Effects of a Self-Determination Theory Based Intervention in University Statistics

Education

Roben Scholten, Kim Stroet, Lasana Harris, and Frank Busing

Leiden University

Author Note

Corresponding author: Roben Scholten. E-Mail: robenscholten@gmail.com

Abstract

We designed and implemented an intervention aiming to improve psychology students' experience of statistics. Eight teachers and 332 students participated in the study. Half of the teachers participated in a training program on need-supportive teaching, while the other half served as control. Based on student reports of perceived need-support, the training program effectively promoted a more need-supportive teaching style. The program however did not successfully provide the hypothesized improvements of students' quality of motivation, engagement and academic performance. The teachers did however report to appreciate the training program, and a promising, but non-significant, medium effect of condition on teacher evaluation by students was found. Results imply that even though students seemed to benefit from need-support, the difference the training program made in teachers' motivating style was too small to have noticeably affected students. The study contributes to a growing body of literature on how to apply need-supportive teaching in practice.

*Keywords:* self-determination theory, intervention, motivation, need-support, statistics education

Effects of a Self-Determination Theory Based Intervention in University Statistics

Education

The first author observed, during his job as a teacher of statistics for psychology students, that many of his students had negative associations with statistics. They found it boring, useless, or very difficult; they were unsure of themselves; some students even were afraid of statistics or hated statistics. These observations are supported by other researchers. Dillon (1982) asked in a first lecture in a statistics course what students feel when thinking about taking the course and when looking at a statistical equation. With the exception of a few students, like those with an engineering background, most students respond negatively, feeling unsure, nauseous, panicky, uneasy, sick, worried, and so on. Ruggeri, Dempster, Hanna, and Cleary (2008) used focus groups of first- and second year psychology students to qualitatively describe what their feelings and impressions towards statistics education were. The students reported negative experiences with seeking help from teaching staff, a failure to see value in statistics, a lack of understanding and statistical literacy, negative affect towards taking classes, anxiety towards using computer instruments, anxiety towards statistics as a subject, low confidence in success, and a lack of awareness about the role of statistics in psychology at the start of their study. In sum, students' overall experience of statistics seemed pretty negative.

In these observations we saw an opportunity for improvement. We designed and implemented an intervention aiming to improve students' experience of statistics. We based our intervention on self-determination theory (SDT; Ryan & Deci, 2000), a theory about motivation. This theory has been applied effectively across a range of fields such as education, parenting, and managing to increase quality of motivation and improve a wide range of outcomes. This is the first time the theory has been applied to university statistics education: Its application to university setting is limited, and to date no SDT-based interventions applied specifically to university statistics education have been reported.

There have been several interventions focused on statistics education that were not

based on SDT. For example, Wiberg (2009) made extensive changes to a statistics course in an attempt to improve student outcomes. A textbook was added, an online earning environment was created, and teaching methods were revised to be more data-driven and student-centered. In the revised course, compared to the regular course, students performed better and had more positive attitudes towards statistics. Carnell (2008) added a student-designed data-collection project to a statistics course, but found no significant improvement in attitudes towards statistics. These interventions generally had far-reaching consequences for the statistics courses they were implemented in, which will likely deter course organizers from implementing them. In this study, we focused on a less invasive approach instead. We focused solely on teacher behavior during tutorials, leaving most of the course unchanged. We expected that teacher behavior would be a good target for improving statistics students' outcomes. This expectation is supported Ruggeri et al. (2008), who found that statistics students report that their instructor is the most important factor in determining their attitudes towards statistics, and by A. S. Williams (2010) , who found that statistics teachers' display of immediacy reduced students' anxiety towards statistics. We hoped that by focusing on changing teacher behavior only, we could improve statistics education without making such radical alterations as described in the interventions by Carnell (2008) and Wiberg (2009).

**Self-Determination Theory**

The theoretical framework underlying the design of our intervention is self-determination theory. SDT has been described extensively in literature (e.g. Deci & Ryan, 2002; Ryan & Deci, 2000); here we summarize its main points relevant to this study. In the theory two kinds of motivation are distinguished: autonomous and controlled motivation. Autonomously motivated people act with an internal perceived locus of causality: They perceive they act from within. They act because they find it enjoyable (*intrinsic regulation*), value the outcomes (*identified regulation*), or because it is consistent

with their personal values (*integrated regulation*). Controlled motivated people act with an external perceived locus of causality: They perceive they act because of external reasons. They act because of external pressure or incentives (*external regulation*) or because of internal pressure such as guilt, shame, or ego-involvement (*introjected regulation*).

The theory proposes that when the environment is supportive of three basic psychological needs, people will become, through a process called internalization, more autonomously and less controlled motivated. These needs are the needs for autonomy, competence, and relatedness. The need for autonomy is met when people perceive they are the origin of behavior: when they perceive a lack of pressure and a sense of volition and choice. The need for competence is met when people feel effective in their actions and have opportunities to express their capacities. It is feeling competent, rather than being competent. And finally, the need for relatedness is met when people feel connected to others, have a sense of belongingness to others, and care for and are being cared for by others.

**Self-Determination Theory Applied to Educational Settings**

We focused on SDT because of the range of benefits students experience when they are autonomously motivated. They perform better (Tessier, Sarrazin, & Ntoumanis, 2010; Vansteenkiste, Simons, Lens, Sheldon, & Deci, 2004), are more engaged (Tessier et al., 2010), are less anxious (Black & Deci, 2000), are more likely to continue in the subject they are autonomously motivated for (G. C. Williams, Saizow, Ross, & Deci, 1997), perceive themselves as more competent (Black & Deci, 2000), process materials more deeply (Vansteenkiste et al., 2004), and are less orentied on grades and more on learning (Black & Deci, 2000).

The educational field has shown considerable interest in SDT (Reeve, 2002); numerous SDT-based interventions have been developed that aimed to capitalize on these beneficial effects by increasing students' autonomous motivation (e.g. Cheon, Reeve, &

Moon, 2012). In most of these studies the teaching style of teachers was targeted. Teachers were trained to become more *need-supportive*: a way of teaching that support students' need for autonomy, competence, and relatedness, which promotes internalization and autonomous motivation (Tessier et al., 2010). These types of studies showed that it is possible to train teachers to become more need-supportive. For example, in an early study by Reeve (1998), teachers that were given an informational booklet about autonomy-supportive teaching later indicated that they used a more autonomy-supportive motivating style, compared to control. Teachers also tend to retain need-supportive motivating styles over longer periods of time. Cheon and Reeve (2013) followed-up their 2012 study, in which they trained teachers to become more need-supportive, and found that even after not receiving any further training for one year, the teachers still showed more need-support compared to control. What also has become clear is that students benefit from teachers' use of need-support. For example, Cheon et al. (2012) found that students of teachers who followed a training program in autonomy support were more engaged, developed more skills, had more positive future intentions, performed better, were more autonomously motivated and were less amotivated compared to control. All these results were promising: There seemed to be a style of teaching that teachers could be trained in and that is beneficial for students. In our intervention we therefore chose to train teachers in need-supportive teaching; this is the first time need-supportive teaching has been applied in statistics education.

While supporting the need for autonomy has received most attention, recently researchers have begun integrating support for all three needs in teaching-style interventions (e.q. Aelterman, Vansteenkiste, Van den Berghe, De Meyer, & Haerens, 2014; Tessier et al., 2010). In our intervention, we followed this trend, and focused on training teachers to support all three needs. In the following three sections, we describe which teacher behaviors we targeted in order to support the three needs and what studies they were based on.

**Satisfying the Need for Autonomy: Autonomy-Supportive Teaching**

A number of studies have focused on how students' need for autonomy can be supported. In an influential study by Deci, Eghrari, Patrick, and Leone (1994), subjects worked on a boring computer task. In a 2x2x2 design, the experimenter orally varied the presence of three contextual factors that were hypothesized to be autonomy-supportive: a rationale for doing the activity (by saying that this task can improve concentration), acknowledgement of possible negative feelings of the subjects (by saying that it is OK if subjects find the task boring, uninteresting and not fun), and wording of instructions in a non-controlling way (by conveying choice and volition). Afterwards they measured the time subjects voluntarily spent working on the boring computer task when the experiment was over (free-choice engagement time) and self-report measures of perceived choice, usefulness, interest and enjoyment. They found that the three factors increased total motivation of the subjects to work on the task: They increased free-choice engagement time. By looking at the correlation between free-choice engagement time and the self-report measures, they found that the three contextual factors contributed to internalization of the behavior; when two or three contextual factors were present (opposed to zero or one), free-choice engagement time was positively correlated with the self-report measures, indicating that the subjects were more autonomously motivated (they felt a sense of choice, usefulness, interest and enjoyment when acting). With only zero or one contextual factors present, this correlation was negative, indicating a more controlled form of motivation. The three autonomy-supportive contextual factors have since been used in many educational interventions (e.g. Edmunds, Ntoumanis, & Duda, 2008; McLachlan & Hagger, 2010).

A set of studies focused on identifying what autonomy-supportive teachers do differently compared to controlling teachers. Deci, Spiegel, Ryan, Koestner, and Kauffman (1982) and Flink, Boggiano, and Barrett (1990) manipulated the context of a teaching task to pressure teachers into becoming either more or less controlling and used raters to measure what the teachers were doing. Reeve, Bolt, and Cai (1999) identified

autonomy-supportive teachers by their score on the Problems In School Questionnaire and correlated that score with a set of hypothesized autonomy-supportive behaviors. Reeve and Jang (2006) measured perceived autonomy in students and correlated that with a set of hypothesized autonomy-supportive behaviors of their teachers. When taken together, these studies show that autonomy-supportive teachers (a) are less directly controlling students behavior (they use less directives and commands; use less should, must, or have to statements; ask less controlling questions; and seem less demanding and controlling), (b) stimulate independent thinking in students (they give less solutions, allow more time for independent work, hold instructional materials less, listen more, and are more responsive to student-generated questions), (c) are interpersonally supportive (they criticize less, use more encouragements, use more emphatic-perspective taking statements, and give more self-disclosure statements), (d) provide students with organizational and procedural control (by asking more questions of what the student wants), and (e) support intrinsic motivation and internalization. Although not all behaviors reached significance in all studies, and there were some conflicting findings regarding the use of hints and praise, the convergence of these three studies is impressive and imply that these behaviors are strongly supported.

Another line of researchers approached autonomy-support from a different angle. Stefanou, Perencevich, DiCintio, and Turner (2004) made an important distinction between three different types of autonomy support: organizational, procedural, and cognitive autonomy support. Organizational and procedural autonomy support aim at student ownership of *behavior*, while cognitive autonomy support aims at student ownership of *learning*. Examples of organizational autonomy support are allowing students to choose group members, choose seating arrangements, and create classroom rules. Examples of procedural autonomy support are allowing students to choose between projects, handle materials, and discuss their wants. Examples of cognitive autonomy support are allowing students to find and justify their own solution, re-evaluate their own errors, and talk more instead of listening to the teacher. Stefanou et al. (2004) argued that it is cognitive

autonomy support that truly leads to autonomous motivation for learning in students. While organizational autonomy support may make students more comfortable with the environment of the classroom, and procedural autonomy support may generate initial engagement with learning tasks, cognitive autonomy support may make learning intrinsically rewarding to students, encouraging psychological investment and deep-level thinking. They supported this proposition by a qualitative description of four lessons given by four different fifth- and sixth grade teachers. Only when cognitive autonomy support was high, students where actively engaged with the lessons. In contrast, when only organizational and procedural autonomy support were high, students showed little sign of engagement.

Based on the description of cognitive autonomy support, it might seem like it is another form of minimally guided instruction. Minimally guided instruction comes in many forms, such as constructivist, discovery, problem-based, experiential, and inquiry-based teaching. These approaches are characterized by minimal provision of information: "...learners, rather than being presented with essential information, must discover or construct essential information for themselves" (Kirschner, Sweller, & Clark, 2006, p. 1). Kirschner et al. (2006) gave sharp critique on this approach by arguing that this search for information is too cognitively taxing, hindering effective learning. We believe however that cognitive autonomy support is different from minimally guided instruction, since it's not characterized by withholding information. For example, a cognitive autonomy-supportive teacher can provide information by directing students to written materials, or by providing informational feedback (e.g. "this is correct"). Instead, cognitive autonomy supports' focus is about shifting the perceived locus of causality, when it comes to learning, from teacher to student. An illustrative example of this is the role of explaining. Intuitively, explaining seems like a good thing for teachers to do. A student is having problems solving an assignment, and the teacher comes and explains how the assignment should be solved. However, even though the student might come to understand the solution this way, it

might also reinforce the dynamic that the student is dependent on the teacher to solve assignments. In contrast, when the teacher directs the student to written materials instead, or guides the student into solving the assignment themselves, the dynamic of the student as independent problem-solver is likely to be reinforced.

## Satisfying the Need for Competence: Provision of Structure

Support for students' need for competence in need-supportive teaching is relatively recent, compared to autonomy support, which has received most attention. In the SDT-based educational intervention literature, the need for competence is generally targeted by provision of *structure* (e.g. Aelterman et al., 2014; Tessier et al., 2010). Provision of structure in this context is defined as setting clear and consistent goals, rules, and expectations; using encouragements; providing optimally challenging learning activities that are not too hard and not too easy; and providing feedback that is informative, positive and skill-building, instead of criticizing and competence-thwarting (see also Reeve, Jang, Carrell, Jeon, & Barch, 2004).

At first sight it might seem that provision of structure is in conflict with providing autonomy. However, high provision of structure does not automatically mean low autonomy-support: They are conceptually different and seen as complementary to each other (Jang, Reeve, & Deci, 2010; Vansteenkiste et al., 2012). Vansteenkiste et al. (2012) found that learning outcomes were the most positive with teachers who provided both autonomy-support and structure, compared to providing only one of the two or none at all. Jang et al. (2010) found a positive correlation between autonomy-support and provision of structure in teachers and found that both positively predicted student engagement.

Teachers are able to learn to provide more structure through training. For example, Aelterman et al. (2014) trained teachers for six hours in becoming more autonomy-supportive and a better provider of structure and found an increase in use of those two types of behaviors compared to a control group.

**Satisfying the Need for Relatedness: Becoming Involved**

SDT-based educational interventions that target student need for relatedness have been relatively rare; exceptions are studies by Tessier et al. (2010) and Edmunds et al. (2008). In these studies, need for relatedness is targeted by increasing teacher *involvement.* Involvement in this context is defined as having a high-quality interpersonal relationship with students, created by spending time, energy, and resources on students; knowing students' names and personal histories; being physically close to students; having a sympathetic, warm, affectionate, and caring attitude; and showing enthusiasm and humor. Its opposite is being hostile, neglectful, cold, distant, sarcastic, judgmental, and strict.

Tessier et al. (2010) trained three physical education teachers in being more autonomy-supportive, more involved, and a better provider of structure. The training consisted of a four hour informational session supplemented by three moments of individualized guidance. After training, teachers showed an increase in autonomy-support, involvement and structure. In the study by Edmunds et al. (2008), an exercise instructor implemented need-supportive behaviors in an intervention class and instructed as usual in a control class. Over the course of 10 weeks, exercise participants in the intervention class perceived a larger linear increase in autonomy support, structure and involvement over time than those in the control group, confirming the experimental manipulation. They also reported a larger linear increase in relatedness and competence need satisfaction, a larger increase in positive affect, and better attendance. Taylor and Ntoumanis (2007) showed that when teachers provide involvement, student need for relatedness is satisfied and students show higher levels of autonomous motivation. Taken together, these studies show that involvement is trainable and is beneficial for students.

**Training Teachers to Become More Need-Supportive**

The educational techniques described above are the contents of need-supportive teaching. This answers the question of what a need-supportive teaching style *is*, but not

*how it can be trained.* Several studies help answering this second question. In an effort to find what makes an intervention program to increase teachers' autonomy-support effective, Su and Reeve (2011) performed a meta-analysis on 19 studies that reported the results of such an intervention. Even though studies included in this review focused mainly on autonomy-support, it seems likely that the same principles are important when designing an intervention that targets support of all three needs. They found that the more effective interventions (a) had medium-long training lengths (for example, a 90 minute training plus independent study of reading materials), (b) focused more on building skills instead of knowledge, (c) used both electronic media and reading materials, and (d) focused on multiple elements of autonomy-support, with non-controlling language having the largest effect and offering choices the lowest. We applied these findings in the design of our intervention.

Other findings about how to effectively train teachers to become more need-supportive come from Gorozidis and Papaioannou (2014) and Lam, Cheng, and Choy (2010). They argued that in the same way that autonomous motivation is important for student learning, it is also important for teacher learning. Gorozidis and Papaioannou (2014) found that teachers who were autonomously motivated towards being trained in and implementing an educational innovation had stronger intentions to do so, while controlled motivated teachers had not. Lam et al. (2010) found that when the school environment supported the three basic needs of competence, relatedness, and autonomy for teachers, they were more autonomously motivated to implement an educational innovation and to persist in it. This has important implications for designing an effective educational intervention. An effective intervention should be designed in such a way to support autonomous motivation of the teachers. This means that training should be given in a need-supportive way, using the same principles as used when supporting the needs of students, which we did in our intervention.

**Need-Support in University Settings**

Since our intervention was applied in an university course, we also looked a at findings from previous SDT-based interventions at universities. While SDT-based interventions to increase students' autonomous motivation have been applied numerous times in high school and elementary school contexts, its application in university-level education is relatively rare. This is unfortunate, since autonomous motivation is related to positive emotions and higher academic achievement for university students (González, Paoloni, Donolo, & Rinaudo, 2012; Hill, 2013).

In an attempt to fill the void of university-level SDT-based interventions, McLachlan and Hagger (2010) gave a brief training in autonomy-supportiveness to postgraduate university tutors. The study was moderately successful: The tutors showed significant improvements in two of the 14 targeted autonomy-supportive behaviors. As the authors noted, this study suffered from a few methodological limitations: (a) Their small sample size of nine tutors likely contributed to the lack of significance in many of the targeted behaviors, however, this sample size is fairly typical in SDT-based educational interventions (see Su & Reeve, 2011). (b) Little time was given to the tutors to implement the autonomy-supportive behaviors (only two weeks). More effective interventions allowed teachers more time to develop autonomy-supportive behaviors (Su & Reeve, 2011). (c) The training itself was very brief, taking only 40 minutes. As shown by Su and Reeve (2011), longer training times tend to be more effective. (d) There was a lack of continuous support and follow-up activities. As shown by Su and Reeve (2011), most effective interventions include some form of continuous support and follow-up activities. (e) Tutors were asked to implement all 14 behaviors at once, possibly overloading the tutors. (f) Need for competence and need for relatedness were not directly targeted. In addition, the effects on students were not measured.

One goal of this study was to expand on this study by McLachlan and Hagger (2010) on whether it is possible to train university tutors to become more need-supportive and

address many of its methodological limitations.

**Research Questions and Hypotheses**

The aim of this study was to answer three related questions: (a) Can a training program make university statistics teachers more need-supportive? (b) Does the training program improve students' educational outcomes? (c) Do teachers following the training program benefit?

To answer these questions, we designed a training program that aimed to increase the use of need-supportive behaviors by university teachers. The behaviors targeted in this program are based on the research on what constitutes a need-supportive teaching style, as described above. Specifically, targeted autonomy-supportive behaviors were "acknowledging negative affect", "providing rationales for requests", "avoiding controlling behavior", and "providing cognitive autonomy support" (Deci et al., 1994, 1982; Flink et al., 1990; Reeve et al., 1999; Reeve & Jang, 2006; Stefanou et al., 2004). The targeted competence-supportive behavior was "providing positive feedback", which is an aspect of provision of structure (Vansteenkiste et al., 2012). We chose to target only this single aspect of structure, rather than the full range of structuring behaviors, because other aspects were either beyond reach of the teachers to implement (e.g. providing optimally challenging activities) or were already part of regular teacher training, unrelated to this study (e.g. setting clear goals and expectations). Targeted relatedness-supporting behaviors were "spending time, energy and attention on students", "knowing students well", "having a warm, open attitude", and "getting physically close", based on the definition of involvement by Reeve et al. (2004). The full list of targeted need-supportive behaviors can be found in table 1.

This list of behaviors is not meant as an exhaustive list of need-supportive behaviors. Rather, we focused only on behaviors that we deemed practical for the teachers to implement. In the targeted statistics course, all teachers had to use the same teaching

materials and lesson structure. This limited the range of need-supportive behaviors that could be implemented. The behaviors we targeted are minimally-invasive: They can be implemented without making any changes to teaching methodology (e.g. which assignments are used, how students are graded or how lessons are organized).

We hypothesized that (a) after training, teachers would increase usage of the targeted need-supportive behaviors; (b) this would improve students' educational outcomes, specifically it would improve students' quality of motivation for statistics, students' classroom engagement, and students' academic performance; and (c) teachers would appreciate the training program and benefit by being evaluated better by their students.

## Methods

### Participants

**Teacher participants.** We focused on an introductory statistics course given to bachelor students of psychology. This course was given by one main lecturer, who also organized the course, and 10 supporting teachers, who led tutorials. The first author was one of the supporting teachers. We invited all nine remaining supporting teachers to participate. Eight teachers agreed to participate (six female, two male; age: $M = 28.4$, $SD = 5.85$, range = 23-41 years). Six teachers were relatively inexperienced, having zero to two years of teaching experience, and two teachers were relatively experienced, having eight and 16 years of teaching experience. The teachers gave two to three tutorials per week (class size: $M = 21.7$, $SD = 4.63$, range = 7-24 students).

**Student participants.** Student participants were students enrolled in the course and following the tutorials of the participating teachers. They were invited to participate in the study at the beginning of the first tutorial at the start of the course, during the first wave of data collection. The response rate was high. There were 409 students enrolled for the tutorials and 355 students agreed to participate, which is a response rate of 86.8%. This number is not entirely accurate, since in practice the actual number of students present at

the start of a course tends to slightly deviate from the number of enrolled students, however
it is safe to say that almost all students agreed to participate when they were asked.

We performed a second wave of data collection during the last tutorial at the end of
the course. During that wave, 335 students filled in our questionnaire. This sample
consisted of 312 original students, who also participated in the first wave, and 23 new
students, who missed the first wave but agreed to participate during the second wave (a
retention rate of $312/355 = 87.9\%$). We dropped all data from students who switched
between tutorials between the first and second wave. This was the case for 3 students,
leading to a final sample size at the second wave of 332 students.

The sample of students at the first wave consisted of 270 females and 85 males with
mean age of 19.6 ($SD = 3.47$, range $= 16$-$55$). The sample consisted of three distinct types
of students: 249 regular Dutch students, 61 international students, and 45 "pre-master"
students. Regular Dutch students followed the course as part of a Dutch bachelor in
psychology, international students as part of an international bachelor in psychology, and
pre-master students as an entrance requirement for their master. The sample of students at
the second wave consisted of 240 females and 70 males, with mean age of 19.7 ($SD = 3.04$,
range $= 16$-$43$), of which 223 were regular Dutch students, 54 were international students,
and 45 were pre-master students.

**Procedure**

Two weeks before the start of the course, the teachers were invited to participate in
the study. The teachers were told about the general outline of the study (e.g. the schedule,
the investment of time required, and that it entailed an informational workshop on
educational techniques) without anything being mentioned about the contents of the
intervention (e.g. motivation, SDT, or need-supportive behaviors). This was done to
prevent leaking contents of the intervention to the control group.

One participating teacher was unable to follow the training program due to scheduling

issues and was assigned to the control group. The two male teachers were distributed equally over the intervention and control condition. The remaining teachers were assigned to the conditions at random. Teachers in the intervention condition participated in a program to train them in becoming more need-supportive towards their students (described below). Teachers in the control condition delivered their lessons as usual.

**Need-Supportive Training Program**

The training program consisted of a workshop and two supplemental meetings. It ran during the introductory statistics course, which took nine weeks and consisted of weekly lectures followed by two hour tutorials a few days later. The training program started with a workshop, given about one week before the first lecture. About one week after the workshop, teachers met their students for the first time during the first tutorials. At the start of these first tutorials (T1) they distributed the first wave of questionnaires. About one week later, when all teachers had given their tutorials at least once, the teachers in the intervention condition met for the first supplemental meeting. About three weeks after that, those teachers met again for the second supplemental meeting. During the last week of the course (T2), the teachers distributed the second wave of questionnaires. After the course was over, the teachers in the intervention condition completed a questionnaire about the training program. See figure 1 for the timeline of the program.

The workshop took about one hour and 10 minutes. It started with a 30-minute PowerPoint presentation in which we explained to the teachers the difference between autonomous and controlled motivation, the advantages of autonomous motivation for students, that satisfaction of the three basic psychological needs promotes autonomous motivation, and what need-supportive teaching is and how it can be implemented in practice. This was the starting point for a group discussion, which lasted about 30 minutes. The first goal of the group discussion was to give the teachers a chance to ask questions and voice their doubts and concerns. The second goal was to involve the teachers

into the practical implementation of the need-supportive techniques. The teachers were stimulated to give feedback on the practicality of the techniques, share any obstacles they might see, suggest improvements, and suggest additional ways of implementing them. After the group discussion, 10 minutes were spent on closing the workshop. The teachers were invited to try a few techniques during their first lesson and to reflect on it. To provide structure for this reflection, we gave forms to the teachers with three questions: "What did you do?", "How was it for you?", and "How did the student(s) react?". The teachers were encouraged to fill in at least three of those forms before the first supplemental meeting. Then, teachers were handed a booklet containing all information presented in the workshop. Finally, the teachers were asked not to discuss any of the contents of the training program with either the teachers in the control group or with students. All materials used in the workshop can be found in the appendix.

A week after the workshop, the teachers attended a 45-minute supplemental meeting. At this point the teachers had given their tutorials two to three times to students and had a chance to try some need-supportive techniques and reflect on it. The purpose of the supplemental meetings was to provide expert guidance and collegial support. First, five minutes were spent on presenting the goals of the meeting and creating a safe environment for the tutors. The teachers were encouraged to be honest about their experiences and opinions and were reassured that they would not be judged on their performance. Then, 35 minutes were spent on teachers sharing their experiences with implementing need-supportive behavior. We noticed that the reflection forms were barely used for this and the teachers reflected from memory instead. Since the supplemental meetings were held shortly after the lessons, the teachers' memories were likely still accurate and we think this had little effect on the training program. During reflection, expert support was given by the first author and collegial support was stimulated, for example by stimulating that teachers share ideas and suggestions for improvement with each other. Finally, five minutes were spent on closing the meeting and encouraging the teachers to keep trying to use more

need-supportive behaviors. This supplemental meeting was repeated after three weeks.

After the course, we thanked the teachers for their participation and presented the results of the study. We presented the results anonymously, however, upon teacher request, they could view class-level results of the tutorials that they taught. At this moment we offered the workshop again to the teachers in the control condition, to avoid creating an unfair difference between teachers. We thanked and debriefed all students as well through e-mail.

We presented the whole training program in a need-supportive manner. We satisfied teachers' need for autonomy by not demanding that teachers use specific behavior, but instead explaining the value of the educational techniques and suggesting that teachers adapt the techniques to their own educational style. We also involved the teachers in the whole process and gave them opportunities to voice negative affect. We met teachers' need for competence by provision of structure from the trainer and by continuous expert guidance through the supplemental meetings. Finally, we met teachers' need for relatedness by involvement from the trainer and by collegial support through the supplemental meetings. Overall, we framed the study as an opportunity for gaining valuable information and self-reflection, instead of as a training aimed at changing their teaching practices.

**Measures**

To assess whether the training program successfully made the teachers more need-supportive, we measured students' perceived need-support. To asses whether the students benefited from the program, we measured students' quality of motivation, academic performance and classroom engagement. And finally, to assess whether the teachers benefited from the program, we measured teachers' appreciation of the training program and students' evaluation of their teacher. Only quality of motivation was measured at both T1 and T2, while the other variables were measured only at T2, since only for quality of motivation it makes sense to measure it before the course has started

(T1).

The first author translated all questionnaires from English to Dutch. The second author, unaware of the English form of the questionnaires, translated the Dutch questionnaires back to English. The two English versions were compared and any discrepancies were resolved.

Originally, we intended to measure students' attitudes towards statistics at T1 and T2. We included the Survey of Attitudes Towards Statistics (SATS; Schau, Stevens, Dauphinee, & Vecchio, 1995) in the T1 questionnaire, however, when designing the T2 questionnaire, we decided that the 28 items of the SATS would make the questionnaire too long. We therefore did not include the SATS in the T2 questionnaire and did not analyze the T1 results of the SATS.

Some of the scales we used (e.g. the Perceived Locus of Causality scale) consisted of a number of subscales. If we had analyzed each subscale separately, there would be 12 student-assessed dependent variables. Analyzing this number of dependent variables would be complex and difficult to interpret. To reduce this complexity, we combined some subscales into one variable. In this section we also provide some justification for combining scales.

**Students' perceived need-support.** There are three ways to measure the provision of need-support by teachers: by teacher self-reports, by student reports, and by observations by trained raters. Each method has its up- and downsides and their results do not always converge (Aelterman et al., 2014). We chose to use student reports after weighing the pros and cons. Specifically, in-class observations were deemed to disruptive to the normal flow of the tutorials and teacher self-reports were deemed too sensitive to bias such as demand characteristics and social desirability. The advantage of student reports over in-class observations is that students experience their teacher over longer periods of time, while in-class observations are more similar to a snapshot. Students reports also make it possible to analyze perceived need-support at the student-level.

Since we did not find a measure of perceived need-support in literature that satisfactorily matched our set of targeted need-supportive behaviors, we developed a new nine-item measure. This measure is directly based on the behaviors targeted in our intervention. For each targeted behavior, students were asked to rate whether it is true that their teacher uses that behavior. Example items are: "Our teacher understands it when we find something difficult or boring", "Our teacher explains the value of what we do in class", and "Our teacher knows us well". To make the items more clear, we gave a few examples along with each behavior. The items were answered on a 5-point Likert-type scale (1 = *not at all true* and 5 = *totally true*).

We performed a principal component analysis on the nine perceived need-support items. A one-component solution emerged[1] (eigenvalue = 3.39, 38% of the total variance, all loadings positive, loadings: $M$ = .60, $SD$ = .15, range = .32-.74). We also performed a confirmatory factor analysis on these items. A one-factor model fit the data reasonably well, $\chi^2(152)$ = 62.5, $p <$ .001, CFI = .94, NFI = .91, RMSEA = .063, SRMR = .047. Both findings support that the nine items measure one construct. We therefore formed a perceived-needsupport scale by taking the mean of the nine items. The scale showed adequate internal consistency ($\alpha$ = .78).

**Students' quality of motivation.** To assess students' quality of motivation, we used the Perceived Locus of Causality for Physical Education scale (Goudas, Biddle, & Fox, 1994; Ryan & Connell, 1989; Standage, Duda, & Ntoumanis, 2006). We adapted the items slightly to conform to the context of statistics education (e.g. "I take part in this PE class..." became "I worked on this statistics course..."). The scale consists of 20 items, which all start with a stem (e.g. "I worked on this statistics course...") and end in a reason for working on a statistics course. These reasons are based on the different types of motivational regulation. The scale has five 4-item subscales: intrinsic motivation (e.g. "...because statistics if fun"), identified regulation (e.g. "...because I want to learn

---

[1] The eigenvalue of the second component was 1.10

statistics skills"), introjected regulation (e.g. "…because I want the teacher to think I'm a good student"), external regulation (e.g. "…because I'll get into trouble if I don't"), and amotivation (e.g. "…but I don't really know why"). Cronbach's alpha coefficients of the subscales were .87, .67, .69, .69, and .82, respectively, at T1, and .89, .79, .73, .71, and .83, respectively, at T2. The items were answered on a 7-point Likert-type scale (1 = *strongly disagree* and 7 = *strongly agree*).

The Perceived Locus of Causality scale consists of five subscales. Some researchers combined four of the five subscales into one scale, the relative autonomy index, using the formula "2*intrinsic motivation + identified regulation - introjected regulation - 2*external regulation" (e.g. Cheon et al., 2012; Standage et al., 2006; Taylor & Ntoumanis, 2007). However, we found some results contradicting the use of this formula. First, when inspecting the correlation matrix of the four subscales, we found positive correlations between introjected regulation and intrinsic motivation ($r = .25$, $p < .001$) and between introjected regulation and identified regulation ($r = .29$, $p < .001$). According to the formula these should be negative. Second, a one-factor model, applied to the items forming the four subscales and with item parameters fixed according to the weights in the formula (e.g. parameters of the intrinsic motivation items fixed at two) fit the data very poorly, $\chi^2(119) = 2810$, $p < .001$, CFI = .32, NFI = .31, RMSEA = .18, SRMR = .22. This indicates that the formula does not fit the data well.

Other researchers combined the intrinsic motivation and identified regulation subscales into one autonomous motivation scale and the introjected regulation and extrinsic regulation subscales into one controlled motivation scale (e.g. Haerens, Aelterman, Vansteenkiste, Soenens, & Van Petegem, 2015; Stroet, Opdenakker, & Minnaert, 2015). We assessed whether this method was appropiate. A two-factor model, with items from the intrinsic motivation and identified regulation subscales loading on one factor and items from the introjected regulation and extrinsic regulation subscales loading on another factor, and with all item parameters fixed at one, fit the data poorly, $\chi^2(117) = 1571$, $p < .001$,

CFI $= .63$, NFI $= .61$, RMSEA $= .13$, SRMR $= .14$. This indicates that the method of forming two scales out of the four subscales does not fit the data well. The model did, however, fit significantly better than the one-factor model, $\chi^2(2) = 1240$, $p < .001$.

We decided to follow the second line of researchers and form autonomous motivation and controlled motivation scales. Even though some information is lost using this approach, this loss is less compared to the first method, while still reducing the number of dependent variables by two, and thus reducing the complexity of the analysis. We considered this to be a good trade-off.

**Students' academic performance.**    We used the students' score on the final course exam as a measure of academic performance. This score ranges from 1 to 10.

**Students' classroom engagement.**    To measure students' classroom engagement, we used a method that was used successfully before in a similar intervention study by Cheon et al. (2012). They conceptualized classroom engagement as consisting of four dimensions: *behavioral engagement* (whether students pay attention, show persistence and exert effort), *emotional engagement* (whether students feel good in class), *cognitive engagement* (whether students use advanced learning strategies in class), and *agentic engagement* (whether students take initiative and communicate desires and opinions to their teacher). To measure these four dimensions, we follow Cheon et al. (2012). We assessed behavioral and emotional engagement using the Engagement Versus Disaffection With Learning scale (Skinner, Kindermann, & Furrer, 2009). This scale consists of 10 items, five of which measure behavioral engagement (e.g. "I try hard to do well in this class"; $\alpha = .73$) and five of which measure emotional engagement (e.g. "When I'm in this class, I feel good"; $\alpha = .77$). These 10 items were answered on a 5-point Likert-type scale (1 $= not\ at\ all\ true$ and 5 $= totally\ true$). We assessed cognitive engagement using the Metacognitive Strategies Questionnaire (Wolters, 2004). This scale consists of four items (e.g. "When I study for this class, I try to connect what I am learning with my own experiences"; $\alpha = .68$). Finally, we assessed agentic engagement using the Agentic

Engagement scale (Reeve, 2013). This scale consists of 4 items (e.g. "I let my teacher know what I need and want"; $\alpha$ = .76). The items of the cognitive engagement and agentic engagement subscales were answered on a 5-point Likert-type scale (1 = *strongly disagree* and 5 = *strongly agree*).

In previous studies, researchers combined the four types of engagement into one engagement scale (e.g. Cheon & Reeve, 2015; Reeve et al., 2014; Tessier et al., 2010). We inspected the validity of this approach. The four engagement types were positively intercorrelated (all $r$s > .20, all $p$s < .001). This indicates that they measure a related construct, which is supportive of the approach. However, a one-factor model, with all engagement items loading on one factor, fit the data poorly, $\chi^2(152) = 1169$, $p < .001$, CFI = .48, NFI = .45, RMSEA = .14, SRMR = .12. A four-factor model, in which items of each engagement type loaded onto a separate factor, fit the data significantly better $\chi^2(6)$ = 615, $p < .001$, however, it also did not fit the data very well, $\chi^2(146) = 554$, $p < .001$, CFI = .79, NFI = .74, RMSEA = .092, SRMR = .096.

We chose to follow the previous studies and combine the scores of the four different engagement types into one engagement scale. We think this method is acceptable, since the four types were positively correlated, however, as the confirmatory factor analyses indicated, we note that some information is lost using this approach.

**Teachers' evaluations by students.** Evaluations of teachers by students were obtained from the regular course-evaluation, unrelated to this study, that happened at the end of the statistics course. After the final exam, students filled in an evaluation form, of which one question was: "What is your evaluation of your workgroup- or practical tutor?". This question was answered on a 5-point Likert-type scale (1 = *very poor* and 5 =*very good*). This evaluation form was filled in anonymously, so we could not link these results to individual students' T1 or T2 measurements. The form was filled in by 272 students.

**Teachers' appreciation of training program.** To evaluate whether teachers' appreciated the training program, we used the following method, which was used

successfully before in a similar intervention study (Cheon & Reeve, 2015). Teachers from the intervention group answered the following four items on a 7-point scale: (a) "Did your participation in the training program help produce a positive significant change in your classroom motivating style?" (1 = *not at all significant* and 7 = *extremely significant*); (b) "Was your participation in the training program important to you?" (1 = *not at all important* and 7 = *extremely important*); (c) "How satisfied with the training program were you?" (1 = *not at all satisfied* and 7 = *extremely satisfied*); and (d) "Was the training program useful to you?" (1 = *not at all useful* and 7 = *extremely useful*). They also answered the following open-ended question: "Were you satisfied with the training program overall? If so, why? If not, why not?".

## Results

### Preliminary Analyses

**Data preprocessing.**  Five students were substantially older (i.e. 35, 35, 37, 42, and 55 years old) compared to the rest of the students (age: $M = 19.5$, $SD = 2.20$ years). Students of this age typically are part-time students in a later phase in their career. Because of their high age, they were highly influential on the relationship of age with the other variables. To reduce this unwanted influence, we decided to recode their age to the maximum age of the remaining students, which was 29 years. This procedure is called *topcoding* (Gelman & Hill, 2007, Chapter 25).

**Missing values.**  There are different methods for dealing with missing values, ranging from simple (e.g. listwise deletion or mean substitution) to complex (e.g. multiple imputation; Schafer & Graham, 2002). In this study missing values were rare: Less than 1% of all possible responses at T1 and T2 were missing. In this case, the differences between methods of dealing with missing values are small (Saunders et al., 2006). Considering this, we chose a method that we deemed a good trade-off between complexity and accuracy.

We used iterative random regression imputation (Gelman & Hill, 2007) to impute

missing values. In brief, the method works as follows. First, a set of interrelated variables is identified. Missing values on these variables are imputed using a crude method (e.g. simple random imputation). Then, each variable in the set is regressed on the other variables in the set, and the resulting regression equations are used to predict and replace the crudely imputed missing values. These two steps (regression and replacement) are iterated until convergence. Finally, some random error is added to the imputed values, based on the residual variances of the regressions they were based on. This last step prevents artificial inflation of the associations between the variables in the set that otherwise would occur. For details about the implementation, see Gelman and Hill (2007, Chapter 25).

Random regression imputation has been shown to be unbiased when missing values are at least Missing At Random (MAR; Schafer & Graham, 2002). When data is MAR, the probability of obtaining a missing value depends only on observed values and not on any unobserved value of any measured variable. Part of our missing data was likely MAR. Some missing values at T1 were likely caused by students accidentally skipping questions, as there was only 0.15% missing values at the electronic questionnaire at T2 (which reminded participants of skipped questions) compared to the 1.64% of missing values at T1. It seems likely that this skipping occurred at random, unrelated to the variables of interest. Some of the missing values on age and gender were caused by an error during data collection: These two items were missing for 22 questionnaires, causing 22 of the 26 missing age values and 22 of the 22 missing gender values. One variable of which missing values likely do not meet this criteria and are thus Missing Not At Random (MNAR), is academic achievement. Missing values on academic achievement (1.93%) were caused by students not taking the final exam, which in general tend to be poor-performing students. Our main interest however is in the difference between the experimental and control group, and it is likely that this process affected both groups equally. Also, a chi-square test revealed that the percentage of missing values did not differ over conditions, $\chi^2(1, N = 332) = 0.12$, $p = .73$. In sum, we expect that our missing values are mostly at least MAR and that random

regression imputation provided mostly unbiased results.

The variables academic achievement and gender showed little association with other variables, which made iterative random regression imputation difficult. For these two variables, we chose to apply listwise deletion instead, and note that for these variables some bias has to be accepted.

**Translation of questionnaires.** To provide a rudimentary check of whether the scales were translated properly, we compared Cronbach's alphas of the Dutch scales to those of the original English scales. The differences in internal consistency were small (less than 3%) and the Cronbach's alphas of the Dutch versions were within the 95% confidence intervals of those of the English versions (see table 2).

## Multi-level modeling

The context in which the data was collected naturally created a hierarchical data structure: Students were nested in classes and classes were nested in teachers. This structure created dependencies in the data. For example, scores of students in the same class were likely more similar to each other than scores of students of different classes. This causes error terms to be correlated, which violates the assumptions of many ordinary statistical methods. When this depencency is present, ordinary statistical methods that do not take this into account will underestimate standard errors and overestimate significance (Hox, 2002).

One often-used method that does take this dependency into account is multilevel modeling (Hox, 2002). This technique can be seen as an extension of classical multiple regression with explicit modeling of hierarchical structures in the data. This method has been applied extensively in educational research (e.g. Reeve et al., 2014; Taylor & Ntoumanis, 2007). We used multilevel modeling to test our hypotheses about student-assessed variables.

**Distribution of variance.**   A three-level hierarchical structure was present in all our student-assessed dependent variables (students nested within classes nested within teachers), except for quality of motivation, which was measured at T1 and T2 and thus formed a four-level hierarchical structure (waves nested within students nested within classes nested within teachers). To inspect how variance was distributed across these levels, we calculated the intraclass correlation (ICC) coefficient for each level. An ICC can be seen as the correlation of scores between members of a higher-level group (e.g. the correlation between grades of students within the same class) (Hox, 2002). High ICCs mean that most variance resides at the group-level, while low ICCs mean that most variance resides at the lower, individual level. The ICCs for the student-assessed variables can be seen in table 3.

As can be seen in table 3, the majority of variance resided at the student level. There was little variance at the class level, as the mean ICC at that level was 1.9%. There was a relatively large portion of variance at the teacher level for the more teacher-focused variables perceived need-support (ICC = 20.6%) and teacher evaluation (ICC = 12.6%). For the other, more student-focused variables, there was little variance at the teacher level (mean ICC = 0.7%). For some variables, variance at class or teacher level was almost non-existent (e.g. there was no class-level variance of academic achievement). Overall, these low ICCs indicate that the hierarchical structure was too complex for the data. In order to reduce the complexity of the multilevel models and make the results more meaningful, we decided to drop the class level from the analysis. This is akin to assuming that there are no between-classes effects within teachers. We think this is acceptable because the variance at class level was very low, and therefore the violation of the assumption of independent errors will be very small. The distribution of variance across the remaining levels is shown in table 4.

In each multi-level model, the intercept was allowed to vary over teachers. For quality of motivation, which was measured at two time points, the intercept was also allowed to vary over students. All other effects were fixed.

**Student demographics.**   Before conducting the main analyses, we checked for possible associations of student demographics with six student-assessed dependent measures. Teacher evaluation was not included in this analysis, since this data was obtained anonymously and therefore could not be linked to the covariates. Whenever a significant association between a covariate and a dependent variable was detected, that covariate was included in the multi-level model for that dependent variable.

Gender was associated with controlled motivation, with males ($M = 3.76$, $SD = 0.88$) having lower controlled motivation compared to females ($M = 4.13$, $SD = 0.90$), $t(259) = 4.57$, $p < .001$. Age was positively associated with engagement, $r(304) = .12$, $p = .038$ and autonomous motivation, $r(654) = .21$, $p < .001$, and negatively associated with controlled motivation, $r(654) = -.08$, and amotivation, $r(654) = -.09$, $p = .021$. Student type was associated with all student-assessed variables: Pre-master students were generally more engaged, more motivated, and performed better in the course and international students perceived more need-support in general compared to Dutch or pre-master students. For a detailed look at the effects of student type, see table 5.

**Assumptions.**   There are three assumptions underlying multi-level models (Singer & Willet, 2003). The first assumption is that residuals should be normally distributed at all levels of the model. We checked this assumption by visual inspection of QQ-plots. The second assumption is homoscedasticity: Equality of variance across all levels of all predictors. For categorical predictors, we checked this using Levene's tests, and for continuous predictors, we checked this using Breush-Pagan tests. The third assumption is that predictors and dependent variables should be linearly related, which we checked by visual inspection. There were no strong indications that the assumptions were violated for any model, except for a model that included amotivation. The distribution of amotivation was strongly positively skewed, leading to non-normal distribution of residuals. This was solved by log-transforming amotivation.

**Analyses of Hypotheses**

There is some discussion regarding which procedure is appropriate when testing fixed effects in a multi-level regression model (Hox, 2002, Chapter 3). In light of this discussion, we tested our fixed effects using two methods. The first method is a Wald-type test, which means that the test statistic is the estimate of the effect divided by the standard error of the estimate of the effect. The test statistic is assumed to be t-distributed, with degrees of freedom approximated by the Satterthwaite approximation. We refer to this method by the "Satterthwaite approximation". The second method is a likelihood ratio test. It is based on the difference in deviance between two models: one with the fixed effect included, and one without. This difference is assumed to be chi-square distributed. We refer to this method by LRT. In one simulation study, it was shown that both methods perform reasonably well, showing a small amount of inflation of Type-I error rate (Type-I error rate was generally between 0.05 and 0.075; Manor & Zucker, 2004). We used Full Maximum Likelihood estimation in all cases. All $p$-values are one-tailed, unless otherwise noted.

There is currently no consensus regarding which effect sizes are most appropriate to report in a multi-level context (Peugh, 2010). In light of this, we report a Cohen's $d$-type measure: The estimate of the fixed effect devided by the standard deviation in unstandardized scale units. We use the criteria by Cohen (1992) to get a rudimentary feel for the size of the effect, which is that 0.20 is a small effect, 0.50 is a medium effect, and 0.80 is a large effect.

**Perceived need-support.** Students in the intervention condition reported higher perceived need-support compared to students in the control condition. This was a medium sized effect ($\beta = 0.24$; scale: range = 1-5, $SD = 0.49$; $d = .49$). The effect was significant according to the Satterthwaite approximation, $t(7.51) = 1.91$, $p = .047$, and marginally significant according to the LRT, $\Delta\chi^2(\Delta1) = 2.65$, $p = .052$.

**Engagement.** Students in the intervention condition showed higher engagement compared to students in the control condition, however, this effect was very small ($\beta =$

0.042, scale: range = 1-5, $SD$ = 0.43, $d$ = 0.10) and nonsignificant, $t(301)$ = .232, $p$ = .23. Note that this variable was analyzed without using ordinary regression, without including the teacher-level, because the multi-level model had problems with the variance at teacher level being exactly zero for this variable.

**Academic achievement.**   Students in the intervention condition showed higher academic achievement compared to students in the control condition, however, this effect was very small (β = 0.029, scale: range = 1-10, $SD$ = 1.53, $d$ = 0.02) and nonsignificant according to both the Satterthwaite approximation, $t(10.7)$ = 0.15, $p$ = .44, and the LRT, $\Delta\chi^2(\Delta1)$ = 0.02, $p$ = .44.

**Teacher evaluation.**   Students in the intervention condition evaluated their teacher better compared to students in the control condition. This was a small to medium sized effect (β = 0.30; scale: range = 1-5, $SD$ = 0.82; $d$ = .37). The effect was nonsignificant according to the Satterthwaite approximation, $t(8.16)$ = 1.42, $p$ = .097, and nonsignificant according to the LRT, $\Delta\chi^2(\Delta1)$ = 1.79, $p$ = .091.

**Autonomous motivation.**   At T1, students in the intervention condition showed more autonomous motivation compared to students in the control condition, however, this effect was small (β = 0.078; scale: range = 1-7, $SD$ = 0.95; $d$ = 0.08) and nonsignificant according to both the Satterthwaite approximation, $t(472.3)$ = 0.74, $p$ = .460 (two-tailed), and the LRT, $\Delta\chi^2(\Delta1)$ = 0.547, $p$ = .460 (two-tailed). From T1 to T2, students generally increased in autonomous motivation. This effect was medium-sized (β = 0.21, $d$ = 0.22) and significant according to both the Satterthwaite approximation, $t(320.6)$ = 4.58, $p$ < .001 (two-tailed), and the LRT, $\Delta\chi^2(\Delta1)$ = 20.44, $p$ < .001 (two-tailed). Finally, the crucial wave*condition interaction showed that students in the intervention condition gained more autonomous motivation from T1 to T2 compared to students in the control condition, however, this effect was small (β = 0.12, $d$ = 0.12) and nonsignificant according to both the Satterthwaite approximation, $t(320.7)$ = 1.27, $p$ = .103 (one-tailed), and the LRT, $\Delta\chi^2(\Delta1)$ = 1.60, $p$ = .103 (one-tailed).

**Controlled motivation.**   At T1, students in the intervention condition showed more controlled motivation compared to students in the control condition, however, this effect was very small ($\beta$ = 0.017; scale: range = 1-7, $SD$ = 0.90; $d$ = 0.019) and nonsignificant according to both the Satterthwaite approximation, $t(472.6) = 0.16$, $p = 0.870$ (two-tailed), and the LRT, $\Delta\chi^2(\Delta 1) = 0.027$, $p = .871$ (two-tailed). From T1 to T2, students generally increased in controlled motivation, however, this effect was very small ($\beta$ = 0.045, $d$ = 0.050) and nonsignificant according to both the Satterthwaite approximation, $t(328.2) = 1.01$, $p = 0.315$ (two-tailed), and the LRT, $\Delta\chi^2(\Delta 1) = 1.01$, $p = 0.315$ (two-tailed). Finally, the crucial wave*condition interaction showed that students in the intervention condition unexpectedly gained more controlled motivation from T1 to T2 compared to students in the control condition, however, this effect was small ($\beta$ = 0.14, $d$ = 0.16) and nonsignificant according to both the Satterthwaite approximation, $t(328.2) = 1.58$, $p = .943$ (one-tailed), and the LRT, $\Delta\chi^2(\Delta 1) = 2.49$, $p = .943$ (one-tailed).

**Amotivation.**   At T1, students in the intervention condition showed more amotivation compared to students in the control condition, however, this effect was almost non-existent ($\beta$ < 0.001; scale: range = 1-7, $SD$ = 0.99; $d$ < 0.001) and nonsignificant according to both the Satterthwaite approximation, $t(532.1) = 0.01$, $p = 0.995$ (two-tailed), and the LRT, $\Delta\chi^2(\Delta 1) = 0.08$, $p = .779$ (two-tailed). From T1 to T2, students generally decreased in amotivation. This effect was small ($\beta$ = -0.12, $d$ = 0.12) and significant according to both the Satterthwaite approximation, $t(336.3) = -5.20$, $p < .001$ (two-tailed), and the LRT, $\Delta\chi^2(\Delta 1) = 26.1$, $p < .001$ (two-tailed). Finally, the crucial wave*condition interaction showed that students in the intervention condition unexpectedly gained more amotivation from T1 to T2 compared to students in the control condition, however, this effect was almost non-existent ($\beta$= 0.007, $d$ = 0.007) and nonsignificant according to both the Satterthwaite approximation, $t(336.3) = 0.281$, $p = .611$ (one-tailed), and the LRT, $\Delta\chi^2(\Delta 1) = 0.08$, $p = .611$ (one-tailed).

**Teacher appreciation of training program.**    The teachers in the intervention group reported (on a 1-5 scale) that they found the training program to be important ($M = 3.75$, $SD = 0.5$) and valuable ($M = 3.75$, $SD = 0.5$). They indicated that they were satisfied with the training program ($M = 4.25$, $SD = 0.25$) and that it produced a positive change in their motivating style ($M = 3.75$, $SD = 0.5$).

Teachers all described the training program in positive terms when answering the open-ended question. The most extensive answer was (translated):

"I was very satisfied about the training program in general. I already did many of the things we discussed, however it is good to get confirmation that those things apparently improve the motivation of the students. I became aware of some things during the program and actively worked on them. It was enjoyable to notice that it helped and that the students also indicated this when giving feedback on the tutorials. I would definitely recommend to discuss this with the remaining teachers".

The answers of the other teachers were similar to this answer.

## Supplemental Analyses

Besides checking the main hypotheses, we also checked whether perceived need-support was related to the student-assessed dependent variables, regardless of condition. We checked this in reaction to the non-significance of the effects of condition on the student outcome variables, so this is more exploratory in nature. We used multi-level models, with perceived need-support and the covariates as predictors and the student-assessed variables as dependent variables, to answer this question. For this analysis, we looked at T2 data only. We used the Satterthwaite approximation to test for significance, using two-tailed tests. Effects are reported in unstandardized scale-units. Perceived need-support was positively associated with autonomous motivation, $\beta = 0.46$, $t(306) = 4.15$, $p < .001$ and engagement, $\beta = 0.32$, $t(306) = 6.68$, $p < .001$, and negatively associated with amotivation, $\beta = 0.37$, $t(332) = -3.22$, $p = .001$. It was not significantly

associated with controlled motivation, $\beta = 0.05$, $t(306) = 0.43$, $p = .668$, and academic achievement, $\beta = -0.14$, $t(306) = -0.74$, $p = .458$.

## Discussion

### Main findings

The contribution of this study is twofold. First, we translated literature on SDT and need-supportive teaching to a set of concrete, practical techniques, that can be used by statistics teachers during tutorials and that are minimally disruptive to the normal flow of the lesson. This is the is the first time this has been done and is valuable on its own, as a set of guidelines or best practices for statistics teachers. Second, we designed and implemented a training program to train statistics teachers to use these need-supportive techniques, and studied its effects. In the next three sections, we discuss these effects of the training program by answering the three research questions.

**Did the training program make the teachers more need-supportive?** We first discuss the effects of the training program on teachers' use of need-supportive techniques. Based on our experiences during the training program, it seemed like we definitely made a positive difference in the teachers' motivating styles. During the supplemental meetings, teachers gave concrete examples of situations where they implemented need-supportive techniques, which they would have handled differently before the training program. For example, one teacher regularly made competence-thwarting remarks to students and made an effort to prevent those moments. Another teacher often gave solutions right away, and tried to make students think for themselves first instead. Some teachers introduced themselves more personally to the students this year compared to last year, using pictures of their private life, to make a more personal connection with their students. One teacher noticed physical distance during class, and made an effort to mix with the students more. Some teachers focused on giving more reasons for requests and some focused on calling students by first name. In sum, based on teachers reports, we

expected to have made a definite positive change in teachers need-support during class.

These reports however should be interpreted tentatively, since it was socially desirable for the teachers to respond positively about the training program. The first author, who led the training program, was also a colleague of the participating teachers, and even though we repeatedly emphasized that it was important that the teachers were honest, it is likely that this skewed their reports about the training program.

In an effort to study this research question in a more scientific way, we asked students to report on perceived need-support using a questionnaire. There was a marginally significant medium effect between the conditions: Students in the intervention group reported higher perceived need-support than those in the control group. This supported our observations made during the program. We therefore make the conclusion that the training program likely provided at least some positive need-supportive change in the teachers motivating style, however, because the sample consisted of only eight teachers, this effect was difficult to detect.

**Did the training program provide benefits for students?**   Our second interest was whether the training program also provided benefits for students. We hypothesized that the training program would improve students' quality of motivation, engagement and academic achievement. We found no effect of condition on grade and engagement, a very small positive effect on autonomous motivation and very small negative effects on controlled motivation and amotivation, and all of these effects were non-significant. In sum, we did not find any convincing evidence that the training program was successful in providing any benefits for students, at least not captured by our dependent variables.

This was surprising, since other need-supportive intervention studies did prove beneficial for students (e.g. Cheon et al., 2012). To explain this finding, we performed an additional analysis and checked whether perceived need-support was associated with student benefits, regardless of condition. We found that students who rated their teacher as more need-supportive were more autonomously motivated, less amotivated, and were

more engaged; there was no effect of need-support on grade and controlled motivation. This indicates that need-support *was* beneficial for students (at least for a subset of student outcomes) while the training program showed no direct beneficial effects. This raises two questions. The first question is why there was no effect of need-support on grade and controlled motivation. This might have happened because all teachers had to use the same teaching materials and lesson structure in this course. This limited the range of need-supportive techniques that could be implemented. It could be that this subset of need-supportive behaviors simply has no effect on grade and controlled motivation. Some changes in the contents and structure of lessons might have to be allowed in order to improve in those two areas.

The second question is that if need-support increases autonomous motivation, engagement and decreases amotivation, then why was this effect not visible between conditions? One obvious answer is that the effects of the intervention simply were not strong enough. This relatively short intervention is limited in the effects it has on teachers educational style, and teachers educational style is only partly responsible for students outcomes (see for example the low variance at teacher level for the student outcomes), so in this small sample of eight teachers the effects on students might have been too small to be detectable. This implicates that in a team of eight teachers, a training program like this just might just not make a noticeable difference.

Another explanation might be that students who are more engaged, less amotivated and more autonomously motivated perceive their teacher as more need-supportive, and not the other way around. For example, it could be that better-performing students entice the teacher to be more need-supportive towards them while need-support does not make students perform better. This is at least plausible considering that the majority of variance in perceived need-support is at student-level, instead of at teacher level. This would imply that need-support might not be a one-size-fits-all solution; it would be interesting for future research to explore individualized need-support.

**Did the training program provide benefits for teachers?** During the training program, the teachers regularly voiced that they enjoyed and appreciated the training program. After the training program, they indicated on a questionnaire that they found the training program valuable, important, satisfying and responsible for a positive change in their motivating style. From these reports it seems that the training program was beneficial for teachers. However, even though we repeatedly emphasized to teachers that the questionnaire was anonymous and that it was important that they gave honest answers, social desirability might have also played a role here. Again it is likely that the teachers avoid being critical of a colleagues work. In light of this, results of the teacher appreciation of training program questionnaire should be interpreted tentatively.

Another possible teacher benefit was improved evaluation by students. From the first author's experience as a teacher, student evaluations are an important source of satisfaction over one's work as a teacher, and are also important performance indicators used by teacher's employers. High student evaluations are beneficial for teachers. We found a medium effect of condition on teacher evaluations; teachers in the intervention group were evaluated better than teachers in the control group. While these effects however did not reach statistical significance, they are still promising: The effect on teacher evaluation is relatively large. It would be interesting for future researchers to explore the effects of need-support on teacher evaluation more, for example by conducting interviews with students about their opinion of their teacher.

## Limitations

One limitation of the study is that we used only student reports to assess teachers use of motivating style. Student reports of need-support do not always converge with teacher reports of need-support or in-class video observations (Aelterman et al., 2014), so the use of multiple informants might have given a more complete image of what happened in class.

Another limitation of the study is our focus on quantitative variables while having a

small sample size of teachers. Even though a sample size of eight teachers is relatively normal in SDT-based intervention research (Su & Reeve, 2011), it does make detecting quantitative effects difficult. We encountered this problem when many of the effects were either marginally- or non-significant. Increasing sample size is not always trivial in educational research, because teachers usually have a busy schedule, the interventions often require a relatively large investment of time from teachers, and larger groups of teachers might make the intervention exponentially harder to plan and implement. We therefore recommend future researchers to use quantitative measurements, such as teachers' descriptions of teaching experiences, student descriptions of classroom experience, or qualitative analysis of in-class video observations, to supplement quantitative findings. Education is both an art and a science, and often time the qualitative findings are the most valuable.

A final limitation of this study is that we did not have equal contact time with the control group, which makes it difficult to distinguish whether this training program provided more benefits compared to regular (not guided by SDT) contact time (e.g. regular group discussions and reflections about teaching). We chose for this approach because we wanted to focus more on finding the benefits of implementing the training program compared to not implementing it, instead of finding the benefits of implementing this training program compared to other programs. Some other studies however focused more on this second question and provided the control group with some general training, not based on SDT, and found that SDT-guided training does provide more benefits compared than general training (e.g. Chatzisarantis & Hagger, 2009).

### Future directions and recommendations

Designing educational interventions is an applied science, where theory meets the practical world. We noticed that in SDT-based intervention literature, the practical side is often under-represented. When educational techniques are described, they are often

described in very general terms (e.g. "acknowledge negative affect" or "provide a meaningful rationale"). Occasionally one or two brief examples are given, but rarely are the techniques described in enough detail to implement them in practice. When we designed our intervention, we spent major effort on translating SDT-guided techniques into concrete, practical examples that are easily applied by teachers in their day-to-day teaching. This translation of theory into practice is one of the core challenges of SDT-based educational interventions and we therefore think that this deserves more attention in literature. We therefore recommend that future researchers focus more on the practical side of implementing interventions (e.g. what practical issues they encountered) and make available all materials used to implement their intervention (e.g. slides, booklets, exact descriptions of intervention procedures). We think that this will improve the speed and quality of design and implementation of educational interventions.

In light of this we have some practical recommendations for future intervention designers. We found it very helpful that the first author had extensive experience teaching the same class as the participating teachers did. This allowed us to explain in a concrete and realistic way to the teachers how need-supportive techniques could be implemented. We recommend that future researchers have or gain intimate knowledge about teaching practices of target teachers before implementing an intervention. We also found it helpful that the contents of our supplemental meetings were not precisely pre-determined but instead partly guided by input of teachers. This way we could adapt to the experiences of teachers when they applied the techniques in practice. We therefore advice intervention planners to leave some room in their intervention to adapt based on teachers input.

It would also be interesting for future researchers to focus on applying need-supportive teaching in other courses besides statistics for psychology students. Within the study of psychology, statistics plays an unusual role: It is a formal science in a field of mostly social sciences. It would be interesting to explore what the effects are when applied to different courses in the psychology curriculum, or even in other studies. Since the

techniques described in this study are relatively context-independent, they could be easily translated to other domains.

Another interesting future direction is giving the training program to teachers who have much more contact time with their students' compared to the two hours a week that the participating teachers saw their students. It might be that because the teachers saw their students so little, what they did had little effect on students' outcomes. So even though the learned need-supportive behaviors might be beneficial for students, the effects are so small since the teachers play such a minor role in the course. It would be interesting to explore what the effects are when teachers play a major role in the course.

Besides researchers, educators might have interest in implementing these techniques. For them, we have the following recommendations. First, the list of need-supportive techniques on its own can be used as best-practices or as a guide for teachers, since they are techniques students seem to appreciate. The training program can be used to train teachers to use these need-supportive techniques. However, when teachers have little contact time with their students, the effect of teacher behavior on student outcomes in general will be low, so the training program will likely not provide much benefits for students. Teachers do however seem to appreciate the program and might be evaluated better by students.

**Bibliography**

References

Aelterman, N., Vansteenkiste, M., Van den Berghe, L., De Meyer, J., & Haerens, L. (2014). Fostering a need-supportive teaching style: Intervention effects on physical education teachers' beliefs and teaching behaviors. *Journal of sport & exercise psychology*, *36*(6), 595–609. doi: 10.1123/jsep.2013-0229

Black, A. E., & Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education*, *84*(6), 740–756. doi: 10.1002/1098-237X(200011)84:6<740::AID-SCE4>3.0.CO;2-3

Carnell, L. J. (2008). The effect of a student-designed data collection project on attitudes toward statistics. *Journal of Statistics Education*, *16*(1).

Chatzisarantis, N. L. D., & Hagger, M. S. (2009). Effects of an intervention based on self-determination theory on self-reported leisure-time physical activity participation. *Psychology & health*, *24*(1), 29–48. doi: 10.1080/08870440701809533

Cheon, S. H., & Reeve, J. (2013). Do the benefits from autonomy-supportive PE teacher training programs endure?: A one-year follow-up investigation. *Psychology of Sport and Exercise*, *14*(4), 508–518. doi: 10.1016/j.psychsport.2013.02.002

Cheon, S. H., & Reeve, J. (2015). A classroom-based intervention to help teachers decrease students' amotivation. *Contemporary Educational Psychology*, *40*, 99–111. doi: 10.1016/j.cedpsych.2014.06.004

Cheon, S. H., Reeve, J., & Moon, I. S. (2012). Experimentally based, longitudinally designed, teacher-focused intervention to help physical education teachers be more autonomy supportive toward their students. *Journal of sport & exercise psychology*, *34*(3), 365–396.

Cohen, J. (1992). Quantitative Methods in Psychology. *Psychological Bulletin*, *112*(1), 155–159. doi: 10.1038/141613a0

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. R. (1994). Facilitating internalization:

The self-determination theory perspective. *Journal of personality*, *62*(1), 119–142. doi: 10.1111/j.1467-6494.1994.tb00797.x

Deci, E. L., & Ryan, R. M. (2002). *Handbook of Self-Determination Research* (E. L. Deci & R. M. Ryan, Eds.). Rochester, NY: University of Rochester Press.

Deci, E. L., Spiegel, N. H., Ryan, R. M., Koestner, R., & Kauffman, M. (1982). Effects of performance standards on teaching styles: Behavior of controlling teachers. *Journal of Educational Psychology*, *74*(6), 852–859. doi: 10.1037/0022-0663.74.6.852

Dillon, K. M. (1982). Statisticophobia. *Teaching of Psychology*, *9*(2), 117. doi: 10.1207/s15328023top0902_23

Edmunds, J., Ntoumanis, N., & Duda, J. L. (2008). Testing a self-determination theory-based teaching style intervention in the exercise domain. *European Journal of Social Psychology*, *38*(2), 375–388. doi: 10.1002/ejsp.463

Flink, C., Boggiano, A. K., & Barrett, M. (1990). Controlling teaching strategies: Undermining children's self-determination and performance. *Journal of Personality and Social Psychology*, *59*(5), 916–924. doi: 10.1037/0022-3514.59.5.916

Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. *Policy Analysis*, 1–651. doi: 10.2277/0521867061

González, A., Paoloni, V., Donolo, D., & Rinaudo, C. (2012). Motivational and emotional profiles in university undergraduates: A Self-Determination Theory perspective. *The Spanish journal of psychology*, *15*(3), 1069–1080. doi: 10.5209/rev_SJOP.2012.v15.n3.39397

Gorozidis, G., & Papaioannou, A. G. (2014). Teachers' motivation to participate in training and to implement innovations. *Teaching and Teacher Education*, *39*, 1–11. doi: 10.1016/j.tate.2013.12.001

Goudas, M., Biddle, S., & Fox, K. (1994). Perceived locus of causality, goal orientations, and perceived competence in school physical education classes. *The British journal of educational psychology*, *64*(3), 453–463. doi: 10.1111/j.2044-8279.1994.tb01116.x

Haerens, L., Aelterman, N., Vansteenkiste, M., Soenens, B., & Van Petegem, S. (2015). Do perceived autonomy-supportive and controlling teaching relate to physical education students' motivational experiences through unique pathways? Distinguishing between the bright and dark side of motivation. *Psychology of Sport and Exercise*, *16*(P3), 26–36. doi: 10.1016/j.psychsport.2014.08.013

Hill, A. P. (2013). Motivation and university experience in first-year university students: A self-determination theory perspective. *Journal of Hospitality, Leisure, Sport & Tourism Education*, *13*, 244–254. doi: 10.1016/j.jhlste.2012.07.001

Hox, J. (2002). *Multilevel Analysis: Techniques and Applications.* Mahwah, N.J.: Lawrence Erlbaum Associates.

Jang, H., Reeve, J., & Deci, E. L. (2010). Engaging students in learning activities: It is not autonomy support or structure but autonomy support and structure. *Journal of Educational Psychology*, *102*(3), 588–600. doi: 10.1037/a0019682

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86. doi: 10.1207/s15326985ep4102_1

Lam, S., Cheng, R. W., & Choy, H. C. (2010). School support and teacher motivation to implement project-based learning. *Learning and Instruction*, *20*(6), 487–497. doi: 10.1016/j.learninstruc.2009.07.003

Manor, O., & Zucker, D. M. (2004). Small sample inference for the fixed effects in the mixed linear model. *Computational Statistics & Data Analysis*, *46*(4), 801–817. doi: 10.1016/j.csda.2003.10.005

McLachlan, S., & Hagger, M. S. (2010). Effects of an autonomy-supportive intervention on tutor behaviors in a higher education context. *Teaching and Teacher Education*, *26*(5), 1204–1210. doi: 10.1016/j.tate.2010.01.006

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School*

*Psychology*, *48*(1), 85–112. doi: 10.1016/j.jsp.2009.09.002

Reeve, J. (1998). Autonomy support as an interpersonal motivating style: Is it teachable? *Contemporary Educational Psychology*, *23*(3), 312–330. doi: 10.1006/ceps.1997.0975

Reeve, J. (2002). Self-Determination Theory applied to educational settings. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 183–203). Rochester, NY: University of Rochester Press.

Reeve, J. (2013). How students create motivationally supportive learning environments for themselves: The concept of agentic engagement. *Journal of Educational Psychology*, *105*(3), 579–595. doi: 10.1037/a0032690

Reeve, J., Bolt, E., & Cai, Y. (1999). Autonomy-supportive teachers: How they teach and motivate students. *Journal of Educational Psychology*, *91*(3), 537–548. doi: 10.1037/0022-0663.91.3.537

Reeve, J., & Jang, H. (2006). What teachers say and do to support students' autonomy during a learning activity. *Journal of Educational Psychology*, *98*(1), 209–218. doi: 10.1037/0022-0663.98.1.209

Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion*, *28*(2), 147–169. doi: 10.1023/B:MOEM.0000032312.95499.6f

Reeve, J., Vansteenkiste, M., Assor, A., Ahmad, I., Cheon, S. H., Jang, H., . . . Wang, C. K. J. (2014). The beliefs that underlie autonomy-supportive and controlling teaching: A multinational investigation. *Motivation and Emotion*, *38*(1), 93–110. doi: 10.1007/s11031-013-9367-0

Ruggeri, K., Dempster, M., Hanna, D., & Cleary, C. (2008). Experiences and expectations: The real reason nobody likes stats. *Psychology Teaching Review*, *14*(2), 75–83.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology*, *57*(5), 749–761. doi: 10.1037/0022-3514.57.5.749

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American psychologist*, *55*(1), 68–78. doi: 10.1037/0003-066X.55.1.68

Saunders, J. a., Morrow-Howell, N., Spitznagel, E., Dore, P., Proctor, E. K., & Pescarino, R. (2006). Imputing Missing Data: A Comparison of Methods for Social Work Researchers. *Social Work Research*, *30*, 19–31. doi: 10.1093/swr/30.1.19

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, *7*(2), 147–177. doi: 10.1037/1082-989X.7.2.147

Schau, C., Stevens, J., Dauphinee, T. L., & Vecchio, a. D. (1995). The development and validation of the Survey of Attitudes toward Statistics. *Educational and Psychological Measurement*, *55*(5), 868–875. doi: 10.1177/0013164495055005022

Singer, J., & Willet, J. (2003). *Applied Longitudinal Data Analysis.* New York: Oxford University Press.

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, *69*(3), 493–525. doi: 10.1177/0013164408323233

Standage, M., Duda, J. L., & Ntoumanis, N. (2006). Students' motivational processes and their relationship to teacher ratings in school physical education: A self-determination theory approach. *Research quarterly for exercise and sport*, *77*(1), 100–110. doi: 10.5641/027013606X13080769704046

Stefanou, C. R., Perencevich, K. C., DiCintio, M., & Turner, J. C. (2004). Supporting autonomy in the classroom: Ways teachers encourage student decision making and ownership. *Educational Psychologist*, *39*(2), 97–110. doi: 10.1207/s15326985ep3902_2

Stroet, K., Opdenakker, M., & Minnaert, A. (2015). What motivates early adolescents for

school? A longitudinal analysis of associations between observed teaching and
motivation. *Contemporary Educational Psychology*, *42*, 129–140. doi:
10.1016/j.cedpsych.2015.06.002

Su, Y. L., & Reeve, J. (2011). A meta-analysis of the effectiveness of intervention programs
designed to support autonomy. *Educational Psychology Review*, *23*(1), 159–188. doi:
10.1007/s10648-010-9142-7

Taylor, I. M., & Ntoumanis, N. (2007). Teacher motivational strategies and student
self-determination in physical education. *Journal of Educational Psychology*, *99*(4),
747–760. doi: 10.1037/0022-0663.99.4.747

Tessier, D., Sarrazin, P., & Ntoumanis, N. (2010). The effect of an intervention to improve
newly qualified teachers' interpersonal style, students motivation and psychological
need satisfaction in sport-based physical education. *Contemporary Educational
Psychology*, *35*(4), 242–253. doi: 10.1016/j.cedpsych.2010.05.005

Vansteenkiste, M., Sierens, E., Goossens, L., Soenens, B., Dochy, F., Mouratidis, A., . . .
Beyers, W. (2012). Identifying configurations of perceived teacher autonomy support
and structure: associations with self-regulated learning, motivation and problem
behavior. *Learning and Instruction*, *22*(6), 431–439. doi:
10.1016/j.learninstruc.2012.04.002

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating
learning, performance, and persistence: The synergistic effects of intrinsic goal
contents and autonomy-supportive contexts. *Journal of Personality and Social
Psychology*, *87*(2), 246–260. doi: 10.1037/0022-3514.87.2.246

Wiberg, M. (2009). Teaching statistics in integration with psychology. *Journal of Statistics
Education*, *17*(1).

Williams, A. S. (2010). Statistics Anxiety and Instructor Immediacy. *Journal of Statistics
Education*, *18*(2), 1–18.

Williams, G. C., Saizow, R., Ross, L., & Deci, E. L. (1997). Motivation underlying career

choice for internal medicine and surgery. *Social Science and Medicine*, *45*(11), 1705–1713. doi: 10.1016/S0277-9536(97)00103-2

Wolters, C. A. (2004). Advancing achievement goal theory: Using goal structures and goal orientations to predict students' motivation, cognition, and achievement. *Journal of Educational Psychology*, *96*(2), 236–250. doi: 10.1037/0022-0663.96.2.236

Table 1

*Targeted need-supportive behaviors during the training program*

| Target SDT-concept | Target teaching behavior | Practical examples |
|---|---|---|
| Need for autonomy | Acknowledge negative affect | • Say "Having a rough day?" when you spot a disinterested looking student.<br>• "Yes this can be tricky", "This can be a tough one".<br>• "I didn't understand it myself when I first saw this"<br>• "It is O.K. to not understand this right away".<br>• Mention that it is O.K. to make mistakes. |
| | Provide rationales for requests | • "It is important that you pay attention, because…".<br>• "Doing this assignment is valuable, because…". |
| | Avoid controlling behavior | • Avoid using commands, threats, deadlines.<br>• Avoid using controlling language. |
| | Provide cognitive autonomy support | • Ask what the student thinks before answering a question.<br>• Talk less and listen more to students.<br>• Let students correct their own errors.<br>• Give hints or encouragements instead of answers.<br>• Stimulate that students look at written materials. |
| Need for competence | Provide positive feedback | • Mention how much a student has improved.<br>• Acknowledge what goes right, when something goes wrong: "Up to this point your logic really makes sense".<br>• "Well done!", "You really performed well!".<br>• Avoid competence-thwarting feedback, like "You should have known this weeks ago!". |
| Need for relatedness | Have a warm, open attitude | • Greet students when they enter the classroom.<br>• Smile and make jokes during lessons.<br>• Send e-mail to students wishing them good luck on the exam. |
| | Spend time, energy and attention on students | • Initiate conversation with students before or after class?<br>• Invite students to meet after class if they have questions? |
| | Get physically close | • Take a seat close to students when talking to them.<br>• Use remote instead of standing behind computer. |
| | Know students | • Call students by first name.<br>• Know which students did good/bad on the exam. |

Table 2

*Cronbach's alphas for Dutch and English versions of scales*

| | Perceived need-support | | Engagement | | Autonomous motivation | | Controlled motivation | | Amotivation | |
|---|---|---|---|---|---|---|---|---|---|---|
| | α | 95% CI | α | 95% CI | α | 95% CI | α | 95% CI | α | 95% CI |
| English | .74 | [.60, .88] | .82 | [.77, .92] | .85 | [.78, .93] | .75 | [.66, .85] | .83 | [.70, .95] |
| Dutch | .76 | [.70, .82] | .82 | [.78, .86] | .85 | [.82, .89] | .76 | [.71, .80] | .83 | [.77, .88] |

*Note*: We combined the T1 and T2 results for autonomous motivation, controlled motivation, and amotivation to keep the table consise.

Table 3

*Distribution of variance across four levels of hierarchy*

|  | Grade | Perceived need-support | Autonomous motivation | Controlled motivation | Amotivation | Engagement | Teacher evaluation |
|---|---|---|---|---|---|---|---|
| Between waves |  |  | 40.3% | 38.9% | 44.6% |  |  |
| Between students | 98.7% | 76.3% | 54.9% | 61.1% | 54.2% | 97.5% | 83.2% |
| Between classes | 0% | 3.12% | 2.55% | 0% | 1.19% | 2.47% | 4.27% |
| Between teachers | 1.25% | 20.6% | 2.21% | 0% | 0% | 0% | 12.6% |

Table 4

*Distribution of variance across three levels of hierarchy*

| | Grade | Perceived need-support | Autonomous motivation | Controlled motivation | Amotivation | Engagement | Teacher evaluation |
|---|---|---|---|---|---|---|---|
| Between waves | | | 40.2% | 38.9% | 44.6% | | |
| Between students | 98.7% | 77.9% | 56.3% | 61.1% | 55.2% | 100% | 85.4% |
| Between teachers | 1.25% | 22.1% | 3.50% | 0% | 0.2% | 0% | 14.6% |

*Note:* The computed intra-class correlations of engagement and controlled motivation were exactly zero. To check whether this was problematic for the analysis, we also checked the results for those variables with ordinary linear models. This gave the same conclusions.

Table 5

*Effect of student type on student-assessed variables*

| | Academic achievement | Perceived need-support | Engagement | Autonomous motivation | Controlled motivation | Amotivation |
|---|---|---|---|---|---|---|
| $F$ | 3.04 | 13.7 | 3.24 | 29.0 | 5.12 | 11.5 |
| $df^1$ | 320 | 329 | 329 | 684 | 684 | 684 |
| $p$ | .049 | < .001 | .04 | < .001 | .006 | < .001 |
| Pre-master - Dutch[2] | 0.49 | -0.06 | 0.17* | 0.71*** | -0.30* | -0.49*** |
| International - Dutch[2] | -0.26 | 0.36*** | 0.04 | 0.43*** | -0.18 | -0.28* |
| International - Pre-master[2] | -0.76* | 0.41*** | -0.13 | -0.29 | 0.12 | 0.21 |

*Note*: We combined the T1 and T2 results for autonomous motivation, controlled motivation, and amotivation to keep the table consise.

[1]Degrees of freedom of the error term. The between-groups degrees of freedom is two in all cases. [2]The difference between means of the respective student types, expressed in unstandardized scale scores.

* $p < .05$. ** $p < .01$. *** $p < .001$.
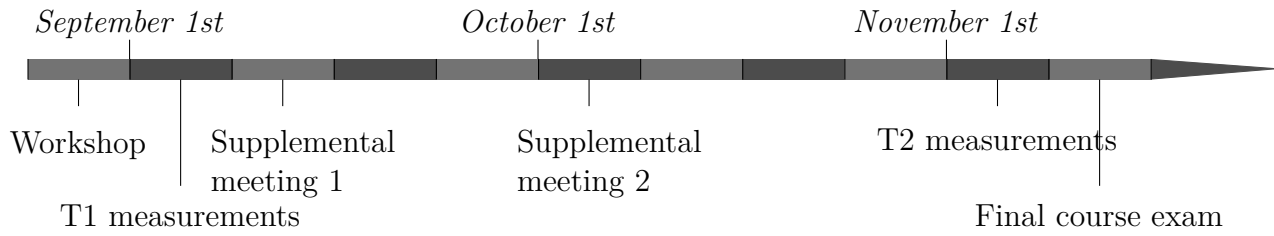
*Figure 1*. Timeline of the training program. The different colored blocks on the timeline represent different weeks. The locations of events on the timeline are approximate.

Appendix A
Perceived Need Support Survey (English Version)

**Our teacher...**

**...understands it when we find something difficult or boring.**
- Example: Says "It is O.K. to make mistakes" or "I can imagine this must be difficult".

**...is bossy.**
- Example: Makes demands by saying "Now go to work!" or "Be quiet!".
- Example: Threathens with punishment.

**...explains what the value is of what we do in class.**
- Example: Says "This is important because. . . " or "Here we learn. . . ".

**...stimulates us to think for ourselves.**
- Example: Does not give solutions right away, but asks first "What do you think?".
- Example: Asks for student input during class discussions.

**...gives us the feeling that we can do it.**
- Example: Makes compliments and is encouraging.
- Example: Avoids remarks like "This doesn't make any sense" or "You should have known this by now!".

**...has a warm, open attitude.**
- Example: Is caring, affectionate.
- Example: Shows that he/she is having fun teaching.
- Example: Is not hostile, cold or distant.

**...is available when we need him/her.**
- Example: Is willing to spend effort on students.
- Example: Is quick to help you when you have a question.

**...is physically close.**
- Example: Takes a seat near you to talk to you.
- Example: Does not stay behind desk during classical discussions.
- Example: Walks through class a lot.

**...knows us well.**
- Example: Calls students by their first name.
- Example: Knows what's going on in the class.

Appendix B
Perceived Need Support Scale (Dutch Version)

**Onze docent...**

**...begrijpt het als we iets moeilijk of saai vinden.**
  - Voorbeeld: Zegt "Fouten maken is niet erg" of "Ik kan me voorstellen dat dit lastig is".
**...is bazig.**
  - Voorbeeld: Geeft veel bevelen zoals "Ga eens aan het werk!" of "Wees eens stil".
  - Voorbeeld: Dreigt met straf.
**...vertelt wat het nut is van wat we doen in de les.**
  - Voorbeeld: Zegt "Dit is belangrijk omdat. . . " of "Hier leren we. . . ".
**...stimuleert ons om <u>zelf</u> na te denken.**
  - Voorbeeld: Geeft niet meteen de oplossing, maar vraagt eerst "Wat denk je zelf?".
  - Voorbeeld: Vraagt om inbreng van studenten tijdens klassikale bespreking.
**...geeft ons het gevoel dat we het kunnen.**
  - Voorbeeld: Geeft complimentjes en aanmoedigingen.
  - Voorbeeld: Vermijdt opmerkingen als "Daar klopt echt niks van!" of "Dat had je al lang moeten weten".
**...heeft een warme, open houding.**
  - Voorbeeld: Is zorgzaam, liefdevol.
  - Voorbeeld: Straalt plezier in het lesgeven uit.
  - Voorbeeld: Is niet vijandig, koud of afstandelijk.
**...is beschikbaar als wij hem/haar nodig hebben.**
  - Voorbeeld: Is bereid moeite te doen voor studenten.
  - Voorbeeld: Komt je snel helpen als je een vraag hebt.
**...komt fysiek dichtbij.**
  - Voorbeeld: Gaat bij je zitten om met je te praten.
  - Voorbeeld: Blijft niet achter bureau tijdens klassikale uitleg.
  - Voorbeeld: Mengt zich tussen de studenten in de klas.

**...kent ons goed.**
  - Voorbeeld: Roept studenten bij hun voornaam.
  - Voorbeeld: Weet wat er speelt in de klas.

Appendix C
Quality of Motivation scale (English version)

At T1: I will work on this statistics course...
　　At T2: I worked for this statistics course...

*Intrinsic motivation*
...because statistics is fun.
...because I enjoy learning new statistics skills.
...because statistics is exciting.
...because of the enjoyment that I feel while learning new statistics skills/techniques.

*Identified regulation*
...because I want to learn statistics skills.
...because it is important to me to do well in statistics.
...because I want to improve in statistics.
...because I can learn statistics skills which I could use in other areas of my life.

*Introjected regulation*
...because I want the teacher to think I'm a good student.
...because I would feel bad about myself if I didn't.
...because I want the other students to think I'm skillful.
...because it bothers me when I don't.

*External regulation*
...because I'll get into trouble if I don't.
...because that's what I am supposed to do.
...so that the teacher won't get mad at me.
...because that's the rule.

*Amotivation*
...but I don't really know why.
...but I don't see why we should have statistics.
...but I really feel I'm wasting my time in statistics.
...but I can't see what I'm getting out of statistics.

Appendix D
Quality of Motivation scale (Dutch version)

At T1: Ik zal aan dit statistiekvak werken...
At T2: Ik heb aan dit statistiekvak gewerkt...

*Intrinsic motivation*
...omdat statistiek leuk is.
...omdat ik ervan geniet nieuwe statistische vaardigheden te leren.
...omdat statistiek boeiend is.
...om het plezier dat ik ervaar tijdens het leren van nieuwe statistiekvaardigheden/technieken.

*Identified regulation*
...omdat ik statistiekvaarigheden wil leren.
...omdat het belangrijk voor me is goed te zijn in statistiek.
...omdat ik me wil verbeteren in statistiek.
...omdat ik statistiekvaardigheden kan leren die van pas komen in andere gebieden van mijn leven.

*Introjected regulation*
...omdat ik wil dat de docent denkt dat ik een goede student ben.
...omdat ik me slecht over mezelf zou voelen als ik dat niet zou doen.
...omdat ik wil dat de andere studenten denken dat ik vaardig ben.
...omdat het me stoort als ik dat niet doe.

*External regulation*
...omdat ik in de problemen zal komen als ik dat niet doe.
...omdat het de bedoeling is dat ik dat doe.
...zodat de docent niet boos op me wordt.
...omdat dat de regel is.

*Amotivation*
...maar ik weet niet echt waarom.
...maar ik zie niet echt in waarom we statistiek krijgen.
...maar ik heb eigenlijk het gevoel dat ik mijn tijd verspil met statistiek.
...maar ik zie niet in wat statistiek mij oplevert.

Appendix E
Classroom Engagement scale (English version)

*Behavioral engagement*
I try hard to do well in this class.
In this class, I work as hard as I can.
When I'm in this class, I participate in class discussions.
I pay attention in this class.
When I'm in this class, I listen very carefully.

*Emotional engagement*
When I'm in this class, I feel good.
When we work on something in this class, I feel interested.
This class is fun.
I enjoy learning new things in this class.
When we work on something in this class, I get involved.

*Cognitive engagement*
When I study for this class, I try to connect what I am learning with my own experiences.
I try to make all the different ideas fit together and make sense when I study for this class.
When doing work for this class, I try to relate what I'm learning to what I already know.
I make up my own examples to help me understand the important concept I study for this class.

*Agentic engagement*
I let my teacher know what I need and want.
During this class, I express my preferences and opinions.
When I need something in this class, I'll ask the teacher for it.
During class, I ask questions to help me learn.
I let my teacher know what I am interested in.

Appendix F
Classroom Engagement scale (English version)

*Behavioral engagement*
Ik doe erg mijn best om het goed te doen in deze lessen.
In deze lessen werk ik zo hard als ik kan.
Tijdens deze lessen doe ik mee met klassikale discussies.
Ik let op in deze lessen.
Tijdens deze lessen luister ik erg aandachtig.

*Emotional engagement*
Tijdens deze lessen voel ik me goed.
Als we aan iets werken in deze lessen voel ik me geïnteresseerd.
Deze lessen zijn leuk.
Ik geniet ervan nieuwe dingen te leren in deze lessen.
Als we aan iets werken in deze lessen doe ik daar aan mee.

*Cognitive engagement*
Als ik studeer voor dit vak probeer ik wat ik aan het leren ben met mijn eigen ervaringen
te verbinden.
Ik probeer alle verschillende ideeën te verbinden en te begrijpen als ik studeer voor dit vak.
Als ik aan het werk ben voor dit vak probeer ik wat ik leer te relateren aan wat ik al weet.
Ik verzin mijn eigen voorbeelden om me te helpen de belangrijke concepten te begrijpen die
ik studeer voor dit vak.

*Agentic engagement*
Ik laat mijn docent weten wat ik nodig heb en wil.
Tijdens deze lessen uit ik mijn voorkeuren en meningen.
Als ik iets nodig heb in deze lessen dan zal ik de docent daarom vragen.
Tijdens deze lessen stel ik vragen die me helpen met leren.
Ik laat mijn docent weten waarnaar ik geïnteresseerd ben.