Also conspicuous is that the difference between the mean scores and its standard deviation is relatively small. Especially for the total scores for fragment 2, where the mean of the control group is 3.860 and its standard deviation is 2.584. This indicates that 68% of the scores lie between 1,276 and 6,444, on a scale of 0-16 points.

Table 5

Percentages of Scores of Fragment 1 for the Control- and Experimental Group

Score	Relationship		Competence		Autonomy		Prevalence	
	Contr. Group	Exp. Group	Contr. Group	Exp. Group	Contr. Group	Exp. Group	Contr. Group	Exp. Group
0	10,8%	12,1%	13,5%	15,2%	8,1%	6,1%	0,0%	0,0%
1	51,4%	39,4%	35,1%	39,4%	40,5%	57,6%	29,7%	54,5%
2	24,3%	36,4%	21,6%	18,2%	40,5%	21,2%	43,2%	30,3%
3	13,5%	12,1%	29,7%	24,2%	8,1%	15,2%	18,9%	12,1%
4	0,0%	0,0%	0,0%	3,0%	2,7%	0,0%	8,1%	3,0%

Table 6

Percentages of Scores of Fragment 2 for the Control- and Experimental Group

Competence		Autonomy		Prevalence	
Contr. Group	Exp. Group	Contr. Group	Exp. Group	Contr. Group	Exp. Group
59,5%	33,3%	45,9%	45,5%	2,7%	3,0%
27,0%	45,5%	32,4%	24,2%	70,3%	78,8%
8,1%	18,2%	16,2%	27,3%	21,6%	15,2%
5,4%	3,0%	5,4%	3,0%	2,7%	3,0%
0,0%	0,0%	0,0%	0,0%	2,7%	0,0%

Table 7

Average Scores per Item for the Control- and Experimental Group

	Scoring Item	Contr. Group		Exp. Group		Δ	p
		M	SD	M	SD		
Fragment 1	Relation	1.41	0.87	1.49	0.87	-0.08	.70
	Competence	1.68	1.06	1.61	1.12	0.07	.79
	Autonomy	1.57	0.87	1.46	0.83	0.11	.58
	Prevalence	2.05	0.91	1.64	0.82	0.42	.05*
	Total Score	6.70	3.05	6.18	3.13	0.52	.48
Fragment 2	Relation	1.14	0.75	1.18	0.85	-0.05	.81
	Competence	0.60	0.87	0.91	0.81	-0.31	.12
	Autonomy	0.81	0.91	0.88	0.93	-0.07	.76
	Prevalence	1.32	0.71	1.18	0.53	0.14	.35
	Total Score	3.86	2.58	4.15	2.41	-0.29	.63

*Significant Difference between Control- and Experimental group with a significance level of $\alpha=.05$, 2-tailed

Table 8

Intraclass Correlation Coefficients

	Contr. Group	Exp. Group
ICC Fragment 1	.15	-.01
ICC Fragment 2	.23	.05
ICC Fragment 1 & 2	.31	.14

Table 9

Levels of Clinical Significance of the ICC (Cicchetti, 1994)

<.40	Poor
.40-.59	Fair
.60-.74	Good
>.74	Excellent

3.4.2 Calculation of the IRR

Based on the scores of the participants in the control- and experimental group, the IRR of the original- and the experimental rubric are computed and compared. For this analysis, each score from each scoring item from each participant is compared to the score for the same scoring item from the other participants from the same condition. This is done for the scores of fragment 1, fragment 2 and for fragment 1 and 2 together, as can be seen in Table 8. The highest ICC that has been observed in this analysis is .308 for all 8 scoring items of fragment 1 and 2, assessed by the control group. However, according to Cicchetti (1994) with respect to the intraclass correlation statistic, when the ICC is below .40, the level of clinical significance is poor (see Table 9). Therefore, it can be concluded that the ICC of both the original- and the experimental rubric is poor. The ICC of the experimental group is repeatedly lower than the ICC of the control group. For fragment 1, the ICC even is negative, an effect which is mathematically possible, but theoretically impossible and can only be interpreted as poor inter-rater agreement (Giraudeau, 1996).

However, since ICC calculations can be influenced by bias between assessors and by the distribution of the data in complex ways (Byrt, Bishop, & Carlin, 1993), further data analysis is required. It is not possible to calculate quantitative indicators of bias and prevalence as described by Byrt et al. (1993) since the data of this study concerns multiple assessors scoring two cases, instead of two assessors scoring multiple cases. But when looking at the assumptions required for calculating the ICC, it can be concluded that none of the results of each scoring item of both conditions are normally distributed. The normality assumption has been violated. Therefore, Fleiss' Kappa has been calculated for each condition and each case, of which the results can be seen in Table 10. The highest kappa resulting from this analysis is that of the control group for both fragment 1 and 2 ($K=.0919$) and the lowest kappa is that of the control group for fragment 1 ($K=-.0254$). This sheds no new light on the case, the results indicate a poor IRR since the distribution of the scores can also be attributed to chance.

3.4.3 Results transferred in Dichotomous Data with regard to High-Stakes Assessment

Since the rubric is used as a high-stakes assessment instrument, as described in the introduction, dichotomous data is computed from the total scores for both of the conditions and for both fragments. The total scores are computed by adding the scores for all four scoring items. Since 11 points is minimal for this assessment to be sufficient, all scores below 11 points are labeled as "Fail" and all scores of 11 or more points are labeled as "Pass". In the control group 32 of 37 assessors fail the

student of the first fragment, as can be seen in Table 11. That is approximately 86%. In the experimental group that is 29 out of 33 assessors failing the student, approximately 88%. The percentage of assessors failing the student of fragment 1 are almost identical for the control- and experimental group. For fragment two 36 of 37 assessors from the control group fail the student, as can be seen in Table 12. That is approximately 97%. For the experimental group that is 32 out of 33 assessors failing the student, approximately 97%. This does not indicate a difference between the control- and experimental group, when examining the dichotomous data. To substantiate this indication, Fisher's Exact test is calculated for the dichotomous data to compare the control group with the experimental group.

Table 10

Fleiss' Kappa

	Contr. Group	Exp. Group
Fleiss' Kappa Fragment 1	-.025	-.004
Fleiss' Kappa Fragment 2	.048	.089
Fleiss' Kappa Fragment 1 & 2	.092	.052

Table 11

Frequencies of Pass- and Fail-Judgements of Fragment 1

	Fail	Pass	Total
Condition Control Group	32	5	37
Experimental Group	29	4	33
Total	61	9	70

Table 12

Frequencies of Pass- and Fail-Judgements of Fragment 2

	Fail	Pass	Total
Condition Control Group	36	1	37
Experimental Group	32	1	33
Total	68	2	70

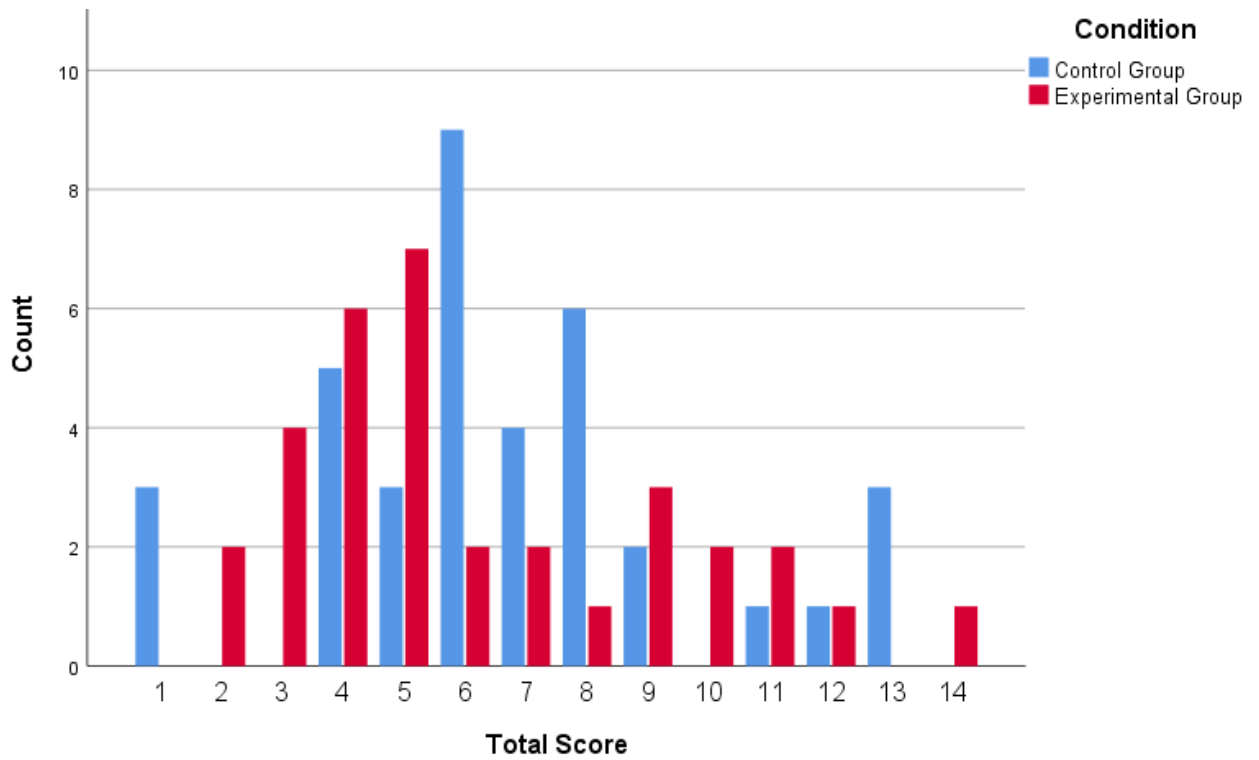


Figure 3. Frequencies of Total Score for Fragment 1

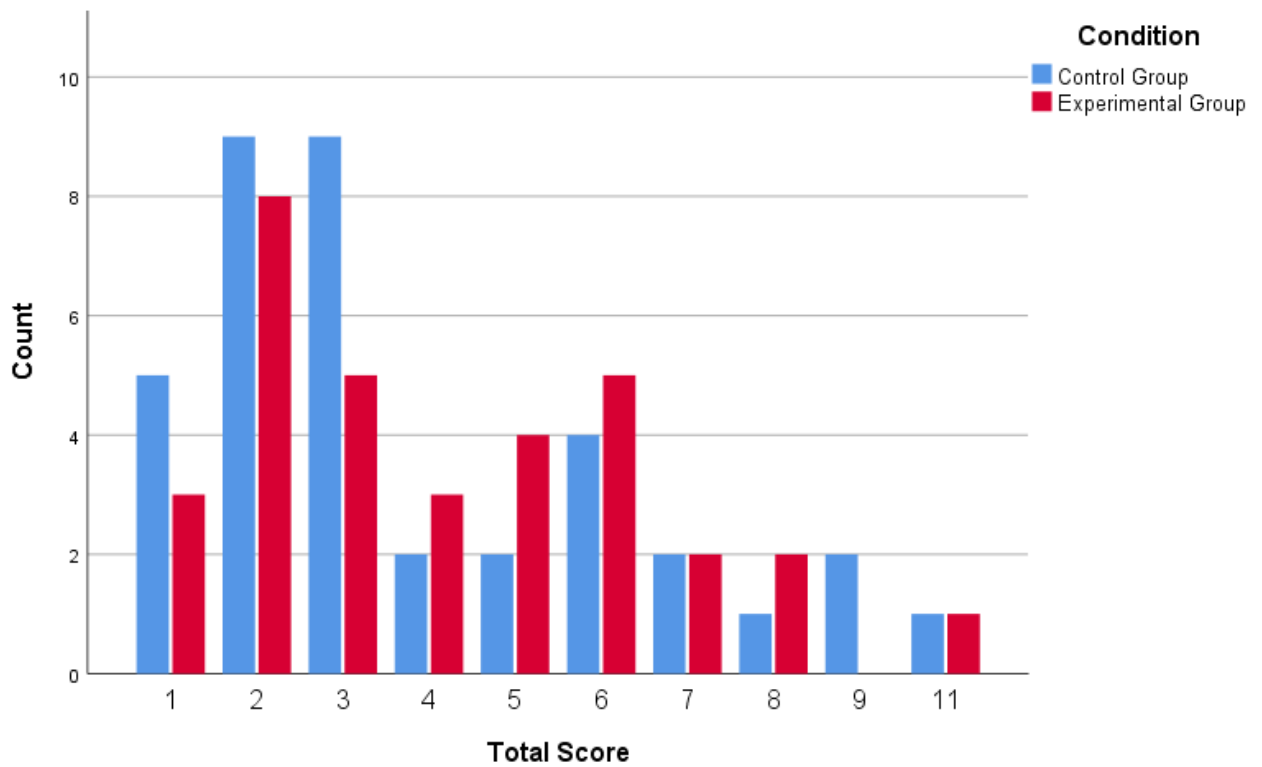


Figure 4. Frequencies of Total Score for Fragment 2

In line with earlier indications based on the percentages of assessors failing the students of fragment 1 and 2, Fisher's Exact test does not indicate a significant difference between both conditions on passing or failing the students for both fragment 1, $p > .05$, and fragment 2, $p > .05$. The null hypothesis of independence is therefore not rejected. Since this study investigates the influence of the different rubrics on the agreement between assessors, it cannot be concluded that the use of the original- or experimental rubric had a significant influence on passing or failing a student.

To analyze the total scores furthermore, the total scores are visualized for both the control- and the experimental group in Figure 3 for fragment 1 and Figure 4 for fragment 2. Both distributions of the total score show a high variance conforming the high standard deviation relative to the mean, as discussed earlier. Both distributions are skewed to the right conforming a low mean, which is in accordance with the low number of participants that pass the student. Finally the ICC is calculated for these total scores. The ICC of the experimental group is still poor (ICC=.278) and the ICC of the control group now is scarcely fair (ICC=.449). However, this result is in contrast with the hypothesis that a rubric with a higher validity, more specific descriptions and added examples, leads to a higher IRR.

4. CONCLUSIONS AND DISCUSSION

4.1 Discussion

The purpose of this study was to investigate the construct validity and the IRR of a rubric used to assess students. More specifically, the influence of a rubric's construct validity on the IRR was studied. An experimental rubric was designed by adding clarifications and examples to the original rubric, based on the input of two focus-groups. For the first focus group CHE-teachers participated and for the second focus group several ICO's and a mentor took part. Studying the construct validity and IRR of an assessment instrument can give insight in how to enhance these attributes or simply identify problems that arise when attempting to achieve this improvement.

During these focus groups it became evident that some construct that are included in the descriptions of the rubric's scoring items needed clarifications. The participants sometimes lacked comprehension of these constructs or had a different view on what the construct comprised. The participants also gave input on how to clarify these constructs by adding explanations or examples of these constructs. Based on this input, an experimental rubric was designed, which was intended to yield a higher construct validity. This rubric was improved with regard to the rubric design as described in Table 1. It was checked on correct formulations (1.d), appraised several times (1.c), specified (1.f) and examples were added (1.g). With regard to the observation protocol, several general scoring guidelines were added that can be included in an assessment manual (2.a). The other recommendations described in Table 1 were not applied in the design of the experimental rubric for this study.

However, this had no visible influence on the IRR of the scores from the experimental rubric. The results of the experiment show a low IRR for both conditions, for both lesson-fragments separately and for both fragments combined. The scores from the experimental rubric did not have a significantly higher IRR than the scores from the original rubric. Further analyzation of the scores indicate that both the control- and the experimental group do not use the whole scale for all of the scoring items. For some items, a possible score is absent for almost all assessors. This indicates a discrepancy between the scoring item and the classroom interactions, which apparently requires revision of those scoring items.

An explanation of the low IRR could be that each quartile another rubric is being used on the CHE, as opposed to a number of other teacher educational programs, who use the same assessment instrument manifold. As stated in the theoretical framework, effective rubrics must be used repeatedly in order to develop a shared meaning (Darling-Hammond, 2006, p. 10). It is questionable whether it is feasible to use a different assessment instrument for each quarter. This setup makes it unworkable to organize rater training for all 16 different instruments. Moreover, as discussed in the theoretical framework, Van Merriënboer and Tjiam (2013) claim that complex skills such as teaching need to be trained as a whole. This can lead to one assessment tool that entails all elements of this complex skill. When one assessment tool is used for all quartiles in the educational program, it is used more repeatedly by the assessors and it can be appraised more frequently by the teacher educators.

Another factor that complicates the assessment of students through rubrics is that a specific rubric cannot include the complexity of the classroom reality, as described in the theoretical

framework. For instance, what if it is unwanted for a student to give the pupils autonomy simply because they cannot handle such a degree of freedom? How to score items that are (deliberately) absent? Perhaps specifying rubrics decreases the reliability because the assessor experiences a greater discrepancy between the formulation of the rubric and the complexity of the classroom reality. It can be argued that a rubric can again be complemented to further specify those questions, for instance by stating what the assessor should do when a scoring item is absent. Nevertheless, it is questionable whether this solves the problem. More elaborate and specific descriptions can also lead to more questions. E.g. when can I conclude that a scoring item is absent? And is this the same for each scoring items or does every scoring item need its own explanation on when it is absent. Specifications can lead to more questions and can further complicate matters for the assessor. It is the challenge to find the right balance between specifying scoring items and leaving the interpretation of the item to the assessor, while at the same time preserve the reliability of the scoring instrument.

A hypothesis was formulated that a more specific rubric that is enriched with examples leads to a higher construct validity and also a higher IRR. However, the IRR of the control group tends to be slightly higher than the IRR of the experimental group. It is questionable whether this is due to the fact that a higher construct validity does not necessarily lead to a higher IRR, or that it can be explained by the fact that the construct validity was not significantly increased by designing the experimental rubric. As discussed before, most recommendations of Table 1 were not applied in the design of the experimental rubric for this study. In the result section, it was made plausible that the construct validity of the experimental rubric was higher than that of the experimental rubric. Nevertheless, the construct validity cannot be qualitatively measured and its quality can only be established by argumentative reasoning.

Regardless of the construct validity of the rubric, both the original and the experimental rubric resulted in scores with a low IRR. Kane (2006) states that when ambitious claims are made based on the scores of an instrument, extensive appraisal is called for. That is, both argumentative and empirical evidence on the instruments validity and reliability must be robust in order to use its scores for high stakes assessment. When assessing students with low interrater agreement, the results can be highly dependent on the characteristics of the assessors and the way the rubric is used by this assessor.

4.2 Limitations and Recommendations

After the invitation-process of many ICOs and the CHE-teachers, only few were present during the focus groups. Moreover, some participants had little or no experience on assessing students using rubrics. Therefore, both samples of the focus groups were not representative. This is another reason to question whether the input from the focus groups that is processed in the experimental rubric enhanced its construct validity. When a similar study is conducted, it is recommended to invite more participants for the focus groups, and to insure the presence of at least 12 participants since it is expected that participants cancel at the last moment.

It is suggested to find a replacement for the assessment instruments that has a higher reliability. Another option is to invest in training the assessors and the use of rubrics in order to

increase the reliability, since studies have shown that training the assessors is one of the foremost condition for reliable assessment (Eppich et al., 2015; Feldman et al., 2012 Lovorn & Rezaei, 2011). It is suggested to investigate the continua model of Grainger and Weir (2016) as discussed in the theoretical framework. This model gives clear standards for a certain score, but also gives assessors the liberty to nuance between those standards. This way the assessors can take into account the complexity of the daily educational practice of a classroom in their assessment. Finally, a recommendation to this study's institution is to use a reliability threshold for the assessment of students. To illustrate, new assessors could begin to follow a short training program after which three fragments of students' lessons need to be assessed. For all three fragments a standard could be developed for the agreement between the resulting scores of the new assessors and pre-determined scores that correspond to the fragments. This way a bar can be raised to include only assessors in the assessment process with a certain level of reliability.

Another recommendation is to use one assessment instrument for all quartiles, as discussed before. This creates opportunities with regard to assessment training and use of a reliability threshold. An assessment instrument can be selected that is already investigated on reliability and validity. Assessment training will be required during the implementation. A possible assessment instrument is the International Comparative Analysis of Learning and Teaching (ICALT) observation protocol (van de Grift, 2014). This is an instrument that is developed for the Dutch educational system and it has been investigated in multiple studies.

For further research it is recommended to include more changes to the rubric in order to increase the construct validity. In the experimental rubric still procedural knowledge is missing, such as when to fill in the assessment form (during or after the observation), how to act during the observation and how to score items that are absent in the observations. This information can be summarized in one general assessment manual in order to increase the IRR (Evers et al., 2010). Not only possible changes to the rubric need to be reviewed with regard to the scoring inference, but also the generalization-, extrapolation- and implication inference, as described by Kane (2006). This way it can be investigated whether the assessment instruments and protocol that is used are appropriate for the implications that are being made based on these assessments.

Finally, the recommendation is made that further qualitative research will be done by reviewing other film fragments of students' lessons with ICOs so that the researcher can specifically ask how the assessor transfers the sample of observations into the observed scores. This way the researcher can gain more insight in the scoring inferences that are currently present among ICOs. Especially how the complexity of the classroom is transferred into a rubric score. This may clarify the problem furthermore.

4.3 Conclusion

For this study the following research question was formulated:

How can the construct validity of a rubric be improved, and to what degree can this improvement increase the inter-rater reliability of the rubric's scores?

This study discussed several ways to improve the construct validity of a rubric. In the development of the experimental rubric, additional explanations and examples indented to increase this validity, along with the addition of general guidelines regarding the scoring rules. However, this can also be done by further clarifying the scoring rules and protocol, for instance in a scoring manual. Several recommendations have been made for further research in order to increase the construct validity of observation instruments.

The conclusion regarding the reliability is that the IRR is poor and unacceptable for both the original and experimental rubric. The ICC that resulted from this study is below 0.308, which is much lower than the boundary of 0.70 from which it can be stated that the ICC is fair. The conclusion regarding the construct validity is that it is plausible that the construct validity of the experimental rubric has been increased compared to the original rubric by adding clarifications and examples. However, even when it is assumed that the construct validity was increased, this did not lead to an increase of IRR. Therefore, this study shows no influence of the construct validity on the IRR. It is questionable whether adding information to the rubric clarifies its descriptions or just overcomplicates it for the assessor. Therefore the research question can be answered by stating that this study found no eminent evidence that improving a rubrics construct validity by adding clarifications and examples leads to an increase in the IRR. Nevertheless, this does not mean that improving the construct validity cannot lead to an increase of the IRR of scores, since limitations of this research show that more could have been done to increase the construct validity.

REFERENCE LIST

- Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational leadership*, 57(5), 13-19.
- Bresciani, M. J., Oakleaf, M., Kolkhorst, F., Nebeker, C., Barlow, J., Duncan, K., & Hickmott, J. (2009). Examining design and inter-rater reliability of a rubric measuring research quality across multiple disciplines. *Practical Assessment, Research & Evaluation*, 14(12), 1-7.
- Bryant, C. L., Maarouf, S., Burcham, J., & Greer, D. (2016). The examination of a teacher candidate assessment rubric: A confirmatory factor analysis. *Teaching and Teacher Education*, 57, 79-96.
- Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of clinical epidemiology*, 46(5), 423-429.
- Caughlan, S., & Jiang, H. (2014). Observation and teacher quality: Critical analysis of observational instruments in preservice teacher performance assessment. *Journal of Teacher Education*, 65(5), 375-388.
- Creemers, B., Kyriakides, L., & Antoniou, P. (2012). *Teacher professional development for improving quality of teaching*. Springer Science & Business Media.
- Darling-Hammond, L. (2006). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. *Journal of teacher education*, 57(2), 120-138.
- Dixon, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into practice*, 55(2), 153-159.
- Dooley, K. (2001). Social research methods. In 4th ed. Upper Saddle River, NJ.
- Duckor, B., Castellano, K. E., Téllez, K., Wihardini, D., & Wilson, M. (2014). Examining the internal structure evidence for the performance assessment for California teachers: A validation study of the elementary literacy teaching event for Tier I teacher licensure. *Journal of Teacher Education*, 65(5), 402-420.
- Eliot & Associates. (2005). Guidelines for conducting a focus group. Chapel Hill, NC: Author. Retrieved from https://datainnovationproject.org/wp-content/uploads/2017/04/4_How_to_Conduct_a_Focus_Group-2-1.pdf
- Eppich, W., Nannicelli, A. P., Seivert, N. P., Sohn, M. W., Rozenfeld, R., Woods, D. M., & Holl, J. L. (2015). A rater training protocol to assess team performance. *Journal of Continuing Education in the Health Professions*, 35(2), 83-90.
- Erdosy, M. U. (2003). Exploring variability in judging writing ability in a second language: A study of four experienced raters of ESL compositions. *ETS Research Report Series*, 2003(1).
- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). COTAN Beoordelingssysteem voor de Kwaliteit van Tests (geheel herziene versie)[COTAN Rating system for test quality (completely revised edition)]. *Amsterdam: NIP*.
- Falotico, R., & Quatto, P. (2015). Fleiss' kappa statistic without paradoxes. *Quality & Quantity*, 49(2), 463-470.
- Feldman, M., Lazzara, E. H., Vanderbilt, A. A., & DiazGranados, D. (2012). Rater training to support high-stakes simulation-based assessments. *Journal of Continuing Education in the Health Professions*, 32(4), 279-286.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5), 378.
- Garcia-Ros, R. (2011). *Analysis and validation of a rubric to assess oral presentation skills in university contexts*. *Electronic Journal in Educational Psychology*, 9(3), 1043-1062.
- Giraudeau, B. (1996). Negative values of the intraclass correlation coefficient are not theoretically possible. *Journal of clinical epidemiology*, 49(10), 1205.
- Grainger, P., & Weir, K. (2016). An alternative grading tool for enhancing assessment practice and quality assurance in higher education. *Innovations in Education and Teaching International*, 53(1), 73-83.
- Grossman, P., Cohen, J., & Brown, L. (2014). Understanding instructional quality in English Language Arts. In T. Kane, K. Kerr, & R. Pianta (Eds.), *Designing Teacher Evaluation Systems: New guidance from the measures of effective teaching project* (pp. 303–331). San Francisco: John Wiley & Sons.
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1), 23.
- Halpern, D. F. (2013). A is for assessment: The other scarlet letter. *Teaching of Psychology*, 40(4), 358-362.

- Harris, D., & McCaffrey, D. (2010). Value-added: Assessing teachers' contributions to student achievement. In M. M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook* (pp. 251-282). San Francisco, CA: Jossey-Bass.
- Hattie, J., & Marsh, H. W. (1996). The relationship between research and teaching: A meta-analysis. *Review of educational research*, 66(4), 507-542.
- Hernon, P., & Schwartz, C. (2009). Reliability and validity. *Library and Information Science Research*, 31(2), 73-74. doi:10.1016/j.lisr.2009.03.001
- Jeong, H. (2015). What is your teacher rubric? Extracting teachers' assessment constructs. *Practical Assessment, Research & Evaluation*, 20.
- Joe, J. N., Tocci, C. M., Holtzman, S. L., & Williams, J. C. (2013). Foundations of Observation: Considerations for Developing a Classroom Observation System That Helps Districts Achieve Consistent and Accurate Scores. MET Project, Policy and Practice Brief. *Bill & Melinda Gates Foundation*.
- Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment & Evaluation in Higher Education*, 39(7), 840-852.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational research review*, 2(2), 130-144.
- K-12 Education Team. (2010a). *The MQI Protocol for Classroom Observations*. Retrieved from <http://k12education.gatesfoundation.org/resource/the-mqi-protocol-for-classroom-observations/>
- K-12 Education Team. (2010b). *The PLATO Protocol for Classroom Observations*. Retrieved from <http://k12education.gatesfoundation.org/resource/the-plato-protocol-for-classroom-observations/>
- Klein, S. P., Stecher, B. M., Shavelson, R. J., McCaffrey, D., Ormseth, T., Bell, R. M., ... & Othman, A. R. (1998). Analytic versus holistic scoring of science performance tasks. *Applied Measurement in Education*, 11(2), 121-137.
- Klenowski, V. (2014). Towards fairer assessment. *The Australian Educational Researcher*, 41(4), 445-470.
- Krueger, R. A. (1994) *Focus Groups: A Practical Guide for Applied Research* (London: Sage Publications)
- Landers, R. (2015) Computing Intraclass Correlations (ICC) as Estimates of Interrater Reliability in SPSS, *The Winnower*. DOI:10.15200/winn.143518.81744
- Leventhal, L., Perry, R. P., & Abrami, P. C. (1977). Effects of lecturer quality and student perception of lecturer's experience on teacher ratings and student achievement. *Journal of Educational Psychology*, 69(4), 360.
- Lovorn, M. G., & Rezaei, A. R. (2011). Assessing the assessment: Rubrics training for pre-service and new in-service teachers. *Practical Assessment, Research & Evaluation*, 16(16), n16.
- Margolis, J., & Doring, A. (2013). National assessments for student teachers: Documenting teaching readiness to the tipping point. *Action in Teacher Education*, 35(4), 272-285.
- Nederlands-Vlaamse Accreditatieorganisatie (NVAO). (2015). *Hbo-bachelor Opleiding tot leraar basisonderwijs: systeembrede analyse*. Retrieved from <https://www.nvao.net/actueel/publicaties/systeembrede-analyse-lerarenopleidingen>
- Pavlina, K., Zorica, M. B., & Pongrac, A. (2011). Student perception of teaching quality in higher education. *Procedia-Social and Behavioral Sciences*, 15, 2288-2292.
- Pecheone, R., & Chung, R. R. (2006). Evidence in teacher education: The performance assessment for California teachers. *Journal of Teacher Education*, 57(1), 22-36.
- Penny, J., Johnson, R. L., & Gordon, B. (2000). Using rating augmentation to expand the scale of an analytic rubric. *The Journal of Experimental Education*, 68(3), 269-287.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational researcher*, 38(2), 109-119.
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & evaluation in higher education*, 35(4), 435-448.
- Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing writing*, 15(1), 18-39.
- Sato, M. (2014). What is the underlying conception of teaching of the edTPA?. *Journal of Teacher Education*, 65(5), 421-434.
- Slater, S. C., & Boulet, J. R. (2001). Predicting holistic ratings of written performance assessments from analytic scoring. *Advances in health sciences education*, 6(2), 103-119.
- Stuhlman, M. W., Hamre, B. K., Downer, J. T., & Pianta, R. C. (2010). A practitioner's guide to conducting classroom observations: What the research tells us about choosing and using

- observational systems. *Charlottesville: University of Virginia Center for Advanced Study of Teaching and Learning*.
- Thaler, N., Kazemi, E., & Huscher, C. (2009). Developing a rubric to assess student learning outcomes using a class assignment. *Teaching of Psychology, 36*(2), 113-116.
- van de Grift, W. J. C. M. (2014). Measuring teaching quality in several European countries. *School effectiveness and school improvement, 25*(3), 295–311. doi:10.1080/09243453.2013.794845
- van Merriënboer, J. J., & Tjiam, I. (2013). Development and teaching of complex skills in invasive procedures. In *Catheter-Based Cardiovascular Interventions* (pp. 173-186). Springer, Berlin, Heidelberg.
- Vos, H. J. (2009). Social research methods. *D. Dooley, Social Research Methods, 89*.
- Womack, S. A. (2011). Measuring Mathematics Instruction in Elementary Classrooms: Comprehensive Mathematics Instruction (CMI) Observation Protocol Development and Validation.
- Zeichner, K. (2010). Rethinking the connections between campus courses and field experiences in college-and university-based teacher education. *Journal of teacher education, 61*(1-2), 89-99.

Appendix A: Rubric VM1.2 Extern

KBS-leeruitkomst VM-cursus 1.2		De student is zich bewust van factoren die het leerklimaat in de klas beïnvloeden, kan een leidinggevende rol aannemen binnen kleine pedagogische relevante situaties in de klas en in lesontwerpen zichtbaar aansluiten bij de basisbehoeften van kinderen (relatie, competentie, autonomie).						
Naam student:		Beoordeeld(e) product/uitgevoerde handeling:			WPL-school en -groep:			
Prestatiecriteria	thema	0	1	2	3	4	Uitkomst	
Beroepsproduct								
Leerklimaat - BASISBEHOEFEN <i>n.b.</i> <i>Slechts door één van de drie beoordelaars in te vullen (bij voorkeur mentor of ico)</i>	Onge-structureerde observatie	Student neemt onbewust waar, zonder plan , beschrijft subjectief , trekt geen conclusie	Student observeert bewust , maar zonder plan , en beschrijft subjectief en/of ongenuanceerd (verslag met conclusie)	Student stelt observatieplan op, voert die uit , maar beschrijft subjectief en/of ongenueanceerd (verslag met conclusie)	Student stelt observatieplan op, voert die uit en beschrijft objectief en genuanceerd , onderbouwd met argumenten (die interpretatie van gegevens betrouwbaar maken)	Student stelt observatieplan op volgens 5XW+H , voert die uit en beschrijft objectief en genuanceerd, onderbouwd met argumenten (die interpretatie van gegevens betrouwbaar maken) en beschrijft bruikbaar advies aan zichzelf/leerkracht		
Beroepshandelen	thema	0	1	2	3	4	Uitkomst	
Leerklimaat - BASISBEHOEFEN <i>n.b.</i> <i>Door drie beoordelaars in te vullen.</i>	Relatie: 'Ik mag er zijn'	Student maakt geen contact met kinderen	Student maakt contact en sluit aan bij belevingswereld kind (<i>als 'technische' handeling</i>).	Student maakt contact en sluit op inhouds- en betrekkningsniveau (<i>natuurlijke belangstelling en interactie</i>) aan bij belevingswereld kind.	Student maakt contact, sluit op inhouds- en betrekkningsniveau aan bij belevingswereld kind en communiceert (in)direct 'je mag er zijn' verbaal en non-verbaal .	Student maakt contact, sluit op inhouds- en betrekkningsniveau aan bij belevingswereld kind, communiceert (in)direct 'je mag er zijn' verbaal en non-verbaal en stimuleert de relatie tussen de kinderen onderling .		
	Competentie: 'Ik kan het'	Student maakt kinderen niet competent .	Student maakt kinderen competent door zicht te geven op eigen kunnen .	Student maakt kinderen competent door zicht te geven op en vertrouwen in eigen kunnen .	Student maakt kinderen competent door zicht te geven op en vertrouwen in eigen kunnen .	Student maakt kinderen competent door zicht te geven op en vertrouwen in eigen kunnen en hierbij rekening houdt met individuele behoeften van het kind .	Student maakt kinderen competent door zicht te geven op en vertrouwen in eigen kunnen en hierbij rekening houdt met individuele behoeften van het kind en de actuele (evt. onvoorziene) situatie in de klas .	
	Autonomie: 'Ik kan het zelf'	Student geeft kinderen geen autonomie (geen ruimte voor eigen inbreng en keuzes, benadert hen als 'houd-je-mond-en-luisterklas').	Student geeft autonomie, maar sluit niet aan bij kind en de actuele (evt. onvoorziene) situatie in de klas, omdat het kind slechts uitgenodigd wordt tot procedurele vragen .	Student geeft autonomie door aan te sluiten bij kind en de actuele (evt. onvoorziene) situatie in de klas .	Student geeft autonomie door aan te sluiten bij (meer en minder initiatiefrijke) kinderen en de actuele (evt. onvoorziene) situatie in de klas .	Student geeft autonomie door aan te sluiten bij (meer en minder initiatiefrijke) kinderen en de actuele (evt. onvoorziene) situatie in de klas en de onderlinge interactie stimuleert .		
Leerklimaat - LEIDERSCHAP <i>n.b.</i> <i>Door drie beoordelaars in te vullen.</i>	Overzicht & overzicht <i>Grip op de groep</i>	Student ziet geen ongewenst gedrag en/of handelt niet .	Student ziet ongewenst gedrag maar handelt niet passend bij situatie.	Student ziet ongewenst gedrag en handelt passend bij situatie.	Student ziet ongewenst gedrag en handelt (verbaal en non-verbaal) passend bij situatie.	Student ziet ongewenst gedrag en handelt passend bij de situatie verbaal en non-verbaal, zowel proactief als reactief .		

Beoordelingstotaal en cijfer		Totaal punten:	Cijfer:
Beoordelaar	Functie/relatie tot student	Handtekening	Datum
Cesuur rubric extern VM1.2 <u>excl.</u> beoordeling beroepsproduct alleen beroepshandelen beoordeeld (leiderschap en basisbehoeften, geen observatie)	1 punt = cijfer 1 2 ptn = 1,5 3 ptn = 2 5 ptn = 3 6 ptn = 3,5	7 ptn = 4 8 ptn = 4,5 9 ptn = 5 10 ptn = 5,5 11 ptn = 6	12 ptn = 6,5 13 ptn = 7 14 ptn = 8 15 ptn = 9 16 ptn = 10
Cesuur rubric extern VM1.2 <u>incl.</u> beoordeling beroepsproduct (observatie)	1-3 punten = cijfer 1 4-6 ptn = 2 7-9 ptn = 3 10 ptn = 4 11 ptn = 4,5	12 ptn = 5 13 ptn = 5,5 14 ptn = 6 15 ptn = 6,5 16 ptn = 7	17 ptn = 7,5 18 ptn = 8 19 ptn = 9 20 ptn = 10

Appendix B:

Toestemmingsverklaringformulier (informed consent)

Titel onderzoek: INCREASING RELIABILITY OF CLASSROOM OBSERVATION INSTRUMENTS FOR PRE-SERVICE TEACHERS

Verantwoordelijke onderzoeker: Arno Broos

In te vullen door de deelnemer

Ik verklaar op een voor mij duidelijke wijze te zijn ingelicht over de aard, methode, doel en [indien aanwezig] de risico's en belasting van het onderzoek. Ik weet dat de gegevens en resultaten van het onderzoek alleen anoniem en vertrouwelijk aan derden bekend gemaakt zullen worden. Mijn vragen zijn naar tevredenheid beantwoord.

Ik begrijp dat film-, foto, en videomateriaal of bewerking daarvan uitsluitend voor analyse en/of wetenschappelijke presentaties zal worden gebruikt. Ik stem geheel vrijwillig in met deelname aan dit onderzoek. Ik behoud me daarbij het recht voor om op elk moment zonder opgaaf van redenen mijn deelname aan dit onderzoek te beëindigen.

Naam deelnemer:

Datum: Handtekening deelnemer:

Appendix C: Form Descriptive Characteristics Participants Focus Group

Formulier Karakteristieken Deelnemers Focus Group

Datum: _____ Tijd: _____ Plaats: _____

Wat is uw geslacht? m/v

Wat is uw functie: _____

Wat is uw rol met betrekking tot de CHE:

- Vakdocent
- Docentbegeleider
- Relatiebeheerder
- ICO
- Mentor
- anders, namelijk _____

In relatie tot de vorige vraag, hoeveel ervaring heeft u in deze rol:

- minder dan 5 jaar
- 5 tot 10 jaar
- meer dan 10 jaar

Hoeveel stagebeoordelingen geeft u gemiddeld per jaar:

- geen
- 0 tot 5 beoordelingen
- 5 tot 10 beoordelingen
- 10 tot 20 beoordelingen
- meer dan 20 beoordelingen

Appendix D: Script regarding the focus groups

Script regarding the focusgroup with the CHE-teachers

Welkom

Bedankt voor jullie deelname aan deze focus group. Wij waarderen jullie betrokkenheid en bereidheid.

Introductie van onderzoeker en assistent onderzoeker

Doel van de focus groep

Wij zijn gevraagd door de CHE om onderzoek te doen naar de betrouwbaarheid van de rubrics die gebruikt worden voor stagebeoordelingen. Met deze focus groep willen wij enkele elementen uit de externe rubric van VM1.2 verduidelijk en/of specificeren. We hebben jullie input nodig en vragen jullie dan ook eerlijk en open jullie gedachten te delen.

Basisregels

- *We vragen ieder volledig deel te nemen.*
- *Er zijn geen foute antwoorden. Alle inbreng is van belang. Laat het merken alss je het eens bent, en als je het niet eens bent. We horen graag een verscheidenheid aan meningen.*
- *Wat hier gezegd wordt blijft hier. Ook gevoelige punten moeten comfortabel gedeeld kunnen worden.*
- *Het geschreven werk wordt gebruikt voor het ontwikkelen van een vernieuwde rubric in een experiment. Alle gegevens worden anoniem verwerkt.*

Script regarding the focus group with the ICO's

Welkom

Bedankt voor jullie deelname aan deze focus group. Wij waarderen jullie betrokkenheid en bereidheid.

Introductie van onderzoeker en assistent onderzoeker

Doel van de focus groep

Wij zijn gevraagd door de CHE om onderzoek te doen naar de betrouwbaarheid van de rubrics die gebruikt worden voor stagebeoordelingen. Met deze focus groep willen wij ontdekken hoe wij vinden dat rubrics gebruikt moeten worden en waar verbetering mogelijk is. We hebben jullie input nodig en vragen jullie dan ook eerlijk en open jullie gedachten te delen.

Basisregels

- *Jullie praten. We vragen iedereen deel te nemen. Daardoor kan het zijn dat ik je inbreng vraag als ik wat langer niet van je hoor.*
- *Er zijn geen foute antwoorden. Alle inbreng is van belang. Vertel het als je het eens bent, en als je het niet eens bent. We horen graag een verscheidenheid aan meningen.*
- *Wat hier gezegd wordt blijft hier. Ook gevoelige punten moeten comfortabel gedeeld kunnen worden.*
- *Het gesprek wordt opgenomen. Maar ook anoniem verwerkt.*

Prestatiecriterium Relatie

“Maakt contact met de kinderen” – niveau 1

Wanneer maakt een student contact met kinderen? Wat is wel contact maken, en wat niet?
Probeer bovenstaand punt te verduidelijken en/of te specificeren.

“Sluit aan bij de belevingswereld op inhouds- en betrekkningsniveau” - niveau 2

Wanneer sluit een student aan bij de belevingswereld? En wanneer niet?
Probeer bovenstaand punt te verduidelijken en/of te specificeren.

Prestatiecriterium Competentie

“Zicht en vertrouwen geven in eigen kunnen” – niveau 2

Wat is het verschil tussen zicht geven op competentie en vertrouwen geven in competentie?

Wanneer is dit wel bereikt, wanneer niet?

Probeer bovenstaand punt te verduidelijken en/of te specificeren.

“...en de actuele (evt. onvoorziene) situatie in de klas” – niveau 4

Hoe kan de student bij het competent maken van kinderen, rekening houden met de actuele (evt. onvoorziene) situatie in de klas. Wanneer is dit niveau behaald? Wanneer niet?

Probeer bovenstaand punt te verduidelijken en/of te specificeren.

Prestatiecriterium Autonomie

“Procedurele vragen” - niveau 1

In het kader van autonomie, wanneer is iets een procedurele vraag? En wanneer niet?
Probeer bovenstaand punt te verduidelijken en/of te specificeren.

“...en de onderlinge interactie stimuleert” - niveau 2

Wanneer is de onderlinge interactie gestimuleerd? En wanneer niet?
Probeer bovenstaand punt te verduidelijken en/of te specificeren.

Prestatiecriterium Overzicht & Overwicht

“Ongewenst gedrag” – niveau 1

Wat is ongewenst gedrag? En wat niet?

Probeer bovenstaand punt te verduidelijken en/of te specificeren.

“Handelt passend”

Wanneer handelt een student passend op ongewenst gedrag? En wanneer niet?

Probeer bovenstaand punt te verduidelijken en/of te specificeren.

Appendix F: Questions for the focus groups with ICO's

Categorieën	Vragen
Openingsvraag	Wat zijn uw ervaringen met het beoordelen van studenten?
Introducerende vraag	Bij welke vakken heeft u rubrics gebruikt om studenten te beoordelen?
Transitionele vraag	Wat zijn de voordelen van rubrics bij het beoordelen van studenten op de stage?
Hoofdvragen	Welke onderdelen/begrippen van deze rubric (VM1.2) zijn nog onduidelijk voor u? (onderstreep deze) Hoe zou dit verduidelijkt kunnen worden? Met welke zinnen of zinsdelen bent u het oneens? (haal deze hoor) Hoe zou dit aangepast kunnen worden? Wat vindt u van de algemene richtlijnen die voor alle rubrics gelden?
Afsluitende vraag	Is er iets wat nog niet gezegd is en waarvan u sterk overtuigd bent dat dit nog gezegd moet worden, omtrent het gebruik van rubrics bij stagebeoordelingen?

Appendix G: Data Focus-Group 1, with CHE-Teachers

Deel van de rubric	Commentaar	Reacties op commentaar
Relatie: "Maakt contact met de kinderen"	-Hoe begroet je de kinderen? -Oogcontact: non-verbaal -> knipoog, hand op schouder, duim aai -Oogcontact: verbaal -> complimenten, belangstelling tonen, d.m.v. gesprekje o.i.d, feedback gerichte vraag, "thermometervragen" naar welbevinden, grenzen aangeven gericht op handelen/gedrag, onderscheid in "zijn" en "handelen" -Buiten, schoolplein, meedoen, belangstelling tonen -Huisbezoek -Kindgesprek: doorvragen naar weekend, hobby NIET: alles wat niet positief bijdraagt aan het opbouwen van een gezonde relatie, bijvoorbeeld reageren vanuit allergie, focus op eigen les, prestatiedrang, ongecontroleerde emotie	Mee eens
		Heel volledig
		Deels mee eens, is er voldoende discriminatie met "Ik kan het" en "Ik kan het zelf"
		NIET: negeren van het kind NIET: ingaan op een reactie (blik of opmerking)
Relatie: "Sluit aan bij de belevingswereld op inhouds- en betrekkingniveau"	-Op het moment dat je weet aan te sluiten op zowel: x Leefwereld (bv. woonwijk) x Belevingswereld (bv. games) x Actualiteit (bv. koningsdag) x Cultuur (bv. Armenië, land herkomst) -Intrinsieke motivatie aanboren NIET: als je puur bezig bent met lesdoel -> extrinsieke motivatie, bijv. als je methodegericht bent, als je uitgaat van het "gemiddelde kind"	Mee eens
		Deels mee eens, Is dit een beschrijving van niveau 2, of wordt ook niveau 3 aangeduid?
		Op betrekkingniveau: aansluitend bij het ontwikkelingsniveau en bij de persoonlijkheid (bijv. ASS Spectrum)
Competentie: "Zicht en vertrouwen geven in eigen kunnen"	Zicht op competentie: je bevestigt kinderen in wat ze kunnen en probeert ze een stapje verder te brengen. Verbaal of non-verbaal stimuleren. Vertrouwen geven in competentie: laten ervaren wat ze kunnen. Een opdracht laten uitvoeren	Mee eens
		Mee eens Bij vertrouwen geven toevoegen: ze ook leren dat ze fouten mogen maken
		Mee eens
		Zicht geven op: ook zelf kunnen benoemen. Dit is eerder cognitief. Vertrouwen is ook emotioneel -> durf
Competentie: "...en de actuele (evt. onvoorziene) situatie in de klas"	Improviseren op spontane gebeurtenissen. Kinderen inschakelen bij klusjes, elkaar helpen.	Oneens Kinderen gebruik laten maken van hun competenties op momenten die zich daarvoor lenen

	NIET: als alleen wat voorbereid is gedaan wordt	<p>Mee eens Aanvulling: inspringen en meebewegen met ideeën die door zowel kinderen + leerkracht + extern (directeur) spontaan opborrelen</p> <p>Gedeeltelijk mee eens Aanvullen met: wanneer de student kinderen in hun kracht zet door ze actief bij zaken te betrekken NIET: wanneer student rol alleen vervult en niets uit handen geeft</p> <p>Weet slim van de gelegenheid gebruik te maken om competentie te bevorderen</p>
Autonomie: "Procedurele vragen"	Vragen over de uitvoering van een opdracht. NIET: vragen over de inhoud van de opdracht. Autonomie: vragen over alternatieven, of je het ook anders zou kunnen aanpakken	<p>Mee eens</p> <p>Mee eens Aanvulling: ruimte geven voor eigen inbreng/manieren</p> <p>Niet mee eens mbt wanneer niet: Ook een vraag over inhoud kan procedureel zijn: op het moment dat kind met een alternatieve invulling/onderwerp komt voor inhoud</p>
Autonomie: "...en de onderlinge interactie stimuleert"	De leraar draagt een aangedragen alternatief voor voor de hele groep, met de uitnodiging om het hier over te hebben.	<p>Mee eens Helemaal als leraar vraagt om het alternatief nader toe te lichten Aanvulling: waardeer de opmerkingen + argumenten + bijdragen van kinderen. Laat kinderen ervaren dat iedereen mee kan/mag praten</p> <p>Mee eens Aanvulling: samenwerken</p> <p>Toevoeging: De leerkracht opent de les (bijv.) door leerlingen zelf over een onderwerp van alles naar voren te laten brengen</p>
Overzicht & Orde: "ongewenst gedrag"	Gedrag dat niet past bij de actuele situatie in de klas (bijv. praten als op dat moment stilte nodig is). Gedrag dat de les verstoort.	<p>Mee eens Aanvulling: als het gedrag andere kinderen niet tot opbouw is. Als het anderen naar beneden haalt</p> <p>Mee eens Aanvulling: gedrag wat niet afgesproken is</p> <p>Toevoeging: Gedrag dat kwetsend kan zijn voor de ander(en)</p>

		Ongewenst gedrag is niet: gedrag dat wordt beoordeeld als negatief wegens "allergie" van de leerkracht t.o.v. de leerling
Orde & Overzicht: "Handelt passend"	Duidelijk benoemd wat het probleem van het gedrag is; niet op de persoon gespeeld maar op het gedrag	Mee eens
		Mee eens Toevoeging: als student er een "ik" of "wij" boodschap van maakt ipv een "jij" boodschap
		Gedeeltelijk mee eens Vóór het benoemen kan ook gevraagd worden aan leerlingen naar motieven van zijn gedrag. Niet passend: als boodschap "jij ook altijd..." wordt gegeven.

1. Algemene Instructies:

- Probeer de tekst in de rubric zo letterlijk mogelijk te nemen.

- Neem alleen de inhoud van het criterium mee in je beoordeling.

Voorbeeld: Bij een criterium over "Overzicht & Overwicht" hoort niet "materiaalgebruik" "correcte spelling", "didactiek", "relatie" etc.

- Alle voorgaande elementen moeten aanwezig zijn, wil een student bijbehorende hoeveelheid punten scoren.

Voorbeeld: Wanneer een element bij 3 punten afwezig is, scoort de student dus maximaal 2 punten.

- Bij twijfel tussen 2 scores mogen er geen halve punten gegeven worden.

- Geef eerst alle punten, reken daarna pas het cijfer uit. Dit mag niet meer veranderd worden.

VM1.2 Rubric beoordeling EXTERN (t.b.v. 360-graden-feedback)

U gaat dadelijk een filmfragment bekijken en deze beoordelen met behulp van onderstaande rubric en bijgevoegde instructies.

Lees eerst de algemene instructies links, vervolgens de toegevoegde instructies onder en rechts van de rubric. U vindt hier extra uitleg voor de gekleurde woorden uit de rubric. Het eerste criterium met betrekking tot het beroepsproduct hoeft u niet te beoordelen.

Fragment nr.: _____

KBS-leeruitkomst VM-cursus 1.2							
De student is zich bewust van factoren die het leerklimaat in de klas beïnvloeden, kan een leidinggevende rol aannemen binnen kleine pedagogische relevante situaties in de klas en in lesontwerpen zichtbaar aansluiten bij de basisbehoeften van kinderen (relatie, competentie, autonomie).		Naam student: n.v.t.		Beoordeeld(e) product/uitgevoerde handeling: n.v.t.		WPL-school en -groep: n.v.t.	
Prestatiecriteria	thema	0	1	2	3	4	Uitkomst
Beroepsproduct							
Leerklimaat - BASISBEHOEFEN <i>Slechts door één van de drie beoordelaars in te vullen (bij voorkeur mentor of ico)</i>	Onge-structureerde observatie	Student neemt onbewust waar, zonder plan , beschrijft subjectief , trekt geen conclusie	Student observeert bewust , maar zonder plan , en beschrijft subjectief en/of ongenuanceerd (verslag met conclusie)	Student stelt observatieplan op, voert die uit , maar beschrijft subjectief en/of ongenuanceerd (verslag met conclusie)	Student stelt observatieplan op, voert die uit en beschrijft objectief en genuanceerd , onderbouwd met argumenten (die interpretatie van gegevens betrouwbaar maken)	Student stelt observatieplan op volgens 5XW+H , voert die uit en beschrijft objectief en genuanceerd , onderbouwd met argumenten (die interpretatie van gegevens betrouwbaar maken) en beschrijft bruikbaar advies aan zichzelf/leerkracht	
Beroepshandelen							
Leerklimaat - BASISBEHOEFEN <i>Door drie beoordelaars in te vullen.</i>	Relatie: 'Ik mag er zijn'	Student maakt geen contact met kinderen	Student maakt contact en sluit aan bij belevingswereld kind (als "technische" handeling).	Student maakt contact en sluit op inhouds- en betrekkningsniveau (natuurlijke belangstelling en interactie) aan bij belevingswereld kind.	Student maakt contact, sluit op inhouds- en betrekkningsniveau aan bij belevingswereld kind en communiceert (indirect 'je mag er zijn') verbaal en non-verbaal.	Student maakt contact, sluit op inhouds- en betrekkningsniveau aan bij belevingswereld kind, communiceert (indirect 'je mag er zijn') verbaal en non-verbaal en stimuleert de relatie tussen de kinderen onderling.	
	Competentie: 'Ik kan het'	Student maakt kinderen niet competent.	Student maakt kinderen competenter door zicht te geven op eigen kunnen.	Student maakt kinderen competenter door zicht te geven op en vertrouwen in eigen kunnen.	Student maakt kinderen competenter door zicht te geven op en vertrouwen in eigen kunnen en hierbij rekening houdt met individuele behoeften van het kind.	Student maakt kinderen competenter door zicht te geven op en vertrouwen in eigen kunnen en hierbij rekening houdt met individuele behoeften van het kind.	
	Autonomie: 'Ik kan het zelf'	Student geeft kinderen geen autonomie (geen ruimte voor eigen inbreng en keuzes, benadert hen als 'houd-je-mond-en-luisterklas').	Student geeft autonomie , maar sluit niet aan bij kind en de actuele (evt. onvoorziene) situatie in de klas, omdat het kind slechts uitgenodigd wordt tot procedurele vragen.	Student geeft autonomie door aan te sluiten bij kind en de actuele (evt. onvoorziene) situatie in de klas.	Student geeft autonomie door aan te sluiten bij meer en minder initiatiefrijke kinderen en de actuele (evt. onvoorziene) situatie in de klas.	Student geeft autonomie door aan te sluiten bij meer en minder initiatiefrijke kinderen en de actuele (evt. onvoorziene) situatie in de klas en de onderlinge interactie stimuleert.	
Leerklimaat - LEIDERSCHAP <i>Door drie beoordelaars in te vullen.</i>	Overzicht & overwicht <i>Grip op de groep</i>	Student ziet geen ongewenst gedrag en/of handelt niet.	Student ziet ongewenst gedrag maar handelt niet passend bij situatie.	Student ziet ongewenst gedrag en handelt passend bij situatie.	Student ziet ongewenst gedrag en handelt (verbaal en non-verbaal) passend bij situatie.	Student ziet ongewenst gedrag en handelt passend bij de situatie verbaal en non-verbaal, zowel proactief als reactief.	
Beoordelingstotaal en cijfer				Totaal punten:		Cijfer:	
Beoordelaar	Functie/relatie tot student	Handtekening		Datum			
Cesuur rubric extern VM1.2 <i>excl. beoordeling beroepsproduct</i> <i>alleen beroepshandelen, geen observatie</i>	1 punt = cijfer 1 2 ptn = 1,5 3 ptn = 2	5 ptn = 3 6 ptn = 3,5 7 ptn = 4	8 ptn = 4,5 9 ptn = 5 10 ptn = 5,5	11 ptn = 6 12 ptn = 6,5 13 ptn = 7	14 ptn = 8 15 ptn = 9 16 ptn = 10		

Handelt passend

Indien nodig, eerst om verduidelijking vragen. Daarna duidelijk benoemen wat het probleem van het gedrag is. Niet op de persoon, maar op het gedrag.

Ongewenst gedrag

Gedrag dat niet past bij de actuele situatie in de klas. Gedrag dat de les verstoort. Gedrag dat niet afgesproken is. VOORBEELD: praten als op dat moment stilte nodig is

Student maakt contact

WEL: verbaal/non-verbaal, reageren, doorvragen, interactie

NIET: regelmatig negeren van kinderen, focus slechts op eigen les

Sluit aan bij de belevingswereld op inhouds- en betrekkningsniveau

Inhoudsniveau: welk onderwerp of lesinhoud wordt gekozen?

Betrekkningsniveau: Sluit de student aan bij leeftijd en persoonlijkheid?

Zicht en vertrouwen geven in eigen kunnen

Zicht geven: kinderen bevestigen in wat ze kunnen en het hen zelf laten benoemen

Vertrouwen geven: Laten ervaren wat ze kunnen. Heeft vaak betrekking op emotie

...en de actuele (evt. onvoorziene) situatie in de klas

WEL: Inspringen en meebewegen met ideeën buiten eigen inbreng en hierbij kinderen inschakelen, passend bij hun competentie

NIET: alle initiatief zelf vasthouden

Het kind wordt slechts uitgenodigd tot procedurele vragen

WEL: Vragen over het hoe en wat van de opdracht waarbij de leerkracht alle vragen beantwoord.

NIET: Vragen over het hoe en wat van de opdracht waarbij de kinderen eigen inbreng ontvangen

"...en de onderlinge interactie stimuleert"

De leerkracht gaat in op een aangedragen alternatief van een leerling en geeft ruimte om dit met de leerkracht of onderling te bespreken. Het kind ervaart dat het mee mag praten.

Welkom

Bedankt voor jullie deelname aan dit onderzoek. Wij waarderen jullie betrokkenheid en bereidheid.

Introductie van onderzoeker

Doel van het onderzoek

Wij zijn gevraagd door de CHE om onderzoek te doen naar de betrouwbaarheid van de rubrics die gebruikt worden voor stagebeoordelingen. Voor het onderzoek vergelijken we twee soorten rubrics. Hiermee willen we informatie verkrijgen over de betrouwbaarheid van rubrics en waar eventueel verbetering mogelijk is. U gaat zo twee lessen van studenten beoordelen aan de hand van twee filmfragmenten met behulp van de rubric van VM1.2. Om de resultaten zo betrouwbaar mogelijk te laten zijn vragen we u om...:

- *...niet te praten, te overleggen of te kijken bij een ander. Pas wanneer de laatste rubrics zijn ingeleverd mag u over uw ingevulde scores en overwegingen spreken. Indien u voor die tijd (al dan niet spontaan) reageert, is het onderzoek niet meer betrouwbaar en moet het opnieuw uitgevoerd worden met alle gevolgen van dien.*
- *Het is niet toegestaan om vragen te stellen. Door het onderzoeksontwerp is dit niet nodig, als het goed is. Als iets onduidelijk is, ga dan door zoals u denkt dat het beste is.*
- *Er zijn geen foute antwoorden. Alle inbreng is van belang. We zijn juist benieuwd naar hoe u een bepaalde student in een bepaalde situatie op een criterium zou scoren. Bovendien worden de gegevens worden anoniem verwerkt.*
- *Het geheel zal ongeveer een halfuur duren. We vragen u om gedurende deze periode gefocust te blijven.*
- *Ten slotte vragen we u vooraf een formulier voor toestemmingsverklaring in te vullen, zodat de scores die u invult gebruikt kunnen worden voor het onderzoek. De formulieren worden direct weer ingenomen om anonimiteit van de rest van het onderzoek te garanderen.*

De ingevulde toestemmingsverklaringen worden opgehaald.

Zo dadelijk gaat u het eerste fragment beoordelen. Pak daarvoor de rubric waar boven staat "Fragment 1". U krijgt nu 5 minuten om het beoordelingsformulier goed door te lezen. We vragen u het geheel in stilte grondig te lezen, zodat u goed voorbereid bent op het fragment.

Na 5 minuten

Goed, u heeft 5 minuten de tijd gehad het beoordelingsformulier door te nemen. Nu start het eerste fragment.

Eerste fragment wordt gestart op de beamer

U heeft zojuist het eerste fragment gezien. We vragen u uw het huidige beoordelingsformulier van fragment 1 omgekeerd op uw tafel te leggen. Zodra u dat gedaan heeft, vragen we u uw eerste beoordeling niet meer aan te passen om zo de betrouwbaarheid van dit onderzoek te kunnen garanderen.

Pak nu de rubric erbij waar boven staat "Fragment 2". Deze is identiek aan de volgende, daarom kunnen we nu direct starten met het tweede fragment.

Tweede fragment wordt gestart op de beamer

U heeft zojuist het tweede fragment gezien. We vragen u opnieuw uw huidige beoordelingsformulier omgekeerd neer te leggen en zodra u dit gedaan hebt hier niets meer op aan te passen.

U krijgt nu als laatste een korte vragenlijst wat betreft uw geslacht, leeftijd, functie et cetera. Opnieuw vragen we u die in te vullen, dit zal slechts 1/2 minuten kosten. Als u klaar bent vragen we u alle drie de formulieren weer terug in het hoesje te stoppen.

Deelnemers vullen descriptieve vragenlijst in. Ondertussen wordt van elke set beoordelingsformulieren één formulier ingenomen.

Hartelijk dank voor uw deelname aan dit onderzoek. Indien u op de hoogte wilt blijven van het onderzoek kunt u contact opnemen met de gegevens die op het beoordelingsformulier staan.

Fragment 1

KBS-leeruitkomst VM-cursus 1.2		De student is zich bewust van factoren die het leerklimaat in de klas beïnvloeden, kan een leidinggevende rol aannemen binnen kleine pedagogische relevante situaties in de klas en in lesontwerpen zichtbaar aansluiten bij de basisbehoeften van kinderen (relatie, competentie, autonomie).						
Naam student: n.v.t.		Beoordeeld(e) product/uitgevoerde handeling: n.v.t.			WPL-school en -groep: n.v.t.			
Prestatiecriteria	thema	0	1	2	3	4	Uitkomst	
Beroepsproduct								
Leerklimaat – BASISBEHOEFEN <i>Slechts door één van de drie beoordelaars in te vullen (bij voorkeur mentor of ico)</i>	Onge-structureerde observatie	Student neemt onbewust waar, zonder plan , beschrijft subjectief , trekt geen conclusie	Student observeert bewust , maar zonder plan , en beschrijft subjectief en/of ongenuanceerd (verslag met conclusie)	Student stelt observatieplan op, voert die uit , maar beschrijft subjectief en/of ongenuanceerd (verslag met conclusie)	Student stelt observatieplan op, voert die uit en beschrijft objectief en genuanceerd , onderbouwd met argumenten (die interpretatie van gegevens betrouwbaar maken)	Student stelt observatieplan op volgens 5XW+H , voert die uit en beschrijft objectief en genuanceerd , onderbouwd met argumenten (die interpretatie van gegevens betrouwbaar maken) en beschrijft bruikbaar advies aan zichzelf/leerkracht		
Beroepshandelen	thema	0	1	2	3	4	Uitkomst	
Leerklimaat - BASISBEHOEFEN <i>Door drie beoordelaars in te vullen.</i>	Relatie: 'Ik mag er zijn'	Student maakt geen contact met kinderen	Student maakt contact en sluit aan bij belevingswereld kind (als 'technische' handeling).	Student maakt contact en sluit op inhouds- en betrekkningsniveau (natuurlijke belangstelling en interactie) aan bij belevingswereld kind.	Student maakt contact, sluit op inhouds- en betrekkningsniveau aan bij belevingswereld kind en communiceert (in)direct 'je mag er zijn' verbaal en non-verbaal .	Student maakt contact, sluit op inhouds- en betrekkningsniveau aan bij belevingswereld kind, communiceert (in)direct 'je mag er zijn' verbaal en non-verbaal en stimuleert de relatie tussen de kinderen onderling .		
	Competentie: 'Ik kan het'	Student maakt kinderen niet competentier .	Student maakt kinderen competentier door zicht te geven op eigen kunnen .	Student maakt kinderen competentier door zicht te geven op en vertrouwen in eigen kunnen .	Student maakt kinderen competentier door zicht te geven op en vertrouwen in eigen kunnen en hierbij rekening houdt met individuele behoeften van het kind .	Student maakt kinderen competentier door zicht te geven op en vertrouwen in eigen kunnen en hierbij rekening houdt met individuele behoeften van het kind .		
	Autonomie: 'Ik kan het zelf'	Student geeft kinderen geen autonomie (geen ruimte voor eigen inbreng en keuzes, benadert hen als 'houd-je-mond- en luisterklas').	Student geeft autonomie, maar sluit niet aan bij kind en de actuele (evt. onvoorziene) situatie in de klas, omdat het kind slechts uitgenodigd wordt tot procedurale vragen .	Student geeft autonomie door aan te sluiten bij kind en de actuele (evt. onvoorziene) situatie in de klas .	Student geeft autonomie door aan te sluiten bij (meer en minder initiatiefrijke) kinderen en de actuele (evt. onvoorziene) situatie in de klas.	Student geeft autonomie door aan te sluiten bij (meer en minder initiatiefrijke) kinderen en de actuele (evt. onvoorziene) situatie in de klas.	Student geeft autonomie door aan te sluiten bij (meer en minder initiatiefrijke) kinderen en de actuele (evt. onvoorziene) situatie in de klas en de onderlinge interactie stimuleert .	
Leerklimaat - LEIDERSCHAP <i>Door drie beoordelaars in te vullen.</i>	Overzicht & overzicht <i>Grip op de groep</i>	Student ziet geen ongewenst gedrag en/of handelt niet .	Student ziet ongewenst gedrag maar handelt niet passend bij situatie.	Student ziet ongewenst gedrag en handelt passend bij situatie.	Student ziet ongewenst gedrag en handelt (verbaal en non-verbaal) passend bij situatie.	Student ziet ongewenst gedrag en handelt passend bij de situatie verbaal en non-verbaal, zowel proactief als reactief .		
Beoordelingstotaal en cijfer					Totaal punten:	Cijfer:		
Beoordelaar	Functie/relatie tot student			Handtekening		Datum		
Cesuur rubric extern VM1.2 <i>excl. beoordeling beroepsproduct</i> <i>alleen beroepshandelen, geen observatie</i>	1 punt = cijfer 1		5 ptn = 3		8 ptn = 4,5		11 ptn = 6	
	2 ptn = 1,5		6 ptn = 3,5		9 ptn = 5		12 ptn = 6,5	
	3 ptn = 2		7 ptn = 4		10 ptn = 5,5		13 ptn = 7	
Cesuur rubric extern VM1.2 <i>incl. beoordeling beroepsproduct</i>	1-3 punten = cijfer 1		10 ptn = 4		13 ptn = 5,5		16 ptn = 7	
	4-6 ptn = 2		11 ptn = 4,5		14 ptn = 6		17 ptn = 7,5	
	7-9 ptn = 3		12 ptn = 5		15 ptn = 6,5		18 ptn = 8	

Fragment 2

KBS-leeruitkomst VM-cursus 1.2		De student is zich bewust van factoren die het leerklimaat in de klas beïnvloeden, kan een leidinggevende rol aannemen binnen kleine pedagogische relevante situaties in de klas en in lesontwerpen zichtbaar aansluiten bij de basisbehoeften van kinderen (relatie, competentie, autonomie).						
Naam student: n.v.t.		Beoordeeld(e) product/uitgevoerde handeling: n.v.t.			WPL-school en -groep: n.v.t.			
Prestatiecriteria	thema	0	1	2	3	4	Uitkomst	
Beroepsproduct								
Leerklimaat – BASISBEHOEFEN <i>Slechts door één van de drie beoordelaars in te vullen (bij voorkeur mentor of ico)</i>	Onge-structureerde observatie	Student neemt onbewust waar, zonder plan , beschrijft subjectief , trekt geen conclusie	Student observeert bewust , maar zonder plan , en beschrijft subjectief en/of ongenuanceerd (verslag met conclusie)	Student stelt observatieplan op, voert die uit , maar beschrijft subjectief en/of ongenuanceerd (verslag met conclusie)	Student stelt observatieplan op, voert die uit en beschrijft objectief en genuanceerd , onderbouwd met argumenten (die interpretatie van gegevens betrouwbaar maken)	Student stelt observatieplan op volgens 5XW+H , voert die uit en beschrijft objectief en genuanceerd , onderbouwd met argumenten (die interpretatie van gegevens betrouwbaar maken) en beschrijft bruikbaar advies aan zichzelf/leerkracht		
Beroepshandelen	thema	0	1	2	3	4	Uitkomst	
Leerklimaat - BASISBEHOEFEN <i>Door drie beoordelaars in te vullen.</i>	Relatie: 'Ik mag er zijn'	Student maakt geen contact met kinderen	Student maakt contact en sluit aan bij belevingswereld kind (als 'technische' handeling).	Student maakt contact en sluit op inhouds- en betrekkningsniveau (natuurlijke belangstelling en interactie) aan bij belevingswereld kind.	Student maakt contact, sluit op inhouds- en betrekkningsniveau aan bij belevingswereld kind en communiceert (in)direct 'je mag er zijn' verbaal en non-verbaal .	Student maakt contact, sluit op inhouds- en betrekkningsniveau aan bij belevingswereld kind, communiceert (in)direct 'je mag er zijn' verbaal en non-verbaal en stimuleert de relatie tussen de kinderen onderling .		
	Competentie: 'Ik kan het'	Student maakt kinderen niet competentier .	Student maakt kinderen competentier door zicht te geven op eigen kunnen .	Student maakt kinderen competentier door zicht te geven op en vertrouwen in eigen kunnen .	Student maakt kinderen competentier door zicht te geven op en vertrouwen in eigen kunnen en hierbij rekening houdt met individuele behoeften van het kind .	Student maakt kinderen competentier door zicht te geven op en vertrouwen in eigen kunnen en hierbij rekening houdt met individuele behoeften van het kind .		
	Autonomie: 'Ik kan het zelf'	Student geeft kinderen geen autonomie (geen ruimte voor eigen inbreng en keuzes, benadert hen als 'houd-je-mond- en luisterklas').	Student geeft autonomie, maar sluit niet aan bij kind en de actuele (evt. onvoorziene) situatie in de klas, omdat het kind slechts uitgenodigd wordt tot procedurele vragen .	Student geeft autonomie door aan te sluiten bij kind en de actuele (evt. onvoorziene) situatie in de klas .	Student geeft autonomie door aan te sluiten bij (meer en minder initiatiefrijke) kinderen en de actuele (evt. onvoorziene) situatie in de klas.	Student geeft autonomie door aan te sluiten bij (meer en minder initiatiefrijke) kinderen en de actuele (evt. onvoorziene) situatie in de klas.	Student geeft autonomie door aan te sluiten bij (meer en minder initiatiefrijke) kinderen en de actuele (evt. onvoorziene) situatie in de klas en de onderlinge interactie stimuleert .	
Leerklimaat - LEIDERSCHAP <i>Door drie beoordelaars in te vullen.</i>	Overzicht & overzicht <i>Grip op de groep</i>	Student ziet geen ongewenst gedrag en/of handelt niet .	Student ziet ongewenst gedrag maar handelt niet passend bij situatie.	Student ziet ongewenst gedrag en handelt passend bij situatie.	Student ziet ongewenst gedrag en handelt (verbaal en non-verbaal) passend bij situatie.	Student ziet ongewenst gedrag en handelt passend bij de situatie verbaal en non-verbaal, zowel proactief als reactief .		
Beoordelingstotaal en cijfer					Totaal punten:	Cijfer:		
Beoordelaar	Functie/relatie tot student			Handtekening		Datum		
Cesuur rubric extern VM1.2 <i>excl. beoordeling beroepsproduct</i> <i>alleen beroepshandelen, geen observatie</i>	1 punt = cijfer 1 2 ptn = 1,5 3 ptn = 2		5 ptn = 3 6 ptn = 3,5 7 ptn = 4		8 ptn = 4,5 9 ptn = 5 10 ptn = 5,5		11 ptn = 6 12 ptn = 6,5 13 ptn = 7 14 ptn = 8 15 ptn = 9 16 ptn = 10	
Cesuur rubric extern VM1.2 <i>incl. beoordeling beroepsproduct</i>	1-3 punten = cijfer 1 4-6 ptn = 2 7-9 ptn = 3		10 ptn = 4 11 ptn = 4,5 12 ptn = 5		13 ptn = 5,5 14 ptn = 6 15 ptn = 6,5		16 ptn = 7 17 ptn = 7,5 18 ptn = 8 19 ptn = 9 20 ptn = 10	

1. Algemene Instructies:

- Probeer de tekst in de rubric zo letterlijk mogelijk te nemen.

- Neem alleen de inhoud van het criterium mee in je beoordeling.

Voorbeeld: Bij een criterium over "Overzicht & Overwicht" hoort niet "materiaalgebruik" "correcte spelling", "didactiek", "relatie" etc.

- Alle elementen moeten aanwezig zijn, voor de bijbehorende hoeveelheid punten.

Voorbeeld: Wanneer een onderdeel bij 3 punten afwezig is, scoort de student dus maximaal 2 punten.

- Bij twijfel tussen 2 scores mogen er geen halve punten gegeven worden.

- Geef eerst alle punten, reken daarna pas het cijfer uit. Dit mag dan niet meer veranderd worden.

Fragment 1

VM1.2 Rubric beoordeling EXTERN (t.b.v. 360-graden-feedback)

U gaat dadelijk een filmfragment bekijken en deze beoordelen met behulp van onderstaande rubric en bijgevoegde instructies.

Lees eerst de algemene instructies links, vervolgens de toegevoegde instructies onder en rechts van de rubric. U vindt hier extra uitleg voor de gekleurde woorden uit de rubric. Het eerste criterium met betrekking tot het beroepsproduct hoeft u niet te beoordelen.

KBS-leeruitkomst VM-cursus 1.2							
De student is zich bewust van factoren die het leerklimaat in de klas beïnvloeden, kan een leidinggevende rol aannemen binnen kleine pedagogische relevante situaties in de klas en in lesontwerpen zichtbaar aansluiten bij de basisbehoeften van kinderen (relatie, competentie, autonomie).		Naam student: n.v.t.		Beoordeeld(e) product/uitgevoerde handeling: n.v.t.		WPL-school en -groep: n.v.t.	
Prestatiecriteria	thema	0	1	2	3	4	Uitkomst
Leerklimaat – BASISBEHOEFEN <i>Slechts door één van de drie beoordelaars in te vullen (bij voorkeur mentor of ico)</i>	Onge-structureerde observatie	Student neemt onbewust waar, zonder plan , beschrijft subjectief , trekt geen conclusie	Student observeert bewust , maar zonder plan , en beschrijft subjectief en/of ongenuanceerd (verslag met conclusie)	Student stelt observatieplan op, voert die uit , maar beschrijft subjectief en/of ongenuanceerd (verslag met conclusie)	Student stelt observatieplan op, voert die uit en beschrijft objectief en genueanceerd , onderbouwd met argumenten (die interpretatie van gegevens betrouwbaar maken)	Student stelt observatieplan op volgens 5XW+H , voert die uit en beschrijft objectief en genueanceerd , onderbouwd met argumenten (die interpretatie van gegevens betrouwbaar maken) en beschrijft bruikbaar advies aan zichzelf/leerkracht	
Beroepshandelen	thema	0	1	2	3	4	Uitkomst
Leerklimaat - BASISBEHOEFEN <i>Door drie beoordelaars in te vullen.</i>	Relatie: 'Ik mag er zijn'	Student maakt geen contact met kinderen	Student maakt contact en sluit aan bij belevingswereld kind (als "technische" handeling, nog onwennig, zonder inlevingsvermogen).	Student maakt contact en sluit op inhouds- en betrekkningsniveau (natuurlijke belangstelling en interactie) aan bij belevingswereld kind .	Student maakt contact, sluit op inhouds- en betrekkningsniveau aan bij belevingswereld kind en communiceert (indirect 'je mag er zijn') verbaal en non-verbaal .	Student maakt contact, sluit op inhouds- en betrekkningsniveau aan bij belevingswereld kind, communiceert (indirect 'je mag er zijn') verbaal en non-verbaal en stimuleert de relatie tussen de kinderen onderling .	
	Competentie: 'Ik kan het'	Student maakt kinderen niet competent .	Student maakt kinderen competenter door zicht te geven op eigen kunnen .	Student maakt kinderen competenter door zicht te geven op en vertrouwen in eigen kunnen .	Student maakt kinderen competenter door zicht te geven op en vertrouwen in eigen kunnen en hierbij rekening houdt met individuele behoeften van het kind .	Student maakt kinderen competenter door zicht te geven op en vertrouwen in eigen kunnen en hierbij rekening houdt met individuele behoeften van het kind en de actuele (evt. onvoorziene) situatie in de klas .	
	Autonomie: 'Ik kan het zelf'	Student geeft kinderen geen autonomie (geen ruimte voor eigen inbreng en keuzes, benadert hen als 'houd-je-mond-en luisterklas').	Student geeft autonomie, maar sluit niet aan bij kind en de actuele (evt. onvoorziene) situatie in de klas, omdat het kind slechts uitgenodigd wordt tot procedurele vragen .	Student geeft autonomie door aan te sluiten bij kind en de actuele (evt. onvoorziene) situatie in de klas .	Student geeft autonomie door aan te sluiten bij meer en minder initiatiefrijke kinderen en de actuele (evt. onvoorziene) situatie in de klas.	Student geeft autonomie door aan te sluiten bij meer en minder initiatiefrijke kinderen en de actuele (evt. onvoorziene) situatie in de klas en de onderlinge interactie stimuleert .	
Leerklimaat - LEIDERSCHAP <i>Door drie beoordelaars in te vullen.</i>	Overzicht & overwicht <i>Grip op de groep</i>	Student ziet geen ongewenst gedrag, of ziet het wel maar handelt niet .	Student ziet ongewenst gedrag maar handelt niet passend bij situatie.	Student ziet ongewenst gedrag en handelt passend bij situatie.	Student ziet ongewenst gedrag en handelt (verbaal en non-verbaal) passend bij situatie.	Student ziet ongewenst gedrag en handelt passend bij de situatie verbaal en non-verbaal, zowel proactief als reactief .	
Beoordelingstotaal en cijfer				Totaal punten:		Cijfer:	
Beoordelaar	Functie/relatie tot student	Handtekening		Datum			
Cesuur rubric extern VM1.2	1 punt = cijfer 1	5 ptn = 3	8 ptn = 4,5	11 ptn = 6	14 ptn = 8		
excl. beoordeling beroepsproduct	2 ptn = 1,5	6 ptn = 3,5	9 ptn = 5	12 ptn = 6,5	15 ptn = 9		
alleen beroepshandelen, geen observatie	3 ptn = 2	7 ptn = 4	10 ptn = 5,5	13 ptn = 7	16 ptn = 10		

Handelt passend

Indien nodig, eerst om verduidelijking vragen. Daarna duidelijk benoemen wat het probleem van het gedrag is. Niet op de persoon, maar op het gedrag.

Ongewenst gedrag

Gedrag dat niet past bij de actuele situatie in de klas. Gedrag dat de les verstoort. Gedrag dat niet afgesproken is. VOORBEELD: praten als op dat moment stilte nodig is

"...en de onderlinge interactie stimuleert"

In het kader van autonomie: Ingaan op een aangedragen alternatief van een leerling (of hiernaar vragen), én ruimte geven om dit onderling te bespreken. Het kind ervaart dat het mee mag praten. NIET: De kinderen laten samenwerken

Student maakt contact

WEL: verbaal/non-verbaal, reageren, doorvragen, interactie

NIET: regelmatig negeren van kinderen, focus slechts op eigen les

Cesuur rubric extern VM1.2

Sluit aan bij de belevingswereld op inhouds- en betrekkningsniveau

Inhoudsniveau: welk onderwerp of lesinhoud wordt gekozen?

Betrekkningsniveau: Sluit de student aan bij leeftijd en interesse van de groep

Zicht en vertrouwen geven in eigen kunnen

ZICHT GEVEN: kinderen bevestigen in wat ze kunnen en het hen zelf laten benoemen
VERTROUWEN GEVEN: Laten ervaren wat ze kunnen. Heeft vaak betrekking op emotie en doorzettingsvermogen.

En hierbij rekening houdt met individuele behoeften van kind.

WEL: een kind competent laten voelen in dat wat het kind kan leren (zone naaste ontwikkeling)
NIET: wanneer de student een kind overtuigt van een competentie, terwijl het kind die competentie nog niet heeft en nog niet kan leren

...en de actuele (evt. onvoorziene) situatie in de klas

WEL: Inspringen en meebewegen met ideeën buiten eigen inbreng. *Als er geen onvoorziene situatie is, vragen naar inbreng v.d. kinderen*
NIET: alle initiatief zelf vasthouden

Het kind wordt slechts uitgenodigd tot procedurele vragen

WEL: Vragen over het hoe en wat van de opdracht waarbij de student alle vragen beantwoordt.
NIET: Vragen over het hoe en wat van de opdracht waarbij de kinderen eigen inbreng ontvangen

Appendix M: Experimental Rubric for VM1.2. Fragment 2

1. Algemene Instructies:

- Probeer de tekst in de rubric zo letterlijk mogelijk te nemen.

- Neem alleen de inhoud van het criterium mee in je beoordeling.

Voorbeeld: Bij een criterium over "Overzicht & Overwicht" hoort niet "materiaalgebruik" "correcte spelling", "didactiek", "relatie" etc.

- Alle elementen moeten aanwezig zijn, voor de bijbehorende hoeveelheid punten.

Voorbeeld: Wanneer een onderdeel bij 3 punten afwezig is, scoort de student dus maximaal 2 punten.

- Bij twijfel tussen 2 scores mogen er geen halve punten gegeven worden.

- Geef eerst alle punten, reken daarna pas het cijfer uit. Dit mag dan niet meer veranderd worden.

Fragment 2

VM1.2 Rubric beoordeling EXTERN (t.b.v. 360-graden-feedback)

U gaat dadelijk een filmfragment bekijken en deze beoordelen met behulp van onderstaande rubric en bijgevoegde instructies.

Lees eerst de algemene instructies links, vervolgens de toegevoegde instructies onder en rechts van de rubric. U vindt hier extra uitleg voor de gekleurde woorden uit de rubric. Het eerste criterium met betrekking tot het beroepsproduct hoeft u niet te beoordelen.

KBS-leeruitkomst VM-cursus 1.2							
De student is zich bewust van factoren die het leerklimaat in de klas beïnvloeden, kan een leidinggevende rol aannemen binnen kleine pedagogische relevante situaties in de klas en in lesontwerpen zichtbaar aansluiten bij de basisbehoeften van kinderen (relatie, competentie, autonomie).							
Naam student: n.v.t.		Beoordeeld(e) product/uitgevoerde handeling: n.v.t.			WPL-school en -groep: n.v.t.		
Prestatiecriteria	thema	0	1	2	3	4	Uitkomst
Beroepsproduct							
Leerklimaat – BASISBEHOEFEN <i>Slechts door één van de drie beoordelaars in te vullen (bij voorkeur mentor of ico)</i>	Onge-structureerde observatie	Student neemt onbewust waar, zonder plan , beschrijft subjectief , trekt geen conclusie	Student observeert bewust , maar zonder plan , en beschrijft subjectief en/of ongenuanceerd (verslag met conclusie)	Student stelt observatieplan op, voert die uit , maar beschrijft subjectief en/of ongenuanceerd (verslag met conclusie)	Student stelt observatieplan op, voert die uit en beschrijft objectief en genueanceerd , onderbouwd met argumenten (die interpretatie van gegevens betrouwbaar maken)	Student stelt observatieplan op volgens 5XW+H , voert die uit en beschrijft objectief en genueanceerd , onderbouwd met argumenten (die interpretatie van gegevens betrouwbaar maken) en beschrijft bruikbaar advies aan zichzelf/leerkracht	
Beroepshandelen	thema	0	1	2	3	4	Uitkomst
Leerklimaat - BASISBEHOEFEN <i>Door drie beoordelaars in te vullen.</i>	Relatie: 'Ik mag er zijn'	Student maakt geen contact met kinderen	Student maakt contact en sluit aan bij belevingswereld kind (als "technische" handeling, nog onwennig, zonder inlevingsvermogen).	Student maakt contact en sluit op inhouds- en betrekkningsniveau (natuurlijke belangstelling en interactie) aan bij belevingswereld kind.	Student maakt contact, sluit op inhouds- en betrekkningsniveau aan bij belevingswereld kind en communiceert (indirect 'je mag er zijn') verbaal en non-verbaal .	Student maakt contact, sluit op inhouds- en betrekkningsniveau aan bij belevingswereld kind, communiceert (indirect 'je mag er zijn') verbaal en non-verbaal en stimuleert de relatie tussen de kinderen onderling .	
	Competentie: 'Ik kan het'	Student maakt kinderen niet competent .	Student maakt kinderen competenter door zicht te geven op eigen kunnen .	Student maakt kinderen competenter door zicht te geven op en vertrouwen in eigen kunnen .	Student maakt kinderen competenter door zicht te geven op en vertrouwen in eigen kunnen en hierbij rekening houdt met individuele behoeften van het kind .	Student maakt kinderen competenter door zicht te geven op en vertrouwen in eigen kunnen en hierbij rekening houdt met individuele behoeften van het kind en de actuele (evt. onvoorziene) situatie in de klas .	
	Autonomie: 'Ik kan het zelf'	Student geeft kinderen geen autonomie (geen ruimte voor eigen inbreng en keuzes, benadert hen als 'houd-je-mond-en luisterklas').	Student geeft autonomie , maar sluit niet aan bij kind en de actuele (evt. onvoorziene) situatie in de klas, omdat het kind slechts uitgenodigd wordt tot procedurele vragen .	Student geeft autonomie door aan te sluiten bij kind en de actuele (evt. onvoorziene) situatie in de klas .	Student geeft autonomie door aan te sluiten bij meer en minder initiatiefrijke kinderen en de actuele (evt. onvoorziene) situatie in de klas.	Student geeft autonomie door aan te sluiten bij meer en minder initiatiefrijke kinderen en de actuele (evt. onvoorziene) situatie in de klas en de onderlinge interactie stimuleert .	
Leerklimaat - LEIDERSCHAP <i>Door drie beoordelaars in te vullen.</i>	Overzicht & overwicht <i>Grip op de groep</i>	Student ziet geen ongewenst gedrag, of ziet het wel maar handelt niet .	Student ziet ongewenst gedrag maar handelt niet passend bij situatie.	Student ziet ongewenst gedrag en handelt passend bij situatie.	Student ziet ongewenst gedrag en handelt (verbaal en non-verbaal) passend bij situatie.	Student ziet ongewenst gedrag en handelt passend bij de situatie verbaal en non-verbaal, zowel proactief als reactief .	
Beoordelingstotaal en cijfer				Totaal punten:		Cijfer:	
Beoordelaar	Functie/relatie tot student	Handtekening		Datum			
Cesuur rubric extern VM1.2 <i>excl. beoordeling beroepsproduct</i> <i>alleen beroepshandelen, geen observatie</i>	1 punt = cijfer 1 2 ptn = 1,5 3 ptn = 2	5 ptn = 3 6 ptn = 3,5 7 ptn = 4	8 ptn = 4,5 9 ptn = 5 10 ptn = 5,5	11 ptn = 6 12 ptn = 6,5 13 ptn = 7	14 ptn = 8 15 ptn = 9 16 ptn = 10		

Handelt passend

Indien nodig, eerst om verduidelijking vragen. Daarna duidelijk benoemen wat het probleem van het gedrag is. Niet op de persoon, maar op het gedrag.

Ongewenst gedrag

Gedrag dat niet past bij de actuele situatie in de klas. Gedrag dat de les verstoort. Gedrag dat niet afgesproken is.
VOORBEELD: praten als op dat moment stilte nodig is

"...en de onderlinge interactie stimuleert"

In het kader van autonomie: Ingaan op een aangedragen alternatief van een leerling (of hiernaar vragen), én ruimte geven om dit onderling te bespreken. Het kind ervaart dat het mee mag praten.
NIET: De kinderen laten samenwerken

Student maakt contact

WEL: verbaal/non-verbaal, reageren, doorvragen, interactie

NIET: regelmatig negeren van kinderen, focus slechts op eigen les

Cesuur rubric extern VM1.2

Sluit aan bij de belevingswereld op inhouds- en betrekkningsniveau
Inhoudsniveau: welk onderwerp of lesinhoud wordt gekozen?

Betrekkningsniveau: Sluit de student aan bij leeftijd en interesse van de groep

1-3 punten = cijfer 1

Zicht en vertrouwen geven in eigen kunnen
ZICHT GEVEN: kinderen bevestigen in wat ze kunnen en het hen zelf laten benoemen
VERTROUWEN GEVEN: Laten ervaren wat ze kunnen. Heeft vaak betrekking op emotie en doorzettingsvermogen.

10 ptn = 4

En hierbij rekening houdt met individuele behoeften van kind.
WEL: een kind competent laten voelen in dat wat het kind kan leren (zone naaste ontwikkeling)
NIET: wanneer de student een kind overtuigt van een competentie, terwijl het kind die competentie nog niet heeft en nog niet kan leren

13 ptn = 5,5

...en de actuele (evt. onvoorziene) situatie in de klas
WEL: Inspringen en meebewegen met ideeën buiten eigen inbreng.
Als er geen onvoorziene situatie is, vragen naar inbreng v.d. kinderen
NIET: alle initiatief zelf vasthouden

16 ptn = 7

Het kind wordt slechts uitgenodigd tot procedurele vragen
WEL: Vragen over het hoe en wat van de opdracht waarbij de student alle vragen beantwoordt.
NIET: Vragen over het hoe en wat van de opdracht waarbij de kinderen eigen inbreng ontvangen

Appendix N: Form Descriptive Characteristics Participants Experiment

Formulier Karakteristieken Deelnemers Onderzoek

Datum: _____ Tijd: _____ Plaats: _____

1. Wat is uw geslacht? m/v

2. Wat is uw vooropleiding?

VMBO

MAVO

HAVO

VWO

Anders, namelijk _____

3. Wat is uw functie op uw basisschool: _____

4. Heeft u al eens training van de CHE of andere instanties gehad voor het invullen van een rubric? Ja/Nee

5. Heeft u de rubric voor VM1.2 (Relatie, Competentie, Autonomie, Overzicht & Orde) al eens eerder gebruikt? Ja/Nee

6. Hoeveel studenten heeft u al ongeveer begeleid als ICO?

minder dan 5 studenten

5 tot 10 studenten

10 tot 20 studenten

20 of meer studenten

Appendix O:Transcription Focus Group with ICO's

Nr.	Quotation	Code
1	I1: Okee, super, ik heb hier dus aangepaste rubrics, ik ga die eerst een uitdelen en dan ga ik eens wat vragen over stellen. Uhm, de eerste vragen zijn wellicht nog wat algemeen, niet zozeer over deze rubric, maar ik deel ze wel alvast even uit, dan kan je ze bekijken. Aantekeningen maken op het formulier is toegestaan, al heb ik niet in het onderzoek verwerkt dat ik daar naar kijk, dus als je iets vind dan wil ik je wel vragen om dat ook daadwerkelijk te zeggen zodat ik het terug vind in de opname. Goed, voordat jullie het formulier helemaal induiken, eerst nog wat algemene vragen. Een is eigenlijk, wat is uw ervaring met het beoordelen van studenten. Daar heb ik het net eigenlijk ook al even over gehad. Zou je kunnen aangeven, in uhm 2, 3 woorden zelfs maar, eigenlijk gewoon even heel kort van hoeveel beoordelingen je per jaar geeft. Of, nou, gemiddeld een beetje. Zullen we, vind je het goed om bij jou te beginnen, W2?	Introduction Question
2	W2: Ja hoor, ik denk dat het er nu zo'n vier studenten, 16 ongeveer zijn. 16 van die rubrics heb ik het dan over.	Assessment Experience
3	I1: Ja bedankt. A3?	
4	A3: Ja, ik denk dat ik er nu 5 heb ingevuld ofzo?	Assessment Experience
5	I1: Ja, oké.	
6	M5: Ja, ik begeleid dan van twee pabo studenten en van de CHE had ik een eerstejaars waar ik een aantal van deze heb ingevuld, een stuk of 3, nou dat meisje is gestopt nu. De Marnix beoordeel ik op een hele andere manier. En ik heb nu een derdejaars student, die doet nog de oude, of de vorige opleidingsvariant zonder de rubrics omdat ze studievertraging had, dus daar vul ik geen rubrics voor in maar die beoordeel ik wel.	Assessment Experience
7	I1: Ja, de ouderwetse SP formulieren	
8	M5: Ja, dat klopt	
9	M6: Nu dit jaar dus niet, maar in het verleden heb ik wel rubrics ook ingevuld voor eerstejaars studenten. Uh, twee keer over een eerstejaars student dus wat is dat op jaarbasis?	Assessment Experience
10	I1: Een stuk of 8?	
11	M6: Ja zoiets, en dat was toen het eerste jaar dat ze ook werkten met rubrics	Assessment Experience
12	I1: Ja, S4?	
13	S4: Ja ik dus dit jaar dus ook heel weinig, tenminste wel van mijn eigen student die ik in de klas heb, maar verder is er school breed niet eh, niet heel veel. Maar er zijn ook jaren geweest waarin ik er eh wel 10 in moest	Assessment Experience

	vullen. Dus eigenlijk heel situatie eh... Afhankelijk van de situatie. Dus dit jaar niet zo veel.	
14	I1: Volgende vraag, uhm. Wat zijn de voordelen van rubrics bij het beoordelen van studenten op stage? Ja, ik heb er van tevoren over nagedacht, ik zit ook even, ik lees hem voor. Laten we hem wel beoordelen, eh, beantwoorden: wat zijn nou de voordelen van het gebruik van rubrics? M5: Nou er staat wel heel concreet waar je op moet letten. Uhm Ik vind het altijd wel fijn om van te voren de rubric te krijgen van de student. Als je naar de les gaat kijken en je moet ook nog al die vakjes doorlezen, dan ben je met twee dingen tegelijk bezig en dan vind ik niet dat de focus op de student kan liggen. Uhm. Maar je heb wel heel concreet een, een middel waar je, ja precies, als je gewoon gaat observeren, een student, en je schrijft van alles op dan moet je het zelf nog bedenken en dit is, nou dit is eigenlijk wel makkelijk.	Introduction Question
15	W2: Ik denk dat je een bepaalde subjectiviteit voorkomt.	Advantages Rubric Use
16	W2: Ja, ook.	Advantages Rubric Use
17	S4: En je hebt natuurlijk tig doelen en ik vraag ook altijd van tevoren, waar wil je dat ik op let, dus met die doelen, die ga ik ook heel nauwkeurig bekijken. Je kan natuurlijk als je een les bekijkt aan twintig verschillende doelen gaan wer..., of hé je ziet natuurlijk heel erg veel. Dus het geeft je ook houvast, richting.	Advantages Rubric Use
18	M5: En ik vond ook, bij de eerstejaars kregen ze ook een bepaald ding, terwijl je inderdaad veel meer ziet en dat moest je er dan naast bespreken met elkaar zeg maar, en ik vond dit wel heel goed voor de student zelf, van daar sta je, dat kan je wel goed, dat kan je niet zo goed, enzovoort.	Advantages Rubric Use
19	A3: Ja, daar ben ik het wel mee eens. Het is wel heel specifiek. Je let wel echt op dat specifieke doel.	Advantages Rubric Use
20	W2: Het helpt je, maar ja ik heb maar heel kort, dus maar een jaar gedaan, maar het hielp mij ook om steeds weer even toch juist naar iets anders te kijken bij een student. Niet steeds naar eenzelfde punt. Dus altijd naar je organisatie, altijd naar je overzicht. Je wordt door die ding gedwongen om steeds naar iets anders te kijken.	Advantages Rubric Use
21	I1: Oké. Oké	
22	S4: En je hebt te maken met de verschillende niveaus. Dus je kan bijvoorbeeld in niveau 3 zitten, he dat je gaat al wel best wel aardig maar je streeft ernaar om een niveautje hoger te komen. Dat vind ik ook wel mooi in zo'n rubric staan.	Advantages Rubric Use

23	M6: Ja er komt een bepaalde beoordeling zit daaraan vast, je gaat punten ervoor geven. En ik was ook een beetje zo aan het kijken van oké als ik dat nu invul wat voor cijfer zou het dan gaan worden, vind ik dat waard, vind ik het dat waard. Ja.	Rubric Appliance
24	I1: Ja, en dan is het natuurlijk meteen al een interessant punt van het gebruik, hoe gebruik je... en daar gaan we het nu verder over hebben. Zouden jullie voor mij deze rubric eens goed willen bekijken. Wellicht bekend, al ga ik ervan uit van niet want er zijn zoveel rubrics dat het maar de vraag is of je deze al eens langs heb zien komen. En zou je voor mij eens willen onderstrepen, zowel in de rubric die in het midden staat.. ik zal hem even... zowel in de rubric die in het midden staat ,als in de tekst erom willen onderstrepen wat nog onduidelijk is. En dat is best even... vraagt even mentaal wat om dus ook te bedenken stel dat ik deze rubric zou gebruiken in de praktijk met de uitleg, wat zou dan nog onduidelijk zijn en om dat eens te onderstrepen.	Introduction Question
25	M6: en de praktijk bedoel je nu de studenten in de stageschool	
26	I1: Ja je zit achterin de klas en je zit de les te bekijken.	
27	W2: Kun je nog eens zeggen, wat streep ik aan.	
28	I1: Ja, goeie, wat nog onduidelijk is. En nu gaat het alleen nog over die onduidelijkheden en nou is de vraag waar we mondeling verder over verder in gesprek kunnen, hoe zou dit verduidelijkt kunnen worden. Dus dan is natuurlijk eerst even de vraag te delen van wat nog onduidelijk is en hoe zouden we met elkaar na kunnen denken hoe zou dit verduidelijkt kunnen worden. Wellicht heb je niks aangestreept dat zou kunnen natuurlijk, misschien heel veel. Zullen we is beginnen bij S4?	Introduction Question
29	S4: Ja, nou op zich is alles duidelijk omdat het ook gewoon al wat onduidelijk was dat wordt eigenlijk uitgelegd hier in de rubric eromheen.	
30	I1: Oké,	
31	S4: Dus daarop heb ik geen onduidelijkheden. Ik heb nog wel vragen inhoudelijk over.. Mag dat nu al?	
32	I1: Ja daar komen we zo op. Ja dat is inderdaad eh, ben ik ook heel benieuwd naar.	
33	S4: oké, ja, het is voor mij wel helemaal duidelijk	
34	I1: Oké, ja ja M5?	
35	M5: Ik kan me wel voorstellen dat ook mentoren zijn dat als ze dit lezen denken van, wat is $5xW+H$.	
36	I1: Ja, oh, dat is een goeie, ja?	
37	M5: Wat is dat?	

38	I1: En nou zit ik, want voor dit onderzoek gaat vooral om uhm het beroepshandelen omdat het beroepsproduct heb je wat meer de rust om het een keer na te kijken zeg maar.	
39	M5: Ja	
40	I1: En bij het beroepshandelen zit je bij zo'n les achterin te kijken, moet je daadwerkelijk op dat moment beoordelen. Ik omcirkel hem.	
41	M5: En ik had ook onderstreept dat ze kunnen dan na kunnen denken van wat er bedoeld wordt met "Student maakt contact en sluit aan bij belevingswereld kind en dan tussen haakjes, als technische handeling" dat de mentor dan denkt wat moet ik dan zien?	As Technical Act
42	I1: oké, ja goeie, zullen we eens met elkaar nadenken, hoe zou dat verbeterd, eh verduidelijkt kunnen worden?	
43	M5: Ja ik denk dat dit gewoon een afkorting is ergens van? Die zou je misschien ook gewoon kunnen verduidelijken in zo'n blokje ernaast, die 5XW+H	
44	I1: Ja, ik zit voor nu vooral te kijken naar die technische handeling omdat die andere heeft te maken met het beroepsproduct? Hoe zou dat van die technische handeling uhm, en dan vraag ik het even aan de rest van... Herkennen jullie dat het onduidelijkheden zou kunnen opleveren? Hoe zou dat verduidelijkt kunnen worden?	
45	W2: Voorbeeldje van wat je ziet? Zou kunnen zien?	
46	I1: Misschien ook in zo'n blokje?	
47	W2: Ja in zo'n blokje?	
48	I1: En wat verstaan jullie onder...	
49	A3: Ja ik vind het lastig om het een andere naam te geven want technische, ik snap hem wel, dat het gewoon een soort aangeleerd is onnatuurlijk dat het niet echt eigen nog is. Maar... hoe je dat...	As Technical Act
50	W2: Het is een beetje een truckje omdat je het ergens zo gelezen hebt, dat het zo moet doe je dat, maar er zit niet een gevoel achter waarom je het doet. Het is niet vanuit je inlevingsvermogen, dat denk ik.	As Technical Act
51	S1: Maar dat is lastig om te beoordelen	As Technical Act
52	W2: Maar je weet niet wat hij of zij voelt als die het geeft... als die het doet.	As Technical Act
53	M5: Maar misschien kun je wel merken of het serieus bedoeld is, heeft de student echt belangstelling voor dat kind op dat moment.	As Technical Act
54	M6: ja zo kan je het ook zeggen.	
55	M5 : Dus maakt het oogcontact, zo (Ja) bijvoorbeeld.	As Technical Act

56	W2: Maar je ziet ook wel vaak dat studenten uh het gedrag van de mentor kopiëren dus hoe de mentor met de klas omgaat, hoe die met de kinderen omgaat, dat het iets eigen wordt.	As Technical Act
57	M5: iets van haarzelf.	As Technical Act
58	I1: Dus misschien zou je kunnen zeggen, gekopieerd gedrag valt ook onder als technische handeling, kopiëren van de mentor of iemand anders, al is dat dan weer lastig om...	
59	S4: Maar is dat dan meteen fout?	As Technical Act
60	W2: Nee dat vind ik niet, dat vind ik niet fout	As Technical Act
61	I1: Oh ja,	
62	M5: Maar dat ligt er ook aan, heb je een pabo 1, 2, 3, wat voor jaar is het?	
63	I1: Ja, pabo 1	
64	M5: oh, ja, oké.	
65	I1: Dat is een goeie om te zeggen, ja, dat is iets wat ik misschien... Ja, blok 2 zeg maar.	
66	W2: Maar het hoeft perse niet fout te zijn tenzij het helemaal niet past bij wie je bent en iedereen bijna kan zien dat het een kopietje is van wat je doet. Maar pabo 1 studenten doen dat natuurlijk nog wel heel sterk want die hebben nog niet een eigen	As Technical Act
67	S4: Ik zou dan meer, want je heb dan bij 0 student maakt geen contact. Nou dat is dan helemaal. Misschien student maakt moeizaam, of in ieder geval dat het reëel, er kan nog wel iets tussen geen contact en maakt contact, tussen 0 en 2, kan daar nog een tussenvorm zou er nog...	As Technical Act
68	M6: Ja, je hebt studenten waarvan je gewoon kunt merken, bijvoorbeeld havo studenten die nog nooit met kinderen hebben gewerkt, die gewoon heel onwennig voor de klas staan en kinde..., eh, studenten die al een uh onderwijsassistentopleiding gehad hebben gaan op een hele natuurlijke manier om met kinderen.	As Technical Act
69	I1: Ja	
70	W2: Student probeert contact te maken... probeert... Bij 0 is het helemaal geen contact	As Technical Act
71	M6: Of op eigen wijze contact te maken. Of...	As Technical Act
72	W2: Ja...	
73	I1: Want jullie zijn heel goed bezig, en ik mag natuurlijk verder inhoudelijk niks zeggen natuurlijk, ik zit te denken, ik moet.... Uhmm.	

74	W2: Als je zegt, maakt op eigen wijze contact, zit je dan niet al in het volgende hokje?	As Technical Act
75	M6: Ja, misschien ook wel...	As Technical Act
76	M5: Als iets aangeleerd, het is geen automatisme, ja ik denk maar even hardop hoor.	As Technical Act
77	S4: Ja	
78	M6: Ja kijk want bij die tweede staat al die natuurlijke belangstelling en interactie dus dan denk ik inderdaad dat 1 dat dat nog iets onwennigs is voor een student.	As Technical Act
79	I1: Ja	
80	S4: Probeert contact te maken.	As Technical Act
81	W2: Volgens het boekje	As Technical Act
82	I1: Ja ik denk wel helpend, weer even input om er weer even verder op door te gaan. Eh, M5 had je nog andere?	
83	M5: Nee, nee	
84	I1: Uh, M2?	
85	M6: Ik had dus ook inderdaad die formulieren, eh, die formule omcirkelt zo van nou dat weet ik niet precies wat daar	
86	I1: oké	
87	M6: Wat dat is. Wat mij wel opviel ook... dat het lezen van zo'n rubric ook al heel erg veel tijd kost	Rubric Appliance
88	I1: Ja	
89	M6: Dus stel dat die student al met de les bezig is geweest, en ik had dit dan ook nog moeten gaan lezen, dan had ik dus het eerste gedeelte... we zijn zeker 10 minuten aan het lezen geweest denk ik, dan mis je veel van de les. Dus ik denk dat het ook heel goed is om studenten erop te wijzen, van let er op dat je van tevoren dit al geeft, aan je, uh, aan degene die je gaat beoordelen, het liefst een dag van tevoren	Rubric Appliance
90	I1: Juist aan het begin van de les zit heel veel... uh ja, instructie ook en contact.	
91	S4: Ja, het beste is zelfs het bespreken ook dat je allebei goed een beeld hebt, wat er gaat gebeuren.	Rubric Appliance
92	M6: Ja, want het zijn ook best veel dingen waar je op moet letten vind ik hoor... Het is niet zo dat je zegt van nou, ik ga vandaag eens goed letten op, eh, hoe jij rekenen gaat introduceren of weet je wel, of hoe jij dat of dat doel gaat uitleggen aan de kinderen. Ik vind hem heel uitgebreid.	Rubric Appliance

93	I1: verder nog begrippen/onderdelen inhoudelijk die nog onduidelijk zijn? A3?	
94	A3: Nou voor mij was niet helemaal duidelijk, bij student maakt contact, weer even dat stukje, dan staat erbij in dat hokje eronder bij wel en niet, en bij wel nou ja, dus een student maakt contact als je waarschijnlijk wel ziet, uh, dat een kind verbaal/ non-verbaal of nee, dat een student verbaal/ non-verbaal reageert op een kind.	Verbal/non-verbal
95	I1: Even voor de duidelijkheid, helemaal links onder.	
96	A3: Ja, helemaal links onder	
97	S4: Oh.	
98	I1: Die verduidelijking in het blauw, student maakt contact, daar hebben we het over. Ja, dus bij wel verbaal/ non-verbaal.	
99	A3: En dat er interactie is, dat er wordt gereageerd, dat een student doorvraagt bij het kind. Maar dan staat er "niet: regelmatig negeren van kinderen". Ja, ik snap hem nu, nu ik hem hardop uitspreek, inderdaad.	Verbal/non-verbal
100	I1: Zou die verbeterd kunnen worden?	
101	A3: Het gaat niet...	
102	I1: Ja, het gaat niet... eh.	
103	A3: nee ik zat hem zelf eventjes in mijn hoofd te koppelen aan... meer orde. Maar het gaat hier inderdaad over iets anders. Omdat ik dacht van, niet, het regelmatig negeren van kinderen is toch goed?	Verbal/non-verbal
104	I1: Oh	
105	A3: maar het gaat nu inderdaad over dat een student contact maakt, en dan zou het niet goed zijn inderdaad, als een student, uh, het niet regelmatig negeert dus nee ik snap hem, hij is verduidelijkt.	Verbal/non-verbal
106	I1: Ja	
107	A3: Nee, verder had ik inderdaad ook niets. Behalve ook dat 5XW.	
108	I1: Oké, W2?	
109	W2: Nee	
110	I1: Oké, komt nu een volgende interessante vraag. Ja, dit vind ik leuk om te stellen, want uhm... Welke zinnen... en dat gaat dan voor, dat is een goeie, daarna ga ik dat nog een keer vragen voor de algemene instructies. Dus nu alleen inhoudelijk en dan voor het beroepshandelen en de uitleg die erom heen staat. Welke zinnen of zinsdelen bent u het mee oneens? Zou je die willen doorstrepen? Probeer af te ronden.	
111	A3: Zal ik beginnen?	
112	I1: Ja	
113	A3: Ehm, ik had er niet zo heel veel.	

114	I1: Oké	
115	A3: Ehm, ik vind wel de zinnen een soort complex in elkaar zitten dat ik echt zinnen meerdere keren moet lezen voordat ik hem zelf begrijp. Dus zoveel zinnen, of de meeste zinnen kloppen taal technisch wel. Soms verdraaiingen denk ik, maar... Ik vind ze eerder, meer ingewikkeld om te lezen. En eh, de zin waarvan ik dacht, nou dat klopt niet helemaal is bij "zicht en vertrouwen geven in eigen kunnen" bij het derde vakje.	Complexity Competence- Emotion
116	I1: Bij competentie.	
117	A3: Ja, daar staat dan vertrouwen geven, laten ervaren wat ze kunnen heeft vaak betrekking op emotie. Ik dacht, heeft dat altijd betrekking op emotie, ik dacht dat heeft ook wel betrekking volgens mij op willen, van...	Competence- Emotion
118	I1: Ja	
119	A3: Dus ik vind alleen op emotie vind ik te beperkt, en...	Competence- Emotion
120	I1: "Vertrouwen geven, laten ervaren wat ze kunnen heeft vaak betrekking op emotie". Wat, eh... Jij zegt, heeft ook wel te maken met ehm willen...	
121	A3: Ja	
122	W2: Doorzettingsvermogen.	Competence- Emotion
123	A3: Ja, ja wilskracht, hmm, inzet.	Competence- Emotion
124	W2: Maar dan kun je onzekerheid daartegenover zetten. Toch?	Competence- Emotion
125	A3: Ja, dus dan heeft het wel te maken met emotie.	Competence- Emotion
126	W2: Ja, iets gevoelsmatigs.	Competence- Emotion
127	A3: Ja.	
128	W2: Ik denk dat het een samenspel is.	Competence- Emotion
129	A3: Ja zeker, maar daarin vind ik emotie te beperkt.	Competence- Emotion
130	I1: Ja, dan kan die aangevuld worden misschien. Ehh... Ja	
131	S4: Daar ben ik het wel mee eens hoor, want je hebt natuurlijk ook, kijk... Vertrouwen geven, als ik met een kind eh... Vanmorgen nog een meisje die eh heel weinig vertrouwen had maar heel veel heeft geoefend en nu wel heeft... He, dus door inzet is zij, heeft zij vertrouwen gekregen. Ik zei, maar je moet wel eerst daarvoor gaan. Maar is dat niet alleen emotie?	Competence- Emotion

132	M6: Ik vind het ook wel veel gevraagd voor een eerstejaars die er net is. Om dit al te kunnen met kinderen. Vooral als jij als eerstejaars, eh, je komt zo van de HAVO en je staat voor de vijfde, zesde keer voor een groep, dan vind ik het heel veel gevraagd al om een kind vertrouwen te kunnen geven, ervaring eh,... met betrekking tot die emotie, dat vind ik heel veel gevraagd voor een eerstejaars. Ik denk zelf dat eerstejaars nog heel veel aan het lesgeven zijn en nog niet zozeer bezig zijn van kan ik de kinderen als ik deze les geef ook vertrouwen geven...	Demanding
133	I1: Ja	
134	W2: Maar ik denk dat een beetje bij al die dingen die hier staan	Demanding
135	I1: Nou	
136	W2: Om hier ooit aan het eind te komen als eerstejaars	Demanding
137	M6: Ik vind het heel wat hoor	Demanding
138	W2: Vind ik knap, dat gaat je gewoon niet lukken, moet je misschien ook niet willen, ook niet als student, maar ik vind dat wel heel veel gevraagd hoor, voordat je ooit een keer in hokje vier komt.	Demanding
139	I1: Ik heb nu die vraag niet voorbereid, maar ik zit nu te denken, we zouden eens met elkaar kunnen kijken, per punt, van waar zou een eerstejaars moeten zitten voor een voldoende? En dan ga ik daarna, of wij, gaan later weer kijken van, past dit ook bij de beoordeling en hoe de cesuur en alles en de punten. Maar misschien is het interessant om dat nu eens samen proberen vast te stellen en dan eerst eens te kijken bij de... dan gaan we... ja, ik zit even te twijfelen of dat nu... Misschien is het goed dat we eerst die waar je mee oneens ben afmaken, dat we daarna even kijken per puntje van waar zou je, als eerstejaars studenten, welk niveau zou je ongeveer moeten zitten? Dat heeft natuurlijk ook, want als je het daar nu al niet mee eens bent, dan kom je al in de knel met je beoordeling.	
140	M5: Krijgt een eerstejaars per definitie altijd een onvoldoende, als je dit dan eh..	Demanding
141	I1: Als je het zo rechtlijnig zou volgen	
142	W2: Maar dat is wel zo want wij hebben een eerstejaars student, nou die geeft een hele goede les, maar als je de puntjes optelt dan zit ie net krap aan op een zes. Nou dat is sneu om tegen zo'n jongen te zeggen. Van we hebben een top les gezien maar je hebt maar een zes. Heeft ie niks aan.	Demanding
143	I1: Gaan we doen. Ik zit even te denken, A3 was begonnen, had het over competentie, daardoor kwamen we op dit punt. Uhm, we hadden het over die emotie gehad. Had jij nog andere dingen waar je het mee oneens bent?	
144	A3: Nee	
145	I1: oké. Dan is het ook goed, ik bedoel dan nu ook of je het oneens bent, los van waar je uit zou moeten komen of het niveau... Uhm W2?	

146	W2: Ja, ik dacht hier bij het laatste hokje, maar daar hoefde je geloof ik eigenlijk niet naar te kijken naar het leerklimate leiderschap he?	
147	I1: Ja, dat wel	
148	W2: Oké, ziet... student ziet geen ongewenst gedrag en of handelt niet, maar als je geen ongewenst gedrag ziet dan handel je toch sowieso niet? Dus dat laatste stukje hoeft er niet bij... En handelt... Of zie ik dat verkeerd?	Undesirable Behavior
149	I1: Ja, nou ja, ik ga er zo even rustig over nadenken maar ik zou het niet weten.	
150	M5: En dus handelt niet.	Undesirable Behavior
151	W2: Ja, want je ziet het niet, dus doe je niets. Maar dat is misschien...	Undesirable Behavior
152	I1: Ja, nee, wel fijn. Gewoon omdat het niet...	
153	M6: Ja, je bedoelt, als je geen gedrag ziet hoef je ook niet te handelen	Undesirable Behavior
154	W2: Nee maar dan kan je niet handelen want je hebt niets gezien.	Undesirable Behavior
155	M6: Nee, want er zijn ook wel studenten, die willen ongewenst gedrag niet zien en die kijken er maar overheen want dan denken ze misschien van nou....	Undesirable Behavior
156	W2: Maar dat staat er niet.	Undesirable Behavior
157	M5: Dat is misschien nummer 1, daar kan je zien... het blokje verder eraast.	Undesirable Behavior
158	M6: Oh ja, handelt niet passen, oh ja inderdaad	Undesirable Behavior
159	M5: Die kan je dan invullen.	Undesirable Behavior
160	I1: Hmm	
161	S4: Ik had nog eh, even kijken hoor, bij dat betrekingsniveau... Dat komt een aantal keren terug, en dan staat er bij eh, relatie ik mag er zijn, bij nummertje 2, "Student maakt contact en sluit op inhoud en betrekingsniveau, natuurlijke belangstelling interactie, aan bij de belevingswereld kind. Eh, inhoud snap ik. Betrekingsniveau, dat staat dan ook bij betrekingsniveau, sluit de student aan bij leeftijd en persoonlijkheid. Ja, dat vind ik in hele klassen altijd best lastig, want je hebt natuurlijk, het ene kind vind het fantastisch als je het over het geraamte hebt en het andere kind niet ofzo weet je dat... Dus dat betrekingsniveau vind ik ook	Relational Level

	wel weer betrekkelijk omdat dat bij het ene kind wel zo zal zijn en bij het andere kind niet omdat je de hele groep voor je hebt.	
162	I1: Hoe zou dat aangepast kunnen worden zodat het dus wellicht duidelijk kan worden, hoe een student omgaat met de groep inderdaad, want het gaat niet zozeer om de individuele interactie maar dat is natuurlijk nog de vraag. Hoe zou dit wellicht verduidelijkt kunnen worden, betrekkningsniveau?	
163	S4: Hmm...	
164	I1: Of eh... Anders geformuleerd kunnen worden?	
165	S4: Betrek kinderen op een positieve manier bij, zoiets meer, of is eh... hmm... student maakt contact, het gaat om de relatie, ik mag er zijn...	Relational Level
166	W2: Ja... probeert een bepaalde aansluiting...	Relational Level
167	M5: Heeft het ook niet te maken met de belevingswereld van dat kind? Met de groep, de belevingswereld van die groep? Van de gemiddelde leerling uit die groep? Want je hebt altijd een aantal kinderen die... er wat buiten vallen, die een andere interesse hebben, maar... Gemiddelde belevings... wereld, volgens mij bij een bepaalde groep.	Relational Level
168	W2: Maar dat is misschien ook wel zoals je, hoe je kijkt als je ernaar kijkt, heeft het de belangstelling van het grootste gedeelte van de groep, dat zou je... maar hoe ga je dat erin zetten?	Relational Level
169	I1: Ja, want ik zit daarbij te kijken wat het lastige nog maakt is dat het wel toegespitst wordt op de belevingswereld van het kind, en dat dat uitgesplitst wordt in het inhoudsniveau en het betrekkningsniveau.	
170	W2: Ja, dat is waar.	
171	I1: En wat is dan, want eh, die opmerking denk ik... Het grootste gedeelte van de groep... Maar wat is dan het betrekkningsniveau? En daar hebben we dus... Daar heb ik met die CHE-docenten ook mee, met mijn collega's, ook mee zitten worstelen. Ehmm, we hebben hier staan, sluit de student aan bij de leeftijd en de persoonlijkheid...	
172	M5: Ik denk eerder dat je wel kan zeggen misschien bij leeftijd, dat je dat eerder ook kan zien, want ze zitten eh in een klas, in een groep met één gemiddelde leeftijd. Dat springt niet echt heel ver uit elkaar. En wellicht hebben ze een les gekozen uit de methode dus dat is allemaal afgestemd op die leeftijd. Maar persoonlijkheid, ja, dat wordt wat lastiger natuurlijk	Relational Level
173	S4: Ja, ik zie niet zo heel, kijk, belevingswereld, dat vind ik helder, weet je, dat eh. Dat je daarbij aansluit, maar... persoonlijkheid, vind ik dan weer, dat moet je weghalen als je met een groep te maken hebt. Denk ik	Relational Level
174	M5: Ja, denk ik ook.	
175	I1: Hmm, ja	

176	M5: Je kan nooit dertig verschillende persoontjes, eh...	Relational Level
177	M6: nee, en hoe kan een student die er zo nu en dan een week is de persoonlijkheden van al die kinderen leren kennen, dat is een onmogelijke taak.	Relational Level
178	I1: Sowieso, gedifferentieerd lesgeven komt ook pas later in de opleiding maar...	
179	M6: Ja, zeker.	
180	I1: Oké.	
181	M5: En ik had in het, onder het vierde blokje achter "Ik kan het zelf; student geeft autonomie door aan te sluiten bij meer en mindere initiatiefrijke kinderen en de actuele evt. onvoorziene omstandigheden in de klas" en dan dat paarse stukje "en dan de onderlinge interactie stimuleert", dat vind ik discutabel.	Interaction
182	M6: Ja, die had ik ook	
183	M5: Want, eh, als je met onzekere kinderen te maken hebt, durven die ook... En die wel voelen van, ik kan het, die hoeven volgens mij nog niet perse deel te nemen aan onderlinge interactie.	Interaction
184	I1: Vat ik het goed samen als ik zeg; onderlinge interactie is geen voorwaarde voor het geven van autonomie? Ja, dat is misschien...	
185	M5: Hoeft niet, kan wel, hoeft niet...	
186	I1: Ja, en wat maakt, als het nog een keer zou kunnen samenvatten, ehm... Wat maakt dat dat stukje "en onderlinge interactie stimuleert", dat dat een knelpunt oplevert? Ik vatte hem nog niet helemaal.	
187	M5: Ehm...	
188	M6: Ik vind dat dat stukje ook een valkuil kan zijn, want je hebt namelijk ook klassen die juist zoveel onderlinge interactie aangaan, dat ze op een gegeven moment zo'n les van zo'n eerstejaars student gaan overnemen, dat dat een te grote, eh, gedeelte van de les gaat worden.	Interaction
189	M5: Of kinderen die misschien wel zelf, het helemaal niet kunnen, de boel gaan overschreeuwen. Op een andere manier gaan compenseren. Dus onderlinge interactie wil voor mij nog niet zeggen dat een kind denkt "Ik kan het".	Interaction
190	I1: Ja.	
191	S4: Ja	
192	M6: Ja, want het is dus de bedoeling dat wij die student gaan beoordelen op het feit dat er een interactie in haar of zijn groep kan plaatsvinden.	Interaction
193	I1: Hmm.	

194	M6: en of hij of zij dat stimuleert, want soms vind ik het ook wel de kunst; kan een student een onderlinge interactie ook weer in banen leiden zodat je weer naar de doel van je les gaat?	Interaction
195	M5: Ja	
196	I1: Ja...	
197	S4: Ik vind trouwens bij, sowieso bij de competentie "ik kan het", zien jullie die?, vind ik eh.. Ik heb bijna de neiging om bij nummertje 3 of 4 dat laatste stukje weg te laten ofzo. Sowieso bij nummertje 4, he, want dat komt dan bij, nog een extra iets; "en een actuele evt. onvoorziene situatie in de klas", ja dat is zo afhankelijk van... He, dat zou van mij weg mogen, dat vind ik geen meerwaarde ofzo.	Unforeseen Circumstances
198	M6: Ja, want stel dat er niks gebeurt in de klas.	Unforeseen Circumstances
199	S4: Nee, en dan haal je dat punt niet.	Unforeseen Circumstances
200	M6: En de kinderen zijn gewoon harstikke braaf omdat jij als juf ook nog achterin de klas zit en misschien wel de ICO'er, dan doen de kinderen juist helemaal niks en dan kan zo'n student nooit kan laten zien van, maar ik kan dat ook	Unforeseen Circumstances
201	S4: Nee, dat vind ik meer... nee, bij de competentie "ik ben flexibel", bij flexibel horen. En niet bij "ik kan het"	Individual Needs
202	M5: Dat is misschien ook wel dat ik erbij heb gezet, daaronder "discutabel". En ik vind het ook wel, "hierbij rekening houden met de individuele behoeften van een kind", hoe peil je dat? Hoe meet je dat? Hoe meetbaar is dat?	Individual Needs
203	S4: Ja, die heb ik ook tussen... staan...	Individual Needs
204	M5: Hoe kan, en zeker een eerstejaars, hoe...	Individual Needs
205	W2: Dan kom je daar weer he,	
206	S4: ik vind eigenlijk al...	
207	M5: Ja, dan kom je daar weer	
208	W2: Dan kom je daar weer van wat wil je dat ze kunnen.	Demanding
209	M5: Ja, maar goed... hoe...	
210	I1: Ja	
211	M5: ook hoe kan ook een tweede of derdejaars misschien wat meer, echt weten wat de individuele behoefte is van dat kind? Moet je gewoon haast leerkracht van de groep zijn, wil je daaraan kunnen voldoen...	Individual Needs Demanding
212	S4: Ja, weet je bij 3 staat "student maakt kinderen competent door kinderen zicht te geven op, en vertrouwen in eigen kunnen". Dat vind ik al een super	Demanding

	mooi doel als ze dat kunnen halen, en dan zou je als extra stap nog kunnen doen, en doet dat doelbewust. He, dat er een plan achter zit hoe diegene dat doet ofzo, dan denk ik, nou dan heb je als eerstejaars je doel wel gehaald.	
213	W2: Ja, dat denk ik zeker wel ja.	
214	A3: ik zit ook nog te denken, inderdaad, wat je zei M6 dat stukje van die onderlinge interactie stimuleert, ehm, volgens mij heeft dat ook meer te maken met samenwerken als je dat wilt gaan...	Interaction
215	M6: Ja, het coöperatieve	Interaction
216	A3: Ja, precies, als je dat wilt gaan beoordelen van een student, hoe een student dat, ehm...	Interaction
217	M5: Maar ik zit dan ook te denken, ik heb ooit een kind gehad, die was super goed, die kon het; top leerling. Alleen die kon niet samenwerken, omdat het kind Asperger had, bijvoorbeeld.	Interaction
218	A3: Precies...	
219	M5: Maar die was wel overtuigt van eigen kunnen.	Interaction
220	A3: Klopt.	
221	M5: En dat is dan misschien die individuele behoefte maar ik vind dat je dat niet kan vragen van een eerstejaars student.	Demanding
222	I1: Oké.	
223	M5: En dan gaan we echt op de inhoudelijke dingen zitten	
224	I1: Maar dat is op zich ook niet perse verkeerd. W2? Wat zijn jouw gedachten over dit alles? Maar misschien niet zozeer over dit alles, maar zijn er nog punten?	
225	W2: Nou, ik heb geen nieuwe dingen hoor.	
226	I1: Oké,	
227	W2: Ik zit alleen gewoon een beetje, ze moeten zoveel, daar zou ik graag wat aan willen veranderen.	Demanding
228	M5: Ja, hebben ze dat dan niet geprobeerd door één onderdeel te pakken in die rubrics? Die ehm, hoe noemen ze dat ook al weer?	
229	I1: Prestatiecriterium? Bedoel je dat?	
230	M5: Ja, ja, echt één onderdeel en de rest waar hier niet op gefocust is, daar sprak je dan samen naderhand met student.	
231	I1: Ja, dat je van deze vier nog een eruit pakt en daar voor die les op focust...	
232	M5: Nou, nee maar ook andere dingetjes van een les bijvoorbeeld als tips van als je de volgende keer dit, dan dat zus... Dat had dan niet zozeer met dit ene onderdeel te maken.	

233	S4: En dit zijn al best wel dingen die, eh, eigenlijk misschien in het tweede of derde jaar... Eh, dit hoort niet, voor mijn gevoel niet heel erg bij de basis nog, van ik kan het zelf.	Demanding
234	M5: Ja, want puur wat jij ook zei, ze zijn nog zo bezig met, hoe moet ik een les maken, hoe moet ik beginnen, wat eh... Zo zitten die eerstejaars in elkaar	Demanding
235	I1: Ja, ehm, gaan we zo nog even inderdaad over hebben van welk, welk punt	
236	S4: Ik moet alleen zo naar huis, dat moet nu echt, want het is tien voor half 4	
237	I1: Oké, ja, ja. Ehm... Voor nu, ook om dan dat officiële gedeelte af te ronden, de vraag voor M2, heb jij nog onduidelijkheden, of dingen waar je het niet mee eens bent?	
238	M6: Ja, ik had bij nummertje vier, bij "ik mag er zijn", eh, het laatste stukje "stimuleert de relatie tussen kinderen onderling", ik vind dat niet perse iets wat een eerstejaars in een les moeten kunnen laten zien, dat ze ook nog, ondanks dat ze daar al heel erg hun best staan te doen om die les te geven, en op allerlei andere dingen ook nog met de relatie bezig is tussen kinderen. Ik vind dat niet echt nodig en noodzakelijk voor ons om dat te beoordelen.	Demanding
239	I1: Oké, ja. Komen we eigenlijk weer terug op dat punt van, waar zit nou een voldoende voor een eerstejaars. Gezien de tijd zou ik willen vragen, ehm, ik wilde eigenlijk ook nog, was van plan om naar die algemene instructies te kijken. Is het goed om even, want... Het is nu acht voor half 4, we lopen al wat uit... Nou, we lopen trouwens al flink uit. Uhm... Hoeveel tijd heb jij nog?	
240	S4: Nou eigenlijk geen tijd. Nee..	
241	I1: Geen tijd meer. Voor jullie is het natuurlijk ook nooit verkeerd om niet zover mogelijk uit te... Wat ik daarbij zit te denken is om jullie te vragen om simpelweg aan te kruisen, van daar zou bij mij een 6 zitten, van eerstejaars per niveau. Al zou ik ook nog even over die algemene instructies willen hebben, wat daar nog onduidelijk is/duidelijk is. Ehm... Gezien de tijd zou ik zeggen, zou je voor mij willen aankruisen, misschien dat eerst even, van die vier punten; relatie competentie autonomie, overwicht/overzicht, waar zou de grens zitten bij voldoende? Bij welk niveau? Is dat bij niveau 1, niveau 2, niveau 3 of bij niveau 4? Wanneer je zegt, nou, dan vind ik voor een eerstejaars vind ik dat een mooi... een voldoende niveau?	Introduction Question
242	M6: Bij welke wil je, bij?	
243	I1: Bij ehm... de... relatie, competentie, autonomie en overzicht/overwicht.	
244	W2: Oh, alleen bij deze 4, zeg maar	

245	<p>I1: Ja, en dan staan daar natuurlijk 0 tot 4 punten kan je geven (Ja), bij welk vakje zou je zeggen, nou dat is net voldoende?</p> <p>I1: En ik zit dan te denken voor de duidelijkheid, zou je voor mij dat vakje is echt flink willen doorkruisen. Want er zitten nog allemaal andere aantekeningen en ik kan het allemaal wel lezen, dat weet ik zeker. Echt even een flink kruis er doorheen van daar zit net die voldoende.</p>	
246	M5: Nou ga je iets heel subjectiefs doen, terwijl het beoordelen van studenten al zo subjectief is.	
247	I1: Komen we zo, eh, nog op.	
248	A3: Je zegt net een voldoende?	
249	W2: Ja	
250	A3: Ja, dat vind ik ook een beetje ik weet niet zo goed want soms zit ie dan alsnog in het midden.	
251	I1: Ja, lastig is dat he.	
252	S4: Ik moet echt gaan, want anders gaat het niet goedkomen	
253	I1: Ja, helemaal goed, ja, bedankt voor je deelname	
254	S; helemaal goed, nou, jij succes ermee	
255	I1: Wij gaan ook wel eh... Ja, en bedankt voor het aankruisen nog even.	
256	S4: Ja, helemaal goed. Succes he	
257	I1: Nou, ik zit te denken dat je nog, als je nou nog nuance zou willen aanbrengen, zou je nog bij een kruisje even een pijltje kunnen zetten, van nou dat zit eigenlijk net hieronder of net hierboven, naar links of naar rechts. Zo van, eh, ja, dat zit eh... In het vakje ook gewoon maar. Of misschien dat er nog iets tussen zou moeten.	
258	M6: Maar eigenlijk is je vraag, wat zou jij, ehm, goed genoeg vinden voor een eerstejaars student. Kijk want die kruisen die ik nu zet in net voldoende, ik vind als mijn eerstejaars student aan het eind van mijn eerste jaar stage dat gaat doen, dan denk ik, nou weet je, jij gaat met voldoende... met vertrouwen kan ik jou naar de volgende school sturen en ze kunnen mij als ICO niet op aanzien, eh, aanzien dat ik een student heb doorgestuurd die het eigenlijk niet aan zou kunnen.	Demanding
259	I1: Ja	
260	M6: En ik vind sowieso die 3 en 4, ik vind dat heel veel gevraagd voor een eerstejaars student hoor	Demanding
261	M5: Ja ik ook, ik zou bijna in de lijn zo willen zetten, met een enkele uitzondering misschien, heb ik dan tenminste, over dat overzicht en overzicht. Dat zou leuk zijn als ze daar ook non-verbaal kan reageren, ik denk dat is wel haalbaar denk ik, voor een eerstejaars. Maar goed, je kunt het zo wel lezen.	Demanding

262	I1: Ja, ja, ik denk alle kruizen zijn gezet, het niveau is aangegeven. Voor nu zou ik nog willen vragen om nog te kijken naar die algemene instructies. En, uhm, denk je daarbij in, we hebben namelijk al heel veel besproken, stel je voor; alles is duidelijk, je bent het inhoudelijk mee eens en het niveau zit op de goede plek, dat is alle drie aangepast, wat vind je dan van deze algemene instructies om een rubric te gebruiken? Zijn er dan nog dingen onduidelijk, en dan stel ik meteen daarbij de vraag, zijn er dan nog dingen waar je het mee oneens bent? Het zijn vijf punten. Oké, nog onduidelijkheden of dingen waar je het mee oneens bent? Non-verbaal lees ik bij A3 en M4 dat het meevalt, M5?	Introduction Question
263	M5: Nee, niet eigenlijk, ik had wel; een student zou zelf kunnen zeggen als je het hebt over inhoudsniveau, "die les moest ik geven van mijn mentor".	
264	I1: Ja, ja uhm je bedoelt over het algemeen bij het gebruik van rubrics.	
265	M5: Ja.	
266	I1: Ja, oké.	
267	M5: En dat je ze er wel op mag hameren bij studenten; geef ze op tijd, geef ze op tijd.	
268	I1: Uh, W2 nog? Ik zag wat uitroeptekens, wat dingetjes?	
269	W2: Uh, nou dat is meer, dat voorbeeld moest ik heel lang over nadenken of dat dat het wel duidelijker maakte bij de algemene instructies. Dat moest ik heel vaak lezen voordat ik hem eigenlijk snapte, maar dat zegt misschien ook iets over mij.	General Guidelines
270	I1: Oké, oh, en dat is, welk voorbeeld is dat?	
271	W2: Er staat voorbeeld wanneer er een minder drie punten afwezig is scoort de student dus maximaal 2 punten.	General Guidelines
272	I1: Die kan misschien ook gewoon wel iets makkelijker geformuleerd	
273	W2: Die kan wel weg. Want ik moest hem heel vaak lezen voordat ik snapte wat de bedoeling was.	General Guidelines
274	I1: Ja, dat snap ik wel.	
275	W2: Dus als die 3 niet heeft, heeft ie gewoon 2 want hij heeft een ding daaruit niet.	General Guidelines
276	I1: Ja, ik zou het bij wijze van kunnen opschrijven.	
277	W2: Net als de... Dat is toch logisch	General Guidelines
278	A3: Ja, je moet gewoon alles halen	General Guidelines
279	M6: Want even voor mijn beeldvorming he, want ik heb nog weinig met rubrics gewerkt. Je vult dit in, dit is het? Of ga je ook nog voor de student tips en tops opschrijven waar die de volgende les weer mee verder kan.	

280	I1: Ja, ik zit te denken, dat is een hele goede vraag die ik ga beantwoorden. Het is natuurlijk fijn voor mij om die opname te stoppen, want ik ga alles lekker uittypen en dat kost even. Zijn er voor nu nog opmerkingen of vragen... Oh en ik stel hem even netjes volgens hoe ik hem hebt heb opgeschreven; is er iets wat nog niet gezegd is wat nog niet gezegd is waarvan u sterk overtuigt bent dat dit nog gezegd moet worden omtrent het gebruik van rubrics bij stagebeoordelingen? Ja	
281	M6: Ja eigenlijk misschien mijn vraag	
282	I1: Ja, die past daar eigenlijk wel bij. Voor nu kan ik zeggen, uhm... Ja, ik kan natuurlijk heel uitgebreid antwoord op geven..	
283	M5: Mag ik zeggen hoe ik dat deed? Ik deed dat wel, tips en tops opschrijven, want daar hebben ze behoefte aan. Tenminste dat kreeg ik later terug van de student als feedback voor mij, dat ze dat juist heel fijn vonden. Want dan konden ze daarop weer verder gaan met een andere les die ze moesten geven.	
284	I1: En ik zit... Een korte samenvatting; juist, ja graag wel, en uhm, past ook in het lesvoorbereidingsformulier bij de feedback en in de WPL-check up. Dat is het korte antwoord, ja.	
285	M5: Oh ja, die had je ook nog. Zo'n check up formulier, zo'n eindformulier.	
286	I1: Andere vragen, opmerkingen.	
287	W2: Ja, ik ben eigenlijk wel benieuwd hoe deze formulieren zijn ontwikkeld. Is daar...?	
288	I1: Ja	
289	M6: En of studenten zelf de inhoud snappen.	
290	W2: Ja, en past die altijd bij de les die ze geven, de rubrics die je voor je neus krijgt. Want dat vind ik ook niet altijd, maar dat ligt misschien ook aan de student.	
291	I1: Zijn wel mooie vragen, ja. Uhm, omtrent... Ja, want ik zit te denken, voor dit onderzoek gaan we natuurlijk die rubric inhoudelijk aanpassen en hoe die omschreven is en et cetera. Die vraag van "past het altijd bij de les?" is een hele belangrijke om daarbij mee te nemen. De vraag van hoe dat ontwikkeld is, is een vraag die ik zo meteen nog zou kunnen beantwoorden. Passend bij de les, ja die ga ik meenemen in het verder aanpassen van deze rubric en daarbij dus ook in het vervolg traject.	
292	M5: En kijken ze dan ook als ze bijvoorbeeld opdrachten aan studenten geven vanuit de CHE of die dan afgestemd worden op de rubrics.	
293	I1: Ja, ja en uhm, dat uh, is voor ons als docenten uhm, soms ook een gepuzzel, laat ik het zo zeggen, omdat je wel richting wilt geven maar je wilt ook niets voorzeggen. Nou, dat is wat simpel gezegd misschien maar, ja daar wordt zeker naar gekeken, juist. Wij hebben als docenten soms ook	

	nog wel eens de discussie, van zijn we met z'n allen niet teveel juist met die rubrics bezig en te weinig met... Maar dat is weer een heel ander verhaal.	
294	M5: Want vroeger de randjes van het leren zeg maar, die waren voor studenten destijds ook wel heel duidelijk. Dat zijn eigenlijk een beetje de tips- en tops formulieren, die waren best helder.	
295	I1: Ja de vraag daarbij ook inderdaad van als een student hem gaat gebruiken, is die voor de student wel duidelijk.	
296	W2: Studenten die we dit jaar hebben die geven aan dat ze liever je tips en tops hebben dan dat je dat ding ingevuld teruggeeft want je tips en tops zijn vaak veel praktischer voor hen.	
297	I1: Ja	
298	W2: Dan denk ik van ja, dan zit ik heel erg m'n best te doen om dat ding zo....	
299	M5: Maar was dat, was dat voor de CHE destijds niet het punt van, ja die zijn te subjectief want bij de een hoor je dit, bij de ander hoor je dat, dat dat de rede is waarom dit toen bedacht is?	
300	I1: Ja, daar antwoord... Maar dan beïnvloed ik jullie ook weer qua inbreng, kan ik zo antwoord op geven. A3, jij wilde ook nog wat zeggen of vragen?	
301	A3: Ja, nee ik vraag me inderdaad af van zo'n rubric, krijgen eerste-, tweede-, derde-, vierdejaars deze zelfde allemaal en zit daar dan een soort opbouw in, wordt dat uitgelegd aan hun van, aan een eerstejaars van, oh maar weet je als je bij, als je een 1 scoort, dan is het goed, dan mag je tevreden zijn, en uhm... Bij een tweedejaars als je dat scoort, dan is het goed, dan...	
302	I1: Ja, dat kan ik beantwoorden dat in een stagebeoordeling niet terugkomt.	
303	A3: Oké.	
304	I1: Daar wordt natuurlijk eerst mee geoefend in de lessen, is het idee. Vervolgens wordt daar aan het eind drie beoordelingen, drie lessen laten zien waarbij je het niveau hebt behaald door die oefening. Dat is het idee. En dan daarna wordt het afgesloten met voldoende, uiteraard.	
305	A3: Oké.	
306	I1: Andere vragen, opmerkingen die gezegd moeten worden over het, over deze rubric?	
307	M6: Maar ja, ik denk als wij hier al als ervaren ICO's al zo mee zitten worstelen, dan kan ik me voorstellen dat zo'n eerstejaars die dit onder ogen krijgt en denkt, zo hier moet ik op beoordeeld worden, dat dat best wel, dat de lat wel hoog ligt voor zo'n student.	
308	A3: En demotiverend werkt.	
309	M6: Nou, ik denk het ook wel. En ik denk dat een eerstejaars student heel graag lessen gaat geven en ook zo veel mogelijk wil oefenen maar als die	

	weet dat wij achterin de klas zitten met zo'n rubric dat ie wel eens drie keer denkt van...	
310	A3: Kan ik het wel	
311	W2: Maar is het dan niet beter om het geen beoordelingsformulier te noemen maar gewoon een feedbackformulier.	
312	I1; Ja, ik zit te denken, dit zijn meer de algemene opmerkingen die ook heel waardevol zijn. En daarom ook wat mij betreft inhoudelijk over deze rubric is die afgerond.	
313	M6: Nou, ja ik denk dat je dit soort vragen ook mee kunt nemen in een discussie op de pabo zelf, met je collega's.	
314	I1: Ja, en daarom zou ik nu de opname willen stoppen en dan juist daarover het nog even in gesprek gaan en die formulieren uitdelen, wat er allemaal nog meer speelt.	