

Development and validation of a questionnaire on classroom management of Dutch secondary  
education teachers

Maj Lettink - s2360667

Educational Science and Technology, BMS faculty, University of Twente

### **Abstract**

With the constant strive for improving the quality of teaching in Dutch schools, more insight into classroom behavior of teachers is necessary. An essential aspect of teaching quality is classroom management, as it is a prerequisite for effective student learning. However, as far as we know, no Dutch instruments specifically measure it yet. Therefore it is important to attend to this need. The most practical way to evaluate teachers, is by having their students fill out a questionnaire, because they know the teacher's day-to-day functioning best, and a questionnaire is a quick and easy way of detecting areas of improvement. Thus, the question this study answers is: "Which items make up a valid and reliable student perception questionnaire for measuring classroom management in secondary education?"

Through literature review and instrument analysis, a first version of the questionnaire was created. Afterwards, the items in the questionnaire were discussed thoroughly in interviews and focus groups with respectively expert researchers, expert teachers, and students. After every round, data was used to improve the draft questionnaire. The final phase of data collection was a pilot test, in which a large group of students used the questionnaire to evaluate their teacher. This data was then statistically analyzed to determine the reliability and validity of the questionnaire. The results indicated that the Classroom Management Questionnaire had a high construct validity as well as a high reliability, in the sense that classes of students awarded their teachers similar scores. Further research could improve this questionnaire by testing it on a larger scale and in different educational contexts, by zooming in on the development of teachers using the questionnaire for professional improvement, by comparing student perceptions to self-evaluations and evaluations by external observers, and by making this questionnaire appropriate for research purposes.

## Introduction

We live in a time in which continuing professional development and permanent education are becoming more and more important. Learning does not end after obtaining a degree, even after several years in the workplace, there is always room for improvement. In education, therefore, teacher evaluation can be a vital part of maintaining and improving teaching quality.

There are several instruments that measure teaching quality in Dutch secondary education. One of those is Impact!: a digital feedback tool used by students to evaluate their teachers at the end of a lesson (Bijlsma, Visscher, Dobbelaer, & Veldkamp, 2019). This study is conducted on behalf of Impact!, since their questionnaire only measures teaching quality as a whole and does not zoom in on sub skills. An extra focus on sub skills can benefit teachers who struggle in a specific area of teaching, such as classroom management (CM).

CM refers to “the actions teachers take to create a supportive environment for the academic and social-emotional learning of students” (Evertson & Weinstein, 2006, pp. 4-5). If discipline problems and other off-task behavior occur, very little academic learning can happen (Elias & Schwab, 2006). This will have a negative impact on student learning (Wang, Haertel, & Walberg, 1993). Therefore, every teacher should be skilled at CM. Especially novice teachers experience difficulty in controlling their classrooms (Veenman, 1984). However, more experienced teachers can also still experience difficulty, for example in classes with more than average learning and/or behavioral problems (Korpershoek, Harms, de Boer, van Kuijk, & Doolaard, 2014). Therefore, a lot of teachers need support on how to properly manage their classrooms.

There are two main ways in which teaching quality is monitored nowadays: observations by external observers and obtaining student perceptions of teaching quality. Formal inspections, such as classroom observations by trained outsiders can be very helpful in identifying strengths and areas of improvement (Lasagabaster & Sierra, 2011). However, an outsider only bases his/her judgement on a snapshot of reality. The lesson they observed may have been an outlier, compared to the average lesson of a teacher. To gain a complete view of the teacher’s performance, multiple observations would have to take place, which is very time consuming and costly (Muijs, 2006).

A more practical way to evaluate the day-to-day teaching skills of a teacher, is by asking the students about the quality of teachers’ teaching, since they have the most experience with the teacher in all types of classroom situations (Peterson, Wahlquist, & Bone, 2000).

Unfortunately, relying on student perceptions also has its disadvantages. Student ratings can be biased (Ferguson, 2012; Van der Lans, Van de Grift, & Van Veen, 2016). Students cannot like a teacher, not like the subject, or be influenced by a bad grade or how the teacher treats them personally. Other research, however, shows that student perceptions can be reliable (Fauth, Decristan, Rieser, Klieme, & Büttner, 2014; Kane & Staiger, 2012) and valid (Balch, 2012; Peterson et al., 2000).

Within student perception evaluations, there are multiple options as well. The most meaningful strategy is to interview students (Hoban & Hastings, 2006). Time, however, is scarce in education, so usually questionnaires are the most popular method.

A student perceptions questionnaire can be of great help here, since it can serve as a way to accurately detect areas for improvement. That means that teachers will be able to continue their professional development in a very targeted manner.

### **This study**

As mentioned earlier, the goal of this study is to create a valid and reliable questionnaire that measures classroom management skills of teachers in secondary education. This leads to the following research question: “Which items make up a valid and reliable student perception questionnaire for measuring classroom management in secondary education?”. To answer this question, first, the following sub questions need answering:

- A. Which items make up a reliable and valid student perception questionnaire for measuring classroom management in secondary education, based on relevant literature and validated instruments?
- B. What revisions should be made to the classroom management questionnaire, based on interviews with expert researchers?
- C. What revisions should be made to the classroom management questionnaire, based on focus groups with expert teachers?
- D. What revisions should be made to the classroom management questionnaire, based on interviews with students?
- E. What revisions should be made to the classroom management questionnaire, based on the statistical analyses of the psychometric quality of the questionnaire?

The development of the instrument can be helpful to (novice) teachers, who experience difficulty in managing their classrooms. The results of the questionnaire can indicate where exactly there is room for improvement for a teacher, after which specific support can be given to improve their classroom management skills. There is no Dutch

questionnaire on student perceptions of classroom management in secondary education yet, so developing such an instrument is of both scientific and practical relevance.

### **Theoretical framework**

In preparation of creating a questionnaire on classroom management (CM), defining the construct requires notable attention. However, because of the broadness of CM and the varying opinions on what aspects do and do not belong to the construct, it is difficult to formulate one single definition. Wubbels (2011) presents six different approaches to CM in order to demonstrate the broadness of CM and to provide clarity concerning the different views on the topic. Although each approach focuses on specific elements linked to CM, actual CM programs commonly integrate elements of multiple approaches. This shows that there is no need to comply to one single paradigm.

#### **Approaches to classroom management**

**The behavioral approach.** The behavioral approach is based on Skinner's five learning principles: Positive reinforcement, negative reinforcement, positive punishment, negative punishment, and extinction (Landrum & Kauffman, 2006). The goal of these principles, and the corresponding CM approach, is to shape desired student behavior. This can be manifested as praising or modelling desired behavior and ignoring or reprimanding undesired behavior, perhaps using techniques such as reward systems. An important characteristic of the behavioral approach is that there are clear rules and expectations, imposed by the teacher (Brophy, 2006).

**The internal control approach.** Contrary to the behavioral approach, the second approach is directed at internal rather than external control, resulting in a student-centered perspective. The internal control approach focuses on teaching social emotional skills, by engaging in positive relationships with students, setting fair boundaries, and creating a sense of shared responsibility. The goal of this approach is to internally motivate students to behave prosocially. This then results in a strong sense of community, creating a caring and safe learning environment.

**The ecological approach.** The third approach focuses on the classroom as a natural habitat, meaning that it is an environment characterized by a large number of unpredictable events happening simultaneously, quickly and publicly, with a collective history that shapes norms and values (Doyle, 2006). An important factor in this teacher-centered approach is the private and fleeting nature of interventions. When interventions are done publicly, they can be disruptive and destabilizing to the classroom situation. Therefore, preferred techniques are

constantly monitoring and making students aware of that, or withitness, and organizing lessons in the most efficient way, so few chaotic and possibly disruptive situations can arise.

**The discourse centered approach.** Fourth, the discourse centered approach focuses on the discourse patterns of teachers and students, and how they are influenced by communication, teacher-student relationships, and constructivism. The latter meaning that context is constructed during interaction, resulting in rules and expectations being deduced from communicative experiences. The discourse centered approach uses strategies similar to the ecological approach: Active listening to understand the social processes at hand, and establishing a clear set of rules and routines early in the school year.

**The curriculum approach.** In the fifth approach, students' academic interest in the curriculum is used to intrinsically motivate them to participate, consequently reducing misbehavior. Strategies used in the curriculum approach come down to organizing classroom activities and instructional materials in such a way that it invites students to participate in on-task discourse, leaving less opportunity and desire for students to engage in disruptive behavior.

**The interpersonal approach.** The final approach is directed at social climate and teacher leadership styles, and building a productive relationship between teachers and students. The teacher-student relationship can be characterized using two characteristics of teacher behavior: The level of influence the teacher has, and the level of proximity to the students a teacher has. Successful teachers are both high in influence and proximity, something they achieve in both verbal and non-verbal communication.

### **Definitions of classroom management**

In defining CM, there are both narrow and broad definitions. Examples of more narrow definitions are those of Brophy (2006), Burden (2000), and Wubbels, Brekelmans, den Brok, & van Tartwijk (2006).

Brophy (2006) defines the concept as actions taken to create and maintain a learning environment that will attain to successful instruction, namely arranging the physical environment, establishing rules and procedures, and maintaining students' attention and engagement in class. This is to be achieved using behaviorist principles, such as punishment and reward.

Burden (2000) emphasizes the role of motivation in CM. Especially intrinsically motivating students to learn, using interesting content and activating instructional strategies. According to Burden, this will make them more engaged and less disruptive.

Wubbels et al. (2006) view the teacher-student relationship as a starting point to understanding a teacher's CM. The way a teacher communicates and interacts with their students is very indicative of their management style, which, in turn, is connected to the teacher's level of success.

Most researchers present broader, more inclusive definitions. Wolff, van den Bogert, Jarodzka, & Boshuizen (2015) argue that CM goes far beyond constructs as compliance and discipline. Although they are important, CM is about knowing how to interact effectively within the full spectrum of classroom events.

Sieberer-Nagler (2015) provides a similar definition, although putting more emphasis on the organization aspect of classroom events, saying that CM is about providing students with the best opportunities by optimally organizing them, space, time, and materials, and dealing with unexpected events using effective strategies.

The broadest interpretation of CM is the widely used definition by Evertson and Weinstein (2006), which lists five aspects of CM:

1. Developing caring and supporting relationships with students.
2. Organizing and implementing instruction in ways that optimize students' access to learning.
3. Encouraging students' engagement in academic tasks, by establishing rules and classroom procedures.
4. Making students responsible for their own behavior.
5. Using appropriate interventions to assist students with behavior problems.

This list was created by reviewing the available literature on CM, and combining multiple views, aspects, and definitions into one thorough and inclusive definition.

### **Connecting definitions and approaches**

The more narrow definitions are easy to connect to Wubbels' (2011) approaches, since they are all directed towards their own specific key aspect: Brophy's (2006) definition, focuses on behaviorist management strategies, and therefore belongs in the behavioral approach. Burden (2002) looks at ways to shape instructional content and activities to make learning more appealing to students, resulting in them being more engaged and compliant. The curriculum approach has the same goal and strategy. In the definition of Wubbels et al. (2006), the teacher-student relationship is put on center stage. The definition originates from the same theoretical background as the interpersonal approach.

The broader definitions overlap in some ways. Both Wolff et al. (2015) and Sieberer-Nageler (2015) speak about the importance of viewing classrooms as a complex accumulation of different events, which coincides with the ecological approach. Wolff et al. (2015) mentions the importance of discipline and compliance more explicitly, which connects to the behavioral approach, while Sieberer-Nageler (2015) focuses more on students developing internalized behavioral patterns, which coincides with the internal control approach.

When comparing Wubbels' (2011) approaches to Evertson and Weinstein's definition, the following becomes apparent: The definition can be split into two important statements. One, CM includes actions teachers take to create a supportive environment for academic learning, and two, actions they take to create a supportive environment for social-emotional learning. The first half of the definition, on academic learning, is similar to the ecological approach, since this approach focuses on the importance of organizing the classroom as efficiently as possible to generate learning, by carefully planning, organizing, and preparing classroom activities. The second half of the definition, on social-emotional learning, matches the internal control approach, which emphasizes the importance of teaching social-emotional skills to internalize positive behavior.

The five key aspects that Evertson and Weinstein present, seem to go beyond both the ecological and internal control approaches. The first and fourth aspect fit the description of the internal control approach (developing caring and supporting relationships with students and making them responsible for their own behavior) and the second and third aspect match with the ecological approach (organizing and implementing instruction to optimize students' access to learning and encouraging them to engage in academic tasks by establishing rules and procedures). Although a key element of the ecological approach is to intervene appropriately when students behave undesirably, the fifth aspect specifically mentions that the interventions are meant to assist students with behavior problems. This concerns more severe classroom disruptions than the minor issues that the ecological approach wants you to solve using non-verbal signals. Therefore, this aspect seems to fit the behavioral approach better. This approach also centers establishing clear rules, which matches the third aspect as well. Finally, the first aspect can, besides the internal control approach, also be connected to the interpersonal approach, which is all about the communication between students and teachers and the relationship that it creates. The remaining two approaches do not specifically match with certain key elements.

Using a broader definition seems better than using one that only complies with one theoretical approach, since more aspects of CM are included, and a more complete view of a

teacher's CM skills can be painted with a broader questionnaire. The most common approaches were the ecological approach, the internal control approach, the behavioral approach, and the interpersonal approach. The curriculum approach focuses more on making instructional content and instruction appealing for students in order for them to be engaged and therefore less disruptive, instead of using management strategies like the other approaches. Therefore, the choice was made not to follow the curriculum approach in designing the questionnaire for this study. The discourse centered approach focuses at the communication patterns between teacher and students and at how they can be used to manage classrooms. Communication is a necessary means in every key aspect. However, analyzing discourse patterns to achieve the goal is not. Therefore, the discourse centered approach is deemed less relevant to this study than the other approaches.

### **Existing instruments that measure CM**

In addition to reflections on the nature and content of the construct, scientifically developed and validated instruments can be used to operationalize and assess the construct in classrooms. By looking at the way CM is defined and assessed in each instrument, more insight can be given into how CM could be defined and assessed in this questionnaire. In this study, the following instruments are reviewed to provide more clarity into their definition and assessment of CM:

- The Classroom Assessment Scoring System – Upper Elementary (CLASS) (Pianta, Hamre, & Mintz, 2012)

CLASS is a Dutch classroom observation instrument designed to assess “the quality of teachers’ social and instructional interactions with students, as well as the intentionality and productivity evident in classroom settings” (pp. 1, Pianta et al., 2012). This instrument consists of three domains (Emotional Support, Classroom Organization, Instructional Support) that each consist of a few dimensions. CLASS connects to elements from internal control, behavioral, ecological, and interpersonal approaches.

- The International Comparative Analysis of Learning and Teaching (ICALT) (Van de Grift, 2007)

ICALT is a Dutch classroom observation instrument, used to evaluate the quality of learning and teaching in primary and secondary schools. The ICALT instrument consists of seven constructs: Safe and stimulating learning climate, Efficient

classroom management, Clarity of instruction, Activating learning, Adaptive teaching, Teaching learning strategies, Student engagement.

- The Questionnaire for the Assessment of Classroom Disruptions (QACD) (Scherzinger & Wettstein, 2019)

The QACD is an English questionnaire that measures classroom management, and especially classroom disruptions. It consists of three main constructs: Classroom disruptions, Classroom management, Student-teacher relationships. It is used by students, teachers, and external observers. It mostly connects to the behavioral approach and the interpersonal approach.

- The Classroom Management Questionnaire (CMQ) (Díaz, González, Jara-Ramírez, & Muñoz-Parra, 2018)

The CMQ is an English questionnaire that measures how teachers perceive their own CM. The constructs used in this questionnaire are based on both scientific literature (Marzano, Foseid, Foseid, Gaddy, & Marzano, 2005; Scrivener, 2012; Webster-Stratton, 2012) and the validated Questionnaire on Classroom Management in Early Childhood Education (Nault, 1994). It consists of three dimensions: Discipline, Teaching and learning, Personal. This questionnaire contains elements of all approaches.

- The Questionnaire on Teacher Interaction (QTI) (Wubbels, Brekelmans, den Brok, & van Tartwijk, 2006)

The QTI is a Dutch questionnaire that measures the student-teacher relationship according to the Model of Interpersonal Teacher Behaviour (Wubbels, Créton & Hooymayers, 1985) from a student perspective. That model consists of eight patterns, each representing a different type of teacher behavior. The eight patterns are placed on an axis that represents the level of influence (either dominance or submission) and an axis that represents the level of proximity (either cooperation or opposition). The results for the questionnaire reflect the type of teacher behavior and the type of classroom environment that students experience with a teacher. The QTI stems from the interpersonal approach.

These instruments were selected to be examined for various reasons. The first two instruments, CLASS and ICALT, were chosen because they are widely used in both research and practice, and because they measure CM as a subconstruct of general teaching quality. It

seemed relevant to review CM measured in such a broad instrument, because this way, constructs that relate to CM were also taken into account.

Two other instruments, QACD and CMQ were found in a Google Scholar search for validated instruments measuring only CM. Next to two instruments that measure CM as a subconstruct, it also seemed relevant to look at two instruments that only measure CM. An advantage of these instruments is that they are both questionnaires, because the aim of this study is to develop a questionnaire as well. Looking at how items were constructed in these questionnaires could aid item construction for this questionnaire for this study.

Finally, the QTI was added to the list of instruments used in this study. All the other instruments mention the teacher-student relationship as an important element of teaching quality and/or classroom management. It seemed relevant to add an instrument that measured this construct alone, so it could become clearer what it entails and whether it will be compatible with the definition of CM in this study. The QTI is a highly validated instrument that is already included in Impact!, so this seems like the right instrument to consult.

There are both similarities and differences between the instruments in how they define and operationalize CM and what kind of teacher behavior they group under which subconstruct. The instruments were compared and afterwards the overlapping elements and subconstructs were grouped into categories. The comparison can be seen in Table 1. The goal of this comparison is to map the different ways in which CM is defined and operationalized, and to use that overview to make educated choices on what to include and what not to include in this questionnaire.

Table 1

*Comparison of constructs in CM instruments*

<b>Constructs</b>	<b>CLASS</b>	<b>ICALT</b>	<b>QACD</b>	<b>CMQ</b>	<b>QTI</b>
Efficiency; Routines; Organization of lessons	x	x	x	x	
Discipline; Rules; Dealing with undesired behavior	x		x	x	x
Climate; Positive/Negative; Safe/Disruptive	x	x	x	x	
Student-teacher relationship	x		x	x	x
Student engagement	x	x			
Instructional support	x			x	

**Similarities between instruments.** The first thing that stands out, is that all instruments except the QTI mention the importance of maximizing on-task learning time. QACD has a construct named ‘classroom management’, which focuses both on efficient use of time to increase the amount of actively-used learning time and disciplinary measures. The latter having more resemblance to the second group of similarities.

The second group of overlapping elements is about the disciplinary nature of CM. Elements that keep returning are monitoring, intervening when undesired behavior occurs, and making sure students are familiar with rules and expectations.

Another similarity is the explicit presence of the student-teacher relationship in all instruments but ICALT. ICALT does, however, mention aspects of positive teacher-student relationships (e.g. humor, listening to students, complimenting students) in their category on classroom climate, possibly implying a connection between good teacher-student relationships and positive classroom climate.

The final similarity that attracts attention, is the focus all instruments have on the climate or ambiance in the learning environment. The amount of negative or positive events, affects the students’ learning process. All instruments agree that a safe, warm, and non-disruptive classroom enhances learning.

**Differences between instruments.** The instruments differ however, in how they measure the type of classroom climate. CLASS aims at both positive and negative climate, positive reflecting a warm and respectful emotional connection between teachers and students, negative reflecting the overall level of negativity among teachers and student in the classroom. ICALT and CMQ only look at a positive climate, emphasizing a learning environment be safe and stimulating, and promoting positive social values. QACD focuses on the undesired negativity that can arise in the classroom: both aggressive and non-aggressive disruptions to the teaching-learning process.

Another difference occurs in the disciplinary department: CMQ and QACD speak only of preventing and reprimanding bad behavior, whereas CLASS also mentions part of managing behavior is encouraging good behavior, by complimenting students that behave desirably. QTI does not mention this type of behavior specifically, but it would seem that a teacher who uses this strategy leans more towards the cooperation side of the axis. ICALT does not mention anything on the topic of behavior management and discipline, implying that good time-management, planning and routines all are that is necessary for efficient classroom management.

Furthermore, in CLASS and ICALT, student engagement is mentioned as a subconstruct. It is not present in the CMQ and QACD, suggesting it is probably an important element in general teaching quality, but not a part of CM.

The same can probably be said for instructional support, which is present in CLASS and ICALT, and not in QACD. CMQ, however, has a construct called ‘interaction during the lesson’, which concerns teacher dialogue during instruction. But, content related instructional support is not relevant to CM as it is about instruction directly, and not about its prerequisite of creating a supportive learning environment.

### **Conclusion and final definitions**

One of the most recurring and important elements of CM seems to be the teacher’s effort to effectively spend as much time as possible on education and as little time possible on off-task activities. For maximizing on-task behavior, it is important that teachers organize their classroom activities in the utmost efficient way. This sentiment strongly agrees with the ecological approach as presented by Wubbels (2011).

However, in order to carry out an efficient lesson, certain conditions have to be met: Students need to feel safe and supported in the classroom, and there should be as few disruptions as possible. The importance of a positive classroom climate for CM is highlighted in the internal control approach. Although the ecological approach does tell how and why the number of classroom disruptions should be limited, there is some uncertainty on how to solve disruptions when they cannot be solved non-verbally. This leaves a need that can be filled by reviewing the behavioral approach. Namely, this approach does have a strong focus on explicitly tackling misbehavior.

Ultimately, there does not seem to be one single fitting theoretical approach. Using the four relevant approaches for combining the strengths from different perspectives seems more valuable. Based on these approaches and the information extracted from the instrument analysis, a categorization of three subconstructs of CM was made.

The first subconstruct is about classroom efficiency, linking the ecological approach to the instrument-derived importance of maximizing instructional time using routines, time-management, and preparation. The definition of the subconstruct classroom efficiency therefore is: The teacher’s ability to manage a lesson without wasting instructional time, using routines and effective time management.

The second subconstruct incorporates the ecological approach’s monitoring and redirecting minor disruptions, the behavioral approach’s reprimanding of more severe

disruptions, and both approaches' rule clarity into one aspect of CM: Classroom discipline. It is the teacher's effort in preventing, redirecting, and reprimanding misbehavior by setting clear rules and expectations, monitoring, and adequately reacting to undesired behavior.

The third subconstruct is classroom climate, which fuses the social-emotional aspect of the internal control approach and the interpersonal approach and the instruments' need for a safe and supporting learning environment and positive student-teacher relationships. The definition of classroom climate therefore is: Classroom climate is the teacher's effort in creating and enforcing a positive and supportive environment in which students experience warmth, safety, and respect.

To make one single definition to capture CM comprehensively in this study, the three aforementioned subconstruct definitions were combined. Classroom management is the maximalization of productive instruction time, by preventing and correcting the loss of time, attention, and undesired behavior by means of rules, routines, and corrective and encouraging teacher behavior.

## **Method**

### **Design**

In this study, the questionnaire that measures classroom management of secondary education teachers was developed. For that, a design research approach was used. Both qualitative (interviews) and quantitative (questionnaire) measures were used. First, literature analysis was conducted and based on that questionnaire items were designed. These items were evaluated in four phases: Phase 1, with expert researchers (in interviews). Phase 2, with teachers (in focus groups). Phase 3, with students (in interviews). Based on their feedback, the items were further developed. Finally, in phase 4 the psychometric quality of the questionnaire was determined and evaluated. The final questionnaire was aimed to consist of around 15 items, so evaluation did not take up much learning time. The questionnaire needed to be valid, and practical, which explained the initial choice for approximately five items per subconstruct and fifteen items in total. This made it easy to fill in during class, while still being lengthy enough to include all necessary items. Just like the other questionnaires available in the Impact! tool, the items were measured by means of a four-point Likert scale, visualized by one or two thumbs up or down. All the items were phrased in the same, positive direction, so students would not be confused. The items were also phrased in past-tense, relating to the lesson they had just participated in. This way, the teacher could prepare for that lesson and gain specific feedback on their classroom management in that lesson.

## Respondents

**Phase 1: Expert researcher interviews.** The first group consisted of five Dutch researchers with expertise on CM. They were selected by expert sampling. Respondents were asked by email if they would like to participate.

**Phase 2: Expert teacher focus groups.** The second group consisted of three<sup>1</sup> experienced secondary education teachers: Two mathematic teachers and one science teacher, all with at least fifteen years of experience teaching in secondary education. Experienced teachers are better than novice teachers at monitoring relevant classroom events, and understanding (disruptive) classroom situations (Wolff, Jarodzka, Van den Bogert, & Boshuizen, 2016). Therefore, they have more time to focus on learning, as opposed to novice teachers, who focus mainly on maintaining discipline (Wolff et al., 2015). Expert teachers were selected based on a colleague's recommendation. These teachers were selected using homogeneous sampling, by contacting schools within the researchers' networks, and asking for volunteer participants with experience teaching in secondary education.

**Phase 3: Student interviews.** The third group consisted of eight secondary education students. Three third year VWO students, three third year HAVO students, and two fifth year HAVO students. Teachers were asked to send their linguistically weaker students, because if they understood the items, than the stronger students would automatically understand them as well. The students were selected through the expert teachers from the second group, by asking them if they would like to participate with their students as well.

**Phase 4: Pilot testing.** The fourth group also consisted of secondary education students ( $N = 438$ ) from 19 different classes, and seven different mathematic teachers from one Dutch secondary school. These students were selected in the same way, through one of the teachers that participated earlier.

Both student groups were selected using homogeneous sampling, considering they were all Dutch students in secondary school. Their ages ranged from 11 to 18, and their grades and levels varied between first and sixth grade, and VMBO and VWO. The exact distribution can be seen below in Table 2.

---

<sup>1</sup> The goal was to have twelve experienced teachers, in two separate focus groups, but due to the Covid-19 pandemic, most teachers did not have time to participate.

Table 2

*Distribution of pilot test students amongst grades and levels*

	VMBO	HAVO	HAVO/VWO	VWO
First grade	18	0	0	0
Second grade	59	0	118	25
Third grade	0	43	-	50
Fourth grade	0	49	-	0
Fifth grade	-	0	-	29
Sixth grade	-	-	-	47

### Procedure

**Phase 1: Expert researcher interviews.** First, expert researchers were interviewed using the first version of the questionnaire. They judged whether the construct of classroom management and its sub-constructs were lacking or superfluous, and gave feedback to improve the questionnaire, so a second version could be made. The interview took place one on one, via Skype, and was conducted using semi structured interview questions (see Instruments). The interview started by explaining the goal and plan of this study, and by explaining the constructs that were used so far. On average the interviews lasted 37 minutes.

**Phase 2: Expert teacher focus groups.** Secondly, one focus group with expert teachers took place. They were also asked to review the latest questionnaire version, alongside with naming missing categories/items that should be included in the next version. The focus group was lead using the designed interview questions (see Instruments). Before it started, the study and the used constructs were explained. In leading the focus group, it was vital to let people discuss freely, but also to steer the conversation back to the original question if it ran off course. The focus group took place via Microsoft Teams. Additionally, a Padlet was used to gather everyone's input on the matter. The focus group lasted 85 minutes.

**Phase 3: Student interviews.** In the third phase, the students were asked one on one, to read the adapted questionnaire and judge whether they understood each item. They were also asked to explain each item, to make sure that they interpreted it correctly. This was done using the think-aloud method, with some additional interview questions if more clarity was needed. The average interview lasted 12 minutes.

Before the interviews and focus groups started, participants consented that they took part voluntarily and that they could quit anytime they wanted to. They also gave permission to

record the conversations and use the data in the development of the questionnaire.

Additionally, the participants were guaranteed beforehand that everything they did and said was confidential and that results would be published anonymously.

**Phase 4: Pilot testing.** In the final phase, data collection encompassed that students used the latest version of the questionnaire to evaluate their teacher's classroom management skills. In an introduction, the study and the relevance of their participation was explained. The students were asked to rank all the statements of the questionnaire honestly and individually. The questionnaire was shared digitally by means of the Impact! tool, so they could fill them out on their own devices. This group received an instruction first, in which the study was explained again, confidentiality and anonymity was guaranteed, and voluntary participation was agreed on. Together, the instruction and the data collection took up five to ten minutes per class.

### **Instruments**

**Phase 1: Expert researcher interviews.** Because a semi structured interview technique was used in this phase, a few leading questions needed to be constructed beforehand. The following questions were used to guide the interview:

- Do you agree with the classification of CM into these three subconstructs: classroom efficiency, classroom discipline, and classroom climate?
- In your professional opinion, is the operationalization of the items valid?
- Is there anything missing in this questionnaire?
- Do you have additional suggestions for improvement of this questionnaire?

**Phase 2: Expert teacher focus groups.** The data collection in the focus group was structured around a Padlet in which all the questionnaire items were displayed. With each item, two questions needed answering:

- Is the item relevant to CM in secondary education classrooms?
- Is the item well-phrased, so that students will interpret it correctly?

Afterwards, the teachers were asked whether they had any more questions, comments, or concerns regarding the questionnaire to address topics that had not been covered by the two leading questions.

**Phase 3: Student interviews.** The following questions were asked about each questionnaire item:

- What do you think that this question means?
- If correct: Is there an easier way of asking this question? If so, what would you suggest?
- If incorrect, first explain what the question means, then ask: What would you ask if you wanted to know this?

**Phase 4: Pilot testing.** The latest version of the CM questionnaire was tested in the final phase. The questionnaire, however, was not a research instrument, but the research topic. The questionnaire used in the pilot test can be seen in Appendix E.

### **Statistical data analysis**

**Construct validity.** To assess the construct validity of the questionnaire statistically, the fit of a combined Item Response Theory (IRT) and Generalizability Theory (GT) model was examined by testing for local independence of the items on the total scale ‘classroom management’ and of the sub scales. The assumption posits that the relation between student responses is entirely due to the latent variable and that responses are therefore uncorrelated after accounting for the latent variable. Van den Wollenberg (1982), Yen (1984) and Glas (1999) pointed out that this assumption could be tested by evaluating the observed and expected association between two item responses. With the associated observed and expected scores, average absolute differences between the scores were calculated and examined. The average absolute difference should be no more than 0.1 (Glas, 2016). If so, the model measured a unidimensional scale or subscale which would support the construct validity of the questionnaire. To examine whether the subscales were associated with each other, correlations between the subscales were calculated.

**Reliability.** To test whether students’ ratings were reliable measurements of CM, it is important to notice that the students are nested in a combination of a specific teacher giving a specific lesson. So the reliability coefficient pertains to the reliability of the students as judges of a specific teacher giving a specific lesson, and teachers cannot be compared regarding their CM ability. For the computation of the reliability of a score averaged over students, the variance of the teacher-by-lessons combinations was divided by the total variance. That is, the reliability coefficient is the ratio of the variance attributable to lessons of teachers and the total variance resulting from the IRT model. In this study, the reliability coefficient showed to what degree students in the same class rated their teacher similarly. Given the fact that each combination of a teacher and a lesson was rated by many raters, that is, many students, this is not unexpected.

## Results

### *A. Which items make up a valid and reliable student perception questionnaire for measuring classroom management in secondary education, based on relevant literature and validated instruments?*

#### **Classroom efficiency**

The definition of classroom efficiency in this study is “the teacher’s ability to manage a lesson without wasting instructional time, using routines and effective time management”. This means that items on routines and time management had to be created. This led to three items on time management (at the beginning of a lesson, during the lesson, and at the end of a lesson) and two on routines (whether students know what to do and whether there is a clear structure in the lesson).

#### **Classroom discipline**

Classroom discipline is defined in this study as “the teacher’s effort in preventing, redirecting, and reprimanding misbehavior by setting clear rules and expectations, monitoring, and adequately reacting to undesired behavior”. Following this definition, two items were created on clear rules and expectations concerning student behavior, two items were created on responding adequately to undesired behavior, and one item was created on monitoring the classroom.

#### **Classroom climate**

Finally, the definition of classroom climate as phrased in this study is “the teacher’s effort in creating and enforcing a positive and supportive environment in which students experience warmth, safety, and respect”. This led to the creation of five items, two on warmth, two on safety, and one on respect.

The complete first version of the questionnaire can be found in Appendix A.

### *B: What revisions should be made to the classroom management questionnaire, based on interviews with expert researchers?*

#### **General remarks**

**Overlap with QTI.** The remark that was most critical towards the categorization of CM into Classroom Efficiency, Classroom Discipline, and Classroom Climate was, that because the Questionnaire on Teacher Interaction (QTI) is also used in the Impact! tool, there was no need to measure anything that was already measured in the QTI again in the new questionnaire on CM. Classroom Climate and partly Classroom Discipline showed overlap with the QTI, which made them less relevant. Therefore, the QTI was compared to the CM

questionnaire, to pinpoint where exactly they overlapped, and what thus had to be eliminated from the CM questionnaire.

The most obvious commonalities were seen in what this study categorizes as classroom climate. The items that questioned whether the teacher has patience, a sense of humor, and a pleasant ambiance in the classroom, fit seamlessly to the classroom climate definition used in our study, which referred to the teacher's effort to create and maintain a positive and supportive learning environment.

The overlap with classroom discipline was less apparent. In this study classroom discipline was defined as the teacher's effort in preventing, redirecting, and reprimanding misbehavior by setting clear rules and expectations, monitoring, and adequately reacting to undesired behavior. The QTI measured these types of teacher behavior by looking at the amount of control the teacher enforces. This resulted in a different focus in the items. The QTI phrased items that questioned whether the teacher had authority and good leadership, whether they were lenient with students, and whether they decided when students were allowed to speak. The CM items on discipline coincided more with another QTI item: Whether the teacher kept strict discipline. The CM items did not focus on the level of control, but on the actions teachers take to keep that discipline: Setting rules and expectations, monitoring, and responding to misbehavior.

Finally, the remaining items in the QTI did not match any items in the CM questionnaire, since they asked about more emotional behavior of teachers: Whether the teacher seems insecure or hesitant, whether they are dissatisfied, sombre, or quick-tempered, and whether they can get angry.

Adapting the CM questionnaire to this feedback resulted in completely eliminating classroom climate and putting a larger emphasis on use of behavior management strategies in classroom discipline. Since classroom efficiency does not have any overlap with the QTI, that construct could stay as it was.

**Internal consistency issues.** Another general remark that applied to multiple items and subconstructs was that it would be better for the internal consistency of a questionnaire to have the same number of questions on all topics. For example, in classroom efficiency there were three items on time management and two on routines. The quality of the questionnaire would improve if they were more balanced.

Additionally, the internal consistency could be influenced by the use of multiple referents. While most items were about the teacher, some items asked about the students themselves. It was also important to decide whether items should ask about the student's

individual experience, or about the experience of the class in general. Thus, a choice had to be made as to what referent to use for all items.

### **Comments on Classroom Efficiency**

The definition spoke of ‘instructional time’, but ‘lesson time’ would be better phrased, since the items were not just about instruction, but about the whole time students were in class.

What lacked in this subconstruct were items on classroom transitions, because transitions between different classroom activities are usually (very) chaotic. It would be relevant to know how a teacher handles those situations.

Finally, a comment was made on the possibility to measure routines. It can be difficult to make good items on routines, because they are very implicit. Students can usually sense their presence pretty well, but it is a matter of sharply phrasing the items. The items on routines should therefore be handled with extra caution.

### **Comments on Classroom Discipline**

To make sure this category is also balanced, a distinction can be made between preventive and reactive management strategies. Items six through eight involve preventive strategies, whereas items nine and ten concern reactive strategies. The reactive strategies are quite harsh, since teachers usually give out a warning before they intervene. Another comment relates to this, saying that a lot of steps precede an intervention, such as gestures or other non-verbal warnings. These so-called micro corrections should be included in the questionnaire.

Another expert stated that there is a subtle difference between rules and agreements. Many rules are created with the class and not solely imposed by the teacher, making them agreements. This difference should be taken into account when phrasing the items.

Lastly, the category lacked an item on fairness or justice concerning the way the teacher intervenes in disruptive situations.

### **Comments on Classroom Climate**

Something that was present in many definitions, but not in this one, was student engagement. The reason it lacked here, was probably because this category focused on climate (meaning on warmth, support, and respect), and not on learning climate (which is more about engaging students and having high expectations). Engagement could be a separate subconstruct or the category could be renamed ‘Positive Learning Climate’ and be adapted correspondingly.

Another comment was that there were not enough items on respect, compared to the importance of that topic for a positive classroom climate.

### **Comments about specific items**

All item-specific feedback is displayed in Appendix B.

### **New questionnaire version**

All the feedback above was taken into account when revising the items that make up the questionnaire. From multiple rounds of adaptation, including a short second review with two of the experts who gave their feedback earlier, the following new questionnaire version was made (see Appendix C for the specific items). Items one through four are about time management and items five through eight are about routines. Together they make up 'Classroom Efficiency'. Items nine through eleven are about preventive strategies and items twelve through sixteen are about reactive strategies. Together they make up 'Classroom Discipline'. Some items have two options, these options will be tested with teachers and students to make sure the best possible item is chosen.

### ***C. What revisions should be made to the classroom management questionnaire, based on focus groups with expert teachers?***

#### **Comments on Time Management**

Both options for item one were flawed, the comments were that students would not always be able to recognize that an activity is in fact relevant to the lesson at hand and that 'good' is a too vague descriptor. An alternative that was offered was 'This teacher uses all available time for the lesson'.

An addition to the second item was to include something about students' working independently after instruction.

For the third item, it was deemed important to be more specific about what exactly is useful. Explicitly mentioning the subject or teaching material may improve the item.

The fourth item was deemed relevant and well phrased. However, there was some debate about whether to change 'this teacher *has* time to answer questions' to 'this teacher *takes* time to answer questions'.

#### **Comments on Routines**

Item five was not believed to be relevant, because students in secondary education usually do not complete their work in the lesson.

Between the two options for item seven, option b was preferred by all teachers. However, this item could be made more understandable by explicitly asking about the different elements of a structured lesson, for example ‘This teacher always includes a start, middle, and end in each lesson’.

There were no comments on items six and eight.

### **Comments on Preventive Strategies**

According to the teachers, the difference between options a and b in item 9, were that option a is more about implicit rules and expectations, whereas option b is more about explicit rules and expectations. In relation to a teacher’s actions to provide good CM, the teachers agreed that explicitness was a better fit.

Item ten was met with uncertainty, since teachers do not usually let students know what the consequences of their possible misbehavior will be beforehand. There are rules at a school level that students may be familiar with, but teachers tend to have their own as well.

Items eleven and twelve were deemed relevant and well phrased.

### **Comments on Reactive Strategies**

There were no comments on items thirteen to fifteen.

There was some debate on the relevance of the last item. From a teacher’s point of view, it is not relevant, because in reality you do not always treat students exactly the same. The teacher’s reaction to the student’s behavior is based on their history, which can result in a more strict response to one student, than to another for the exact same behavior. From a student perspective however, it is a relevant item, because students are very heartfelt on being treated fair and just. When they feel like they are not treated equally, this can harm the relationship with the teacher and order in the classroom.

### **General comments**

A final, general suggestion that came up, was that it could be of great help to the students if there were content specific headings between the clusters of questions. This could provide them with more context as to what the questions are about and aid them in answering adequately. This could, for example, provide an explanation to the students that they should answer the questions on their own ‘misbehavior’ hypothetically, since most students behave well and have not experienced these situations personally.

The new and improved version of the questionnaire can be seen in Appendix D.

*D. What revisions should be made to the classroom management questionnaire, based on interviews with students?*

In general, students interpreted the items correctly, with a few exceptions. A misconception that stood out was that, when speaking about ‘learning’, students would almost always interpret that as ‘studying’, while it was intended to include all educational activities in the classroom. Students suggested that it would be better to ask whether the teacher dedicated all lesson time to the educational content.

Additionally, there were three instances where multiple items were shown, and students had to pick the one that they thought resembled the purpose of the item best. The first one was about whether, in the class of this teacher, the students were always busy with relevant educational activities. With the word ‘relevant’ being too difficult for the target population, students were asked to choose between ‘educational content’, ‘the subject’, and ‘learning’. Students immediately ruled out ‘learning’ for the same reason as mentioned above. They found ‘educational content’ and ‘the subject’ to be synonymous. Because they interpreted both options correctly, but similarly, ‘the subject’ was chosen to provide variation in phrasing, since ‘educational content’ was already used in the first item.

The second choice was about an item that was supposed to give insight in the time management and planning skills of the teacher. This could be assessed by checking whether the teacher always started with a clear introduction or finished with a clear ending. Students were more able to identify a clear ending, with the teacher summarizing, concluding, and giving homework. They also interpreted a teacher’s introduction as the teacher telling them what they are going to do in that lesson. This is so similar to the fifth item, that it seemed more relevant to add the item about the ending of the lesson.

The third dilemma was how to assess whether a teacher is just or fair. Both options were found to be quite vague, which is why all students preferred the phrasing ‘this teacher treats all students equally’.

The fourth version of the questionnaire can be seen in Appendix E.

*E. What revisions should be made to the classroom management questionnaire, based on the statistical analyses of the psychometric quality of the questionnaire?*

To directly answer the question, no revisions needed to be made based on the statistical analyses of the psychometric quality of the questionnaire. The construct validity of the questionnaire as a whole, the construct validity of the subconstructs, and the reliability of the questionnaire were all high.

**Questionnaire statistics**

The mean score of the results of all teachers was 3.34 ( $SD = 0.28$ ). Two teachers were evaluated once. The other five teachers were evaluated multiple times by different classes, ranging from two to five times.

**Construct validity**

**Classroom Management.** To assess the validity of the questionnaire, an IRT-GT model was used. This model provided estimated outcomes for each item of the questionnaire. The estimated outcomes exemplify a questionnaire that is perfectly valid. Therefore, the observed data from the pilot test should be as close to the estimated data as possible, for the questionnaire to have a high validity. The average absolute differences (AADs) between the estimated data and the observed data are less than 0.1 on all items (see Table 3, column AADs CM), with the highest AAD being 0.09 for item 1 and the lowest being 0.02 for item 2. The mean of all AADs is 0.044 ( $SD = 0.018$ ). This signifies that the validity of the Classroom Management Questionnaire is good.

Table 3

*Average absolute differences between estimated and observed data*

Item	AADs CM	AADs subscales
1	0.09	0.06
2	0.02	0.04
3	0.03	0.01
4	0.04	0.05
5	0.08	0.06
6	0.03	0.01
7	0.04	0.02
8	0.05	0.08

9	0.03	0.03
10	0.04	0.02
11	0.05	0.03
12	0.04	0.01
13	0.03	0.02
14	0.03	0.04
15	0.05	0.06
16	0.06	0.06

**The four subconstructs.** The AAD's between the estimated and the observed data are also less than 0.1 for all individual subscales (see Table 3). The mean of all AAD's for time management (items 1 through 4) is 0.04 ( $SD = 0.019$ ). For routines (items 5 through 8), the mean is 0.043 ( $SD = 0.029$ ). For subconstruct preventive strategies (items 9 through 12) it is 0.023 ( $SD = 0.008$ ), and the mean of all AAD's for reactive strategies (items 13 through 16) is 0.045 ( $SD = 0.017$ ). Since the observed data is so close to the ideal situation, the estimated data from the IRT-GT model, each individual subconstruct has good validity as well.

Additionally, the correlations between subscales are high (see Table 4). This could indicate that the four subconstructs time management, routines, preventive strategies, and reactive strategies are related. It would make sense for the subconstructs to be related, since a teacher that is skilled in one aspect of CM, is more likely to also be skilled in other aspects of CM. However, this could also mean that the subconstructs were not distinct enough from each other, resulting in them all measuring one identical construct, which in this case would be CM.

Table 4

*Correlations between subscales*

	<b>Time management</b>	<b>Routines</b>	<b>Preventive strategies</b>	<b>Reactive strategies</b>
Time management	1.0	0.94	0.86	0.79
Routines		1.0	0.92	0.86
Preventive strategies			1.0	0.91
Reactive strategies				1.0

Nine out of sixteen items loaded better on their respective subscale than on the total scale of CM. This is a good indication that the subconstructs were distinct enough. However, five out of sixteen items loaded better on the total scale than on the subscale, which would indicate the opposite. Preventive strategies was the only subconstruct of which all items loaded better or equally on the subscale. Only for this subconstruct could be said with certainty that it is distinct from the other subconstructs and the total scale of CM.

### **Reliability**

Finally, the analyses also showed that the reliability of the Classroom Management Questionnaire is very high, with a reliability coefficient of 0.989. In this study, the reliability represented to what extent students rated teachers similarly. Since different students amounted similar scores to different teachers, this resulted in a high reliability coefficient.

## **Conclusion and discussion**

### **Conclusion**

Literature research and analysis of existing instruments on CM and teaching quality provided a solid basis for the Classroom Management Questionnaire. Researchers and teachers then gave their expert opinions on early versions of the questionnaire, shaping it into a cohesive list of items, consistent of four CM components: Time management, routines, preventives strategies, and reactive strategies. The items were then tested with students for correct interpretation. A pilot test revealed that both the questionnaire as a whole and the subconstructs individually are valid. This means that the questionnaire accurately measures CM. Not all subconstructs are, however, very distinct from each other and CM. Only the subconstruct preventive strategies showed sufficient distinction. The questionnaire also has a high reliability, but the data could only prove this in the sense that classes of students scored their teachers similarly. More critical remarks concerning this study are discussed further on.

The main research question of this study was: “Which items make up a valid and reliable student perception questionnaire for measuring classroom management in secondary education?”. Listed below are the items of the complete Classroom Management Questionnaire translated to English. The Dutch version can be seen in Appendix E.

1. This teacher dedicates all lesson time to educational content.
2. In this teacher’s class, I get started quickly after the instruction.
3. In this teacher’s class, I am constantly involved with the subject.

4. This teacher has the time to answer questions during the lesson.
5. In this teacher's class, I constantly know what to do.
6. In this teacher's class, I know what I need to do if I need help.
7. This teacher finishes every lesson with a clear ending.
8. This teacher has clear rules for working individually.
9. This teacher has clear rules about how I need to behave.
10. In this teacher's class, I know how to behave during the lesson.
11. This teacher always notices what is happening in the classroom.
12. This teacher controls what is happening during the lesson.
13. If I misbehave, this teacher gives me a warning first.
14. If I do not pay attention, this teacher makes sure that I will.
15. If I disrupt the lesson, this teacher makes sure that I will stop doing that.
16. This teacher treats all students equally.

### **Discussion**

The high construct validity could be explained by the pursuance of face validity in the first three research phases. Face validity is the extent to which a test is subjectively viewed as covering the concept it intends to measure. The considerable contribution of subject matter experts (researchers with expertise on CM, experienced teachers in secondary education, and secondary education students) amounted to a high level of face validity. Because of their expertise, a good theoretical and practical foundation for construct validity was created, that eventually resulted in a valid questionnaire.

Although the reliability is very high, this result should be treated with some caution, because the class sizes were quite large, with an average of 23 students per class and a total of 438 students. As the number of students increases, a regression to the mean occurs, leading to coinciding evaluation scores. This drove up the reliability coefficient. What we did investigate in this study, is whether students in a class judge a teacher similarly or differently. This turned out to be similar, thus reliable. We cannot, however, make a conclusion about other types of reliability. What we did not investigate in this study, is whether students in a class judge different teachers similarly or differently. This type of reliability could be researched in future studies, by having students fill out the questionnaire for multiple teachers.

Another important remark that requires attention is that choices had to be made whilst developing this questionnaire. Choices were always supported by expert opinions or practical concerns, however, this questionnaire could have easily looked different if certain choices had

been justified from a different perspective. For example, not all expert researchers advised to exclude classroom climate as a construct of CM. If we had deemed this construct to be too important to leave out and be measured in a separate questionnaire, this questionnaire would have looked very different. This is true for many choices that were made in the development process. Therefore, it is vital to remember that a certain level of arbitrary and chance had a hand in creating this instrument.

### **Limitations of the study and recommendations for further research**

Finally, this study has some limitations that can be solved by undertaking further research into the Classroom Management Questionnaire. First, some general improvements that could be made to this study. As mentioned before, the reliability of the questionnaire could be analyzed more thoroughly, by having classes of students fill out the questionnaire for multiple teachers, in order to discover whether they judge different teachers differently. Other limitations include the pilot test being conducted at one school, only including math teachers, and not testing amongst all levels and grades equally. It would be interesting to see whether there are differences between schools, between teachers from different subjects, and between levels and grades. A larger test that includes these factors could also provide us with more information on the validity and reliability of the questionnaire.

Second, it would be very valuable to see whether this questionnaire indeed helps teachers identify areas of improvement. This could be investigated using a second or even third measurement moment, so development can be mapped. Another recommendation that relates to this point, is to conduct research into the activities that teachers undergo after they have received the evaluation results. This will establish whether and how teachers used the evaluation outcomes for their own professional improvement.

Third, the focus in this study is on student perception. This is a valuable way of gathering feedback for teachers, but it is not the only way. By also making an appropriate version of this questionnaire available for external observers and for teachers to use as an instrument of self-evaluation, triangulation could be used to improve interrater reliability, and provide teachers with more relevant input. It would also be interesting to investigate if there are differences between student perceptions, external observations, and self-evaluations.

A final recommendation for further research relates to the fact that a questionnaire on CM could not only be helpful in educational practice, but also for research purposes. However, this requires additional analyses. Future studies should correct students' ratings of teaching quality for factors that could influence students' evaluations of teaching quality, such

as class size, grade average, and gender. This can clarify which part of the ratings reflect a teacher's CM skills.

### References

- Balch, R. T. (2012). *The validation of a student survey on teacher practice*. Nashville: Vanderbilt University.
- Bijlsma, H. J., Visscher, A. J., Dobbelaer, M. J., & Veldkamp, B. P. (2019). Does smartphone-assisted student feedback affect teachers' teaching quality?. *Technology, Pedagogy and Education, 28*(2), 217-236.
- Brophy, J. (2006). History of research on classroom management. C. Evertson, C. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues*, 17-43.
- Burden, P. R. (2000). *Powerful classroom management strategies: Motivating students to learn*. Corwin Press.
- Díaz, C., González, G., Jara-Ramírez, L. I., & Muñoz-Parra, J. A. (2018). Validation of a Classroom Management Questionnaire for pre and Inservice Teachers of English. *Revista Colombiana de Educación, (75)*, 263-286.
- Doyle, W. (2006). Ecological management and classroom management. C. Evertson, C. Weinstein (Eds.), *Handbook of classroom management. Research, practice, and contemporary issues*, 97-126.
- Elias, M. J., & Schwab, Y. (2006). From compliance to responsibility: Social and emotional learning and classroom management. C. Evertson, C. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues*, 309-341.
- Evertson, C. M., & Weinstein, C. S. (2006). *Handbook of Classroom Management: Research, Practice and Contemporary Issues*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1-9.
- Ferguson, R. F. (2012). Can student surveys measure teaching quality? *Phi Delta Kappan, 94*(3), 24–28.
- Glas, C. A. (1999). Modification indices for the 2-PL and the nominal response model. *Psychometrika, 64*(3), 273-294.
- Glas, C. (2016). Frequentist model-fit tests. *Handbook of item response theory: Vol. 2. Statistical tools*, 343-361.

- Hoban, G. F., & Hastings, G. (2006). Developing different forms of student feedback to promote teacher reflection: A 10-year collaboration. *Teaching and Teacher Education*, 22, 1006–1019.
- Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Research paper. MET project.* Seattle, WA: Bill & Melinda Gates Foundation.
- Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2014). Effective classroom management strategies and classroom management programs for educational practice. *Journal of Educations*.
- Landrum, T. J., & Kauffman, J. M. (2006). Behavioral approaches to classroom management. C. Evertson, C. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues*, 47-71.
- Lasagabaster, D., & Sierra, J. M. (2011). Classroom observation: desirable conditions established by teachers. *European Journal of Teacher Education*, 34(4), 449-463.
- Marzano, R., Foseid, M., Foseid, M., Gaddy, B., & Marzano, J. (2005). *A handbook for classroom management that works.* Alexandria, VA: Association for Supervision and Curriculum Development.
- Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research and Evaluation*, 12, 53–75.
- Nault, T. (1994). *L'enseignant et la gestion de la classe. Comment se donner la liberté d'enseigner* [The teacher and classroom management. How to give freedom to teach] Montréal: Logiques.
- Peterson, K. D., Wahlquist, C., & Bone, K. (2000). Student surveys for school teacher evaluation. *Journal of Personnel Evaluation in Education*, 14, 135–153.
- Pianta, R. C., Hamre, B. K., & Mintz, S. (2012). *Classroom Assessment Scoring System (CLASS) Manual, Secondary.* Teachstone.
- Scherzinger, M., & Wettstein, A. (2019). Classroom disruptions, the teacher–student relationship and classroom management from the perspective of teachers, students and external observers: a multimethod approach. *Learning Environments Research*, 22(1), 101-116.
- Scrivener, J. (2012). *Classroom management techniques.* Cambridge, UK: Cambridge University Press.
- Sieberer-Nagler, K. (2016). Effective Classroom-Management & Positive Teaching. *English Language Teaching*, 9(1), 163-172.

- Van de Grift, W. J. C. M. (2007). Quality of teaching in four European countries: A review of the literature and application of an assessment instrument. *Educational Research and Evaluation, 49*, 127–152.
- Van den Wollenberg, A. L. (1982). Two new test statistics for the Rasch model. *Psychometrika, 47*(2), 123-140.
- Van der Lans, R. M., Van de Grift, W. J. C. M., & Van Veen, K. (2016). Developing a teacher evaluation instrument to provide formative feedback using students' ratings of teaching acts. *Educational Measurement: Issue and Practice, 34*(3), 18–27.
- Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research, 54*(2), 143-178.
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63*(3), 249-294.
- Webster-Stratton, C. (2012). *The incredible years. Training series for parents, teachers, and children*. Retrieved from <http://www.incredibleyears.com>.
- Wolff, C. E., Jarodzka, H., Van den Bogert, N., & Boshuizen, H. P. (2016). Teacher vision: expert and novice teachers' perception of problematic classroom management scenes. *Instructional Science, 44*(3), 243-265.
- Wolff, C. E., Van den Bogert, N., Jarodzka, H., & Boshuizen, H. P. (2015). Keeping an eye on learning: Differences between expert and novice teachers' representations of classroom management events. *Journal of Teacher Education, 66*(1), 68-85.
- Wubbels, T. (2011). An international perspective on classroom management: What should prospective teachers learn?. *Teaching Education, 22*(2), 113-131.
- Wubbels, T., Brekelmans, M., den Brok, P., & Van Tartwijk, J. W. F. (2006). An interpersonal perspective on classroom management in secondary classrooms in the Netherlands. C. Evertson, C. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues*, 1161-1192.
- Wubbels, T., Créton, H., & Hoymayers, H. (1992). Review of research on teacher communication styles with use of the Leary model. *The Journal of Classroom Interaction, 1*-11.
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*(2), 125-145.

**Appendix A**

Table 5

*The Classroom Management Questionnaire (first version)*

---

Classroom Efficiency
1. De docent begon de les op tijd.
2. De docent verspilte tijdens de les geen tijd.
3. De docent had genoeg tijd om alles te doen wat hij gepland had.
4. De docent zorgde ervoor dat ik tijdens de les steeds wist wat ik moest doen.
5. De docent zorgde voor een duidelijke opbouw van de les.
Classroom Discipline
6. Ik wist tijdens de les wat ik wel en wat ik niet mocht doen.
7. Ik wist tijdens de les wat de docent zou doen als ik iets deed wat niet mocht.
8. De docent hield steeds goed in de gaten wat er in de klas gebeurde.
9. De docent greep goed in als leerlingen niet opletten.
10. De docent greep goed in als leerlingen de les verstoorden.
Classroom Climate
11. De docent zorgde voor een positieve sfeer in de klas.
12. De docent gaf leerlingen complimenten over hun werk.
13. De docent zorgde ervoor dat ik me veilig voelde in de klas.
14. De docent reageerde vriendelijk wanneer iemand een fout maakte.
15. De docent respecteerde mij tijdens de les.

---

## Appendix B

Table 6

*Item-specific feedback from the expert researcher reviews*

Items	Feedback	
<b>Classroom Efficiency</b>		
1	De docent begon de les op tijd.	Item 1 and 2 are very similar, merge them into one item.
2	De docent verspilte tijdens de les geen tijd.	Strong overlap with item 1 and too implicit on maintaining order during transitions between activities.
3	De docent had genoeg tijd om alles te doen wat hij gepland had.	Students may have trouble assessing this. It might be more relevant to know if the lesson was organized in a way that students did not have time to be distracted and show off-task behavior.
4	De docent zorgde ervoor dat ik tijdens de les steeds wist wat ik moest doen.	Specify that student know what they need to do when they have finished a task, and that this is about having and knowing classroom routines.
5	De docent zorgde voor een duidelijke opbouw van de les.	Make it clearer that it is about the structure of the lesson, without using the word structure, because that might be too difficult.
<b>Classroom Discipline</b>		
6	Ik wist tijdens de les wat ik wel en wat ik niet mocht doen.	This item consists of two separate items, that can be confusing for students. Also note more explicitly that this is about student behavior. Referent is different from ‘the teacher’ everywhere else.
7	Ik wist tijdens de les wat de docent zou doen als ik iets deed wat niet mocht.	If-then phrasing can be confusing for students. Also, the referent is different from the other items.
8	De docent hield steeds goed in de gaten wat er in de klas gebeurde.	No comments. Maybe change ‘hield’ to ‘had’.

9	De docent greep goed in als leerlingen niet opletten.	Item 9 and 10 are very similar, make the difference between them more clear. 'Good' is vague, this can both mean 'effective' and 'fair'. In this item focus on effectiveness.
10	De docent greep goed in als leerlingen de les verstoorden.	See feedback item 9. In this item focus on fairness.
Classroom Climate		
11	De docent zorgde voor een positieve sfeer in de klas.	Decide on a focus, perhaps learning climate. Just 'positive ambiance' might be too broad.
12	De docent gaf leerlingen complimenten over hun werk.	Specify where they get compliments on. Is it their work, their behavior, both? Maybe change it to something about rewarding and/or encouraging good behavior.
13	De docent zorgde ervoor dat ik me veilig voelde in de klas.	No comments.
14	De docent reageerde vriendelijk wanneer iemand een fout maakte.	It is not necessarily about getting a friendly response, the goal is that students are allowed to make mistakes. Phrase it like that.
15	De docent respecteerde mij tijdens de les.	The word 'respect' might be too abstract, this could be changed to 'listens to me' or 'takes me seriously'.

### Appendix C

Table 7

*The Classroom Management Questionnaire (second version)*

Classroom Efficiency – Time Management
1a. Deze docent besteedt de les alleen aan dingen die met de les te maken hebben.
1b. Deze docent gebruikt de lestijd goed.
2. Bij deze docent ga ik na de uitleg snel aan de slag.
3a. Bij deze docent ben ik steeds nuttig bezig tijdens de les.
3b. Bij deze docent heb ik steeds iets nuttigs te doen tijdens de les.
4. Deze docent heeft tijdens de les tijd om vragen te beantwoorden.
Classroom Efficiency - Routines
5. Bij deze docent weet ik wat ik moet doen als ik klaar ben met mijn opdrachten.
6. Bij deze docent weet ik wat ik moet doen als ik hulp nodig heb.
7a. Bij deze docent is steeds duidelijk wat er in de les gaat gebeuren.
7b. Bij deze docent is steeds duidelijk hoe de les is opgebouwd.
8. Deze docent heeft duidelijke afspraken over zelfstandig werken.
Classroom Discipline – Preventive Strategies
9a. Bij deze docent weet ik hoe ik mij tijdens de les moet gedragen.
9b. Deze docent heeft duidelijke afspraken over hoe ik mij moet gedragen.
10a. Bij deze docent weet ik wat er gebeurt als ik mij niet goed gedraag.
10b. Bij deze docent weet ik wat de gevolgen zijn als ik mij niet goed gedraag.
11. Deze docent heeft steeds in de gaten wat er in de klas gebeurt.
12. Deze docent bepaalt wat er gebeurt tijdens de les.
Classroom Discipline – Reactive Strategies
13. Als ik me niet goed gedraag, dan geeft deze docent eerst een waarschuwing.
14. Als ik niet oplet, dan zorgt deze docent ervoor dat ik dat wel ga doen.
15. Als ik de les verstoort, dan zorgt deze docent ervoor dat ik daarmee ophoud.
16. Als ik me niet goed gedraag, dan reageert deze docent op een eerlijke manier.

### Appendix D

Table 8

*The Classroom Management Questionnaire (third version)*

Classroom Efficiency – Time Management
1. Deze docent gebruikt alle lestijd voor leren.
2. Bij deze docent ga ik na de uitleg snel aan de slag.
3a. Bij deze docent ben ik tijdens de les steeds bezig met de lesstof.
3b. Bij deze docent ben ik tijdens de les steeds bezig met het vak.
3c. Bij deze docent ben ik tijdens de les steeds bezig met leren.
4. Deze docent heeft tijdens de les tijd om vragen te beantwoorden.
Classroom Efficiency - Routines
5. Bij deze docent weet ik steeds wat ik moet doen.
6. Bij deze docent weet ik wat ik moet doen als ik hulp nodig heb.
7a. Deze docent begint elke les met een inleiding.
7b. Deze docent eindigt elke les met een duidelijke afsluiting.
8. Deze docent heeft duidelijke regels voor zelfstandig werken.
Classroom Discipline – Preventive Strategies
9. Deze docent heeft duidelijke regels voor hoe ik mij moet gedragen.
10. Bij deze docent weet ik hoe ik mij tijdens de les moet gedragen.
11. Deze docent heeft steeds in de gaten wat er in de klas gebeurt.
12. Deze docent bepaalt wat er gebeurt tijdens de les.
Classroom Discipline – Reactive Strategies
13. Als ik me niet goed gedraag, dan geeft deze docent eerst een waarschuwing.
14. Als ik niet oplet, dan zorgt deze docent ervoor dat ik dat wel ga doen.
15. Als ik de les verstoort, dan zorgt deze docent ervoor dat ik daarmee ophoud.
16a. Deze docent is rechtvaardig.
16b. Deze docent is eerlijk.
16c. Deze docent behandelt alle leerlingen gelijk.
16d. Deze docent reageert op (het gedrag van?) alle leerlingen gelijk.

### Appendix E

Table 9

*The Classroom Management Questionnaire (final version)*

Classroom Efficiency – Time Management
1. Deze docent besteedt alle lestijd aan de lesstof.
2. Bij deze docent ga ik na de uitleg snel aan de slag.
3. Bij deze docent ben ik tijdens de les steeds bezig met het vak.
4. Deze docent heeft tijdens de les tijd om vragen te beantwoorden.
Classroom Efficiency - Routines
5. Bij deze docent weet ik steeds wat ik moet doen.
6. Bij deze docent weet ik wat ik moet doen als ik hulp nodig heb.
7. Deze docent eindigt elke les met een duidelijke afsluiting.
8. Deze docent heeft duidelijke regels voor zelfstandig werken.
Classroom Discipline – Preventive Strategies
9. Deze docent heeft duidelijke regels voor hoe ik mij moet gedragen.
10. Bij deze docent weet ik hoe ik mij tijdens de les moet gedragen.
11. Deze docent heeft steeds in de gaten wat er in de klas gebeurt.
12. Deze docent bepaalt wat er gebeurt tijdens de les.
Classroom Discipline – Reactive Strategies
13. Als ik me niet goed gedraag, dan geeft deze docent eerst een waarschuwing.
14. Als ik niet oplet, dan zorgt deze docent ervoor dat ik dat wel ga doen.
15. Als ik de les verstoort, dan zorgt deze docent ervoor dat ik daarmee ophoud.
16. Deze docent behandelt alle leerlingen gelijk.