Learning and Instruction 91 (2024) 101873

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

A Bayesian approach to students' perceptions of teachers' autonomy support

Barbara Flunger[a,*], Anouk Verdonschot[a,**], Steffen Zitzmann[b], Lisette Hornstra[a], Tamara van Gog[a]

[a] Department of Education, Utrecht University, Heidelberglaan 1, Utrecht, the Netherlands
[b] Hector Research Institute of Education Sciences and Psychology, Europastraße 1, University of Tübingen, Germany

ARTICLE INFO

Keywords:
Teachers' autonomy support
Intrinsic motivation
Extrinsic regulation
Doubly latent multilevel models
Bayesian estimation

ABSTRACT

Background: How to best operationalize teachers' autonomy support, an instructional style aiming to satisfy students' psychological need for autonomy, is unclear because teachers can support the whole class and/or individual students. Students might perceive inequalities concerning the autonomy support they receive relative to classmates, which might undermine their motivation and engagement.
Aims: The current study aims to contribute to the conceptualization of autonomy support. We investigated students' perceptions of teachers' autonomy support (individual, class-directed, and perceived differences), concerning choice provision, fostering relevance, stimulating interest, and acknowledging frustration, and associations with students' motivation and engagement.
Sample: 446 Dutch primary school students ($\text{age}_{\text{range}}$ = 9–14) from 22 mathematics classes.
Methods: With Bayesian Multilevel-CFA and -SEM, we examined the factorial structure of students' perceptions of teachers' autonomy support and associations with motivation and engagement. We evaluated whether individual and class-directed autonomy support were distinct constructs, both concerning individual ratings at the student level, and regarding the whole-class-aggregated assessments at the class level.
Results: Individual and class-directed autonomy support was differentiated at the student level. At the class level, one factor (overall autonomy-supportive atmosphere) was found. Regarding perceived differences, we revealed three student-level factors (e.g., relative lack of autonomy support). At the student level, individual and class-directed autonomy support positively predicted intrinsic motivation, effort, and persistence; perceived relative lack of autonomy support positively predicted extrinsic regulation.
Conclusions: Both individual and class-level support should be high to yield optimal results for students' motivation and engagement. Focusing solely on class-directed autonomy support may omit essential information.

Students tend to enjoy mathematics less and less over time (e.g., Garon-Carrier et al., 2016; Jacobs et al., 2002). This is reason for concern because students' enjoyment (i.e., intrinsic motivation, Ryan & Deci, 2000) has been found to be linked with lower risk of drop-out (Hardre & Reeve, 2003) and higher academic achievement (effect sizes for self-reported grades, $\rho = .32$; for objectively measured grades $\rho = 0.13$; see Howard et al., 2021). It is, therefore, important to find ways to promote students' intrinsic motivation in mathematics. According to the theoretical framework of Self-Determination Theory (SDT, Ryan & Deci, 2000), many empirical studies show that teachers can promote students' intrinsic motivation by supporting students' psychological need for autonomy ($\rho = 0.48$, see the meta-analysis by Bureau et al., 2022 considering 47 samples).

Autonomy support "is the adoption of a student-focused attitude and an understanding interpersonal tone that enables the skillful enactment of (…) autonomy-satisfying instructional behaviors" (Reeve & Cheon, 2021, p. 56). Teachers can provide autonomy support in diverse ways, for example, by explaining how a topic connects to an individual student's daily life or through developing interesting lesson materials for the whole class (e.g., Patall et al., 2013). This can promote students'

* Corresponding author.
** Corresponding author.
E-mail addresses: b.flunger@uu.nl (B. Flunger), a.a.n.verdonschot@uu.nl (A. Verdonschot), steffen.zitzmann@uni-tuebingen.de (S. Zitzmann), T.E.Hornstra@uu.nl (L. Hornstra), T.vanGog@uu.nl (T. van Gog).

https://doi.org/10.1016/j.learninstruc.2023.101873
Received 24 March 2023; Received in revised form 20 December 2023; Accepted 27 December 2023
Available online 10 February 2024
0959-4752/

autonomy because they perceive that what they learn corresponds with their interests and values (Assor, 2012). So far, researchers have used measures for both individual and class-directed autonomy support (e.g., “The teacher allows me/us to choose how to do my/our work in the classroom”).

However, when using class-directed or individual measures, researchers need to decide whether to analyze autonomy support at the class or the individual student level.

In theory, teachers' *class-directed* support should be experienced similarly by students within the same classroom, leading to a high level of agreement between students. Student ratings of teachers' class-directed support can be aggregated to the class level, to evaluate students' shared perception of the autonomy-supportive atmosphere in a class (e.g., Morin et al., 2014). Vice versa, student ratings of teachers' *individual* autonomy support can differ within a class, as teachers might adjust their autonomy support to individual students (e.g., some students receive more choices than others). Teachers may differ in the extent to which they adapt their autonomy support to students, and ratings of individual support can vary between classes, depending on the overall level of individual support teachers provide in their teaching. Thus, class-level aggregated measures of individual autonomy support can inform about the differences between classes and teachers concerning how much teachers differentiate their support. Researchers' choices on the measurement and analysis not only have consequences for what construct is being measured exactly but also for the associations with (and thus, the conclusions drawn about) students' motivation and engagement.

The first major objective of the current study was to contribute to the conceptualization of autonomy support. Therefore, a multilevel design was employed to shed light on the question of whether it is crucial to operationalize autonomy support with the “us/class" versus “me/I″ distinction. Second, this study aimed to provide more insight into the associations of different autonomy-supportive strategies teachers can use with students' motivation and behavioral engagement. Third, we also examined the multilevel structure of *perceived differences* in teachers' autonomy support and their associations with students' motivation and behavioral engagement. This is important because when teachers differentiate their autonomy support, students might perceive inequalities regarding the amount of support they receive compared to their classmates, which can undermine their motivation (Chatzisarantis et al., 2019) due to perceived unfairness.

1. Students' motivation and engagement

According to SDT, students' motivation is defined by distinct regulation styles determining learning behavior and engagement (e.g., Ryan & Deci, 2000). The most autonomous regulation style is defined as intrinsic regulation because the reason students want to perform the task is completely self-determined (e.g., Ryan & Deci, 2020): Intrinsically motivated students want to engage in activities because they experience them as inherently interesting and enjoyable. Distinct types of extrinsic motivation can be distinguished. The least autonomous form of motivation is extrinsic regulation. Extrinsic regulation can become internalized and turn to introjected, identified, or integrated regulation, depending on the degree of autonomy experienced (Ryan & Deci, 2020). Students motivated by extrinsic and introjected regulation perform tasks to achieve or avoid an external outcome such as a reward or punishment (Ryan & Deci, 2000). Extrinsic regulation has been found to be associated with various disadvantageous outcomes, such as procrastination (Mouratidis et al., 2018).

Students' motivation can manifest in their learning behavior through their behavioral engagement (Martin et al., 2010). Behavioral engagement can be understood as investing effort (i.e., working on a task as well as one can) and persisting in working hard when the task becomes difficult (Trautwein & Köller, 2003). Previous research suggests that teachers can use autonomy-supportive teaching strategies to foster students' lesson-specific joy (effect size, Glass est. = 0.96), interest (Glass est. = 0.76), and behavioral engagement (Glass est. = 0.56; Flunger et al., 2019).

Students from distinct cultures may differ in their achievement motivation because they may be motivated by different reasons to invest effort in an activity (Bong, 2003; King & McInerney, 2014; Wentzel, 2020). For example, the decrease in the level of students' motivation across their school career is larger in Europe (Glass $\Delta = -0.189$) than in North America ($\Delta = -0.079$) or Asia ($\Delta = 0.022$; Scherrer & Preckel, 2019). While Dutch students achieve good results, they are generally not among the top performers in international comparisons, compared to Asian countries (OECD, 2016). Moreover, Dutch students' motivation has been found to decline over time, also within primary school education (Hornstra et al., 2016).

According to SDT (Ryan & Deci, 2000), a major reason why students are not motivated and engaged is that their need for autonomy is not supported at school. Students who experience autonomy feel that they can make their own choices and sense that what they learn and do matches their values, goals, and interests (Assor, 2012).

2. Adequately conceptualizing teachers' autonomy support in the classroom context

Teachers can use a variety of autonomy-supportive strategies (e.g., Ahmadi et al., 2023; Reeve & Cheon, 2021), such as offering students choices between several tasks and activities, explaining to students why it is relevant for them to perform a certain task (i.e., providing meaningful rationales), offering students activities that match their interests, and acknowledging students' questions, feelings, and frustrations (e.g., Su & Reeve, 2011). The latter means that teachers recognize that students' negative feelings, frustrations, and requests are valid (Su & Reeve, 2011).

Given the benefits of autonomy support (e.g., Bureau et al., 2022), it is important to understand how teachers can optimally provide it in their classrooms. Specifically, teachers can direct autonomy-supportive strategies to the entire classroom, for instance, when informing the whole class about the rationales of a new assignment, and they can offer autonomy support to individual students. When researchers do not differentiate in their measures between the support targeted at the classroom or individual students, it cannot be determined whether students' responses to an item such as “The teacher provides choices” reflect the choices students feel they receive themselves or the choices they think all students in the classroom receive (e.g., Morin et al., 2014). In case students receiving high individual autonomy support interpret the item as relating to their personal support, and students receiving low individual autonomy support interpret it as referring to the support for the entire class, their responses to the item might show strong agreement, while their individual motivation and commitment could vary greatly. To yield deeper insights into whether students rate their personal or the overall-class-directed support, both measures for individual autonomy support (e.g., “The teacher allows *me* to choose how to do my work in the classroom”) and class-directed autonomy support (e.g., “The teacher asks *us* which topics we would like to study more”) can be implemented (e.g., Patall et al., 2013).

Yet, it is possible that measures of class-directed autonomy support might not effectively differentiate between individual students. By aggregating data for the entire class, such measures are more suitable for evaluating the overall, shared perception of the autonomy-supportive atmosphere in a class. When researchers analyze and interpret variables without considering the multilevel hierarchy, they can commit two mistakes: 1) an ecological fallacy (formulating conclusions about individual students [L1: Level 1] based on the class-level or ‘climate’ construct [L2: Level 2]), and 2) an atomistic fallacy (formulating conclusions about a class-level construct based on data that were analyzed on the student level; Hox et al., 2017).

The degree to which perceptions of individual autonomy support differ between students within the same class might depend on how strongly teachers differentiate their autonomy support between students. If teachers interact with students in their class in similar ways, students within a class may evaluate individual autonomy support alike and agree that individual students in that class generally receive similar types of support. Additionally, teachers may differ from one another in how they differentiate in distinct classes which then may lead to between-classroom differences.

2.1. Class-directed autonomy support

When students are asked about the autonomy support directed towards the class, this measurement can provide valuable information on the overall classroom-specific autonomy-supportive climate at a group level (L2). To measure climate constructs, researchers often use individual students' direct ratings of the L2 variable and aggregate these into class-average ratings (e.g., Hospel & Galand, 2016). To target class-level constructs, the referent to the L2 unit (the class) needs to be made explicit in the items (*"In our class, the teacher gives us the opportunity to work at our own pace"*, Hospel & Galand, 2016). In theory, students' ratings of climate constructs should be interchangeable and differences between students' scores indicate measurement unreliability (e.g., Morin et al., 2014). Yet, it needs to be noted that uniform ratings of a class-level construct by students are unlikely because "one will most probably be measuring subjective perceptions of the construct (…) for which within-cluster differences would be expected" (Jak et al., 2021, p. 148). Therefore, examining class-directed autonomy support through items that allow students to provide varying responses can relate to an individual, L1 variable, but might be most informative as an aggregated L2 variable. The L2 construct reflects the whole-class aggregates of individual students' perceptions of class-directed autonomy support within classrooms.

2.2. Individual autonomy support

When autonomy support is targeted at individual students, it refers to an L1 construct that can be measured with an item referring specifically to the individual student (e.g., *"During math, I am often allowed to work on my own"*). Individual autonomy support can also be transformed into an L2 construct by aggregating the data. In such instances, we refer to it as a contextual construct (Marsh et al., 2012). In contrast to a climate construct, students are not asked to directly rate the class-directed support (L2 construct) but rate the individual autonomy support they perceive to receive, which is subsequently transformed into an L2 construct by averaging the individual perception of autonomy support (e.g., Stapleton et al., 2016). Therefore, the contextual construct reflects the whole-class-average perception of individual autonomy support within classrooms and indicates the degree to which teachers provide similar support to the students in their class.

2.3. An overall autonomy-supportive atmosphere

In SDT, it is often claimed that using a set of autonomy-supportive strategies in combination is most beneficial for student outcomes (e.g., Reeve & Cheon, 2021). However, it is unknown whether the resulting *overall* autonomy-supportive atmosphere ("a cumulative perception," Patall et al., 2013, p. 28) refers to an L1 construct (the autonomy support individual students feel) or an L2 construct (the whole-class shared perception of class-directed and individual autonomy support in the class). It is also unknown whether teachers' class-directed and individual autonomy support is equally effective for student outcomes. On the one hand, both class-directed and individual autonomy support might be markers of the overall need-supportive environment in a teacher's classroom, and if aggregated represent one overall construct (see Fig. 1 for a conceptual model). On the other hand, the autonomy support students experience individually might tell us more about the actual dynamics and atmosphere in classrooms (when aggregated as a contextual variable) than class-directed autonomy support. Moreover, if some students feel they receive less support relative to their classmates, and other students perceive their need for autonomy to be fostered, individual autonomy support might trigger differential outcome patterns in students.

3. Perceiving differences in teachers' autonomy support

Findings from prior research suggest that teachers give some students more autonomy support than others (e.g., Chatzisarantis et al., 2019). Students' perceptions of need support tend to vary substantially within classes: Intraclass coefficients ranged from 0.05 to 0.31 (Domen et al., 2019). Furthermore, there is evidence that teachers differentiate

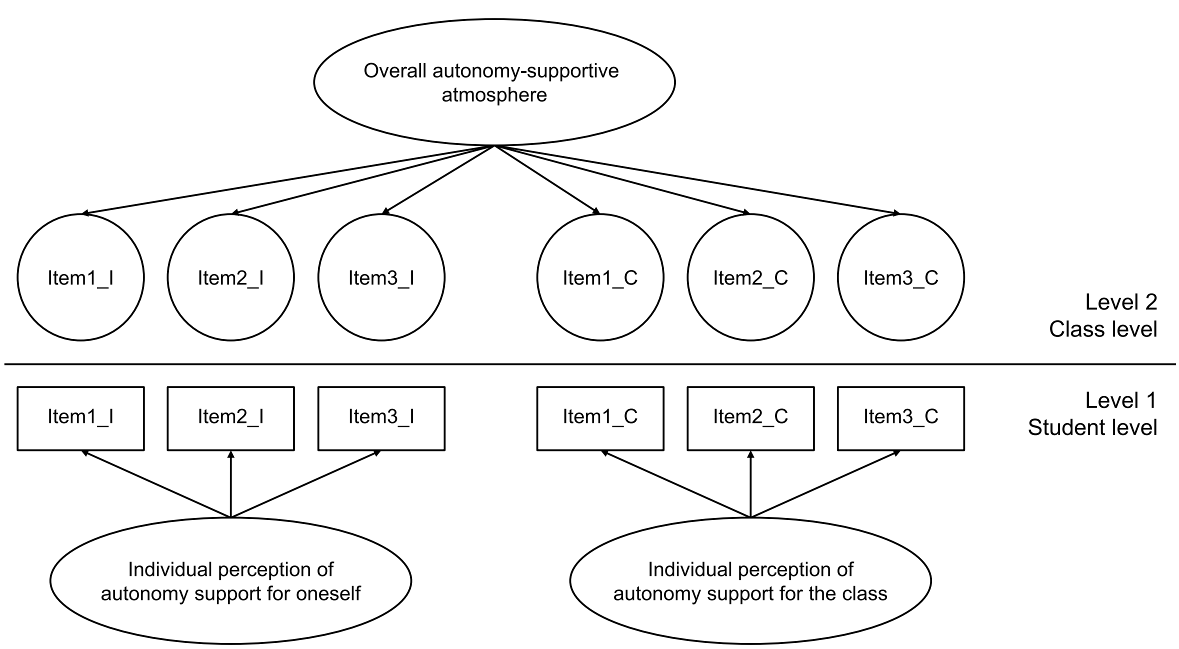

Fig. 1. Students' perceptions of autonomy support as a multilevel construct.

in autonomy support based on various student characteristics, such as their perceptions of student motivation and academic ability (e.g., Bloem et al., 2023; Hornstra et al., 2018).

Students might be aware of this and may perceive themselves as either receiving more support than their peers or comparatively less support. The results of a study by Chatzisarantis et al. (2019) with 359 high school students (grade 9–11) suggest that students who noticed that their classmates received more or less autonomy support than they received themselves, were unhappier and showed less autonomous forms of motivation, lower levels of need satisfaction, and lower academic achievement than students who perceived autonomy support to be equal for all students. In detail, analyzing response surfaces of non-linear regression equations, students who experienced to receive large and equal amounts of autonomy support compared to their classmates could be predicted to report higher autonomous forms of motivation ($M = 3.51$) than students who reported more autonomy support ($M = 1.29$) or less autonomy support ($M = 1.47$) than their classmates (see Chatzisarantis et al., 2019, p. 40). Both favorable autonomy support (receiving more autonomy support than classmates) and unfavorable support (perceiving less autonomy support compared to peers) might trigger perceptions of unfairness, which matter for student outcomes already at an early age (e.g., Helm et al., 2020). Students might also notice that their teachers provide high-performing students more autonomy support than low-performing students (e.g., Hornstra et al., 2018). Accordingly, students might believe that their teachers think they are less competent when they receive less individual autonomy support in comparison to others. As a response, students might work hard to get more autonomy support and to make the teacher think they are smart. Such strivings have been shown to undermine students' intrinsic motivation (Cohen's $d = -0.36$, referring to a small to moderate negative effect of performance orientation on students' self-reported interest and enjoyment, see the meta-analysis by Rawsthorne & Elliot, 1999) and might drive their extrinsic regulation instead.

Students can make comparisons regarding the autonomy support they receive and the support received by their classmates, but they can also compare their classmates' autonomy support to the one received by the rest of the class. Therefore, perceived differences in teachers' autonomy support might also be a multilevel construct. Students' personal experience of receiving less autonomy support from their teacher relative to their classmates (i.e., relative lack of autonomy support, "*my classmates receive more* support *than I do*") may be conceptualized as a student-level construct, whereas the overall degree of perceived differences in autonomy support in the classroom (i.e., "*in this class, some classmates receive more* support *than others*") might be a classroom-level construct.

4. Bayesian approaches

Methodologically, perceived autonomy support belongs to those constructs that are "not directly observable but need to be inferred indirectly, which necessitates the use of latent variable models" (Hofmans et al., 2021, p. 497). One would typically need a large number of classrooms and students to study perceptions of teachers' autonomy support, because multilevel constructs are bound not only to measurement but also to sampling errors, and to account for these errors, doubly latent multilevel analyses are state of the art (e.g., Marsh et al., 2009). However, schools often have limited resources for participating in educational research. In case the sample size at Level 2 is below 25, there is a risk that multilevel models yield "downwardly biased estimates of both the variance components and the fixed effect standard errors, resulting in inflated Type-I error rates for inference about fixed effects" (McNeish, 2017, p. 662). When using smaller samples, a solution lies in the use of Bayesian estimation instead of a conventional, frequentist approach (e.g., Dunson, 2001). Bayesian approaches do not change the model under study, only the estimation procedure differs from frequentist approaches. In Bayesian estimation, knowledge from prior research on the values of specific parameters to be estimated (e.g., concerning the typical factor loadings) can be considered in the estimation (e.g., Wagenmakers et al., 2018). Incorporating such knowledge into the estimation can improve the estimation in small samples (e.g., Zitzmann et al., 2021). Therefore, Bayesian analyses allow for complex multilevel analyses with relatively small sample sizes at Level 2.

5. The present study

The present study investigated four research questions with advanced statistical analyses (Bayesian multilevel analyses), in the context of mathematics classrooms. To yield new evidence on the potential differential impact of students' perceptions of class-directed versus individual autonomy support, we considered a set of four well-substantiated autonomy-supportive strategies (provision of choices, provision of rationales, stimulating interest, acknowledging student frustration, see e.g., Su & Reeve, 2011).

First, we aimed to evaluate the factorial structure of students' perceptions of teachers' class-directed and individual autonomy support (RQ1) and perceived differences in autonomy support (RQ2) via doubly latent multilevel models. At L1, we investigated whether autonomy support provided to students individually versus for the entire class was one or two constructs. Based on findings by Chatzisarantis et al. (2019) and Morin et al. (2014), we expected that students differ in how they perceive the autonomy support provided to them individually and to the whole class. Perceptions of individual and class-directed autonomy support might represent two distinct constructs of teachers' autonomy support at the student level (see Fig. 1), which may shape the overall autonomy-supportive atmosphere in a classroom (representing a shared perception of class-aggregated individual and class-directed autonomy support). At L2, we explored whether the aggregated whole-class assessments of individual autonomy support were distinguishable from the aggregated assessments of class-directed autonomy support. For example, if there is high agreement in students' perceptions that their teachers' individual autonomy support is high, it would mean that teachers tend to differentiate their autonomy support to every student within a class, which is not necessarily the same as the class-directed provision of autonomy support. Thus, conceptually, the aggregated scores of individual and class-directed autonomy support could differ (i.e., represent two constructs that are weakly correlated).

We also studied the associations of students' perceptions of class-directed and individual autonomy support (RQ3), and perceived differences in autonomy support (RQ4) with four important outcome variables: students' intrinsic motivation, extrinsic regulation, effort, and persistence (behavioral engagement). Concerning Research Question 3, based on SDT, it can be expected that students' perceptions of teachers' autonomy-supportive practices are positively associated with intrinsic motivation (see e.g., Ryan & Deci, 2020 for an overview), and effort and persistence (e.g., Flunger et al., 2019). Yet, we could not derive specific hypotheses about the most beneficial form of autonomy support in the classroom. It can be advantageous when teachers tailor their autonomy support to the needs of individual students because students feel that the provided support matches their preferences. However, the *overall* autonomy-supportive atmosphere that students observe in a classroom (Patall et al., 2013) may be more beneficial than individual autonomy support, because a focus on individual support could mean that only a handful of students feel that their need for autonomy is met.

Concerning perceived differences in autonomy support (RQ4), we expected that the higher the discrepancies between the autonomy support students receive individually and the support their classmates receive, the more negative the effects on their motivation might be (Chatzisarantis et al., 2019; Domen et al., 2019), which could imply a positive association between perceived differences in autonomy support and students' extrinsic regulation.

6. Method

6.1. Sample and procedure

The data stemmed from a larger research project focusing on student perceptions of teachers' individual and class-directed autonomy support, considering the perspective of students, teachers, and trained observers in mathematics. This is the first study using these data. Due to the aim to reveal how students perceive their teachers' autonomy support, we focused solely on students' perceptions in this study. Moreover, a recent study revealed that student ratings of teachers' autonomy-supportive behaviors were better predictors of student outcomes than teacher ratings (Flunger et al., 2022). Likewise, the observer or teacher perspective may not explain much additional variance in outcomes compared to only focusing on the student perspective (Donker et al., 2021).

The sample consisted of 473 Dutch primary school students of whom 446 students (48.9% female, $M_{age} = 11.14$; $SD = 0.92$, age range 9–14) from sixth ($N = 166$), seventh ($N = 155$), and eighth ($N = 75$) grade and from classes combining seventh and eighth grade ($N = 77$), from 12 primary schools (22 classes) completed the student questionnaires and were included in the present study. In the Netherlands, the primary school has eight grades, ranging from first grade (4-year-olds) to eighth grade (12-year-olds); the grades are comparable to US fourth, fifth, and sixth grades. A total of 22 teachers (15 women, 7 men; $M_{age} = 39.23$, $SD = 12.90$) participated in the study, from Dutch grade 6 ($N = 9$), 7 ($N = 7$), 8 ($N = 3$) and 7/8 ($N = 3$). The teachers had been working in education for an average of 17.14 years ($SD = 11.00$). The number of students participating per class ranged from 6 to 33 students ($M = 20.65$; $SD = 6.24$); next to unequal class sizes, other reasons for this wide variation are that several students did not have parental consent to participate, and 27 students were absent on the day of the data collection.

The data were collected in class during a lesson period of approximately 45 min by trained master students. The questionnaires were administered on paper. Parents were informed and could opt out of having their children participate in the study (passive informed consent). At the time the data collection was conducted (2017), no official approval was required by the Ethical Review Board of the University. Students received an explanation about the purpose of the study and the voluntary and anonymous character of their participation. Students were instructed that they could omit questions if they did not want to answer them or were undecided, and stop their participation at any time, without facing any consequences.

6.2. Measures

The constructs were rated on a 4-point Likert-type scale, ranging from 1 (completely untrue) to 4 (completely true). Negatively worded items were recoded. All indicator variables were z-standardized to improve convergence and interpretability of the results.

6.2.1. Student perceptions of teachers' autonomy support

Four autonomy-supportive strategies were measured: Providing choices, providing meaningful rationales, acknowledging frustrations, and stimulating interest (three items per strategy, all items are depicted in Table 1). Twelve items were used to measure students' perceptions of individual and class-directed autonomy-support (e.g., "During math class, I can often work in my own way" or "In our class, we are often allowed to work in our own way"). Students were explicitly instructed that these items referred to the support they perceived to receive individually or together as a classroom. The items were developed by Flunger et al. (2022), based on Aelterman et al. (2019) and Flunger et al. (2019).

6.2.2. Perceived differences in teachers' autonomy support

Eight items, which are presented in Table 1, were newly developed to capture *perceived differences* in teachers' autonomy support in the classroom (e.g., "During math class, the teacher often searches for additional strategies to make math more interesting for some students"). Five items were newly developed to capture students' perceptions of *relative lack of autonomy* support (e.g., "During math class, the teacher invests more effort to make math class interesting for other students than for me").

6.2.3. Students' motivation and behavioral engagement

The Dutch translation of the Academic Self-regulation Questionnaire (SRQ-A; Ryan & Connell, 1989; Sempels, 2014) was used to measure intrinsic motivation (three items) and extrinsic regulation (four items). All items, ICC(1), and internal consistencies are presented in Table 1. Students' effort was measured with four items, adapted from Trautwein and Köller (2003). Students' persistence was measured with three items, based on Flunger et al. (2015). The validity of the constructs was confirmed via Bayesian Multilevel CFA, and the model fit was acceptable (see https://osf.io/s87wb/?view_only=1eb3263e49664444b7952f0b61918e92).

6.3. CITO scores

The school administration provided the latest CITO score of students in mathematics, which is a national standardized measure assessing students' general mathematics achievement. Dutch students' achievement in mathematics is measured twice a year in January and June with national, curriculum-based tests from the Dutch National Institute for Educational Measurement (Cito; see Jansen et al., 2010; Scheltens et al., 2020). The CITO scores are used by teachers and parents to evaluate students' mathematics abilities and for educational decisions, such as future secondary school tracking (Scheltens et al., 2020). The raw Cito test scores are transformed to normed ability scores on a scale from A to E, with A reflecting the highest (A-C each 25% of norm group, D = 15%) and E the lowest score (10 % of norm group). The Cito tests have been shown to have good validity and reliability ($\alpha > 0.80$) (e.g., Janssen et al., 2010). There were 118 students with an A-score, 105 students had a B-score, 79 students a C-score, 54 students a D-score, and 81 students an E-score. We recoded the CITO scores to a scale of 1–5 with 1 indicating the lowest and 5 the highest score ($M = 3.28$; $SD = 1.45$).

6.4. Methodological considerations

6.4.1. Doubly latent multilevel models

In multilevel models using manifest factors and manifest aggregation, a construct cannot be measured without error. One type of error is due to "the sampling of items at the individual (L1) and the group (L2) levels" (i.e., measurement error), and another type of error is due to "the sampling of persons in the aggregation of L1 constructs to form L2 constructs" (i.e., sampling error; Marsh et al., 2009, p. 765). To account for both measurement error and sampling error, Marsh et al. (2009) suggested the use of doubly latent multilevel analyses: latent factors are modeled at both the student level and the classroom level to control for measurement error, and latent aggregation to control for sampling error. A disadvantage of conventional multilevel analysis techniques is that these require a considerable number of L2 units (Hamaker & Klugkist, 2011). Using Bayesian estimation instead of a conventional, frequentist approach can offer a solution for smaller samples (e.g., Zitzmann et al., 2016).

6.4.2. A Bayesian approach

If the number of classrooms in the sample is smaller than 50, there is a risk that conventional, frequentist approaches, such as approaches making use of maximum likelihood (ML) estimation, yield biased estimates (Maas & Hox, 2004). Furthermore, when the number of

Table 1
Descriptive statistics (M, SD) of the measures, the items, and the intraclass correlations.

Construct	*M*	*SD*	Ω(L1)	Item number	Item	ICC (1)
Student perceptions of autonomy support						
Class-directed choice provision	2.78	.74	.623	ZSCAuKe1	During our math class, we receive many choices.	0.032
				ZSCAuke3	During our math class, we are often allowed to work in our way.	0.072
Individual choice provision	2.92	.66	.402	ZSIAuKe1	During math class, I receive many choices.	0.040
				ZSIAuke3	During math class, I am often allowed to work in my way.	0.091
Class-directed rationale provision	2.66	.88	.861	ZSCAuRe1	During our math class, the teacher explains to us how important math is in daily life.	0.086
				ZSCAuRe2	During our math class, the teacher often explains to us that we will need math in our future.	0.087
				ZSCAuRe3	During our math class, the teacher encourages us to think about how math can be used in real life.	0.066
Individual rationale provision	2.77	.85	.774	ZSIAuRe1	During math class, the teacher explains to me how important math is in my daily life.	0.062
				ZSIAuRe2	During math class, the teacher often explains to me that I will need math in my future.	0.063
				ZSIAuRe3	During math class, the teacher encourages me to think about how math can be used in real life.	0.059
Class-directed interest stimulation	2.78	.79	.740	ZSCAuIs1	During our math class, the teacher shows us that math is interesting.	0.059
				ZSCAuIs2	During our math class, the teacher searches for new ways to make math class more interesting for us.	0.065
				ZSCAuIs3	During our math class, our teacher makes sure we find math engaging.	0.039
Individual interest stimulation	2.85	.74	.608	ZSIAuIs1	During math class, the teacher shows me that math is interesting.	0.076
				ZSIAuIs2	During math class, the teacher searches for new ways to make math class more interesting for me.	0.059
				ZSIAuIs3	During math class, the teacher makes sure I find math engaging.	0.051
Class-directed frustration acknowledgment	2.38	.68	.553	ZSCAuN1r	During our math class, the teacher easily gets angry when the classroom becomes busy/ noisy.	0.128
				ZSCAuNe2	During our math class, the teacher responds calmly whenever we misbehave.	0.122
				ZSCAuNe3	During our math class, the teacher shows understanding when we are bored.	0.050
Individual frustration acknowledgment	3.23	.74	.658	ZSIAuNe1	If I am sad, nervous, or bored during math class, the teacher asks me if I want to talk about it.	0.032
				ZSIAuNe3	If I am sad, nervous, or bored during math class, my teacher shows understanding.	0.040
Perceived differences in autonomy support						
Differences in relevance instruction	2.74	.76	.688	ZSCDAuR1	For some students, the teacher puts more effort into explaining why math is relevant in math class.	0.039
				ZSCDAuR2	For some students, the teacher explains more often how relevant mathematics is for their future.	0.051
				ZSCDAuI1	For some students, the teacher looks for additional strategies to make math more interesting.	0.033
Differences in autonomy support	2.46	.90	.630	ZSCDAuK2	During math class, some have a greater say in what happens than others.	0.038
				ZSCDAuN1	During math class, our teacher is not aware of the feelings of all students.	0.017
				ZSCDAuK1	During math class, not everyone gets the same number of choices from our teacher.	0.048
Relative lack of autonomy support	1.88	.73	.670	ZSIDAuK1	During math class, other students have a greater say in what happens than I do.	0.078
				ZSIDAuN	When I am stressed or bored during math class, the teacher talks with me about it less often than he/she does with other students.	0.050
				ZSIDAuK2	During math class, some students are often allowed to make their own choices, but I am rarely allowed to do so.	0.035
				ZSIDAuI	During math class, the teacher puts more effort into making math class interesting for other students than for me.	0.076
				ZSIDAuR	During math class, the teacher tries harder to convince other students why math is important than me.	0.072
Motivation and Behavioral Engagement			Ω/α			
Intrinsic motivation	2.87	.90	.853/ .874	ZSIMOIM1	I do my best in math class because it is fun	0.077
				ZSIMOIM2	I do my best in math class because I enjoy doing exercises in math.	0.116
				ZSIMOIM3	I do my best in math class because I like to engage in math.	0.080
Extrinsic motivation	2.70	.86	.726/ .711	ZSIMOEX1	I do my best in math class because I'll get in trouble if I don't.	0.027
				ZSIMOEX3	I do my best in math class because I'll get in trouble with my teacher if I don't.	0.059
				ZSIMOEX4	I do my best in math class because that's the rule I have to follow.	0.026
Effort	3.42	.53	.745/ .775	ZSIINZ1	I am willing to invest effort in math.	0.077
				ZSIINZ2	I try my best in math.	0.075
				ZSIINZ3	I work seriously on math assignments.	0.049
				ZSIINZ4	I finish all math assignments.	0.074
Persistence	3.33	.62	.733/ .761	ZSIPERS1	If I cannot solve an assignment for math in the first attempt, I keep trying.	0.069
				ZSIPERS2	Even if the math tasks are difficult, I try my best.	0.063
				ZSIPERS3	Even with difficult math assignments, I do not give up easily.	0.074

Note. M = Mean of the manifest factor; *SD* = Standard deviation of the manifest factor; ω = McDonald's omega; α = Cronbach's alpha; ICC = intraclass correlation. Concerning level-specific reliabilities, we used the syntax by Geldhof et al. (2014), which only refers to congeneric models in which the factor loadings for L1 and L2 differ. We do not report the level-specific reliabilities for the outcomes, for which the factor loadings were fixed across levels.

classrooms is small, multilevel models may result in estimation problems, such as nonconvergence or negative variance estimates (e.g., Zitzmann et al., 2016).

Bayesian estimation can help to overcome these problems. Bayesians make their existing beliefs about the parameters they estimate explicit. They represent their uncertainty or beliefs about the parameters with a probability distribution for the parameters they want to estimate. For all model parameters (e.g., in a multilevel model: the random intercept, variance of the random effects, et cetera), so-called prior distributions are specified. Researchers can use information they already have about model parameters to form the priors. When choosing prior distributions, the values of a parameter can be restricted to be estimated only within a specific range (e.g., −1 to 1 for correlation coefficients, see Wagenmakers et al., 2018) or to be positive (e.g., 0 to 1 for variance estimates, to prevent negative variance estimates, Fox & Smink, 2023; Hox & McNeish, 2020, p. 221). In the model estimation, the prior information is combined with the data. By combining the prior for the parameters and the likelihood function, a posterior distribution of possible values for the parameters is estimated, often utilizing Markov Chain Monte Carlo simulation (e.g., Muthén, 2010). These posterior distributions tend to exhibit narrower ranges than the prior distributions, indicating that uncertainty is reduced by the data (Hamaker & Klugkist, 2011).

If prior knowledge is limited, researchers can specify priors that are only weakly informative. These priors contain some relevant information, but they should not affect the final parameter estimate much (van de Schoot & Depaoli, 2014). Alternatively, researchers can use uninformative priors, which means that no prior knowledge enters the estimation, and the posterior distribution is determined by the data (Hamaker & Klugkist, 2011). As researchers incorporate more information in their estimation by using priors, the posterior distribution will be more precise, the statistical power will increase, and it is less likely that inference errors will be made (Zondervan-Zwijnenburg et al., 2017).

Results of simulation studies indicated that a Bayesian estimation of multilevel latent contextual models with weakly informative priors yielded less inadmissible solutions and more accurate estimates of the group-level effect than ML when the L2 sample size was small (Zitzmann et al., 2016). Therefore, it is promising to use Bayesian multilevel latent analyses to study individual and class-directed autonomy support and perceived inequalities in teachers' autonomy support in the classroom.

6.5. Analyses

Data were analyzed using M*plus*, version 8.6 (Muthén, & Muthén, 1998–2017), using the Bayesian estimator (for a comparison with frequentist estimation using maximum likelihood estimation, see Supplementary Material). All analysis codes, outputs, and the codebook of the measures used are available at https://osf.io/s87wb/?view_only=1eb3263e49664444b7952f0b61918e92. We used unit variance identification for scaling, to be able to evaluate all factor loadings. In the estimations, we used priors that were approximately uninformative or only weakly informative. More detailed information on model specification and prior selection can be found in the Supplementary Material.

We first assessed the dimensionality of autonomy support (RQ1 and RQ2) by means of Bayesian Multilevel Confirmatory Factor Analyses (ML-CFA). Second, we used Bayesian Multilevel Structural Equation Modelling (ML-SEM) to examine the associations with the outcome variables (RQ3 and RQ4).

A pre-requisite for conducting multilevel analysis is to examine the ICC(1), which indicates whether the variability in students' perceptions can be attributed to class membership. The item ICC(1)s (Table 1) ranged from 0.017 (for the item "During math class, our teacher is not aware of the feelings of all students") to 0.128 ("During our math class, the teacher easily gets angry when the classroom becomes busy/noisy"). Simulation studies by Bliese (see 2000) have revealed that a non-zero ICC(1) value reflects that the student-level responses differ between classes, implying that students within a classroom show greater similarities in their responses than students from different classrooms.

6.5.1. Part 1: Bayesian ML-CFA

Because the multi-dimensionality of autonomy-supportive strategies (e.g., Patall et al., 2013), also for the measure used, had been confirmed previously (Flunger et al., 2022), we first assessed whether we needed to differentiate between individual and class-directed support for each specific autonomy-supportive strategy with ML-CFA. Additionally, we combined the distinct strategies, to check whether the four autonomy-supportive strategies can be distinguished from each other at L1 and L2 (see Supplementary Material). We inspected the correlation matrices of all models to identify strong correlations. Next to establishing the dimensionality of the factors, we assessed the reliability of the latent factors at L1 and L2 and adapted the factor structure if the factors were unreliable.

Models 1: Assessing the Two-Level Structure of Autonomy Supportive Strategies. To investigate research question 1, in step 1, we used Bayesian ML-CFAs to assess the two-level structure of each of the autonomy-supportive strategies (providing choices, providing meaningful rationales, acknowledging frustrations, and stimulating interest) and perceived differences in autonomy support, separately. Although perceived class-directed support may, in theory, be conceptualized as an L2 climate construct, we specified the class-directed items at both levels to control for sampling error (Morin et al., 2014), and because within-class differences in student ratings were likely (Jak et al., 2021). Students' perceptions of individual support can both be an L1 construct and a meaningful L2 contextual construct and were examined at both levels. We compared three models: (a) Model 1.1: two distinct factors (class-directed and individual support) at both levels, (b) Model 1.2: two distinct factors (class-directed and individual support) at L1 and one overall autonomy support factor at L2 (shared perception of overall class-directed and individual support), (c) Model 1.3: one overall factor on both levels. We did not examine a model with one factor at L1 and two factors at L2 because this reflects a misunderstanding of group- and individual-level constructs and does not make sense conceptually.

Models 2: Assessing the Two-Level Structure of Perceived Differences in Autonomy Support. To examine the factor structure of perceived differences in autonomy support (research question 2), we compared three models: (a) Model 2.1: one factor of general perceived differences in autonomy support on both levels, (b) Model 2.2: two distinct factors ('relative lack of autonomy support' vs. 'perceived differences') at L1 and one factor on L2 representing shared perceptions of differences, (c) Model 2.3: two distinct factors ('relative lack of autonomy support' vs. 'perceived differences') on both levels.

To determine the goodness of fit of the Bayesian multilevel CFA models, we engaged in posterior predictive checking, inspecting the posterior predictive *p*-value (PP*p*-value, Kaplan & Depaoli, 2012). The idea behind posterior predictive checking is that the model fits the data well if the observed data reasonably resembles the data generated by the model. A PP*p*-value smaller than 0.05 indicates a bad model fit. Apart from the PP*p*-value, M*plus* generates a 95% confidence interval for the difference between the observed and simulated chi-square values. A positive lower limit suggests poor model fit (Muthén & Asparouhov, 2012). To compare the fit of different models, we consulted the Deviance Information Criterium (DIC) (Spiegelhalter et al., 2002), preferring the model with the smallest DIC value.

6.5.2. Part 2: Bayesian ML-SEM

Using the measurement model identified in Part 1, we estimated the associations of students' perceptions of autonomy support (RQ3) and differences in autonomy support (RQ4) with effort, persistence, intrinsic motivation, and extrinsic regulation with Bayesian ML-SEMs. Because of the complexity of the model, we performed a Bayesian ML-SEM for each outcome separately, resulting in four models for RQ3 and four models for RQ4. M*plus* estimates one-tailed *p*-values but we report two-tailed tests of significance because we could not derive specific hypotheses for all research questions. The latent student-level predictor variables were group mean-centered (e.g., Enders & Tofighi, 2007), which is the default in M*plus* when covariates are specified to predict outcomes both at the student- and at the class level.

6.5.3. Missingness

The percentage of missing data in the item responses ranged from 5.9% to 7.6%, which can be considered small. When missingness is handled with Bayesian Multilevel SEM, all available information from the observed data is included in the estimations (similar to full information maximum likelihood, Asparouhov & Muthén, 2010).

7. Results

Descriptive statistics (Means, *SD*s of the raw scores) and reliabilities of the latent factors of the final models are summarized in Table 1.

7.1. Factorial structure of autonomy-support

7.1.1. Autonomy support as a two-level construct (RQ1)

Table 2 shows the model fit statistics for the evaluation of the two-level structures of each of the four components of autonomy-support (providing choices, providing meaningful rationales, acknowledging frustrations, and stimulating interest) separately.

Results from Step 1 (Models 1.1-1.3). For all autonomy-supportive strategies, Model 1.2, specifying perceptions of individual and class-directed autonomy as two separate factors at L1, and one latent factor of class-average autonomy support at L2, had the best fit. Thus, individual support could be distinguished from class-directed support on L1, but not on L2. This implies that the whole-class aggregates of individual and class-directed autonomy support were best predicted by one overall factor of shared perceptions of overall class-directed and individual support. The ICC(1)s of the items showed differences in students' perceptions of individual autonomy support within classrooms. Thus, while students perceive that their teachers tend to differentiate in their autonomy support for specific students, this teaching style seems to contribute to the overall autonomy-supportive climate in classrooms.

Step 2: Reliability Check and Model Adaptation. When estimating multilevel models, level-specific reliability should be evaluated. Following Geldhof et al. (2014), we estimated the level-specific omegas for congeneric models, which means that, as this reflected our models, the factor loadings at the within-group and between-group levels differed (for more information, see Supplementary Material). The resulting reliability of an L2 construct provides information about whether the items intended to assess the same construct at L2 produce similar scores.

Although the fit of Model 4 was sufficient, the reliability of the factor individual choice provision was too low (ω_{L1} = .402), and the items did not load significantly on this factor. Therefore, we merged the two factors representing perceptions of individual and class-directed choice provision into one factor and compared two further models. In Step 3, Model 1.4-Adapted specified 7 latent factors at L1 and 4 latent factors at L2, and Model 1.5-Adapted specified 7 latent factors at L1 and 1 latent factor at L2. Based on a slightly better model fit, we opted for Model 1.4-Adapted (see Table 2). The factor loadings yielded with Model 1.4-Adapted are displayed in Table 3, for further information see Supplementary Material. The reliability (ω_{L1} = .602) of the resulting L1 factor perceived choice provision was improved but still low.

Table 2
Model fit statistics resulting from Bayesian Multilevel-CFAs assessing different models for perceived strategies of autonomy-support.

Model fit criteria					
Models and variables	DIC	PP*p*	95% CI		
Step 1: Models 1.1–1.3 Analysis of separate strategies					
Variable choice provision					
Model 1.1. Two Factors at L1, Two Factors at L2	4734.668	0.157	−10.962	-	37.615
Model 1.2. Two Factors at L1, One Factor at L2	**4716.488**	**0.415**	**−21.091**	**-**	**24.404**
Model 1.3. One Factor at L1, One Factor at L2	4727.308	0.066	−6.808	-	45.514
Variable rationale provision					
Model 1.1. Two Factors at L1, Two Factors at L2	6206.484	0.000	10.806	-	83.435
Model 1.2. Two Factors at L1, One Factor at L2	**6197.997**	**0.476**	**−26.618**	**-**	**31.664**
Model 1.3. One Factor at L1, One Factor at L2	6295.091	0.000	91.389	-	150.562
Variable interest stimulation					
Model 1.1. Two Factors at L1, Two Factors at L2	6737.003	0.023	0.345	-	67.960
Model 1.2. Two Factors at L1, One Factor at L2	**6732.433**	**0.450**	**−28.265**	**-**	**32.763**
Model 1.3. One Factor at L1, One Factor at L2	6731.148	0.039	−2.896	-	51.291
Variable acknowledging frustration					
Model 1.1. Two Factors at L1, Two Factors at L2	6026.725	0.090	−8.089	-	45.733
Model 1.2. Two Factors at L1, One Factor at L2	**6010.974**	**0.479**	**−24.493**	**-**	**28.152**
Model 1.3. One Factor at L1, One Factor at L2	6067.941	0.000	22.813	-	74.936
Step 2: Models 1.4–1.7 Analysis of strategies combined					
Model 1.4. Eight Factors at L1, Four Factors at L2	**23151.591**	**.324**	**−156.841**	**-**	**255.187**
Model 1.5. Eight Factors at L1, One Factor at L2	23164.571	.308	−156.741	-	253.814
Model 1.6. Two factors at L1, One Factor at L2	23533.022	.000	280.919	-	669.319
Model 1.7. Two factors at L1, Two Factors at L2	23458.079	.000	212.035	-	626.076
Step 3: Model adaptation after reliability check					
Model 1.4-A. Seven Factors at L1, Four Factors at L2	**23155.177**	**.296**	**−147.154**	**-**	**255.179**
Model 1.5-A. Seven factors at L1, One Factor at L2	23165.858	.316	−151.553	-	250.512

Note. L1: Level 1, L2: Level 2. DIC = Deviance Information Criterion; PP*p* = posterior predictive probability; CI = confidence interval; Fit indices of the best-fitting models are highlighted in bold. In Models 1.1, 1.3, and 1.7, the factor loadings were fixed across the student and class levels. For choice provision, two correlated residuals were added; for rationale provision and interest stimulation, three correlated residuals were added in order to be able to evaluate all factor loadings and unit variance identification was used.

Generally, the reliabilities of the L1 factors varied from low to acceptable (ω_{L1} = .553-.861). These results underline that doubly latent ML-CFAs were the adequate choice because they allow for correcting for this type of measurement error (e.g., Marsh et al., 2009). We decided to not eliminate items from the scales, because we considered them to add unique content (although reliability would have increased to ω_{L1} = .641 for the scale with the lowest reliability, class-directed acknowledgment of emotions). Low reliabilities imply that for some scales, student responses varied between different items from the same scale. Thus, students' answers to the item "During our math class, the teacher shows understanding when we are bored" did not correspond well with their

Table 3
Factor loadings of Model 4-adapted: specifying 7 latent Level 1 factors (choice merged) and 4 latent Level 2 factors of autonomy support.

	Level 1 Factor Loadings							Level 2 Factor Loadings			
	4 Level 1 latent factors class-directed autonomy support				3 Level 1 latent factors individual autonomy support			4 Level 2 latent factors			
Item	Choice provision	Rationale provision	Interest stimulation	Frustration acknowledgement	Rationale provision	Interest stimulation	Frustration acknowledgement	Choice provision	Rationale provision	Interest stimulation	Frustration acknowledgement
ZSCAuKe1	**.521** **	.087	.082	−.057	−.004	.016	.076	**.095**	.070	.004	−.025
ZSCAuKe3	**.714** **	−.029	−.059	.054	−.050	−.108	.029	**.309** *	−.032	.020	.057
ZSIAuKe1	**.355** **	−.051	−.007	.032	−.023	**.216** *	−.046	**.162**	.070	−.012	−.016
ZSIAuKe3	**.282** **	−.007	.007	.046	−.032	.053	.050	**.375** *	−.014	−.015	.058
ZSCAuRe1	.003	**.795** ***	.009	−.007	.086	.017	−.026	.022	**.287** *	.018	−.012
ZSCAuRe2	.012	**.830** ***	.020	.011	.070	−.097	.005	−.004	**.259** *	.045	−.039
ZSCAuRe3	.026	**.522** ***	**.147** *	.010	.035	.094	−.021	−.019	**.192** *	.045	.031
ZSCAuIs1	.046	.163 *	**.479** **	.025	.040	.135	−.028	−.021	.065	**.124**	−.027
ZSCAuIs2	.106	.086	**.491** **	.040	−.040	.036	.092	−.043	.012	**.163**	.074
ZSCAuIs3	−.067	−.004	**.622** **	.065	−.034	.028	.039	−.011	.034	**.109**	.035
ZSCAuN1R	.028	−.023	.004	**.474** ***	−.030	−.031	−.007	.038	.011	.016	**.474** *
ZSCAuNe2	−.051	−.020	−.036	**.638** ***	.032	−.010	.044	.051	−.010	.029	**.424** *
ZSCAuNe3	.096	.009	.102	**.358** ***	.002	.038	−.071	−.030	−.046	.025	**.206** *
ZSIAuRe1	−.057	.084	−.094	.081	**.764** *	.020	.014	−.016	**.261** *	−.015	−.037
ZSIAuRe2	−.065	.082	.030	−.045	**.724** *	−.059	.031	.022	**.260** *	−.018	−.006
ZSIAuRe3	.078	−.031	.019	−.010	**.552** *	.096	.035	.004	**.151**	.047	.036
ZSIAuIs1	.121	.066	.063	.022	.092	**.339** *	.019	−.003	.104	**.123**	.008
ZSIAuIs2	.010	−.040	.076	.018	.053	**.613** *	.040	.017	−.035	**.130**	.046
ZSIAuIs3	.033	−.011	.107	.081	−.014	**.311** *	.044	.023	−.014	**.147**	.018
ZSIAuNe1	.042	−.002	.017	.038	−.029	−.003	**.530** *	.007	−.044	.017	**.000**
ZSIAuNe3	.029	−.058	.005	−.028	.094	.013	**.626** **	.073	−.042	.033	**.037**

*$p < .05$. **$p < .01$. ***$p < .001$ (one-tailed).

Note. The model specified seven factors at Level 1 (perceptions of class-directed choice provision, rationale provision, interest stimulation and acknowledgment of frustrations; individual rationale provision, interest stimulation and acknowledgment of frustrations) and four factors at Level 2 (shared perceptions of general choice provision, rationale provision, interest stimulation and acknowledgment of frustrations). The factor loadings of the specified latent factors are highlighted in bold, regardless of their significance. Cross-loadings were specified for all factors on all items (for further information see Supplementary Material).

responses to the items “The teacher easily gets angry when the classroom becomes busy/noisy" and “the teacher responds calmly whenever we misbehave.”

The reliabilities of the L2 factors were all satisfactory (Overall choice provision in class: $\omega_{L2} = .963$; Overall rationale provision: $\omega_{L2} = .984$; Overall interest stimulation: $\omega_{L2} = .947$; Overall frustration acceptance: $\omega_{L2} = .961$). The intercorrelations of the latent factors are displayed in Table 4.

7.1.2. Perceived differences in autonomy support as a two-level construct (RQ2)

Regarding perceived differences in teachers’ autonomy support, when comparing different factor structures at L1 and L2 (Models 2.1-Model 2.3), none of the models with two latent factors had an acceptable fit (see Table 5). Based on the inspection of the factor loadings, we added a third latent factor, defined by items concerning perceived differences in promoting the relevance and interestingness of mathematics. Subsequently, we assessed two additional models: (a) Model 2.4: three latent factors (‘perceived differences,’ ‘differences in fostering relevance’ and ‘relative lack of autonomy support’) on both levels, (b) Model 2.5: three factors ‘perceived differences,’ ‘differences in fostering relevance’ and ‘relative lack of autonomy support’) on L1 and one factor reflecting overall perceived differences of teachers’ autonomy support in classes on L2. Only Model 2.5 had an adequate fit, so we proceeded with this model. The factor loadings are displayed in Table 6.

7.2. Associations between autonomy support and student outcomes

The correlations between the latent factors of perceived autonomy support and students’ outcomes at L1 and L2 are shown in Table 7.

7.2.1. Individual and class-directed autonomy support and associations with outcomes (RQ3)

To investigate the unique associations of distinct individual and class-directed autonomy-supportive strategies teachers are observed to use with students’ outcomes, we conducted ML-SEMs (Model 3.1–3.4) for each of the four strategies of autonomy support (providing choices, providing meaningful rationales, acknowledging frustrations, and stimulating interest). The model fits were appropriate (see Table 8). To reduce model complexity and increase convergence,[1] we used estimates of the loadings in the measurement models to fix the parameters in the ML-SEM models. Table 9 presents the results of the multilevel-SEMs. In the next sections, we describe the statistically significant associations.

Motivation. At L1, intrinsic motivation was positively associated with perceived choice provision ($b = 0.21$; $p < .001$). In addition, intrinsic motivation was positively associated with students’ perceptions of how teachers acknowledged their individual frustrations ($b = 0.13$; $p = .028$). Thus, students who reported that their teacher provided choices in general or acknowledged their individual frustrations during math class also were more likely to report enjoying engaging in math activities.

Concerning students’ perceptions of class-directed support, students’ intrinsic motivation was positively associated with teachers’ class-directed rationale provision ($b = 0.13$; $p = .050$). Thus, students who perceived that their math teacher offered meaningful rationales to their class were more likely to be intrinsically motivated for math class. At L2, the class-average perceptions of rationale provision ($b = 0.28$; $p < .001$)

[1] It would have been interesting to explore if perceived differences in autonomy support explain additional variance in outcomes, over and above individual and class-directed autonomy. Unfortunately, the analyses concerning this question did not yield trustworthy results, see https://osf.io/s87wb/?view_only=1eb3263e49664444b7952f0b61918e92. We observed suppression effects and untrustworthy results, e.g., concerning the standard errors, once we combined the distinct strategies.

Table 4
Intercorrelations of the latent factors of perceptions of individual and class-directed autonomy support and perceived differences in autonomy support.

	Level 1									Level 2			
Latent factor	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(11)	(12)	(13)	(14)
Level 1													
(1) CP													
(2) C-RP	.42*												
(3) I-RP	.31*	.64*											
(4) C-IS	.54*	.59*	.44*										
(5) I-IS	.52*	.45*	.43*	.61*									
(6) C-AF	.46*	.38*	.34*	.49*	.45*								
(7) I-AF	.45*	.37*	.39*	.46*	.41*	.37*							
(8) Dif-RP	.42*	.58*	.51*	.60*	.51*	.31	.37*						
(9) Dif-RL	−.27*	−.09	−.11	−.10	−.15	−.31*	−.40*	.08					
(10) Dif-O	−.28*	−.16	−.12	−.18	−.17	−.47*	−.42*	−.02	.63*				
Level 2													
(11) CP													
(12) RP										−.03			
(13) SI										.02	.07		
(14) AF										.15	−.11	.07	
(15) Dif										−.05	−.03	.02	.03

Note. CP = Choice provision; RP = Rationale provision; IS = Interest stimulation; AF= Accepting frustration; C = Class-directed; I = Individual. Dif = Perceived Differences; Dif-RL = Relative lack of autonomy support; Dif-O = Overall perceived differences. *p (two-tailed) <.05.

Table 5
Model fit statistics Bayesian MCFA for perceived differences in autonomy support.

Models 2.1-2.5	DIC	PP*p*	95% CI		
Model 2.1. One Factor at L1, One Factor at L2	12967.718	.000	185.109	-	299.623
Model 2.2. Two Factors at L1, One Factor at L2	12791.558	.051	−11.023	-	103.329
Model 2.3. Two Factors at L1, Two Factors at L2	12783.701	.005	16.895	-	130.573
Model adaptation based on factor loading patterns					
Model 2.4. Three Factors at L1, Three Factors at L2	12755.188	.042	−6.202	-	107.713
Model 2.5. Three Factors at L1, One Factor at L2	**12767.563**	**.269**	**−46.566**	**-**	**72.250**

Note. DIC = Deviance Information Criterion; PP*p* = posterior predictive probability; CI = confidence interval; Fit indices of the best-fitting models are highlighted in bold. To be able to evaluate all factor loadings, unit variance identification was used.

Table 6
Factor loadings of model 2.5: Specifying 3 latent level 1 factors and 1 latent level 2 factors of perceived differences in autonomy support.

Item	Level 1 Factor Loadings		Level 2 Factor Loadings	
	Perceived differences in relevance instruction	Perceived differences in autonomy support	Relative lack of autonomy support	Overall perceived differences
ZSCDAUR1	**.630*****	−.007	.034	**.206***
ZSCDAUR2	**.769*****	−.027	−.002	**.198**
ZSCDAUI1	**.445*****	.017	−.018	**.160**
ZSCDAUK2	−.038	**.719*****	.065	**.149**
ZSCDAUN1	.078	**.421*****	.035	**.132**
ZSCDAUK1	−.017	**.537*****	.099	**.241***
ZSIDAUK1	−.037	.147*	**.554*****	**.316****
ZSIDAUN	−.029	.094	**.488*****	**.251****
ZSIDAUK2	−.016	−.008	**.648*****	**.194***
ZSIDAUI	−.015	−.022	**.577*****	**.237***
ZSIDAUR	.141**	.000	**.547*****	**.318****

*$p < .05$. **$p < .01$. ***$p < .001$ (one-tailed).
Note. The model specified three factors at Level 1 (perceived differences in relevance instruction, perceived differences in autonomy support, and perceived relative lack of autonomy support) and one factor at Level 1 (shared perceptions of overall perceived differences). The factor loadings of the specified latent factors are highlighted in bold. Cross-loadings were specified for all factors on all items.

positively predicted the overall intrinsic motivation of the class. Furthermore, students' shared perceptions of interest stimulation (b = 0.28; $p < .002$) positively predicted the class-average intrinsic motivation.

Regarding extrinsic regulation, at L1, negative associations with perceived choice provision were revealed ($b = -0.14$; $p = < .001$). In addition, extrinsic regulation was negatively associated with two autonomy-supportive strategies directed at individual students, namely with perceived individual interest stimulation ($b = -0.22$; $p = .006$) and acknowledgment of frustrations ($b = -0.09$; $p = .038$). At L2, students' shared perceptions of acceptance of students' frustration ($b = -0.11$; $p = .030$) were negatively associated with the overall extrinsic regulation of the class.

Behavioral Engagement. At L1, effort was positively associated with perceived choice provision ($b = 0.21$; $p = <.001$). In addition, effort was positively associated with acknowledgment of frustrations students observed to be directed to them individually ($b = 0.14$; $p = .004$). At L2, students' shared perceptions of rationale provision ($b = 0.23$; $p < .001$) and interest stimulation ($b = 0.22$; $p = .002$) positively predicted the average effort of the class.

Concerning persistence, at L1, positive associations with perceived choice provision were found ($b = 0.20$; $p = < .001$). Moreover, persistence was positively associated with perceived individual rationale provision ($b = 0.15$; $p = .030$) and acknowledgment of frustrations ($b = 0.13$; $p = .014$). At L2, students' shared perceptions of rationale provision ($b = 0.22$; $p = .006$) and interest stimulation ($b = 0.20$; $p = .034$) positively predicted the class-average persistence.

7.2.2. Differentiated autonomy support and associations with student outcomes (RQ4)

In Model 3.5, we added the outcomes to Model 2.5, which specified three independent variables at L1 (perceived differences in fostering relevance, perceived differences in autonomy support, and relative lack of autonomy support) and one independent variable at L2 (overall perceived differences in class) as predictors of the outcomes intrinsic motivation, extrinsic regulation, effort, and persistence, respectively (see Table 6).

Motivation. At L1, extrinsic regulation was positively associated with perceived relative lack of autonomy support ($b = 0.17$; $p = .012$). Thus, students who had the impression that they received less autonomy support from their teacher than their classmates were more likely to report engaging in activities for extrinsic reasons.

Behavioral Engagement. Regarding persistence, at L1, positive associations with perceived differences in fostering relevance ($b = 0.11$; $p = .022$) were confirmed. Students who perceived that their math teachers provided additional relevance instruction to specific students were more likely to continue working on math tasks, even if they were difficult.

Incremental validity. To assess the incremental validity of students' perceptions of autonomy support on motivation and engagement over and above students' math abilities, we added the CITO score as a predictor in all analyses, see Table S5. Through the inclusion of the CITO score as a covariate, we aim to better single out the unique associations of teachers' support with student motivation and engagement. When controlling for the CITO score, all previously confirmed associations remained statistically significant, while the CITO score had the theoretically expected associations with all outcomes.

8. Discussion

The present study evaluated the meaning of teachers' autonomy support in the multilevel classroom context from the student perspective. We studied whether there is a difference in students' perceptions of the autonomy support that is aimed at them individually (*individual autonomy* support*)* and at the whole classroom (*class-directed autonomy* support), and how students perceive inequalities in the autonomy support they receive compared to their classmates (*perceived differences in autonomy* support). In addition, we analyzed the associations between students' perceptions of individual, class-directed, and differences in autonomy support and their motivation and behavioral engagement.

First, our results showed that teachers' autonomy support directed at students or the whole class can be perceived by individual students (L1) as two distinct strategies. Both shared perceptions of individual and class-directed support seem to shape the overall autonomy-supportive atmosphere in the classroom (L2). Moreover, we found that primary school students observed differences in how teachers provided autonomy support and relevance instruction to distinct students. Students also perceived inequalities in the amount of autonomy support they received in comparison to their classmates (relative lack of autonomy support).

Overall, our results showed that studying teachers' autonomy support as a multilevel construct enables to uncover distinct associations of several aspects of perceived autonomy support (individual, class-directed, and differences in autonomy support) with qualitatively distinct outcomes (intrinsic versus extrinsic regulation).

Table 7
Correlations of students' perceptions of teachers' autonomy support with aspects of students' motivation and behavioral engagement.

Latent factor	Intrinsic motivation		Extrinsic motivation		Effort		Persistence	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Student level (Level 1)								
Choice provision	.24	<.001	−.23	<.001	.32	<.001	.25	<.001
Class-directed rationale provision	.10	.037	−.09	.071	.12	.019	.02	.389
Individual rationale provision	.01	.427	−.00	.481	.12	.027	.13	.019
Class-directed interest simulation	.35	<.001	−.13	.029	.35	<.001	.20	.001
Individual interest simulation	.34	<.001	−.24	<.001	.32	<.001	.25	.001
Class-directed accepting frustration	.11	.070	−.22	.001	.09	.107	.07	.173
Individual accepting frustration	.24	<.001	−.27	.000	.31	<.001	.24	<.001
Differences in relevance provision	.10	.056	.10	.080	.10	.065	.13	.032
Relative lack of autonomy support	.04	.238	.28	<.001	−.09	.082	−.05	.251
Differences in autonomy support	−.02	.381	.20	.002	−.06	.199	.02	.414
Class level (Level 2)								
Choice provision	.01	.481	−.01	.484	.06	.406	.15	.281
Rationale provision	.29	.111	−.01	.481	.31	.094	.33	.083
Interest simulation	.21	.202	.03	.455	.12	.312	.15	.281
Accepting frustration	.02	.467	.07	.395	−.06	.409	.01	.491
Differences in autonomy support	−.03	.446	.15	.264	.07	.393	.16	.262

Note. r = correlation coefficient; p = one-tailed p-value. The correlations were estimated in distinct analyses per outcome: the respective outcome was added to the measurement model of Model 4-adapted (with fixed loadings) concerning the correlations with perceptions of teachers' autonomy support, and Model 2.5, concerning the correlations with perceived differences in autonomy support.

Table 8
Model Fit Statistics Bayesian Multilevel SEM analyses (Models 3.1–3.5) Depicted per Outcome.

Models 3.1-3.5	DIC	PP*p*	95% CI		
Predicting intrinsic motivation					
Model 3.1 Choice provision	7747.72	0.36	−25.39	-	40.75
Model 3.2 Rationale provision	9214.81	0.50	−43.26	-	42.00
Model 3.3 Stimulating interest	9721.19	0.61	−51.80	-	36.92
Model 3.4 Accepting frustration	9048.89	0.60	−42.43	-	33.54
Model 3.5 Perceived differences in autonomy support	15758.94	0.47	−67.55	-	78.51
Predicting extrinsic motivation					
Model 3.1 Choice provision	8153.57	0.35	−25.51	-	41.17
Model 3.2 Rationale provision	9630.68	0.48	−42.02	-	44.01
Model 3.3 Stimulating interest	10156.39	0.47	−43.46	-	45.18
Model 3.4 Accepting frustration	9453.14	0.30	−28.17	-	48.53
Model 3.5 Perceived differences in autonomy support	16144.55	0.54	−76.09	-	68.80
Predicting effort					
Model 3.1 Choice provision	9242.05	0.25	−24.20	-	53.38
Model 3.2 Rationale provision	10714.05	0.36	−34.76	-	50.36
Model 3.3 Stimulating interest	11243.94	0.24	−30.55	-	65.54
Model 3.4 Accepting frustration	10552.11	0.42	−38.25	-	50.97
Model 3.5 Perceived differences in autonomy support	17260.95	0.38	−65.19	-	92.41
Predicting persistence					
Model 3.1 Choice provision	8107.52	0.50	−31.57	-	35.24
Model 3.2 Rationale provision	9577.46	0.57	−48.29	-	38.48
Model 3.3 Stimulating interest	10109.11	0.69	−59.201	-	32.35
Model 3.4 Accepting frustration	9411.31	0.43	−33.30	-	41.63
Model 3.5 Perceived differences in autonomy support	16117.90	0.38	−54.89	-	88.74

8.1. Investigating teachers' autonomy support in the multilevel classroom context

The first aim of this study was to examine the factorial structure of students' perceptions of teachers' class-directed and individual autonomy support (RQ1) and perceived differences in autonomy support (RQ2) to find out how teachers' autonomy support can best be operationalized. Our findings indicated that the aggregates of the different autonomy-supportive strategies (e.g., providing choice, fostering relevance) directed at either the classroom or at individual students reflected the same classroom-level factor. This class-level factor reflects the shared perception within the classrooms of the autonomy-supportive atmosphere created by the teacher using a strategy for individual and class-directed support. Thus, the overall autonomy-supportive atmosphere in a classroom seems to be shaped by both individual and class-directed autonomy support.

At the student level, however, a class-directed factor could be distinguished from an individual autonomy-supportive factor for the strategies rationale provision, interest stimulation, and acknowledgment of frustration. Only for the strategy of choice provision a 1-factor model at the student- and classroom level seemed to be a more appropriate conceptualization. This finding could reflect teachers' tendencies to inform the whole class about additional options when allowing them for individual students, or it could be a result of students overhearing these possibilities and asking for similar options.

In addition, our study yielded new findings on distinct facets of perceived differences in autonomy support at the student level: differences in autonomy support (individual students' perceptions that some students in the classroom receive more autonomy support than others) and relative lack of autonomy support (individual students' perception that other students receive more autonomy support than they receive themselves). We also revealed a separate component referring to perceived differences in relevance instruction (individual students' perceptions that some students in the classroom receive more relevance-related instruction than others). This component might reflect teachers refining the justifications they provide for different students, e.g., based on what they think a student might find interesting or useful. By comparison, when addressing individual students' boredom or frustration, teachers might react similarly across situations. Thus, it might be easier to adapt relevance instruction to individual students compared to other autonomy-supportive strategies.

In all, our findings regarding the first two research questions highlight the importance of understanding autonomy support as a multifaceted construct and investigating autonomy-supportive processes both at the student and at the class level. Specifically, the findings of the present study suggest that the overall autonomy-supportive climate within a classroom is informed by how students perceive the autonomy-support directed to them individually and to the classroom.

Table 9
Results of Bayesian multilevel SEMs: Predictive effects of different strategies of autonomy support on students' motivation and behavioral engagement.

		Intrinsic motivation			Extrinsic regulation			Effort			Persistence		
Model	Predictors	*b*	*(SE)*	*p*	*b*	*(SE)*	*p*	*b*	*(SE)*	*p*	*b*	*(SE)*	*p*
Model 3.1	Choice provision (L1)	**.21**	**(.05)**	**<.001**	**−.14**	**(.04)**	**<.001**	**.21**	**(.04)**	**<.001**	**.20**	**(.05)**	**<.001**
	Overall choice provision in class (L2)	−.08	(.10)	.424	−.09	(.05)	.098	−.01	(.08)	.910	−.04	(.08)	.658
	Residuals												
	Residual (L1)	.53	(.05)	<.001	.35	(.03)	<.001	.31	(.03)	<.001	.41	(.04)	<.001
	Residual (L2)	.07	(.05)	<.001	.01	(.01)	<.001	.05	(.03)	<.001	.05	(.03)	<.001
Model 3.2	Class-directed rationale provision (L1)	**.13**	**(.06)**	**.050**	−.09	(.05)	.086	.05	(.06)	.378	−.09	(.07)	.144
	Individual rationale provision (L1)	−.10	(.07)	.144	.07	(.05)	.196	.03	(.06)	.616	**.15**	**(.07)**	**.030**
	Overall rationale provision in class (L2)	**.28**	**(.08)**	**<.001**	−.05	(.06)	.392	**.23**	**(.06)**	**<.001**	**.22**	**(.07)**	**.006**
	Residuals												
	Residual (L1)	.56	(.05)	<.001	.37	(.03)	<.001	.35	(.03)	<.001	.44	(.04)	<.001
	Residual (L2)	.02	(.03)	<.001	.02	(.02)	<.001	.01	(.01)	<.001	.01	(.02)	<.001
Model 3.3	Class-directed interest stimulation (L1)	.19	(.10)	.070	.10	(.09)	.214	.13	(.09)	.136	−.01	(.10)	.884
	Individual interest stimulation (L1)	.08	(.11)	.470	**−.22**	**(.09)**	**.006**	.08	(.09)	.354	.17	(.11)	.106
	Overall interest stimulation in class (L2)	**.28**	**(.08)**	**.002**	−.07	(.06)	.200	**.22**	**(.06)**	**.002**	**.20**	**(.08)**	**.034**
	Residuals												
	Residual (L1)	.50	(.04)	<.001	.35	(.03)	<.001	.31	(.03)	<.001	.42	(.04)	<.001
	Residual (L2)	.02	(.03)	<.001	.01	(.01)	<.001	.01	(.02)	<.001	.02	(.04)	<.001
Model 3.4	Class-directed frustration acceptance (L1)	.03	(.06)	.622	−.06	(.04)	.146	.00	(.05)	.940	−.00	(.05)	.954
	Individual frustration acceptance (L1)	**.13**	**(.06)**	**.028**	**−.09**	**(.04)**	**.038**	**.14**	**(.05)**	**.004**	**.13**	**(.05)**	**.014**
	Overall frustration acceptance in class (L2)	.07	(.10)	.472	**−.11**	**(.05)**	**.030**	−.00	(.08)	.988	−.03	(.09)	.700
	Residuals												
	Residual (L1)	.55	(.05)	<.001	.36	(.03)	<.001	.34	(.03)	<.001	.43	(.04)	<.001
	Residual (L2)	.09	(.05)	<.001	.01	(.01)	<.001	.06	(.03)	<.001	.06	(.04)	<.001
Model 3.5	Differences in relevance instruction (L1)	.08	(.05)	.134	.03	(.04)	.476	.08	(.04)	.052	**.11**	**(.05)**	**.022**
	Differences in autonomy support (L1)	−.06	(.09)	.524	−.02	(.07)	.742	.03	(.08)	.722	.09	(.09)	.314
	Relative lack of autonomy support (L1)	.06	(.09)	.522	**.17**	**(.07)**	**.012**	−.09	(.08)	.232	−.12	(.09)	.176
	Overall perceived differences in class (L2)	−.02	(.11)	.860	.10	(.06)	.058	.05	(.08)	.520	.14	(.08)	.104
	Residuals												
	Residual (L1)	.56	(.05)	<.001	.35	(.03)	<.001	.35	(.03)	<.001	.44	(.04)	<.001
	Residual (L2)	.09	(.05)	<.001	.01	(.01)	<.001	.05	(.03)	<.001	.04	(.03)	<.001

Note. b = unstandardized regression coefficient; SE = standard error; p = two-tailed p-value; statistically significant results $p < .05$ are highlighted in bold. L1: Level 1; L2: Level 2.

8.2. *Associations of students' perceptions of class-directed and individual autonomy support and perceived differences with students' motivation*

The second aim of our study was to examine the associations of students' perceptions of class-directed and individual autonomy support (RQ3) as well as perceived differences in autonomy support (RQ4) with students' motivation. Our results showed positive associations between perceived individual and class-directed autonomy support and students' intrinsic motivation and effort. This is in line with earlier research highlighting that single autonomy-supportive strategies may not translate into favorable learning conditions "when used in isolation" (Reeve, 2016, p. 149). Thus, it might not be sufficient to tailor autonomy support to individual students, or to prepare complete lesson materials for the whole class in an autonomy-supportive way: Both individual and class-level support should be high in order to yield the optimal results for students' motivation and engagement.

It is noteworthy that the findings regarding the associations of strategies of individual and class-directed autonomy support with students' motivation and behavioral engagement might mirror effective instructional practices: Teachers may tend to communicate rationales for a topic or limits of behaviors to the whole class and design interesting lesson materials for the whole class. In contrast, acknowledging student frustration may be a strategy teachers employ when interacting with individual students.

8.3. *Perceived differences in teachers' autonomy support*

Another aim of our study was to investigate how perceived differences in autonomy support were associated with students' motivation. We found that perceiving differences in teachers' autonomy support in the classroom may not necessarily be associated with negative outcomes in students: Only students' perceptions that other students receive more support than they do (i.e., a perceived relative lack of autonomy support) were positively associated with extrinsic regulation.

This study adds to earlier research showing that students notice inequalities in the amount of autonomy support provided for different students in the classroom, which may have an impact on their motivation and engagement (e.g., Chatzisarantis et al., 2019). Chatzisarantis et al. (2019) have found that favorable support (perceiving to receive more autonomy support than classmates) is not as beneficial as equal autonomy support (perceiving to receive similar autonomy support than peers), e.g., because students might evaluate extra support as unfair. Even if students who perceive to receive more autonomy support than their peers would thrive more than their peers in terms of their motivation and engagement due to increased autonomy, this effect could drive societal inequalities. Specifically, studies have shown that teachers tend to provide additional support for specific groups of students: Hornstra et al. (2015) found that teachers provided less autonomy support to "at-risk" students, e.g., to students from families with low socioeconomic backgrounds, than to their peers. Likewise, teachers reported relatively lower levels of involvement concerning students from families with low socioeconomic backgrounds compared to students from high socioeconomic backgrounds, and lower autonomy support for lower achieving students (Bloem et al., 2023). Therefore, educational research could benefit from studying the perceptions of favorable and unfavorable autonomy support in the classroom from the perspective of individual students, their peers, and their teachers, and their associations with students' personal characteristics, their family background, and student outcomes.

8.4. Bayesian Multilevel analyses with small sample sizes at the class level

In the present study, we used Bayesian multilevel analyses to study a substantial research problem in the classroom context with a relatively small sample size at the class level. Bayes estimation allowed us to fit the models without convergence issues (which often occur with ML analyses in small samples). Doubly latent multilevel modeling has become increasingly popular in educational research (e.g., Morin et al., 2014) as it enables controlling for both measurement error and sampling error in measures (e.g., Lüdtke et al., 2011), but it usually requires a large number of classrooms or schools to produce stable estimates. Educational researchers often have limited resources for collecting data in schools. Moreover, from an ethical perspective, it would be desirable to be able to address important questions, e.g., how teachers' differentiation is perceived within classrooms by students, with less time and effort investment by schools and students. We capitalized on an advanced statistical methodology and illustrated how Bayesian estimation for doubly latent multilevel models for data with a small L2 sample size (Zitzmann et al., 2016) can be used to investigate the multilevel nature of student perceptions of teachers' autonomy support in the classroom context.

Instead of using Bayesian analyses, users of SPSS, SAS, HLM, and R can apply the restricted maximum likelihood (REML) estimator to estimate multilevel models with a small sample size at Level 2, which is not implemented in M*plus*. REML estimation has the potential to yield results that are identical to Bayesian estimation with uninformative priors, in case the correct *df* are specified (for details see Elff et al., 2021). A notable difference is that REML may result in negative variance estimates (e.g., El Leithy et al., 2016), while the variance can be restricted to be positive when using (uninformative) priors in Bayesian estimation (Fox & Smink, 2023; Hox & McNeish, 2020, p. 221).

8.5. Limitations

Certain limitations should be noted when interpreting our results. First, the reliabilities for some of the scales were relatively low. In order to measure individual, class-directed, and perceived differences in autonomy support with equivalent items, we limited ourselves to a small number of items. Considering the factor loadings, some of the items seemed to mainly capture individual students' perceptions. Given that cross-level invariance did not hold (see Supplementary Material), the respective latent constructs did not have the same meaning across levels (e.g., Stapleton et al., 2016).

Second, self-report data can be imprecise, e.g., because the meaning of items is not well understood (Aiken & West, 1990). Thus, within-class differences for student ratings of class-level constructs are likely, also because students might miss autonomy-supportive instruction at random or are more sensitive to it. In future research, broader scales could be developed to improve the measurement of shared perceptions. The perspective of teachers, classmates, or observers could additionally be evaluated. Our research findings point to the possibility that students' individual experience of autonomy support is the most important referent for this specific construct. Yet, using ambulatory assessments and a focus on within-person differences in the perceptions of autonomy support might help to yield more information on how students perceive classrooms (Fahrenberg, 1996).

Third, our study only referred to one age group, the domain of mathematics, and our sample of teachers and classes can be considered relatively small. Therefore, our study's findings might not be fully representative of the variability present in the population.

Fourth, our study solely referred to general perceptions of differences in teachers' autonomy support within a classroom and students' perceptions of unfavorable support (i.e., receiving less support than classmates), and did not include perceptions of favorable autonomy support (receiving more support than peers). Moreover, we did not measure competence or relatedness support or teachers' controlling instructional styles. Teachers' competence- and relatedness-supportive behaviors might interact with their provision of autonomy support in shaping students' motivation and engagement within or across domains. For example, Steingut et al. (2017) found that rationale provision can undermine students' feelings of competence, and additional competence support might be essential to counteract this effect. It might be worthwhile for future research to study more domains and a greater set of need-supportive and need-thwarting strategies that teachers may use in different situations in a classroom.

As is common with Multilevel SEM analyses, our analyses were limited due to convergence problems. We could not target relevant research questions due to non-convergence or inadmissible solutions (see also Jak et al., 2021), such as considering all predictors and outcomes at once. Therefore, we could not test whether the autonomy-supportive strategies work at best in concert, as is claimed in SDT (e.g., Patall et al., 2013). Moreover, autonomy-supportive strategies may strengthen each other; for instance, choices can strengthen the effects of relevance instruction (Rosenzweig et al., 2019). It is possible that a combination of high individual and high class-directed autonomy support would be most beneficial for students. Unfortunately, we were not able to test a model considering all latent factors simultaneously with the current data, because this led to suppression effects (Beckstead, 2012, see https://osf.io/s87wb/?view_only=1eb3263e49664444b7952f0b61918e92). Future research could study these questions with a bigger sample size of classes.

Finally, it needs to be noted that our data was gathered in 2017 and our study procedure was in line with the prevailing ethical guidelines at the time, which, for a non-invasive questionnaire study, allowed passive informed consent and did not require a standard routine for obtaining ethical approval by a faculty review board. Information letters on the study were sent out to all parents or legal representatives according to the procedures typically used by the schools to contact parents and legal representatives. The researchers checked with the teachers whether the letters were indeed sent out. These letters included detailed information on the study's purpose and whom to contact, as well as an explanation of the procedure to opt out of participation. We were contacted by several parents who did not want their children to participate. At the beginning of the data collection, students were instructed that they could omit items and could stop completing the questionnaire at any time. To obtain demographic information about our sample, we asked for language spoken at home and country of birth, but we did not ask for further personal data (such as religion or health) or ask invasive questions on sensitive topics. We believe that the burden and risk of our research for the participating students were limited and that the opt-out procedure was clear. All our procedures were in line with the guidelines of the local institutional review board and the ethical principles of APA that were in place at the time (standard 3.10, American Psychological Association, 2016).

8.6. Implications for educational research and suggestions for future research

The present study has several important implications for the future measurement of teachers' autonomy support. First, our study can help researchers determine how to measure and model autonomy support. We found that analyzing autonomy support as a multilevel construct in all its facets (individual, class-directed, perceived differences) helps to reveal the predictive value of the various facets of autonomy support students experience in a classroom. Yet, individual autonomy support can have undesirable side effects on students' outcomes if students perceive a mismatch between their individual support and the support their classmates receive.

Second, researchers might consider the conceptual difference between class-directed and individual support (e.g., Ntoumanis, 2023) and adapt their measurement accordingly. That is, it might be useful for researchers to make a conscious choice regarding their target of interest rather than mixing items for individual and class-directed support because these two dimensions might refer to distinct situations and processes in the classroom. As an example, Hospel and Galand (2016) focused on autonomy support as an aggregated class-level construct and found marginally significant effects on students' behavioral engagement, no effects on self-regulation, and statistically significant effects on students' positive emotions. They concluded that teachers' autonomy support only has a "complementary role" in the classroom. Our findings suggest that when researchers study individual or class-directed support in the eyes of individual students (i.e., as an unaggregated construct), the role of autonomy support for students' outcomes may be more pronounced.

Third, in our study, several items assessing individual autonomy support had slightly higher ICC(1)s than the parallel items targeting class-directed support. However, one would expect higher ICC(1)s for the class-directed items because class-directed support should be observed similarly by distinct students, which should lead to consistent ratings within classrooms. Moreover, given that teachers might differ in their class-directed support, the variability between classrooms should be higher. It needs to be examined whether using parallel items, which is the common approach when assessing classroom processes affecting individual students and classrooms (e.g., Downer et al., 2015; Lüdtke et al., 2011), is the most informative approach in classroom research. When measuring class-directed need support, measures could also more explicitly address the situations in which universal support is targeted at the class collective to benefit the entire group of students. For example, measures used to assess whole-class instruction could refer to strategies that aim to encourage collaboration among all students (for relatedness support), to provide feedback regarding concepts that the majority of students struggle with (for competence support), or to speak to examples on how content connects to daily life (for autonomy support).

Fourth, we focused on students' perceptions of teachers' general provision of autonomy support, considering a set of well-acknowledged autonomy-supportive strategies such as their provision of choices (see an overview by Ahmadi et al., 2023, AS1: Allow for student input or choice). However, our assessment does not enable to consider the extent to which students thrive from choices. Indeed, it might be possible that teachers adapt to students, which implies that teachers reduce the number of choices given to a student, once they realize that the student becomes overwhelmed when presented with too many choices.

Future research could potentially benefit by explicitly considering the values and preferences of students in the measurements, which reflects another autonomy-supportive strategy, namely 'teaching in students' preferred way' (Ahmadi et al., 2023, AS2; Jang et al., 2016). This could lead to adapting items to "My teacher takes into account my values and preferences when engaging with me" or "My teacher encourages me to study the way I want, e.g., through providing choices."

8.7. Implications for educational practice

Concerning educational practice, our study sheds light on undesired side effects associated with individual motivational support. Specifically, it may be important to consider how students perceive the way their teachers distribute autonomy support within the classroom, to gain a better understanding of how autonomy support is associated with students' extrinsic regulation.

How should teachers then provide autonomy support within a classroom? Our findings are in line with the conclusion of Martin et al. (2010) that "the bulk of variance in motivation and engagement resides at the student level" (p. 987). Our results indicate that support for autonomy can be associated with individual student motivation and engagement, as long as individual students perceive it as directed toward them, either through one-to-one or classroom instruction. The findings of our study also imply that it might be most beneficial if teachers aim to distribute their autonomy support equally (e.g., Chatzisarantis et al., 2019), and to provide support not only to the students who actively ask for it (e.g., Matos et al., 2018).

We know from earlier research that student characteristics, such as their academic achievement, can bias their teacher ratings (e.g., Marsh & Roche, 1997). Thus, unmotivated students may notice that classmates receive more support than themselves, even if this is not valid. Future research could study ways in which teachers can monitor how much autonomy support students individually perceive, e.g., using digital tools. Focusing on perceived differences in autonomy support can help to identify motivational problems in students.

Finally, the effects of class-directed and individual autonomy support on student outcomes could depend on the interplay with structure and involvement targeted at students and whole classrooms. There might be "demanding but warm" types of teachers (Kleinfeld, 1975; Irvine & Fraser, 1998) who provide certain students with less individual autonomy support but in an involved and caring way, which could potentially buffer against the negative effect of low individual autonomy support.

8.8. Conclusion

As Dunson (2001) highlighted, using Bayesian estimation allows for analyzing a real-life, complex research problem (e.g., how students perceive their teachers' autonomy support in the classroom) with a highly flexible approach. Capitalizing on Bayesian multilevel analysis, we found that individual students' perceptions of their teachers' individual and class-directed autonomy support and resulting discrepancies may play distinct roles in students' motivation. Overall, our results reveal that adapting autonomy support to individual students might not be the crucial element to promote student motivation and engagement in a classroom. Students might thrive most when they perceive their teachers as autonomy-supportive in various classroom situations when teachers interact with the entire classroom and with individual students.

Author note

Barbara Flunger and Anouk Verdonschot contributed equally to this study.

This research was funded in part by funds from the Faculty of Social Sciences, Utrecht University. This research did not receive any further grant from funding agencies in the public, commercial, or not-for-profit sectors.

Transparency and openness

The study materials and analysis codes for this paper can be found at https://osf.io/s87wb/?view_only=1eb3263e49664444b7952f0b61918e92. We do not share the data, because our informed consents at the time of data collection did not inform participants about the possibility of making the data public. Correspondence concerning this article should be addressed to Barbara Flunger, Utrecht University, Heidelberglaan 1, Utrecht, the Netherlands. E-mail: b.flunger@uu.nl.

CRediT authorship contribution statement

Barbara Flunger: Conceptualization, Formal analysis, Methodology, Project administration, Writing – review & editing, Supervision. **Anouk Verdonschot:** Data curation, Formal analysis, Writing – original draft. **Steffen Zitzmann:** Methodology, Writing – review & editing.

Lisette Hornstra: Conceptualization, Investigation, Writing – review & editing. **Tamara van Gog:** Supervision, Writing – review & editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2023.101873.

References

Aelterman, N., Vansteenkiste, M., Haerens, L., Soenens, B., & Fontaine, J. R. J. (2019). Toward an integrative and fine-grained insight in motivating and demotivating teaching styles: The merits of a circumplex approach. *Journal of Educational Psychology, 111*(3), 497–521. https://doi.org/10.1037/edu0000293.supp

Ahmadi, A., Noetel, M., Parker, P., Ryan, R. M., Ntoumanis, N., Reeve, J., Beauchamp, M., Dicke, T., Yeung, A., Ahmadi, M., Bartholomew, K., Chiu, T. K. F., Curran, T., Erturan, G., Flunger, B., Frederick, C., Froiland, J. M., González-Cutre, D., Haerens, L., & Lonsdale, C. (2023). A classification system for teachers' motivational behaviors recommended in self-determination theory interventions. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000783. Advance online publication.

Aiken, L. S., & West, S. G. (1990). Invalidity of true experiments: Self-report pretest biases. *Evaluation Review, 14*(4), 374–390. https://doi.org/10.1177/0193841X9001400403

American Psychological Association. (2016). Revision of ethical standard 3.04 of the "ethical principles of psychologists and code of conduct" (2002, as amended 2010). *American Psychologist, 71*, 900.

Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis using Mplus: Technical implementation, 1–38*. http://statmodel2.com/download/Bayes3.pdf.

Assor, A. (2012). Allowing choice and nurturing an inner compass: Educational practices supporting students' need for autonomy. In S. L. Christenson, C. Wylie, & A. L. Reschly (Eds.), *Handbook of Research on student engagement (issue June 2012* (pp. 421–439). Springer. https://doi.org/10.1007/978-1-4614-2018-7.

Bong, M. (2003). Choices, evaluations, and opportunities for success: Academic motivation of Korean adolescents. In F. Pajares, & T. C. Urdan (Eds.), *International perspectives: 3. Adolescence and education* (pp. 323–345). Greenwich, CT: Information Age.

Beckstead, J. W. (2012). Isolating and examining sources of suppression and multicollinearity in multiple linear regression. *Multivariate Behavioral Research, 47* (2), 224–246. https://doi.org/10.1080/00273171.2012.658331

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein, & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). Jossey-Bass.

Bloem, J., Flunger, B., Stroet, K., & Hornstra, L. (2023). Differences in need-supportive teaching: The role of teachers' attitudes toward a low SES. *Social Psychology of Education, Special Issue: Stereotypes and Prejudice in School*, 1–51. https://doi.org/10.1007/s11218-023-09831-w

Bureau, J. S., Howard, J. L., Chong, J. X., & Guay, F. (2022). Pathways to student motivation: A meta-analysis of antecedents of autonomous and controlled motivations. *Review of Educational Research, 92*(1), 46–72. https://doi.org/10.3102/00346543211042426

Chatzisarantis, N. L., Ada, E. N., Ahmadi, M., Caltabiano, N., Wang, D., Thogersen-Ntoumani, C., & Hagger, M. S. (2019). Differential effects of perceptions of equal, favourable and unfavourable autonomy support on educational and well-being outcomes. *Contemporary Educational Psychology, 58*, 33–43. https://doi.org/10.1016/j.cedpsych.2019.02.002

Domen, J., Hornstra, L., Weijers, D., Veen, I., & Peetsma, T. (2019). Differentiated need support by teachers: Student-specific provision of autonomy and structure and relations with student motivation. *British Journal of Educational Psychology*, 1–21. https://doi.org/10.1111/bjep.12302

Donker, M. H., van Vemde, L., Hessen, D. J., van Gog, T., & Mainhard, T. (2021). Observational, student, and teacher perspectives on interpersonal teacher behavior: Shared and unique associations with teacher and student emotions. *Learning and Instruction, 73*. https://doi.org/10.1016/j.learninstruc.2020.101414

Downer, J. T., Stuhlman, M., Schweig, J., Martínez, J. F., & Ruzek, E. (2015). Measuring effective teacher-student interactions from a student perspective: A multi-level analysis. *The Journal of Early Adolescence, 35*(5–6), 722–758. https://doi.org/10.1177/0272431614564059

Dunson, D. B. (2001). Commentary: Practical advantages of Bayesian analysis of epidemiologic data. *American Journal of Epidemiology, 153*(12), 1222–1226. https://doi.org/10.1093/aje/153.12.1222

El Leithy, H. A., Wahed, Z. A. A., & Abdallah, M. S. (2016). On non-negative estimation of variance components in mixed linear models. *Journal of Advanced Research, 7*(1), 59–68. https://doi.org/10.1016/j.jare.2015.02.001

Elff, M., Heisig, J. P., Schaeffer, M., & Shikano, S. (2021). Multilevel analysis with few clusters: Improving likelihood-based methods to provide unbiased estimates and accurate inference. *British Journal of Political Science, 51*(1), 412–426. https://doi.org/10.1017/S0007123419000097

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Fahrenberg, J. (1996). Ambulatory assessment: Issues and perspectives. In J. Fahrenberg, & M. Myrtek (Eds.), *Ambulatory Assessment: Computer-assisted psychological and psychophysiological methods in monitoring and field studies* (pp. 3–20). Hogrefe & Huber.

Flunger, B., Hollmann, L., Hornstra, L., & Murayama, K. (2022). It's more about a lesson than a domain: Lesson-specific autonomy support, motivation, and engagement in Math and a second language. *Learning and Instruction, 77*. https://doi.org/10.1016/j.learninstruc.2021.101500

Flunger, B., Mayer, A., & Umbach, N. (2019). Beneficial for some or for everyone? Exploring the effects of an autonomy-supportive intervention in the real-life classroom. *Journal of Educational Psychology, 111*(2), 210–234. https://doi.org/10.1037/edu0000284

Flunger, B., Trautwein, U., Nagengast, B., Lüdtke, O., Niggli, A., & Schnyder, I. (2015). The Janus-faced nature of time spent on homework: Using latent profile analyses to predict academic achievement over a school year. *Learning and Instruction, 39*, 97–106. https://doi.org/10.1016/j.learninstruc.2015.05.008

Fox, J.-P., & Smink, W. A. C. (2023). Assessing an alternative for "negative variance components": A gentle introduction to Bayesian covariance structure modeling for negative associations among patients with personalized treatments. *Psychological Methods, 28*(1), 1–20. https://doi.org/10.1037/met0000442

Garon-Carrier, G., Boivin, M., Guay, F., Kovas, Y., Dionne, G., Lemelin, J. P., Séguin, J. R., Vitaro, F., & Tremblay, R. E. (2016). Intrinsic motivation and achievement in mathematics in elementary school: A longitudinal investigation of their association. *Child Development, 87*(1), 165–175. https://doi.org/10.1111/cdev.12458

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*(1), 72–91. https://doi.org/10.1037/a0032138

Hamaker, E. L., & Klugkist, I. (2011). Bayesian estimation of multilevel models. In J. Hox, & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 145–170). https://doi.org/10.4324/9780203848852.ch8. Routledge.

Hardre, P. L., & Reeve, J. (2003). A motivational model of rural students' intentions to persist in, versus drop out of, high school. *Journal of Educational Psychology, 95*(2), 347–356. https://doi.org/10.1037/0022-0663.95.2.347

Helm, F., Arens, A. K., & Möller, J. (2020). Perceived teacher unfairness and student motivation in math and German: An application of the generalized internal/external frame of reference model. *Learning and Individual Differences, 81*. https://doi.org/10.1016/j.lindif.2020.101891

Hofmans, J., Morin, A. J., Breitsohl, H., Ceulemans, E., Chénard-Poirier, L. A., Driver, C. C., Fernet, C., Gagné, M., Gillet, N., González-Romá, V., Grimm, K. J., Hamaker, E. L., Hau, K.-T., Houle, S. A., Howard, J. L., Kline, R. B., Kuijpers, E., Leyens, T., Litalien, D., & Wille, B. (2021). The baby and the bathwater: On the need for substantive–methodological synergy in organizational research. *Industrial and Organizational Psychology, 14*(4), 497–504. https://doi.org/10.1017/iop.2021.111

Hornstra, L., Mansfield, C., Van Der Veen, I., Peetsma, T., & Volman, M. (2015). Motivational teacher strategies: The role of beliefs and contextual factors. *Learning Environments Research, 18*(3), 363–392. https://doi.org/10.1007/s10984-015-9189-y

Hornstra, L., Stroet, K., van Eijden, E., Goudsblom, J., & Roskamp, C. (2018). Teacher expectation effects on need-supportive teaching, student motivation, and engagement: A self-determination perspective. *Educational Research and Evaluation, 24*(3–5), 324–345. https://doi.org/10.1080/13803611.2018.1550841

Hornstra, L., van der Veen, I., & Peetsma, T. (2016). Domain-specificity of motivation: A longitudinal study in upper primary school. *Learning and Individual Differences, 51*, 167–178. https://doi.org/10.1016/j.lindif.2016.08.012

Hospel, V., & Galand, B. (2016). Are both classroom autonomy support and structure equally important for students' engagement? A multilevel analysis. *Learning and Instruction, 41*, 1–10. https://doi.org/10.1016/j.learninstruc.2015.09.001

Howard, J. L., Bureau, J., Guay, F., Chong, J. X., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives on Psychological Science, 16*(6), 1300–1323. https://doi.org/10.1177/1745691620966789

Hox, J., & McNeish, D. (2020). Small samples in multilevel modeling. In R. Van de Schoot, & M. Miočević (Eds.), *Small sample size solutions: A guide for applied researchers and practitioners*. Taylor & Francis, 2020.

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge.

Irvine, J. J., & Fraser, J. W. (1998). Warm demanders. *Education Week, 17*(35), 56–57.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73*(2), 509–527. https://doi.org/10.1111/1467-8624.00421

Jak, S., Jorgensen, T. D., & Rosseel, Y. (2021). Evaluating cluster-level factor models with lavaan and Mplus. *Psyche, 3*(2), 134–152. https://doi.org/10.3390/psych3020012

Jang, H., Reeve, J., & Halusic, M. (2016). A new autonomy-supportive way of teaching that increases conceptual learning: Teaching in students' preferred ways. *The Journal of Experimental Education, 84*(4), 686–701. https://doi.org/10.1080/00220973.2015.1083522

Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2010). *Wetenschappelijke verantwoording van de toetsen LOVS rekenen-wiskunde voor groep 3 tot en met 8 [Scientific justification of the mathematics test for grade 1 until grade 6*. The Netherlands: Cito: Arnhem.

Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 650–673). Guilford Press.

King, R. B., & McInerney, D. M. (2014). Culture's consequences on student motivation: Capturing cross-cultural universality and variability through personal investment

theory. *Educational Psychologist, 49*(3), 175–198. https://doi.org/10.1080/00461520.2014.926813
Kleinfeld, J. (1975). Effective teachers of Eskimo and Indian students. *The School Review, 83*, 301–344.
Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2× 2 taxonomy of multilevel latent contextual models: Accuracy–bias trade-offs in full and partial error correction models. *Psychological Methods, 16*(4), 444. https://doi.org/10.1037/a0024376
Maas, C. J. M., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica, 58*(2), 127–137. https://doi.org/10.1046/j.0039-0402.2003.00252.x
Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist, 47*(2), 106–124. https://doi.org/10.1080/00461520.2012.670488
Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research, 44*(6), 764–802. https://doi.org/10.1080/00273170903333665
Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist, 52*(11), 1187–1197. https://doi.org/10.1037/0003-066X.52.11.1187
Martin, A. J., Malmberg, L. E., & Liem, G. A. D. (2010). Multilevel motivation and engagement: Assessing construct validity across students and schools. *Educational and Psychological Measurement, 70*(6), 973–989. https://doi.org/10.1177/0013164410378089
Matos, L., Reeve, J., Herrera, D., & Claux, M. (2018). Students' agentic engagement predicts longitudinal increases in perceived autonomy-supportive teaching: The squeaky wheel gets the grease. *The Journal of Experimental Education, 86*(4), 579–596. https://doi.org/10.1080/00220973.2018.1448746
McNeish, D. (2017). Small sample methods for multilevel modeling: A colloquial elucidation of REML and the kenward-roger correction. *Multivariate Behavioral Research, 52*(5), 661–670. https://doi.org/10.1080/00273171.2017.1344538
Morin, A. J. S., Marsh, H. W., Nagengast, B., & Scalas, L. F. (2014). Doubly latent multilevel analyses of classroom climate: An illustration. *The Journal of Experimental Education, 82*(2), 143–167. https://doi.org/10.1080/00220973.2013.769412
Mouratidis, A., Michou, A., Aelterman, N., Haerens, L., & Vansteenkiste, M. (2018). Begin-of-school-year perceived autonomy-support and structure as predictors of end-of-school-year study efforts and procrastination: The mediating role of autonomous and controlled motivation. *Educational Psychology, 38*(4), 435–450. https://doi.org/10.1080/01443410.2017.1402863
Muthén, B. (2010). *Bayesian analysis in Mplus: A brief introduction* (Unpublished manuscript). Retrieved from https://www.statmodel.com/download/IntroBayesVersion%203.pdf.
Muthén, B., & Asparouhov, T. (2012). Bayesian SEM : A more flexible representation of substantive theory. *Psychological Methods, 17*(3), 313–335. https://doi.org/10.1037/a0026802
Ntoumanis, N. (2023). The bright, dark, and dim light colors of motivation: Advances in conceptualization and measurement from a self-determination theory perspective. In A. J. Elliot (Ed.), *Advances in motivation science, 9*. Elsevier.
OECD. (2016). *Netherlands 2016: Foundations for the future*. Paris: OECD Publishing.
Patall, E. A., Dent, A. L., Oyer, M., & Wynn, S. R. (2013). Student autonomy and course value: The unique and cumulative roles of various teacher practices. *Motivation and Emotion, 37*(1), 14–32. https://doi.org/10.1007/s11031-012-9305-6
Rawsthorne, L. J., & Elliot, A. J. (1999). Achievement goals and intrinsic motivation: A meta-analytic review. *Personality and Social Psychology Review, 3*(4), 326–344. https://doi.org/10.1207/s15327957pspr0304_3
Reeve, J. (2016). Autonomy-supportive teaching: What it is, how to do it. In J. C. K. Wang, W. C. Liu, & R. M. Ryan's (Eds.), *Motivation in educational research: Translating theory into classroom practice (ch. 7, 129–152)*. Springer.
Reeve, J., & Cheon, S. H. (2021). Autonomy-supportive teaching: Its malleability, benefits, and potential to improve educational practice. *Educational Psychologist, 56*(1), 54–77. https://doi.org/10.1080/00461520.2020.1862657
Rosenzweig, E. Q., Harackiewicz, J. M., Priniski, S. J., Hecht, C. A., Canning, E. A., Tibbetts, Y., & Hyde, J. S. (2019). Choose your own intervention: Using choice to enhance the effectiveness of a utility-value intervention. *Motivation Science, 5*(3), 269. https://doi.org/10.1037/mot0000113
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. https://doi.org/10.1037/0022-3514.57.5.749
Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020
Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*. https://doi.org/10.1016/j.cedpsych.2020.101860
Scheltens, F., Hollenberg-Vos, J., Limpens, G., & Stolwijk, R. (2020). Testing in mathematics education in the Netherlands. In M. Van den Heuvel-Panhuizen (Ed.), *National reflections on The Netherlands didactics of mathematics: Teaching and learning in the context of realistic mathematics education* (pp. 303–330). Springer Open.
Scherrer, V., & Preckel, F. (2019). Development of motivational variables and self-esteem during the school career: A meta-analysis of longitudinal studies. *Review of Educational Research, 89*(2), 211–258. https://doi.org/10.3102/0034654318819127
Sempels, H. (2014). *Self-Determination and support for it in Flanders: Effects of the learners' age and study level, 1–53*.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society - Series C: Applied Statistics, 64*(3), 583–639. https://doi.org/10.1111/1467-9868.00353
Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics, 41*(5), 481–520. https://doi.org/10.3102/1076998616646200
Steingut, R. R., Patall, E. A., & Trimble, S. S. (2017). The effect of rationale provision on motivation and performance outcomes: A meta-analysis. *Motivation Science, 3*(1), 19–50. https://doi.org/10.1037/mot0000039
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. https://doi.org/10.1007/s10648-010-9142-7
Trautwein, U., & Köller, O. (2003). Was lange währt, wird nicht immer gut: Zur Rolle selbstregulativer Strategien bei der Hausaufgabenerledigung [Time investment does not always pay off: The role of self-regulatory strategies in homework execution]. *Zeitschrift für Padagogische Psychologie, 17*(3), 199–209. https://doi.org/10.1024//1010-0652.17.34.199
van de Schoot, R., & Depaoli, S. (2014). Bayesian statistics: Where to start and what to report. *The European Health Psychologist, 16*(2), 74–84.
Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*, 35–57. https://doi.org/10.3758/s13423-017-1343-3
Wentzel, K. R. (2020). The role of family and culture in motivation to learn in K.R. Wentzel. In *(2020). Motivating students to learn* (5th ed.). Routledge. https://doi.org/10.4324/9780429027963.
Zitzmann, S., Lüdtke, O., Robitzsch, A., & Hecht, M. (2021). On the performance of Bayesian approaches in small samples: A comment on Smid, McNeish, Miočević, and van de Schoot (2020). *Structural Equation Modeling, 28*, 40–50. https://doi.org/10.1080/10705511.2020.1752216
Zitzmann, S., Lüdtke, O., Robitzsch, A., & Marsh, H. W. (2016). A Bayesian approach for estimating multilevel latent contextual models. *Structural Equation Modeling, 23*(5), 661–679. https://doi.org/10.1080/10705511.2016.1207179
Zondervan-Zwijnenburg, M., Peeters, M., Depaoli, S., & Van de Schoot, R. (2017). Where do priors come from? Applying guidelines to construct informative priors in small sample research. *Research in Human Development, 14*(4), 305–320. https://doi.org/10.1080/15427609.2017.1370966

Barbara Flunger, Dr., is an assistant professor of Educational Sciences at Utrecht University, studies how students' motivation can be improved.

Anouk Verdonschot was a researcher at Utrecht University.

Steffen Zitzmann, Dr. habil., was an assistant professor of Educational Research Methods at the University of Tübingen, and is now professor at the Medical School Hamburg, develops and evaluates statistical methods.

Lisette Hornstra, Dr., is an associate professor of Educational Sciences at Utrecht University.

Tamara van Gog, Dr., is professor of Educational Sciences at Utrecht University.