The Journal of Experimental Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/vjxe20

The Questionnaire on Teacher Support Adaptivity (QTSA): Reliability and Validity of Student Perceptions

Janneke van de Pol, Nicky de Vries, Astrid M.G. Poorthuis & Tim Mainhard

To cite this article: Janneke van de Pol, Nicky de Vries, Astrid M.G. Poorthuis & Tim Mainhard (2023) The Questionnaire on Teacher Support Adaptivity (QTSA): Reliability and Validity of Student Perceptions, The Journal of Experimental Education, 91:4, 765-797, DOI: 10.1080/00220973.2022.2100732

To link to this article: https://doi.org/10.1080/00220973.2022.2100732

THE JOURNAL OF EXPERIMENTAL EDUCATION
2023, VOL. 91, NO. 4, 765–797
https://doi.org/10.1080/00220973.2022.2100732

Routledge
Taylor & Francis Group

The Questionnaire on Teacher Support Adaptivity (QTSA): Reliability and Validity of Student Perceptions

Janneke van de Pol[a], Nicky de Vries[b], Astrid M.G. Poorthuis[a], and Tim Mainhard[c]

[a]Utrecht University, Utrecht, The Netherlands; [b]Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [c]Education and Child Studies, Leiden, The Netherlands

ABSTRACT

Adaptive teacher support (i.e., support that is tailored to students' understanding) is considered crucial for students' learning. Previous research examined support adaptivity mostly in small-scale, observational studies without distinguishing between theory-based facets of support adaptivity. Yet, students' perceptions of different aspects of teachers' support adaptivity may be important for students' academic functioning. We investigated to what extent students perceive theory-based support adaptivity facets. We developed the Questionnaire on Teacher Support Adaptivity (QTSA) measuring students' perceptions of teacher support adaptivity and examined its reliability and validity in two samples of secondary school students ($N_{study1} = 1682$; $N_{study2} = 1048$). Overall, the QTSA provided reliable and valid scores of students' perceptions of teacher support adaptivity at the student and teacher level. One facet (non-adaptive support with much regulation upon high student understanding) should be used with some caution and needs further research. The QTSA scores converged with but also added to other constructs tapping into teachers' support adaptivity. Students' support adaptivity perceptions were distinguishable from students' perceptions of other teaching indicators (agency and autonomy support). The QTSA can provide both individual and class level feedback for practitioners and help researchers to gain more insight into possible differential effectiveness of different adaptivity facets.

KEYWORDS

Adaptive teaching; measurement invariance; scaffolding; student questionnaire; Teacher support adaptivity; validity

IT IS WIDELY accepted that teachers' academic support, especially if it is adapted to a student's current level of understanding, fosters effective learning (i.e., adaptive or contingent support; Hardy et al., 2019; Krämer & Zimmermann, 2021; Parsons et al., 2018; Van de Pol et al., 2010; Wood et al., 1978). Teacher support adaptivity refers to the degree to which a teacher adapts the regulation or steering of the students' learning process to a student's understanding and thus is a complex construct that takes two situational factors into account (i.e., teacher regulation and student understanding). For example, support is adaptive when a teacher provides an explanation (high regulation) when a student shows a low level of understanding in response to an open question. Support is adaptive if students can understand the question or instruction but are still challenged when working on the assignment, that is, if "the child never succeeds too easily nor fails too often" (Wood et al., 1978, p. 144). Existing evidence of the effectiveness of adaptive

CONTACT Janneke van de Pol j.e.vandepol@uu.nl Department of Education, Utrecht University, PO Box 80.140, 3508 TC Utrecht, The Netherlands.

Supplemental data for this article is available online at https://doi.org/10.1080/00220973.2022.2100732.

support is limited and it is still unclear how and in what circumstances adaptive support enhances student learning (Parsons et al., 2018; Van de Pol et al., 2010). Perhaps because of the complex nature of the construct, support adaptivity has mostly been examined in small samples using qualitative methods (Parsons et al., 2018). However, utilizing student perceptions of their teachers' support adaptivity in larger samples may promote a deeper understanding of this specific aspect of teaching and its effectiveness. For many indicators of teachers' instructional quality (e.g., classroom management, supportive climate, cognitive activation), students are considered experts as they are exposed to a variety of teachers and spend much time with their teachers in class (Lüdtke et al., 2009). Furthermore, it is often assumed that for a student's own academic functioning, especially their personal perception of the classroom context may be crucial (Kunter & Baumert, 2007). Thus, insight into students' perceptions can help to better understand individual students' learning while at the same time, multiple students' perceptions of the same teacher may help to get an insight into a teacher's support adaptivity (as perceived by the whole classroom). Although student perceptions as indicator of a teacher's instructional quality and instruments for feedback are widely and successfully applied in educational research (e.g., Fauth et al., 2014; Wagner et al., 2013), it is unknown whether students are indeed able to identify the adaptivity of their teacher's academic support.

In the current study, we first present a framework that unpacks the nature of and the conditions for teacher support adaptivity. Then, we present the Questionnaire on Teacher Support Adaptivity (QTSA) and examine the reliability and validity of its scores in two independent samples of secondary education students for both the student and the teacher level. An instrument that measures teachers' support adaptivity in a reliable and valid way could help researchers to conduct further, large-scale research on the effectiveness and mechanisms of teacher support adaptivity. Furthermore, such a questionnaire could be used as a feedback tool for teachers to get insight into how individual students and classes as a whole experience teacher support in terms of different aspects of support adaptivity.

A framework for teacher support adaptivity

Teachers' academic support is adaptive if it fits students' current understanding on a moment-to-moment basis (Wood & Middleton, 1975; Wood et al., 1978). Adaptive support is a key feature of scaffolding, which describes a "process that enables a child or novice to solve a problem, carry out a task or achieve a goal which would be beyond his unassisted efforts." (Wood et al., 1976, p. 90). Scaffolding is typically linked to Vygotsky's sociocultural theory and specifically to students' zone of proximal development, indicating the distance between what students can currently do and what they can do with the support of a more capable other, such as a teacher (Vygotsky, 1978). If support is adaptive, it is situated in this zone of proximal development. To be able to be adaptive, teachers needs to know or *diagnose* students' current level of understanding (e.g., Smit et al., 2013; Van de Pol et al., 2011). That is, the level of *regulation* teachers offer needs to be adapted to students' understanding (Pino-Pasternak et al., 2010; Van de Pol et al., 2019). Following Vermunt and Verloop (1999), we define teacher regulation as the steering of students' learning process. Regulation can be seen as a continuum ranging from low regulation (e.g., ask open questions), to medium regulation (e.g., give hints), to high regulation (e.g., provide extensive explanations; Van de Pol & Elbers, 2013; Wood et al., 1978). The level of regulation in itself does not determine whether support is considered adaptive. Adaptivity is determined by the match between a student's understanding and a teacher's response (i.e., an increase or decrease in regulation). An increase in the level of regulation (e.g., going from an open question to a hint) is considered adaptive only in response to low student understanding, not to high student understanding. A decrease in the level of regulation (e.g., going from a hint to an open question) is considered adaptive only in response to high student understanding, not to low student

Table 1. Facets of teacher support adaptivity.

Student's level of understanding	Teacher's response	Adaptivity Facet	Explanation
Low	**Much regulation**	A+	Adaptive support by providing much regulation upon low student understanding
	Little regulation	NA – challenge	Non-adaptive support by providing much challenge upon low student understanding
		NA – no support	Non-adaptive support by providing little regulation by not helping upon low student understanding
High	**Little regulation**	A – challenge	Adaptive support by providing much challenge upon high student understanding
		A – no support	Adaptive support by providing little regulation by not helping upon high student understanding
	Much regulation	NA+	Non-adaptive support by providing much regulation upon high student understanding

Note. 'A' = adaptive, 'NA' = non-adaptive. '–' = indicates little teacher regulation; '+' indicates much teacher regulation.

understanding (cf. Wood et al., 1978). In line with this, four general facets of teacher adaptivity can be identified (cf. Van de Pol & Elbers, 2013): (1) increase of teacher regulation upon low student understanding (Adaptive + or A+), (2) decrease of regulation upon high understanding (Adaptive- or A−), (3) increase of regulation upon high understanding (Non-Adaptive + or NA+), and (4) decrease of regulation upon low understanding (Non-Adaptive- or NA−).

Further, teachers can decrease regulation in different ways. Teachers could provide no support at all (e.g., by letting students try something themselves; cf. Wood et al., 1978), or alternatively could decrease regulation within an instructional interaction by providing much challenge without extra instruction (i.e., providing a more difficult exercise; cf. Van de Pol & Elbers, 2013). The two adaptivity facets including decrease of regulation (A − and NA−) can therefore be sub divided into another two facets of adaptivity: A−/NA − no support and A−/NA − challenge and thus six adaptivity facets can be distinguished (see Table 1).

Previous research on teacher support adaptivity

To date, studies have mostly qualitatively coded teacher support adaptivity in teacher-student interaction both on a holistic level, deciding for an entire interaction whether or not it was adaptive (Van de Pol et al., 2011), or on a turn-to-turn level while determining the provision of teachers' regulation levels in relation to students' understanding (e.g., Hermkes et al., 2018; Van Braak et al., 2021; Van de Pol et al., 2014). These studies assessed whether interactions were adaptive or non-adaptive but did not distinguish between different adaptivity facets. This is unfortunate, because teachers may well be able to, for example, decrease regulation when they interact with a student who does understand a task (A−) but may fail to increase regulation when they interact with a student who does not understand a task (A+).

Indeed, findings from an observational study coding teacher-student interaction on a turn-to-turn level—while distinguishing adaptivity facets (Kupers et al., 2015)—showed that teacher support was mostly adaptive due to decreasing regulation upon high understanding rather than increasing regulation upon low understanding. Furthermore, an observational study of Van de Pol and Elbers (2013), using a similar coding scheme, has shown that neither the mere increase of regulation nor an adaptive decrease of regulation upon high student understanding was effective for students' learning; only *adaptive* increase of teacher regulation upon low student

understanding (i.e., A+) was effective for students' learning. Thus, distinguishing between facets of (non) adaptivity in empirical research does more justice to adaptivity theory and could improve our understanding of when, how, and why facets of adaptive support may affect student learning and what constitutes effective support.

Other studies have assessed teachers' support adaptivity from the teachers' point of view e.g., by interviewing teachers (cf. Parsons et al., 2018 for a review). Using for example (video stimulated recall) interviews, these studies have shown that teachers in innovative teaching environments (e.g., Van de Pol et al., 2011) or expert teachers (De Kleijn et al., 2016; Moallem, 1998) reflect on adapting their support to their students' current understanding. Adaptivity facets were, however, not systematically addressed in these studies.

Questionnaires tapping into support adaptivity

A few student questionnaires have been developed that tap into (aspects of) teachers' support adaptivity. The *Teachers' Adaptation of Instruction Questionnaire* (Bayer et al., 2016; Kuger et al., 2016), focuses on how well students think their teacher adapts the lesson to the understanding of the class and provides individual support to students who struggle. Most items do not connect to one adaptivity facet in particular, but relate to the general idea of adaptivity (i.e., adapting ones teaching to students' understanding). One item taps into to the facet of adaptive increase of regulation upon low student understanding (A + in Table 1).

The *Pressurized Teaching Questionnaire* (Brühwiler & Blatchford, 2011) focuses on pressurized teaching which refers to a teaching method in which the "pace of teaching is too fast; the teacher continues instruction despite students' lack of understanding" (Brühwiler & Blatchford, 2011; p. 99). This relates to non-adaptive decrease of regulation upon low student understanding (NA + in Table 1).

The *Diagnostic Competence Questionnaire* (Baumert et al., 1997), focuses on how well students think their teacher can diagnose the understanding of the class. This questionnaire does not connect to a specific adaptivity facet, but relates to teachers' (correct) diagnoses of student understanding as one important aspect of adaptive support (cf. Smit et al., 2013; Van de Pol et al., 2011).

Finally, the *Adaptive Intervention Questionnaire* (Klimczak et al., 2012) was used in the Conditions and Consequences of Classroom Assessment project (cf. Klimczak et al., 2012). The questionnaire focuses on how well students think their teacher can diagnose their understanding, whether the teacher's support helped the student to complete their work, and whether the teacher's support fitted the problems the student had when working on an assignment. This questionnaire does not tap into one specific adaptivity facet but relates to teachers' (correct) diagnoses and the general support adaptivity.

These questionnaires (except the Adaptive Intervention Questionnaire) address the class level of understanding which may be difficult for students to judge. Further, they do not distinguish between separate adaptivity facets, but mostly focus on adaptivity in general terms or on one aspect of adaptivity (e.g., diagnostic competence or non-adaptive decrease of regulation upon low student understanding (NA + in Table 1)).

The current study

We examined whether student perceptions reflect teacher support adaptivity in its full complexity, as reflected in adaptivity theory (e.g., Murphy & Messer, 2000; Pratt & Savoy-Levine, 1998; Wood, 1988; Wood et al., 1976; 1978). Students' perceptions of teacher support adaptivity may reveal differences between teachers in the degree to which their regulation fits students' level of understanding. Furthermore, individual student's perceptions may provide insight into differences between students and how teacher support adaptivity may affect for example their learning,

motivation, and self-regulation. Being able to assess student perceptions reliably and validly would enable researchers to study support adaptivity on a large scale, which has recently been called for in review articles on adaptive teaching (cf. Hardy et al., 2019; Parsons et al., 2018). If the QTSA provides reliable and valid scores of students' perceived support adaptivity, it could be used as a tool for researchers and practitioners to get a differentiated view of students' perceptions of different aspects of support adaptivity.

In two studies, we investigated student perceptions of facets of teacher support adaptivity, as presented in Table 1. We developed the Questionnaire on Teacher Support Adaptivity (QTSA) measuring students' perceptions of their teachers' support adaptivity during seatwork in a wide variety of subjects and tracks throughout all grade levels of secondary education (age 12–18). In everyday classrooms, seatwork is common (Chia & Lim, 2020; Mullis et al., 2008). The seatwork context presents an ideal setting for teachers to provide adaptive support, because they have the opportunity to diagnose and support students' learning while students work individually at their desk (O'Keefe et al., 2006).

Regarding the validity, we focused on structural, generalizable, and external validity (Messick, 1995) of students' responses to the QTSA. We viewed validity as an integrated evaluation of the degree to which theory and statistical evidence support the appropriateness of the QTSA as a measure for perceived teacher support adaptivity (cf. Messick, 1989). For Study 1, we had the following research questions:

1.1. Can students distinguish between teacher support adaptivity facets?

In doing so, we examined the factor structure of items that targeted the six facets of teacher adaptivity (i.e., structural aspect of validity; cf. Table 1).

1.2. Can the QTSA be generalized across student gender, educational level, and time (generalizability aspect of validity)?

Thus, we investigated whether differences in perceived support adaptivity were due to 'true differences' in support adaptivity and not student differences or change over time (Borsboom et al., 2004).

In Study 2, we addressed the following research questions, using a second sample:

2.1. To what extent does the QTSA provide reliable and valid scores of students' perceptions of teacher support adaptivity?
2.2. To what extent do students' QTSA answers converge with other questionnaires that measure related constructs (i.e., convergent validity)?
2.3. To what extent do students' QTSA answers discriminate between teacher support adaptivity and teacher interpersonal regulation/autonomy support (i.e., teachers' regulation regardless of students' understanding), teachers' ability to diagnose students' social-emotional well-being (instead of understanding), or teacher friendliness (i.e., discriminant validity or external aspects of validity)?

Study 1

Method

Participants

Eighty-two teachers (37.1% male[1]) were recruited via the network of the researchers. The teachers worked at 24 schools that were located throughout the Netherlands. Dutch schools receive a disadvantage score that is based on information about parents' socio-economic status and ethnic

background[2]. Based on this score, schools receive money from the government to address disadvantages. Lower scores indicate that a school has less disadvantaged students. The disadvantage scores for all secondary schools in the Netherlands ranges from 0 to 788.73 ($M = 64.30$, $SD = 99.74$). The schools from the sample of Study 1 had an average disadvantage score of 31.25 (range: $0 - 140.56$; $SD = 49.34$), which is relatively low compared to the average of all Dutch schools, meaning that these schools had relatively few disadvantaged students. Data were gathered between spring 2013 and June 2015 (questionnaires were not administered in September or October as the students may not know their teachers well enough yet at the beginning of the school year). Teachers taught languages[3] (8.5%), gamma subjects[4] (54%), natural sciences/mathematics (24.2%), and other subjects (13.3%). Per teacher, one regular class (so not for example special classes with only gifted children or only children with disabilities) was randomly chosen, resulting in a sample of 1695 students in all grades of Dutch secondary education, which is equal to US grade 7 to grade 12 (47.8% boys; $M_{age} = 13.82$ years, $SD_{age} = 1.40$; prevocational education, 59.2%; higher general education, 14.5%; pre-university education, 24.4%; other, 1.8%; Dutch nationality: 95.8%). On all demographic variables, no more than 2.2% was missing. Values reported are based on the valid n per variable. To investigate whether the QTSA structure was invariant over time (RQ1.2), a subsample of 509 students from 27 classes and 4 schools[5] (48.1% boys; $M_{age} = 13.70$, $SD_{age} = 1.15$) completed the questionnaire twice about the same teacher: Once in the first round of data gathering (T0; December 2014) and again after about three months (T1; March 2015). These schools were located in the Northern and middle part of the Netherlands. Study 1 and 2 were approved by the Ethics committee of Utrecht University.

Rationale and approach for the QTSA item construction

Given that adaptivity theory with clearly circumscribed facets served as a basis of the QTSA, we followed a deductive approach in constructing our questionnaire. Two deductive methods are the construct method (aimed at optimizing construct validity) and the facet method (aimed at optimizing content validity), which were combined in the current study (Guttman, 1954; Jackson, 1971; Oosterveld et al., 2019).

In the concept analysis phase, following the facet method, we identified facets and facet elements of the construct of support adaptivity. Facets, in our questionnaire, refer to the different types of adaptivity (e.g., A+, A − in Table 1) and facet elements to teachers' regulation and students' understanding. Following the construct method, we then described the theoretical framework and definitions "taking into account important variables, and specifying the assumed relations between them" (Oosterveld et al., 2019, p. 2506). That is, for each facet, we determined how the facet elements should be related, based on adaptivity theory and definitions. For example, for the facet *adaptive support with much regulation* (A+), we determined that the regulation level of the teacher in questions of this facet should be high, and students' understanding low.

In the item production phase, items were constructed by using concrete terminology that described high and low teacher regulation and high and low student understanding (facet method), while taking into account the overall definition of the concept (construct method). To find terminology that is comprehensible for students, we interviewed seven secondary education students from one class (grade 9). Before this interview took place, we videotaped a social studies lesson from this class. From this lesson, we selected video clips that included seatwork interactions in which the students worked on their task and that varied—at face value—in the degree of support adaptivity. This resulted in seven video clips containing dyadic teacher-student interactions (one clip per student). Subsequently, we interviewed these seven students. In doing so, we showed the video clip to the student and asked the student to describe what happened in the clip, what their problem or understanding was, and how the teacher helped them. Students' wordings describing their own and the teachers' behavior (e.g., asking questions, explaining) was

Table 2. Descriptives and intraclass correlation per QTSA item.

Factor			Item	N	M	SD	ICC
Adaptive with much regulation (A+)		2.	When I don't know how to continue, this teacher helps me to find the correct answer.	1681	3.99	1.10	.14
		5.	When I get completely stuck with an exercise, this teacher shows me how to do it.	1676	4.16	1.02	.15
		9.	This teacher gives me an example, when I really don't know how to continue with the exercise.	1670	3.94	1.06	.10
		11.	When I really don't understand an exercise, this teacher explains to me how to go about it.	1678	4.11	0.96	.11
		22.	When I get totally confused, this teacher helps me to find a solution.	1676	3.96	1.17	.15
		27.	When I really don't get it, this teacher helps me to find out what I need to write down.	1674	3.78	1.27	.11
Adaptive with little regulation (A−)	Adaptive with little regulation in the form of much challenge (A − challenge)	1.	When I understand something well, this teacher makes it a little bit harder for me.	1675	2.65	1.44	.11
		6.	When I am doing well, this teacher lets me do a difficult exercise.	1672	2.78	1.41	.11
		13.	When I understand it well, I am allowed to do another exercise.	1668	2.94	1.50	.07
		18.	When I know how to do it, I get a more difficult exercise.	1669	2.76	1.39	.09
		20.	When I understand an exercise well, this teacher makes it nice and challenging for me.	1653	2.42	1.21	.80
	Adaptive with little regulation in the form of no support (A− no support)	4.	When I understand an exercise well, this teacher lets me do it on my own.	1679	4.05	1.08	.10
		17.	Whenever an exercise is going well for me, this teacher helps other students.	1672	4.17	0.95	.09
		25.	This teacher lets me work independently, when I'm working well.	1669	4.14	0.99	.09
Non-adaptive with much regulation (NA+)		3.	When I am doing well on an exercise, this teacher still shows me how to do it.	1675	2.78	1.47	.09
		8.	This teacher helps me with things that I already understand.	1666	2.49	1.15	.11
		15.	When I am already able do it, this teacher still helps me.	1665	2.31	1.19	.09
		19.	This teacher's explanations are too easy for me.	1661	2.40	1.04	.06
		21.		1669	2.61	1.14	.05

(*continued*)

Table 2. Continued.

Factor			Item	N	M	SD	ICC
			This teacher asks me things that I already know.				
		23.	This teacher gives me an explanation, even though I already understand it.	1668	2.68	1.16	.09
Non-adaptive with little regulation (NA−	Non-adaptive with little regulation in the form of much challenge (NA − challenge)	14.	When I don't know how to do it yet, I still have to continue with the next exercise.	1665	2.75	1.36	.10
		16.	When I do not yet understand the exercise, this teacher makes it more difficult for me.	1668	1.99	1.09	.07
		24.	When I do not yet understand the exercise, this teacher still makes it more difficult for me.	1673	2.07	1.05	.07
		26.	This teacher asks me things that I do not understand.	1674	2.63	1.20	.08
	Non-adaptive with little regulation in the form of no support (NA − no support)	7.	When I find an exercise very difficult, this teacher still lets me do it on my own.	1675	2.26	1.29	.12
		10.	This teacher tells me to do it on my own, even though I am unable to continue.	1669	2.11	1.21	.11
		12.	When I find an exercise difficult, this teacher barely explains it to me.	1674	1.97	1.15	.15

Note. These items were translated from Dutch using the back translation method. The Dutch items have been used in this study. Average cluster size was 20.5 students per class. Students answered the items on a 5-point scale ranging from 1 (totally disagree) to 5 (totally agree). Items in gray font were deleted from the questionnaire. The final questionnaire including the original Dutch can items be found in Appendix A.

used to formulate the items. Furthermore, experts were asked to formulate items as well. Consistent with the definition of adaptivity (i.e., the match between student understanding and teacher regulation) the student and the teacher served as referents for the items. Thus, we ensured that each item contained both the facet element of teacher regulation and of student understanding (cf. Wood et al., 1978).

An initial version of the questionnaire was piloted in February 2013 with four secondary education students (two from grade 7, two from grade 9; these were not the students who participated in the interview) and revisions to the items were made based on this pilot. In this pilot, we asked the students to complete the questionnaire while thinking out loud to gain information about how they interpreted the questions. The phrase *When I do not understand it* was, for example, revised into *When I do not understand the exercise*. And the phrase *this teacher often explains it* was understood by students as a teacher explaining to the whole class, which was not the intended context. Therefore, this item was revised into *This teacher explains to me how to do it*. The final phase, i.e., the evaluation phase, is described under 'analyses'.

We used a five-point Likert scale ranging from *totally disagree* to *totally agree*. In Table 2, all items are listed and the final version of the questionnaire (Dutch and English[6]) can be found in Appendix A. The Supplementary Material contains the final questionnaire formatted for use in research or practice. General instructions that accompanied the questionnaire were: "The questions that you are about to answer are meant to assess how you think your teacher helps you during seatwork." In addition, on the top of the page on which the questions were printed, it was written: "This questionnaire concerns the way the teacher helps you when working independently on an assignment."

Procedure

Teachers administered the questionnaire in their class and received written instructions that had to be read aloud verbatim, so that each teacher would give the same instructions. Students were informed that the completed questionnaire would only be used by the researchers, and that their teacher and parents would not see the students' answers[7]. Students were instructed to complete the questionnaire individually. Students needed about 10 minutes to complete the questionnaire.

Data preparation and analysis

We analyzed our data in a multilevel framework given the nested structure of our data (i.e., student perceptions nested within teachers within schools), the nature of the construct (i.e., both student and teacher are sources of variance in teacher support adaptivity), and our ambition to assess adaptivity at both the student and the teacher level (Hox et al., 2017). We checked the assumptions for linearity and multicollinearity by visually checking the scatterplots and linear graphs and we did not find violations. Furthermore, one outlier was found for the dataset of T1. We ran the analyses with and without the outlier, yielding similar results (results with outliers are reported). Finally, to account for small violations of the assumption of normal distributions, we used a robust maximum likelihood (MLR) estimator for all analyses (Yuan & Bentler, 2000).

Questionnaires with little variability between items ($SD < .45$) and questionnaires returned empty were excluded (T0: $n = 13$, 0.8%; T1: $n = 9$, 1.6%). Our final sample size for T0 was 1682 students ($1695 - 13$). The final sample size for the subsample of T1 (i.e., to examine invariance over time) was 500 students ($509 - 9$). For the included questionnaires (99.23% at T0; 98.38% at T1), missing data was limited to less than 1.7% per item. Little's MCAR test showed our data to be not missing completely at random, for T0 χ^2 (1526) = 1924.38, $p < .001$ and for T1 χ^2 (891) = 1014.13, $p = .002$. Visual inspection of missing data patterns did not show systematic patterns, such as more missingness toward the end of the questionnaire. Furthermore, a linear regression analysis with age, gender, level, grade, and subject did not show any significant predictors of missingness for both T0 and T1. In our analyses, we used full information maximum likelihood estimation to handle missing data at each time point (i.e., 0.8% at T0, 1.6% at T1; cf. Schreiber et al., 2006).

We performed our analyses for RQ1.1 in Mplus 7, using the five steps of Muthén (1994). The QTSA is directly based on adaptivity theory (e.g., Murphy & Messer, 2000; Pratt & Savoy-Levine, 1998; Wood, 1988; Wood et al., 1976; 1978), that clearly distinguishes between different adaptivity facets. As Levine (2005) states, "CFA is used when the researcher has an expectation of how the items will factor, and CFA is used to test this expectation against the data." (Levine, 2005; p. 336). Given that we departed from adaptivity theory, we did have clear theoretical assumptions about the underlying structure (cf. Brown & Moore, 2012; Levine, 2005). In addition, using CFA's is recommended when using the construct and facet methods (cf. Oosterveld & Vorst, 1996), given that the construct and facet methods are theory deductive methods. Therefore, we explored several theory-based models using CFA's. Thus, in line with the theory-based facets (Table 1), we tested models with 1 factor (general adaptivity), 2 factors (adaptive vs. non-adaptive support), 4 factors (A+, A−, NA+, and NA−), or 6 factors (A+, A − no support, A − much challenge, NA+, NA − no support, NA − much challenge). In addition to these theory-based models, we also checked the fit of an additional two-factor model, consisting of items that focused on either low versus high understanding, as this might be an alternative way to perceive teacher support. For each model, we assessed the model fit based on the same criteria (RMSEA < .06; CFI and TLI > .90; SRMR < .08; cf. Hu & Bentler, 1999). The chi-square value is reported, but not used as a fit measure as for large samples differences for statistic between models will almost always be significant (Bentler & Bonett, 1980). When estimating a multilevel model, SRMR values are provided at the within (i.e., student) and between (i.e., teacher) level. However,

no established guidelines exist for interpreting the $SRMR_{between}$ so we reported these values but did not use them in deciding between models.

Changes in model fit were assessed by changes in the Akaike Information Criterion (AIC; Akaike, 1974) and the Sample Adjusted Bayesian Information Criterion (SABIC; Sclove, 1987), for which smaller values denote better fitting models. If the model with the lowest AIC and SABIC values fitted insufficiently according to the fit indices, factor loadings (> .5; cf. Dunn et al., 2015), R^2 values (> .2; cf. Hooper et al., 2008), and modification indices were inspected to detect problems. Model modifications were only made when reasonable from a substantive point of view, and at least three items per factor were retained. We did not allow for cross-loadings using a strict CFA framework (Brown, 2006). In measurement invariance tests, a more restrictive model was retained, if the decrease in CFI was less or equal to .01 (except the longitudinal invariance) and if increases in RMSEA and SRMR were less or equal to .015 (cf. Cheung & Rensvold, 2002). In the next sections, we outline our procedure for each step.

Step 1: One Level CFA. To get a general understanding of the factor, we first performed one level CFAs for the different models. Because multigroup factor analysis is not available for within (i.e., student) level variables in multilevel models and because our model was too complex for restricted factor analysis (Jak, 2013), we used the one-level model to test for measurement invariance with regard to student gender and educational level, as Kim et al. (2012) proposed as an alternative. For gender and educational level, we tested consecutively for configural, weak factorial, and strong factorial invariance (cf. Brown, 2006).

Step 2: Estimation at the Student Level. We modeled the student variance with a saturated model (i.e., all items correlated) for the teacher level variance. This analysis estimates the student level parameters without distortion by the teacher level variance.

Step 3: Estimation at the Teacher Level. Here, we used the same approach as in the previous step, but now using a saturated model for the student level.

Step 4: Estimation of a Full MCFA and Reliability. To test the factor structure at the student and teacher level in an integrated way, we performed a full MCFA in which the best fitting factor structures from step 2 and 3 were combined. We considered support adaptivity to be a configural construct, meaning that the student and the teacher level were of interest (Stapleton et al., 2016). At the student level, responses pertain to students' idiosyncratic perceptions of the teachers' support adaptivity. The teacher level represents differences between teachers' support adaptivity, as perceived by their classes (class aggregates). For configural constructs, invariance across levels is essential for the validity of our measure of teacher support adaptivity; for cross-level invariance, equal factor structures and loadings at both levels are necessary (Stapleton et al., 2016). If the best fitting factor structures were unequal at the student and teacher level, we tried the two solutions as our full model. The best fitting full model was used to examine cross-level invariance by constraining the factor loadings to be equal at both levels.

Subsequently, we calculated both the multilevel alpha and omega (Geldhof et al., 2014). Values above .70 were considered sufficient for research purposes and above .80 as adequate for individual decisions in applied settings[8] (cf. Brunner & Süβ, 2005; Lance et al., 2006; Nunnally & Bernstein, 1994). Furthermore, we calculated the ICC(2) per factor (Lüdtke et al., 2009) to test the reliability of teacher-mean aggregated ratings. Values of .70 - .85 are seen as acceptable (LeBreton & Senter, 2008, Lüdtke et al., 2007).

Step 5: Longitudinal Measurement Invariance (RQ1.2). We tested longitudinal measurement invariance over two occasions. We only assessed the RMSEA and SRMR as the CFI and TLI assume all latent factors to be uncorrelated (Hooper et al., 2008), which is not the case here.

Table 3. Model fit indices for CFAs.

Step	Model	Df	X^2	RMSEA	CFI	TLI	$SRMR_{(w, b)}$	AIC	SABIC
1 One-level CFA	1-factor	324	4939.611*	.09	.60	0.6	.11	126899.99	127082.31
	2-factor (adaptive vs non-adaptive)	323	4339.594*	.09	.65	0.6	.10	126121.21	126305.79
	2 factor (high vs low understanding)	323	4782.223*	.09	.61	.58	.11	126703.44	126888.01
	4-factor	318	1945.194*	.06	.86	0.8	.09	123214.03	123409.85
	6-factor	309	930.174*	.04	.95	0.9	.04	121904.48	122120.56
2 Estimation student level	1-factor student level	324	31112.518*	.24	.00	−3.28	(.11, .04)	125771.29	126804.44
	2-factor student level (adaptive vs non-adaptive)	323	4624.295*	.09	.72	.40	(.10, .04)	125007.10	126042.50
	2-factor student level (high vs low understanding)	323	27753.546*	.02	.00	−2.82	(.11, .04)	125581.89	128078.65
	4-factor student level	318	16070.230*	.17	.00	−1.23	(.09, .03)	122572.74	123619.39
	6-factor student level	309	3529.352*	.08	.79	.53	(.04, .02)	121479.85	122546.77
	6-factor student level modified[1]	237	769.635*	.04	.96	.91	(.04. .01)	106971.53	107842.62
3 Estimation teacher level	1-factor teacher level[1, 2]	253	1231.228*	.05	.93	.85	(.01, .26)	106625.96	107461.03
	2-factor teacher level[1, 3] (adaptive vs non-adaptive)	253	1244.036*	.05	.93	.85	(.01, .26)	106622.87	107457.94
	2-factor teacher level[1, 3] (high vs low understanding)	253	1248.850*	.05	.93	.85	(.01, .26)	106619.99	107455.06
	4-factor teacher level[1, 4]	250	853.819*	.04	.96	.91	(.01, .20)	106514.13	107355.95
	6-factor teacher level[1, 5]	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
4 Estimation full MCFA	Equal structure[1, 4]	499	2932.979*	.05	.83	.81	(.09, .30)	108017.76	108299.11
	Equal structure (modified)[4, 6]	370	1406.354*	.04	.92	.91	(.05, .12)	94051.68	94306.03
	Equal structure and loadings[4, 6]	387	1447.590*	.04	.92	.91	(.05, .12)	94064.73	94280.81

Note. 1 factor all items; 2 factors: adaptive items (A+/A−) versus non-adaptive items (NA+/NA−); 4 factors (A+, A−, NA+, NA− 6 factors (A+, A − no support, A − much challenge, NA+, NA − no support, NA − much challenge).

[1]items 3, 19, and 26 were deleted from the model (cf. Table 2) [2] residual variance of item 18 fixed at zero (teacher level); [3] residual variances of items 14 and 18 fixed at zero (teacher level); [4] residual variances of items 11, 14, 18, and 23 fixed at zero (teacher level); [5] model did not terminate normally due to highly correlated latent factors at the teacher level; [6] items 3, 4, 17, 19, 25 and 26 (Table 2) were deleted from the model.

*$p < .001$.

Results

Table 2 shows all items per expected factor and their descriptives.

Pre-analyses (Steps 1 to 3)

In Step 1 (i.e., one level CFA), a six-factor structure fitted the data best (see Table 3). Furthermore, the factor structure was invariant for student gender and educational level. In Step 2 (i.e., CFA at student level), a six-factor structure again produced the best fit. Items 3, 19, and 26 (cf. Table 2) had low factor loadings and were excluded from all further analyses. The wording of items 19 and 26 was slightly different than the wording of other items, which might explain why these items' factor loadings were low. Most items were formulated by explicitly stating whether the student has difficulties or not, whereas in items 19 and 26, this is left more implicit. Moreover, these items' ICCs were relatively low. In Step 3 (i.e., CFA at teacher level), a four-factor structure fitted the data best. Note that at the teacher level, the one facet structure (i.e., general support adaptivity) also showed good fit in terms of RMSEA, SRMR, and CFI (TLI was just

slightly below the threshold, i.e., .85; see Table 3). The teacher level aggregated alpha was .944 and omega was .946.

Estimation of the full multilevel CFA and reliability (Step 4)

Based on results in previous steps, we fitted a four-factor model at both levels, which did not fit sufficiently (Table 3). Removing three items with factor loadings below .5 (item 4, 17, and 25 from Table 2, factor *A− no support*) resulted in good fit. The factors represent adaptive support with much regulation upon low understanding (A+), adaptive support with little regulation upon high understanding (with much challenge; A−), non-adaptive support with much regulation upon high understanding (NA+), and non-adaptive support with little regulation upon low understanding (with no support or much challenge: NA−). Factor correlations can be found in Appendix B, Table B1.

The results for cross-level invariance showed that the four factors were valid both at the student and teacher level. Figure 1 shows the full multilevel structure of the QTSA and its parameter results (see Table B2 in Appendix B for residual variances and R^2-values). The correlations between the four factors were small, ranging from −.12 to .06.

At the teacher level, the multilevel alpha and omega were above .80 for all factors, indicating that factor scores can be used as feedback at the teacher level and for research purposes (Table 4). When using the questionnaire, a class size of ≥15 is sufficient for A+, A−, and NA−; to obtain a reliable estimate at the teacher level (i.e., ICC(2)>.70); a class size of ≥20 is needed for NA+ (see Table 4).

At the individual student level, the internal consistency for A + and A − was >.80 and thus suited to be used for feedback at the student level (i.e., an individual student's score gives rather reliable information about this student's perceptions and could be used to inform a teacher) and for research purposes. Internal consistency for NA − was >.70 and this factor is thus suited for research purposes at the individual student level. The reliability for NA + at the individual student level was <.70 (i.e., alpha = .65; omega = . 66) and thus did not reach our criteria for feedback or research purposes. This factor should thus be used with some caution at the individual student level. Further deletion of items did not increase reliability and the factor was thus retained as it was.

Longitudinal measurement invariance (Step 5)

The QTSA structure was invariant over time (Configural invariance: RMSEA: .09 (change = −.05); $SRMR_{within}$: .05 (change = 0), $SRMR_{between}$: .19 (change = −.07); Weak factorial invariance: RMSEA: .05 (change = −.01), $SRMR_{within}$: .06 (change = −.01), $SRMR_{between}$: .18 (change = −.07); Strong factorial invariance: RMSEA = .05 (change = −.01); $SRMR_{within}$ = .06 (change = −.01), $SRMR_{between}$: .19; (change = −.06)). Correlations between measurements per factor were .56/.94 (student/teacher level) for A+, .53/.93 for A−, .59/.29 for NA+, and .56/.80 for NA−.

Discussion and conclusion

Study 1 tested whether students' perceptions of teacher support adaptivity were reflected at the individual student and the aggregated teacher level (RQ1.1). Students were able to distinguish facets of teacher support adaptivity. In terms of generalizability (RQ1.2), student perceptions seemed independent from student gender, educational level, and time.

Facets of teacher support adaptivity

The present study showed that students distinguished between four facets of teacher support adaptivity: Adaptive support with much teacher regulation upon low student understanding (A+) and little teacher regulation upon high student understanding (A−); non-adaptive support with much teacher regulation upon high student understanding (NA+) and little teacher regulation upon low student understanding (NA−).

At the teacher level, scores on all four facets can be reliably used for research purposes and feedback for individual teachers (e.g., assess differences between teachers' support adaptivity). At the student level, individual student perceptions can be reliably used to assess A + and A−. These individual perceptions can be used for feedback for teachers and for research purposes. NA − was suitable to be used for research purposes but should be used with some caution when using these individual perceptions for example as feedback for teachers. NA + should also be used with caution when using it for research purposes or when using these individual perceptions for example as feedback for teachers.

Overall, the results indicated that students were able to distinguish between rather specific facets of teacher support adaptivity, which underlines the potential for the utilization of student ratings (cf. Lüdtke et al., 2009).

Generalizability of student perceptions

Student perceptions were not systematically affected by their gender or educational level. The results also indicated that students' perceptions of their teachers' support adaptivity was stable over time. This is in line with research on college students' perceptions about teaching (Marsh, 2007), which showed that student reports were highly stable over time including individual differences between teachers. These results demonstrate that student perceptions can be used to validly generalize across students regardless of their gender or educational level and occasion (Borsboom et al., 2004; Messick, 1995).

Study 2

In Study 2, we investigated the reliability and validity of student perceptions of teachers' support adaptivity using the QTSA in a new sample (RQ2.1). We checked whether the four-factor structure obtained in Study 1 also fitted the data in a new sample. To further determine whether the QTSA can provide valid scores that measure teachers' support adaptivity (external aspects of validity, Messick (1995)), we checked whether students' QTSA answers were related to their answers on questionnaires on related constructs (convergent validity; RQ2.2). Furthermore, we investigated to what extent students discriminated between teacher support adaptivity (using the QTSA) versus other constructs related to, but conceptually different from support adaptivity (discriminant validity; RQ2.3).

For convergent validity, we selected four questionnaires that address the adaptivity of teachers' instructional activities and/or the degree to which teachers know what their students understand, a prerequisite of adaptive teaching (Table 5). Given that these constructs relate closely to support adaptivity, we expected medium to large correlations with the QTSA. One exception is the Pressurized Teaching Questionnaire (Brühwiler & Blatchford, 2011), which specifically relates to the facet of non-adaptive support with little regulation in the form of much challenge upon low understanding (NA−) of the QTSA. Therefore, we expected large correlations between the NA − facet of the QTSA and the Pressurized Teaching Questionnaire. As for the direction of the correlations, we expected that questionnaires aimed at adaptive support or the prerequisite of knowing what their students understand correlated positively with A + and A − and negatively with NA + and NA− (i.e., the Teachers' Adaptation of Instruction Questionnaire, the Diagnostic

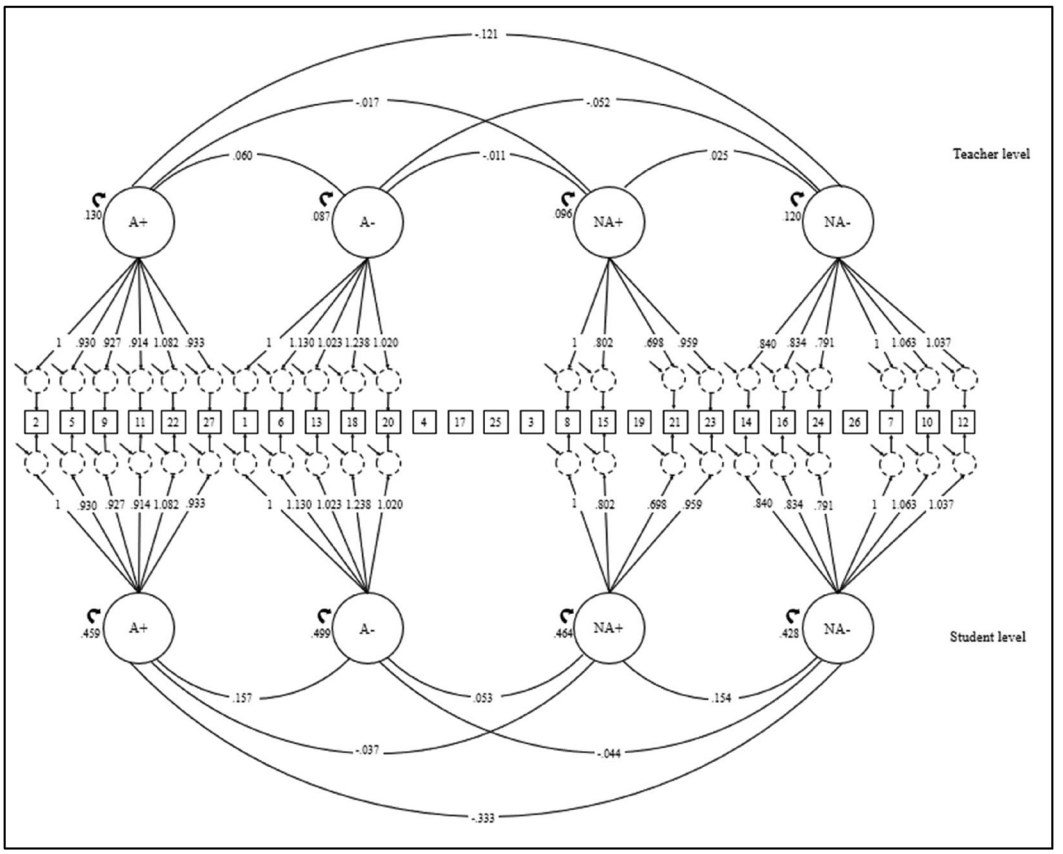

Figure 1. Unstandardized model results (Study 1) full multilevel CFA with cross-level invariance.

Competence Questionnaire and the Adaptive Intervention Questionnaire). As the Pressurized Teaching Questionnaire refers to non-adaptive support, we expected negative correlations with A + and A − and positive correlations with NA + and NA−.

For discriminant validity, we selected four questionnaires to establish whether the QTSA can distinguish between support adaptivity and other constructs. We selected questionnaires that had some relation to (aspects of) support adaptivity, but that were fundamentally different. The autonomy scale of the shortened Teacher as Social Context questionnaire (TASC; Belmont et al., 1988; Sierens et al., 2009) addresses the degree to which teachers give their students choice, give a rationale when choice is constrained, emphasize with students' perspective, and avoid using controlling language (cf. Sierens et al., 2009). It thus addresses the degree of teachers' regulation and although the aim is to promote students' autonomy, autonomy support is strongly related to the provision of structure (cf. Jang et al., 2010; Sierens et al., 2009). Therefore, we expected medium to strong correlations between A+ (positive) and NA− (negative) and the TASC and medium correlations at most between A− (positive) and NA+ (negative) and the TASC.

With the Diagnostic Competence of Students' Social-Emotional Well-Being Questionnaire (DCSW; PISA, 2003), we checked whether students distinguished between adaptivity (including diagnosis) to social-emotional well-being (using the DCSW) and adaptivity (including diagnosis) to subject matter understanding (QTSA). Since some relation can be expected, we expected medium correlations at most between A + and A − and the DCSW to be positive and between NA + and NA − and the DCSW to be negative.

Table 4. Factor reliability for Study 1.

Measure	Level/Class size	A+	A−	NA+	NA−
Alpha and omega	Student (α, ω)	.83, .83	.81, .81	.65, .66	.76, .76
	Teacher (α, ω)	.98, .98	.97, .98	.93, .95	.98, .99
ICC(2)	k = 10	.70	.61	.58	.69
	k = 15	.78	.70	.67	.77
	k = 20	.83	.76	.73	.83
	k = 30	.88	.83	.80	.87

Note. k is class size. A+: Adaptive support with much regulation, A − Adaptive support with little regulation in the form of much challenge, NA+: Non-adaptive support with much regulation, NA−: Non-adaptive support with little regulation in the form of no support or much challenge.

The interpersonal agency dimension of the Questionnaire on Teacher Interaction (QTI, Wubbels & Levy, 1993) enabled us to check whether students can distinguish between more general classroom regulation (QTI) and teacher regulation in relation to their understanding (QTSA). We expected positive and medium correlations at most between the agency dimension and adaptivity with much regulation (A + and NA+) and negative and medium correlations at most between the agency dimension and adaptivity with little regulation (A − and NA−).

Finally, we used the interpersonal warmth or communion dimension of the QTI. This dimension reflects a teacher's warmth and friendliness (Wubbels & Levy, 1993). Although not concerning support adaptivity regarding academic support, students may perceive adaptive teachers as friendlier, because they are better able to help students. However, friendliness or likability may also result in a bias in student ratings. That is, friendly teachers may receive high teaching quality ratings from their students but may not necessarily show high quality teaching. Several studies have shown medium to large positive correlations between students' ratings of teaching quality and teacher likeability variables (Fauth et al., 2014; Shevlin et al., 2000). Very high correlations may point to such bias. For the QTI warmth dimension, we expected medium to large positive correlations with QTSA facets (maximally .90).

Method

Participants

Fifty-five teachers (46.3% male) from the researchers' network were recruited. Teachers worked at schools that were located throughout the Netherlands. The schools from the sample of Study 2 had an average disadvantage score of 42.63 (range: 0 − 28.86; $SD = 95.57$), which is relatively low compared to the average of all Dutch schools, meaning that these schools had relatively few disadvantaged students. The disadvantage scores for all secondary schools in the Netherlands ranges from 0 to 788.73 ($M = 64.30$, $SD = 99.74$). Data were gathered between March and November 2016 (questionnaires were not administered in September or October as the students may not know their teachers well enough yet at the beginning of the school year). Teachers had, on average, 8.9 years of teaching experience ($SD = 9.59$), and taught languages[9] (42.3%), gamma subjects[10] (36.8%), natural sciences/mathematics (15.2%), arts (1.9%), and for two teachers their subject was unknown (3.8%). Per teacher, one regular class (so not for example special classes with only gifted children or only children with disabilities) was randomly selected, resulting in a sample of 1048 students from 54 classes in 18 regular schools of all grades of Dutch secondary education, which is equal to US grade 7 to grade 12 (48.5% boys; $M_{age} = 14.43$, $SD_{age} = 1.45$; prevocational education: 2.7%; higher general education: 65.4%; pre-university education: 14.5%; other: 17.4%; Dutch nationality: 95.5%). On all demographic variables, missing data was $\leq$ 5% (except for age: 8.8%). Values reported are based on the valid n per variable.

Instruments

Questionnaire on teacher support adaptivity (QTSA). The 27 items assessing facets of teacher support adaptivity were used here (see Study 1). Next, we describe the questionnaires used for convergent and discriminant validity (also see Table 7). Non-Dutch instruments were translated using the back-translation[11] procedure. The final questionnaire, formatted for use by researchers and practitioners can be found in the Supplementary Material.

Pressurized Teaching Questionnaire. This three-item questionnaire (Brühwiler & Blatchford, 2011; Eder & Mayr, 2000) uses a scale from 1 (disagree) to 5 (agree). We omitted one item (i.e., "Often exercises cannot be discussed because there is too much subject matter that needs to be discussed") because it does not pertain to adaptivity. The items we used are: "The teacher often explains so fast that we can hardly keep up" and "The teacher often continues during a lesson although (s)he knows that not everyone understands it." Internal consistency was $\alpha = .63$; results regarding this questionnaire should be interpreted with caution (Nunnally & Bernstein, 1994). Prior research provides indications of the questionnaire's criterion validity with regard to student's well-being (e.g., school satisfaction), behavior in class (e.g., collaboration), and psychological stress (e.g., school anxiety) with correlations between $r = .30$ to .50; cf. Eder & Mayr, 2000).

Teachers' Adaptation of Instruction Questionnaire. The three items of this questionnaire (Bayer et al., 2016; Kuger et al., 2016) are answered on a scale from 1 (never) to 4 (always). An example item is: "The teacher gives individual support when a student finds the topic or a task difficult." Internal consistency was $\alpha = .66$; results regarding this questionnaire should be interpreted with caution (Nunnally & Bernstein, 1994).

Diagnostic Competence Questionnaire. The five items of this questionnaire (Baumert et al., 1997) are answered on a scale from 1 (entirely incorrect) to 4 (entirely correct). An example item is: "Our teachers knows what we did not understand." Internal consistency was $\alpha = .84$.

Adaptive Intervention Questionnaire. This questionnaire (Klimczak et al., 2012) focuses on how well students think their teacher can diagnose students' personal understanding (1 item), whether the teacher's support helped the student to complete their work (2 items), and whether the teacher's support fitted the problems the student had when working on an assignment (1 item). The items are answered on a scale from 1 (entirely correct) to 4 (entirely incorrect). An example item is: "I have the feeling that my teacher understood my difficulties with doing the assignments." Internal consistency was $\alpha = .77$.

Questionnaire on teacher interaction. This questionnaire (Wubbels & Levy, 1993) consists of 16 (short version) or 24 (long version) items. Items are answered on a scale from 1 (never) to 5 (always). The questionnaire consists of two dimensions: interpersonal agency (having control and being leading in class) and communion (being warm and friendly). Internal consistency for the communion dimension was $\alpha_{\text{long version}} = .88$; $\alpha_{\text{short version}} = .87$. For the agency dimension, the internal consistency of the long version was $\alpha_{\text{long version}} = .70$. The internal consistency of the short version was $\alpha = .68$; results regarding this questionnaire should be interpreted with caution (Nunnally & Bernstein, 1994). About half of the classes completed the short version, the other half completed the long version. Example items are "This teacher is a good leader" and "This teacher is patient." Previous research with the QTI has indicated good internal validity, see for example Mainhard et al. (2018) for a CFA of the short version of the QTI χ^2 (28) = 30880.36, $p < 0.001$, CFI = 0.99, TLI = 0.99, RMSEA = 0.44. For a more general overview of the validity of the QTI and associations with other teaching constructs see Wubbels et al. (2006).

Table 5. Information about questionnaires used for convergent and discriminant validity.

Questionnaire	Reference	Background	Comparison with the QTSA
Used for convergent validity			
Pressurized Teaching Questionnaire	Brühwiler & Blatchford, 2011; adapted from Eder & Mayr, 2000	Pressurized teaching provides a "pace of teaching is too fast; teacher continues despite students' lack of understanding." (p.99)	The items relate to NA−.
Teachers' Adaptation of Instruction Questionnaire	Bayer et al., 2016; Kuger et al., 2016	Focuses on how well students think the teacher adapts the lesson (structure) to the class' understanding (item 1 and 2) and provides support to struggling students (item 3).	Item 1 and 2 do not correspond to one of the facets, but to the idea of support adaptivity in general. The third item corresponds to A+.[13]
Diagnostic Competence Questionnaire	Baumert et al., 1997	The questionnaire focuses on how well students think their teacher can diagnose the students' understanding.	The items of this questionnaire do not specifically correspond to one of our facets, but to the idea of support adaptivity—with diagnosing students' understanding as a central part of support adaptivity—in general.
Adaptive Intervention Questionnaire	Klimczak et al., 2012	Focuses on whether students think: the teacher can diagnose their own understanding (1 item), the teacher's support helped the student to complete their work (2 items), the teacher's support fitted the problems the student had when working on their assignment (1 item).	
Used for discriminant validity			
Questionnaire on Teacher Interaction (QTI)—Interpersonal agency dimension	Wubbels & Levy, 1993	A high level of agency refers to having control and being dominant.	Regulation in interpersonal sense (QTI-agency) versus regulation in relation to one's own understanding in a didactic context (QTSA).
Teacher Autonomy Support Questionnaire (TASC)	Sierens et al., 2009	Autonomy support involves, amongst others, "providing students with an amount of choice (Katz & Assor, 2007), […] and avoiding the use of controlling language (e.g., 'you should')." (Sierens et al., 2009, p. 61).	Regulation only (TASC) versus regulation in relation to one's own understanding (QTSA).
Diagnostic Competence of Students' Social-Emotional Well-Being Questionnaire (DCSW)	PISA, 2003	This questionnaire focuses on the teacher's ability to diagnose students' social-emotional well-being (e.g., whether students have had a fight, whether a student is sad).	Adaptivity (including diagnosis) regarding social-emotional well-being (DCSW) versus adaptivity (including diagnosis) regarding understanding (QTSA).
Questionnaire on Teacher Interaction (QTI)—Interpersonal warmth (communion) dimension	Wubbels & Levy, 1993	Communion refers to being friendly and showing affiliation.	Teachers' friendliness (QTI-communion) versus teachers' support adaptivity (QTSA).

Autonomy Scale of Teacher as Social Context Questionnaire (TASC). Autonomy support involves, amongst others, "providing students with an amount of choice […] and avoiding the use of controlling language (e.g., 'you should')" (Sierens et al., 2009, p. 61). The autonomy scale

of the TASC contains eight items that are answered on a scale from 1 (completely disagree) to 5 (completely agree). An example item is: "My teacher gives me choices about how to do my school work." Internal consistency was $\alpha = .71$. Previous research using the Dutch and Spanish version of the TASC in secondary education has shown that scores on items from the autonomy questionnaire could be distinguished from scores on items from a related scale on teacher structure, indicating good construct validity (Iglesias García et al., 2020; Sierens et al., 2009). Moreover, research regarding the Spanish version showed the TASC to be invariant for gender (Iglesias García et al., 2020).

Diagnostic competence of students' social-emotional well-being. This questionnaire focuses on teachers' ability to diagnose students' social-emotional well-being (e.g., whether students have had a fight, whether a student is sad). This questionnaire (PISA, 2003) consists of four items that are answered on a scale from 1 (entirely incorrect) to 5 (entirely correct). An example item is "Our teacher notices when someone is sad." Internal consistency was $\alpha = .88$.

Procedure

The procedure of the data collection was similar to Study 1. Students first completed the QTSA and then the other questionnaires.

Data preparation and analysis

Data preparation was similar to Study 1. We checked the assumptions for linearity and multicollinearity by visually checking the scatterplots and linear graphs and we did not find violations. Furthermore, 151 outliers were found. We ran the analyses with and without the outliers, yielding similar results (results with outliers are reported)[12]. To account for small violations of the assumption of normal distributions, we used a robust maximum likelihood (MLR) estimator for all analyses (Yuan & Bentler, 2000).

Responses of 27 students (2.2%) were excluded based on an $SD < .45$ or missing values on all items. For the remaining students, missing data was $\leq$ 2.3%. Data were not missing completely at random (Little's MCAR test, χ^2 (1450) = 1630, p =.001). A regression showed gender to significantly predict the number of missing values; boys had more missing values than girls. Yet, it only explained 0.4% of the variance. As overall missingness was $\leq$ 2.3%, we expected this not to affect results (Luengo et al., 2010). In our analyses, we used full information maximum likelihood estimation to handle missing data (cf. Schreiber et al., 2006).

Factors of teacher support adaptivity. To determine whether the final factor structure in Study 1 fitted the data in the second sample as well (RQ2.1), we conducted a CFA (i.e., four factors). In addition, with a multi-group analysis, we tested for measurement invariance (i.e., population heterogeneity; Brown, 2006). That is, we tested directly whether the items followed the same factor structure *and* whether they had equal factor loadings in the second sample. We used the same approach as in Study 1 with T0 data from Study 1 as one group and data from Study 2 as the second group. We report the same fit measures as in Study 1 (RMSEA $< .06$; CFI and TLI $> .90$; $SRMR_{within} < .08$; cf. Hu & Bentler, 1999).

Convergent and discriminant validity. To test the relations between teacher support adaptivity and the other questionnaires (RQ2.2 and RQ2.3), we computed correlations between all questionnaires at the student level (group mean centered scores) and teacher level (group means). Because of a lack of previous research including instruments addressing two or more of the constructs that we included for convergent and divergent validity, we used Cohen's (1992) classification of correlations (.1 = small; .3 = medium; .5 = large).

Table 6. Factor reliability for Study 2.

Measure	Level	A+	A−	NA+	NA−
Alpha and omega	Student (α, ω)	.80[a]; .80[a]	.79; .80	.63; .64	.72; .73
	Teacher (α, ω)	.98; .99	.97; .98	.80; .86	.95; .97
ICC(2)	k = 10	.76	.73	.42	.65
	k = 15	.83	.80	.52	.73
	k = 20	.86	.84	.59	.79
	k = 30	.91	.89	.68	.85
	k = 33	.91	.90	.70	.86

Note. k is class size.
[a]When running these analyses without outliers, these values were .78 and thus just below the threshold of .80.

Results

Factor structure

The CFA showed that the final factor structure with four adaptivity facets as identified in Study 1 also fitted the data in the second sample well (RMSEA = .04; CFI = .90; TLI = .90; $SRMR_{within}$ = .05; $SRMR_{between}$ = .19; see Table B4 in Appendix B for factor loadings and residual variances).

In addition, a direct comparison of the two samples indicated strong factorial invariance (RMSEA = .04; $SRMR_{within}$ = .04; $SRMR_{between}$ = .09). Thus, not only the factor structure on the student and teacher level, but also factor loadings were comparable between the samples. Factor correlations can be found in Appendix B, Table B5.

The multilevel α and Ω and ICC(2) showed generally the same reliability patterns as in Study 1 (Table 6). At the teacher level, the multilevel alpha and omega were >.80 for all factors, indicating that scores on the four factors can be used for research purposes and feedback for individual teachers. When using the questionnaire, a class size of ≥10 is sufficient for A+, A−, and NA − to obtain a reliable estimate at the teacher level (i.e., ICC (2)> .70); a class size of ≥33 is needed for NA+ (see Table 6).

At the individual student level, the internal consistency for A + and A − was adequate for the factors to be used for feedback at the student level (i.e., an individual student's score gives rather reliable information about this student's perceptions and could be used to inform a teacher) and for research purposes. Internal consistency for NA − was >.70 but <.80: this factor can thus be used for research purposes at the individual student level and should be used with some caution when using these individual perceptions for example as feedback for teachers. The reliability of NA + was <.70 (alpha = .63; omega = .64) at the individual student level and thus did not reach our criteria to be used for feedback or research purposes; this factor should thus be used with some caution at the individual student level. Further deletion of items did not increase reliability; the factor was thus retained as it was.

Convergent validity

Correlations had the expected *direction* (Table 7/B3) except for one correlation which was expected to be negative but turned out to be not significant (between NA + and the student level Adaptive Instruction Questionnaire results, r = .01, 95% CI [−.05, .07). Furthermore, most correlations between support adaptivity facets and the related questionnaires had the expected *strength* (22 of 32; +/- .1). Exceptions mostly regarded NA+ (non-adaptive with much regulation upon high understanding): correlations were smaller than expected at the student (4 correlations) and teacher level (3 correlations). Furthermore, student level correlations of NA− (non-adaptive with little regulation upon low understanding) with the Pressurized Teaching Questionnaire and the Diagnostic Competence Questionnaire were smaller than expected. Student level correlations between A− (adaptive with little regulation upon high understanding) and the Pressurized Teaching Questionnaire were smaller than expected.

Table 7. Correlations (and expected correlation sizes and directions) between QTSA facets (teacher and student level) and other questionnaires.

	Student level				Teacher level			
	A+	A−	NA+	NA−	A+	A−	NA+	NA−
Questionnaires used for convergent validity								
Pressurized Teaching Questionnaire (Brühwiler & Blatchford, 2011)	−.33** (M/L−)	**−.11*** (M/L−)	**.14**** (M/L+)	**.35**** (L+)	−.73** (M/L−)	−.40* (M/L−)	.20 (M/L+)	.76** (L+)
Teachers' Adaptation of Instruction Questionnaire (Bayer et al., 2016; Kuger et al., 2016)	.36** (M/L+)	.23** (M/L+)	**.01** (M/L−)	−.20** (M/L−)	.74** (M/L+)	.70** (M/L+)	**−.15** (M/L−)	−.58** (M/L−)
Diagnostic Competence Questionnaire (Baumert et al., 1997)	.30** (M/L+)	.26** (M/L+)	**−.03** (M/L−)	**−.17**** (M/L−)	.79** (M/L+)	.63** (M/L+)	**.07** (M/L−)	−.70** (M/L−)
Adaptive Intervention Questionnaire (Klimczak et al., 2012)	.44** (M/L+)	.20** (M/L+)	**−.06*** (M/L−)	−.32** (M/L−)	.81** (M/L+)	.58** (M/L+)	**−.18** (M/L−)	−.68** (M/L−)
Questionnaires used for discriminant validity								
Diagnostic Competence of Social − Emotional Well − Being Questionnaire (PISA, 2003)	.24** (≤M+)	.24** (≤M+)	**.02** (≤M−)	−.08* (≤M−)	**.67**** (≤M+)	**.53**** (≤M+)	−.03 (≤M−)	**−.66**** (≤M−)
Teacher Autonomy Support Questionnaire (Sierens et al., 2009)	.40** (M/L+)	.19** (≤M+)	−.17** (≤M−)	−.35** (M/L−)	.74** (M/L+)	.36* (≤M+)	−.33* (≤M−)	−.77** (M/L−)
Interpersonal warmth dimension (communion) (QTI *16 items*) (Wubbels & Levy, 1993)	.29 (M/L+)	**.06** (M/L+)	**.03** (M/L−)	−.25** (M/L−)	.64** (M/L+)	.42* (M/L+)	−.35* (M/L−)	−.66** (M/L−)
Interpersonal warmth dimension (communion) (QTI *24 items*) (Wubbels & Levy, 1993)	.41** (M/L+)	.21** (M/L+)	**−.16**** (M/L−)	−.32** (M/L−)	.79** (M/L+)	.42* (M/L+)	−.41* (M/L−)	−.78** (M/L−)
Interpersonal agency dimension (QTI *16 items*) (Wubbels & Levy, 1993)	.18** (≤M+)	**.10*** (≤M−)	.02 (≤M+)	−.17* (≤M−)	**.67**** (≤M+)	**.34** (≤M−)	**−.05** (≤M+)	**−.56*** (≤M−)
Interpersonal agency dimension (QTI *24 items*) (Wubbels & Levy, 1993)	.20** (≤M+)	**.04** (≤M−)	**−.03** (≤M+)	−.12* (≤M−)	.18 (≤M+)	−.05 (≤M−)	.15 (≤M+)	−.25 (≤M−)

Note. Bold font indicates correlations that were not in line with our expectations. S = small, M = medium, L = Large. A+: adaptive support with much regulation upon low understanding; A−: adaptive support with little regulation upon high understanding; NA+: non-adaptive support with much regulation upon high understanding; NA−: non-adaptive support with little regulation upon low understanding.

* $p < .05$, ** $p < .001$.

Discriminant validity

Generally, our expectations regarding discriminant validity were confirmed as the majority of the correlations had the *direction* (41 of 48) and the *strength* (40 of 48) as expected (Table 7). The most important deviations from our expectations were the following. First, teacher level correlations of A+, A−, and NA − on the one hand and the Diagnostic Competence of Students' Social-Emotional Well-Being Questionnaire on the other hand were somewhat stronger than expected (large instead of ≤ medium). Second, teacher level correlations of A+, A−, and NA − on the one hand and Interpersonal Agency (16-item version) on the other hand were stronger than expected. Third, teacher level correlations with Interpersonal Agency were somewhat different for the two QTI versions. More strongly worded items in the long version may explain these differences, but differences may also be due to smaller teacher level sample sizes (due to the split sample).

Discussion

The four-facet structure (A+, A−, NA+, NA−) of student perceptions of teacher support adaptivity found in Study 1 also fitted the data of our second sample well (RQ2.1). Study 2 further showed that student perceptions of teacher support adaptivity as measured with the QTSA converged with other measures addressing teachers' support adaptivity or diagnostic competence (RQ2.2). QTSA scores generally diverged from student perceptions of constructs distinct from support adaptivity (RQ2.3), such as diagnostic competence of social-emotional well-being, teacher autonomy support, and interpersonal warmth and agency.

Facets of teacher support adaptivity

Similar to Study 1, students in Study 2 clearly distinguished between different facets of teacher support adaptivity. At the teacher level, all four facets could be reliably used for research purposes (e.g., to assess differences between teachers in support adaptivity) and for feedback for individual teachers (e.g., assess differences between teachers' support adaptivity). The only exception was NA + that seems mainly suited for research purposes and should be used with some caution when used for feedback for individual teachers.

At the student level, individual student perceptions could be reliably used to assess A + and A − for research purposes and for feedback for teachers (of individual students). NA − was suitable to be used for research purposes and should be used with some caution when using it as feedback from individual students. NA + should also be used with caution when using it for research purposes and as feedback from individual students.

Convergent and discriminant validity

The strength and direction of the relations between QTSA adaptivity facets and convergent and discriminant validity measures indicated that student perceptions of teacher support adaptivity could be assessed rather validly. Regarding convergent validity, we found that student perceptions of the QTSA facets generally correlated positively with questionnaires assessing aspects of support adaptivity and/or diagnostic competence, but not as high as to suggest multicollinearity. Only NA+ (non-adaptive with much regulation upon high understanding) was not related to these constructs. We discuss this finding in the General Discussion. One other unexpected finding was that student level correlations between NA− (non-adaptive with little regulation upon low understanding) and the Pressurized Teaching Questionnaire were lower than expected (medium, not large). A possible explanation could be that the QTSA uses the individual student as a referent whereas the Pressurized Teaching Questionnaire uses the class as a referent.

Overall, the correlations calculated for the assessment of discriminant validity had the direction and strength as expected. For example, associations between support adaptivity (and more

specifically, adaptive support with much regulation upon low understanding; A+) and interpersonal agency were substantial. Apparently, perceptions of teacher regulation and interpersonal agency overlap. However, given that the correlation is far from perfect, students still seem to make a distinction between high levels of interpersonal regulation (i.e., high agency) and high levels of regulation related to their own understanding of subject matter.

As expected (Fauth et al., 2014; Shevlin et al., 2000), correlations between support adaptivity and interpersonal warmth were medium to high, suggesting that students perceive adaptive teachers as warmer and more friendly, but at the same time, friendliness was not synonymous to adaptivity. Shevlin et al. (2000) argued that a central trait, such as likability, may exist and play a role in teacher ratings. Such a trait may explain part of the variability in QTSA ratings, although a likability *bias* did not seem present in the current study.

Correlations between support adaptivity and diagnostic competence of social-emotional well-being were somewhat stronger than expected. Apparently, teachers who are perceived to give adaptive support are also perceived to be social-emotionally sensitive and well able to diagnose students' social-emotional well-being. Either students confuse these two constructs to some extent, or teachers who are good at providing adaptive support are also good at diagnosing students' social-emotional well-being. It is thus likely that to be adaptive in terms of organizing learning processes, it is also helpful to take students' concurrent socio-emotional states into account. Also, the positive correlations between the QTSA and the warmth/communion dimension of the QTI aligns with this interpretation.

General discussion

The aim of this study was to assess teacher support adaptivity from the students' perspective. We first presented a conceptual framework based on adaptivity theory (Wood et al., 1978) distinguishing different facets of teacher support adaptivity. The Questionnaire on Teacher Support Adaptivity (QTSA) can provide reliable and valid scores of different facets of students' perceived support adaptivity, as we have shown in two independent large samples of secondary school students.

Support adaptivity facets

Students, both at the individual student and aggregated teacher level, distinguished four adaptivity facets in line with adaptivity theory (Wood et al., 1978): A+ (adaptive support with much teacher regulation upon low student understanding), A− (adaptive support with little regulation upon high understanding); NA+ (non-adaptive support with much regulation upon high understanding), and NA− (non-adaptive support with little regulation upon low understanding). Thus, *structural validity* of the QTSA was supported within the two large samples used in this study.

A further distinction between little regulation in the form of no support or much challenge could not be made. Apparently, students perceive teachers to provide little regulation in terms of no support, but only when their understanding is low (NA−). When their understanding is high, they perceive little regulation as providing much challenge. This seems reasonable: when students understand something, teachers probably do not stop giving support but take the next step in teaching a next skill or topic, thus providing much challenge. However, if students struggle, they notice when they get less support or regulation.

At the teacher level, students' aggregated perceptions on the four separate facets could be reliably used for research purposes or to provide feedback to individual teachers about their support adaptivity. At the student level, individual students' perceptions of A + and A − can be used for research purposes or as individual student's feedback for a teacher. NA − perceptions seemed mostly suitable for research purposes at the individual student level and should be used with

caution when using it as feedback from individual students. NA + should also be used with caution (or in combination with other sources such as interviews with students) when using it for research purposes and as feedback from individual students and requires some further investigation. Reports of 10 to 15 students per class are sufficient to reliably measure teachers' support adaptivity for A+, A−, and NA−; for NA+, reports of 20–33 students are needed.

The facets A + and NA − were highly correlated at the student level. This implies that, when a student is struggling, teachers who are inclined to give much regulation (e.g., provide an explanation) will also tend to refrain from support that provides little regulation (e.g., ask an open question or give a difficult assignment). For the current study, we decided to keep the four facet structure as this structure reflects adaptivity theory and showed a good fit in two large, independent samples. Future research could investigate why these two facets are so strongly correlated and could—with new samples—investigate whether the four facet structure is indeed the best structure.

Generalizability and external aspects of validity

There were no structural differences between boys' and girls' perceptions, students of different educational levels, and measurement occasions, indicating that student perceptions of teacher support adaptivity generalize across these aspects. This makes using student perceptions of teacher support adaptivity a potentially useful tool in a broad range of samples.

Overall, our analyses indicated that students' perceptions of support adaptivity as measured with the QTSA converged with—but were not similar to—other measures of teacher support adaptivity and diagnostic competence (i.e., convergent validity). Differences between the QTSA and the other questionnaires may be explained by the fact that the QTSA uses the individual student as a referent instead of the whole class. For students, reporting their own understanding rather than that of the whole class may be easier and more relevant for their own learning process. One exception is the Adaptive Intervention Questionnaire (Klimczak et al., 2012), that does use the individual student as a referent. Yet, the correlations with the QTSA were not higher compared to the correlations between the QTSA and the other questionnaires. This may be due to the fact that the Adaptive Intervention Questionnaire focuses both on students' perceptions of the teacher's ability to diagnose a student's understanding and teachers' adaptivity, whereas the QTSA only focuses on teachers' adaptivity.

Further, the QTSA targets the two constituent parts of teacher adaptivity separately, that is students own understanding and the teachers' regulation, instead of asking students to judge teacher adaptivity directly. In the Adaptive Intervention Questionnaire (Klimczak et al., 2012), for example, students have to judge whether the support fitted their understanding, which may be hard to do for students. Describing what a teacher does in situations in which they do or do not understand the subject matter—as is required in the QTSA—may be easier to report for students and may result in more objective data.

In addition, based on adaptivity theory, the QTSA addresses fine-grained facets of support adaptivity (Wood et al., 1978), providing a comprehensive picture of teacher support adaptivity. Existing questionnaires addressing (aspects) of support adaptivity were not subjected to factor analyses, probably given the number of items per questionnaire (max. 5). In addition, none of the existing questionnaires addresses more than one adaptivity facet.

Student perceptions of support adaptivity were sufficiently different from their perceptions of distinct constructs such as teachers' interpersonal agency, supporting external validity. Also, as expected, students' perceptions of teachers' support adaptivity and interpersonal warmth converged to some extent, but not as to say that students confused support adaptivity with warmth. The relation between students' perceptions of teachers' support adaptivity and diagnostic competence of social-emotional well-being was higher than expected, but not as much as to suggest that

students did not see them as separate teaching aspects. Support adaptivity is possibly easier to achieve for teachers if they also take students' social-emotional well-being into account. Future research could test this hypothesis. Future research could also further investigate the QTSA's external validity by relating student perceptions to observational measures of support adaptivity (e.g., Van de Pol et al., 2014).

Non-Adaptive support with much regulation upon high understanding (NA+)

Although students could distinguish between four facets, the facet of NA + showed, in some regards, unexpected results. For example, although the reliability at the teacher level was sufficient for using this facet for individual decisions and research purposes, the reliability at the student level would preferably be a bit higher (current $\alpha/\Omega = .63/.66$). For now, students' individual perceptions of this facet should be interpreted with caution and the student-level reliability should be closely monitored in future samples. Possibly, this specific combination of good student understanding and more teacher regulation (i.e., NA+) occurs rarely in the perception of students and was therefore hard to judge. Future research could explore whether this facet's internal consistency can be improved e.g., by testing additional items.

Also, with regard to convergent validity, the NA + facet showed deviating results. Students' NA + perceptions showed relatively small correlations with students' scores of the other adaptivity questionnaires that were used to test for convergent validity. Future research could explore reasons for this finding. Again, perhaps providing more regulation upon good student understanding (NA+) occurs rarely, is therefore hard to judge, and does therefore relate only weakly to other measures of support adaptivity.

Limitations

Some limitations should be taken into account when interpreting the results. First, adaptivity theory focuses on how regulation *shifts* in relation to students' understanding. That is, support is adaptive when regulation levels *increase* upon low student understanding or *decrease* upon high understanding. However, QTSA items address much/little regulation in combination with low/high student understanding. The deliberate simplification of this process in our items may have oversimplified teacher adaptive support. However, we deemed that estimating regulation *shifts* would have been too complex for students.

Second, our sample cannot be generalized to the whole population regarding nationality (25.5% of the Dutch population in secondary education has a migration background, probably relating to students' nationality (Centraal Bureau voor de Statistiek (CBS),), 2021) compared to around 5% in our samples). We were also unable to test for measurement invariance regarding grade level due to too a small number of upper grade classes in our sample (Kline, 2005).

Furthermore, although suitable for a wide range of subjects and educational levels or tracks, the QTSA addressed teacher support during seatwork. Although seatwork is a common educational activity (Mullis et al., 2008), future research could explore the applicability of the QTSA to other instructional settings. We chose to apply this focus to be able to assess teacher support adaptivity in interaction with specific, individual students (O'Keefe et al., 2006).

Finally, we have modified our model using statistical criteria (e.g., removing items with low factor loadings). This raises the question of generalizability to the population. The findings of Study 2, where we fitted the same theory-based four factor structure using a new sample and tested whether factor loadings were comparable when using the same restrictions, provide a first indication for the robustness of the factor structure. Further replications are necessary to confirm this structure with other samples.

QTSA score interpretation and meaning

We approached support adaptivity as a configural construct (Stapleton et al., 2016). That is, we were both interested in class aggregated scores referring to support adaptivity at the teacher level and in student scores at the individual level. For teachers, aggregated scores on the QTSA could be seen as an indication of teachers' competence with regard to their support adaptivity (as perceived by their students). Individual student scores can indicate what an individual student needs (according to their own perception) regarding support adaptivity (e.g., if a student scores low on A+, the student may need more help that is of increasing regulation upon low understanding). The factorial invariance analysis showed that perceived support adaptivity can be conceptualized as a four facet construct (A+, A−, NA+, NA−) at both the teacher and the student level (with rather similar factor loadings at both levels). Therefore, we recommend researchers and practitioners who use this questionnaire to calculate four scores, one per facet. Given that the item loadings were rather similar within facets, facet scores can be calculated as means over the facet's items (see Appendix A for an overview of which item belongs to which facet and the Supplementary Material for the actual questionnaire formatted for use by researchers and practitioners).

At the teacher level, a one factor structure—indicating general adaptivity—also showed good fit for all fit indices except the TLI. Therefore, at the teacher level, users of the questionnaire could also calculate one general adaptivity score by computing the average over all items. Note that before doing so, items from the NA scales (NA − and NA+) should be reverse coded (see Supplementary Material). At the student level, other facet structures (e.g., one general adaptivity score) did not show sufficient fit. Therefore, at this level we strongly recommend to calculate four separate facet scores when using student level scores of the QTSA.

Conclusion

This study investigated to what degree teacher support adaptivity can be assessed with, and is reflected in student perceptions using a newly developed questionnaire based on adaptivity theory (Wood et al., 1978). In line with this theory, the QTSA distinguishes between four facets of support adaptivity (A+, A−, NA+, NA−).

Using four facets can help to give a differentiated picture of support adaptivity as perceived by students. Expressing support adaptivity in terms of four facets can help teachers to get detailed feedback from individual students ('how does this particular student experience my support in terms of the four facets of support adaptivity') or from their class as a whole when they want to know which facets of support adaptivity they master or could further improve according to their students. In addition, it can help researchers to gain more insight into possible differential mechanisms or effectiveness of different adaptivity facets.

Notes

1. Demographic information was not available for all teachers ($n_{gender} = 31$).
2. This disadvantage score is based on the level of education and country of birth of the parents, the length of stay of the mother in the Netherlands, whether the parents are taking part in a debt restructuring program, and average level of education of the mother (Statistics Netherlands, 2020).
3. German, French, Dutch, and English.
4. Geography, history, economics, and social studies.
5. The average disadvantage score of these schools was 24.86 ($SD = 37.57$).
6. For the English version, the backtranslation method was used. Only the Dutch items were tested in this study.

7. One student per class collected the questionnaires, put the questionnaires in an envelope and sealed it, to ensure teachers would not see the questionnaires. Students were informed about this procedure beforehand.
8. Nunnally and Bernstein (1994) indicated that a threshold of .80 should be used for basic or applied research whereas a threshold of .90 should be used in applied settings where crucial decisions are being taken based on individual questionnaire outcomes. Given that the QTSA is about perceived support adaptivity, this instrument will/should not be used to make crucial decisions based on individual scores and therefore we adhere to the threshold of .80 here.
9. German, French, Dutch, and English.
10. Geography, history, economics, and social studies.
11. First, items were translated to Dutch. Second, items were translated back to the original language. Third, original and back-translated items were compared and, if needed, adjustments were made to the translations.
12. Only for the reliability analyses, we found two instances in which the alpha/omega were just below instead of on the threshold of .80; these are reported in the footnote of Table 6.
13. Given that it is only one of the three questions that relates to a specific QTSA factor, we did not formulate a different expectation for this scale.

Declaration of interest statement

No potential competing interest was reported by the authors.

References

Akaike, H. (1974). A new look at the statistical model identification. In *Selected papers of Hirotugu Akaike* (pp. 215–222). Springer.

Baumert, J., Gruehn, S., Heyn, S., Köller, O., & Schnabel, K.-U. (1997). *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU) [Educational and psychosocial development in adolescence]*. Max-Planck-Institut für Bildungsforschung.

Bayer, S., Klieme, E., & Jude, N. (2016). Assessment and evaluation in educational contexts. In *Assessing contexts of learning* (pp. 469–488). Springer.

Belmont, M., Skinner, E., Wellborn, J., & Connell, J. (1988). *Teacher as social context: A measure of student perceptions of teacher provision of involvement, structure, and autonomy support* (Technical Report No. 102). University of Rochester.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford Press.

Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural-equation modeling* (pp. 361–379). The Guilford Press.

Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, *21*(1), 95–108. https://doi.org/10.1016/j.learninstruc.2009.11.004

Brunner, M., & Süβ, H. M. (2005). Analyzing the reliability of multidimensional measures: An example from intelligence research. *Educational and Psychological Measurement*, *65*(2), 227–240. https://doi.org/10.1177/0013164404268669

Centraal Bureau voor de Statistiek (CBS). (2021, Januari 14). *Bevolking; kerncijfers*. Retrieved from https://statline.cbs.nl/Statweb/selection/?DM=SLNL&PA=37296NED&VW=T

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chia, H. M., & Lim, C. S. (2020). Characterising the pedagogical practices in mathematics lessons among selected Malaysian primary schools. *The Mathematics Enthusiast*, *17*(1), 307–323. https://doi.org/10.54870/1551-3440.1488

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

De Kleijn, R. A., Bronkhorst, L. H., Meijer, P. C., Pilot, A., & Brekelmans, M. (2016). Understanding the up, back, and forward-component in master's thesis supervision with adaptivity. *Studies in Higher Education*, *41*(8), 1463–1479. https://doi.org/10.1080/03075079.2014.980399

Dunn, E. C., Masyn, K. E., Jones, S. M., Subramanian, S. V., & Koenen, K. C. (2015). Measuring psychosocial environments using individual responses: An application of multilevel factor analysis to examining students in schools. *Prevention Science*, *16*(5), 718–733. https://doi.org/10.1007/s11121-014-0523-x

Eder, F., & Mayr, J. (2000). *Linzer Fragebogen zum Schüle und Klassenklima für die 4-8. Klassenstufe). [Questionnaire on school and class climate for 4th to 8th grade]*. Hogrefe.

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, *29*, 1–9. https://doi.org/10.1016/j.learninstruc.2013.07.001

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

Guttman, L. (1954). An outline of some new methodology for social research. *Public Opinion Quarterly*, *18*(4), 395–404. https://doi.org/10.1086/266532

Hardy, I., Decristan, J., & Klieme, E. (2019). Adaptive teaching in research on learning and instruction. *Journal for Educational Research Online*, *11*(2), 169–191.

Hermkes, R., Mach, H., & Minnameier, G. (2018). Interaction-based coding of scaffolding processes. *Learning and Instruction*, *54*, 147–155. https://doi.org/10.1016/j.learninstruc.2017.09.003

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modeling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*, 53–60. https://arrow.dit.ie/libart/4/

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

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Iglesias García, M. T., Maulana, R., Fernández García, C. M., & García Pérez, O. (2020). Teacher as social context (TASC) questionnaire in the Spanish setting: Teacher version. *Psicología Educativa*, *26*(1), 17–26. https://doi.org/10.5093/psed2019a15

Jackson, D. N. (1971). The dynamics of structured personality tests. *Psychological Review*, *78*(3), 229–248. https://doi.org/10.1037/h0030852

Jak, S. (2013). *Cluster bias: Testing measurement invariance in multilevel data* (Doctoral's dissertation). https://pure.uva.nl/ws/files/2313851/127281_06.pdf

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

Katz, I., & Assor, A. (2007). When choice motivates and when it does not. *Educational Psychology Review*, *19*, 429–442. https://doi.org/10.1007/s10648-006-9027-y

Kim, E., Kwok, O., & Yoon, M. (2012). Testing factorial invariance in multilevel data: A Monte Carlo study. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(2), 250–267. https://doi.org/10.1080/10705511.2012.659623

Klimczak, M., Kampa, M., Bürgermeister, A., Harks, B., Rakoczy, K., Besser, M., Klieme, E., Blum, W., & Leiss, D. (2012). *Dokumentation der Befragungsinstrumente der Interventionsstudie im Projekt "Conditions and Consequences of Classroom Assessment" (Co2CA)*. DIPF.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). The Guilford Press.

Krämer, S., & Zimmermann, F. (2021). Students with emotional and behavioral disorder and teachers' stereotypes–effects on teacher judgments. *The Journal of Experimental Education*, 89, 1–22. https://doi.org/10.1080/00220973.2021.1934809

Kuger, S., Jude, N., Klieme, E., & Kaplan, D. (2016). Adaptation of instruction - Schüler [Fragebogenskala: Version 1.0]. In: *Programme for International Student Assessment (2015)—Fragebogenerhebung Chapter 4—Field Trial (PISA 2015) [Skalenkollektion: Version 1.0]. Datenerhebung 2014*. Forschungsdatenzentrum Bildung am DIPF. http://dx.doi.org/10.7477/150:0:1

Kunter, M., & Baumert, J. (2007). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, *9*(3), 231–251. https://doi.org/10.1007/s10984-006-9015-7

Kupers, E., van Dijk, M., & van Geert, P. (2015). Within-teacher differences in one-to-one teacher–student interactions in instrumental music lessons. *Learning and Individual Differences*, *37*, 283–289. https://doi.org/10.1016/j.lindif.2014.11.012

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, *9*(2), 202–220. https://doi.org/10.1177/1094428105284919

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*(4), 815–852. https://doi.org/10.1177/1094428106296642

Levine, T. R. (2005). Confirmatory factor analysis and scale validation in communication research. *Communication Research Reports*, *22*(4), 335–338. https://doi.org/10.1080/00036810500317730

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology*, *34*(2), 120–131. https://doi.org/10.1016/j.cedpsych.2008.12.001

Lüdtke, O., Trautwein, U., Kunter, M., & Baumert, J. (2007). Reliability and agreement of student ratings of the classroom environment: A reanalysis of TIMSS data. *Learning Environments Research*, *9*(3), 215–230. https://doi.org/10.1007/s10984-006-9014-8

Luengo, J., García, S., & Herrera, F. (2010). A study on the use of imputation methods for experimentation with Radial Basis Function Network classifiers handling missing attribute values: The good synergy between RBFNs and Event Covering method. *Neural Networks*, *23*(3), 406–418. https://doi.org/10.1016/j.neunet.2009.11.014

Mainhard, T., Oudman, S., Hornstra, L., Bosker, R. J., & Goetz, T. (2018). Student emotions in class: The relative importance of teachers and their interpersonal relations with students. *Learning and Instruction*, *53*, 109–119. https://doi.org/10.1016/j.learninstruc.2017.07.011

Marsh, H. W. (2007). Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *Journal of Educational Psychology*, *99*(4), 775–790. https://doi.org/10.1037/0022-0663.99.4.775

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*(2), 5–11. https://doi.org/10.3102/0013189X018002005

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. https://doi.org/10.1002/j.2333-8504.1994.tb01618.x

Moallem, M. (1998). An expert teacher's thinking and teaching and instructional design models and principles: An ethnographic study. *Educational Technology Research and Development*, *46*(2), 37–64. https://doi.org/10.1007/BF02299788

Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 International Mathematics report*. TIMSS & PIRLS International Study Center.

Murphy, N., & Messer, D. (2000). Differential benefits from scaffolding and children working alone. *Educational Psychology*, *20*(1), 17–31. https://doi.org/10.1080/014434100110353

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*(3), 376–398. https://doi.org/10.1177/0049124194022003006

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

O'Keefe, C., Xu, L. H., & Clarke, D. (2006). Kikan-Shido: Between desks instruction. In D. Clarke, J. Emanuelsson, E. Jablonka, and I. A. C. Mok (Eds.), *Making connections: Comparing mathematics classrooms around the world* (pp. 73–105). Sense Publishers.

Oosterveld, P., & Vorst, H. C. M. (1996). Methoden voor vragenlijstconstructie [Questionnaire Design Methods]. *Nederlands Tijdschrift voor de Psychologie*, *51*, 11–27.

Oosterveld, P., Vorst, H. C., & Smits, N. (2019). Methods for questionnaire design: A taxonomy linking procedures to test goals. *Quality of Life Research*, *28*(9), 2501–2512. https://doi.org/10.1007/s11136-019-02209-6

Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. G., Pierczynski, M., & Allen, M. (2018). Teachers' instructional adaptations: A research synthesis. *Review of Educational Research*, *88*(2), 205–242. https://doi.org/10.3102/0034654317743198

Pino-Pasternak, D., Whitebread, D., & Tolmie, A. (2010). A multidimensional analysis of parent-child interactions during academic tasks and their relationships with children's self-regulated learning. *Cognition and Instruction*, *28*(3), 219–272. https://doi.org/10.1080/07370008.2010.490494

PISA. (2003). Dokumentation der Erhebungsinstrumente. http://library.mpib-berlin.mpg.de/toc/ze_2006_549.pdf

Pratt, M. W., & Savoy-Levine, K. M. (1998). Contingent tutoring of long-division skills in fourth and fifth graders: Experimental tests of some hypotheses about scaffolding. *Journal of Applied Developmental Psychology*, *19*(2), 287–304. https://doi.org/10.1016/S0193-3973(99)80041-0

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*(3), 333–343. https://doi.org/10.1007/BF02294360

Shevlin, M., Banyard, P., Davies, M., & Griffiths, M. (2000). The validity of student evaluation of teaching in higher education: love me, love my lectures? *Assessment & Evaluation in Higher Education*, *25*(4), 397–405. https://doi.org/10.1080/713611436

Sierens, E., Vansteenkiste, M., Goossens, L., Soenens, B., & Dochy, F. (2009). The synergistic relationship of perceived autonomy support and structure in the prediction of self-regulated learning. *The British Journal of Educational Psychology*, *79*(Pt 1), 57–68. https://doi.org/10.1348/000709908X304398

Smit, J., A. A. van Eerde, H., & Bakker, A. (2013). A conceptualisation of whole-class scaffolding. *British Educational Research Journal*, *39*(5), 817–834. https://doi.org/10.1002/berj.3007

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

Statistics Netherlands, (2020). https://www.cbs.nl/nl-nl/maatwerk/2021/05/achterstandsscores-per-school-2020.

Van Braak, M., van de Pol, J., Poorthuis, A. M., & Mainhard, T. (2021). A micro-perspective on students' behavioral engagement in the context of teachers' instructional support during seatwork: Sources of variability and the role of teacher adaptive support. *Contemporary Educational Psychology*, *64*, 101928. https://doi.org/10.1016/j.cedpsych.2020.101928

Van de Pol, J., & Elbers, E. (2013). Scaffolding student learning: A micro-analysis of teacher–student interaction. *Learning, Culture and Social Interaction*, *2*(1), 32–41. https://doi.org/10.1016/j.lcsi.2012.12.001

Van de Pol, J., Mercer, N., & Volman, M. (2019). Scaffolding student understanding in small-group work: Students' uptake of teacher support in subsequent small-group interaction. *Journal of the Learning Sciences*, *28*(2), 206–239. https://doi.org/10.1080/10508406.2018.1522258

Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, *22*(3), 271–296. https://doi.org/10.1007/s10648-010-9127-6

Van de Pol, J., Volman, M., & Beishuizen, J. (2011). Patterns of contingent teaching in teacher–student interaction. *Learning and Instruction*, *21*(1), 46–57. https://doi.org/10.1016/j.learninstruc.2009.10.004

Van de Pol, J., Volman, M., Oort, F., & Beishuizen, J. (2014). Teacher scaffolding in small- group work – An intervention study. *Journal of the Learning Sciences*, *23*(4), 600–650. https://doi.org/10.1080/10508406.2013.805300

Vermunt, J. D., & Verloop, N. (1999). Congruence and friction between learning and teaching. *Learning and Instruction*, *9*(3), 257–280. https://doi.org/10.1016/S0959-4752(98)00028-0

Vygotsky, L. S. (1978). *Mind in society—The development of higher psychological processes*. Harvard University Press.

Wagner, W., Göllner, R., Helmke, A., Trautwein, U., & Lüdtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: Dimensionality and generalizability of domain-independent assessments. *Learning and Instruction*, *28*, 1–11. https://doi.org/10.1016/j.learninstruc.2013.03.003

Wood, D. (1988). *How children think and learn*. Blackwell.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem-solving. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Wood, D., & Middleton, D. (1975). A study of assisted problem-solving. *British Journal of Psychology*, *66*(2), 181–191. https://doi.org/10.1111/j.2044-8295.1975.tb01454.x

Wood, D., Wood, H., & Middleton, D. (1978). An experimental evaluation of four face-to-face teaching strategies. *International Journal of Behavioral Development*, *1*(2), 131–147. https://doi.org/10.1177/016502547800100203

Wubbels, T. & Levy, J. (Eds.) (1993). *Do you know what you look like: Interpersonal relationships in education*. Falmer Press.

Wubbels, T., Brekelmans, M., den Brok, P., & van Tartwijk, J. (2006). An interpersonal perspective on classroom management in secondary classrooms in the Netherlands. In: *Handbook of classroom management* (pp.1171–1202). Routledge.

Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*(1), 165–200. https://doi.org/10.1111/0081-1750.00078

Appendix A

Questionnaire on teacher support adaptivity (final version)

General instruction: This questionnaire concerns the way the teacher helps you when working independently on an assignment [In Dutch: Deze vragenlijst gaat over hoe deze docent jou helpt als je zelfstandig aan een opdracht werkt.]

No.	Facet	Item in English	Item as used in this study (in Dutch)
1.	A−	When I understand something well, this teacher makes it a little bit harder for me.	Als ik het goed snap, maakt deze docent het ietsje moeilijker voor mij.
2.	A+	When I don't know how to continue, this teacher helps me to find the correct answer.	Als ik echt niet verder kan, helpt deze docent mij met het goede antwoord vinden.
3.	A+	When I get completely stuck with an exercise, this teacher shows me how to do it.	Als ik helemaal vastloop met een opdracht laat deze docent mij zien hoe het moet.
4.	A−	When I am doing well, this teacher lets me do a difficult exercise.	Als ik goed bezig ben, mag ik een moeilijke opdracht doen van deze docent.
5.	NA−	When I find an exercise very difficult, this teacher still lets me do it on my own.	Als ik een opdracht heel moeilijk vind, laat deze docent het mij toch zelf uitzoeken.
6.	NA+	This teacher helps me with things that I already understand.	Deze docent helpt mij met dingen die ik al lang begrijp.
7.	A+	This teacher gives me an example, when I really don't know how to continue with the exercise.	Deze docent geeft mij een voorbeeld als ik echt niet verder kan met de opdracht.
8.	NA−	This teacher tells me to do it on my own, even though I am unable to continue.	Deze docent zegt dat ik het zelf moet doen, ook al kom ik echt niet verder.
9.	A+	When I really do not understand an exercise, this teacher explains to me how to go about it.	Als ik een opdracht echt niet begrijp, legt deze docent mij uit hoe ik het aan kan pakken.
10.	NA−	When I find an exercise difficult, this teacher barely explains it to me.	Als ik een opdracht moeilijk vind, geeft deze docent mij nauwelijks uitleg.
11.	A−	When I understand it well, I am allowed to do another exercise.	Als ik het goed snap, mag ik een andere opdracht doen.
12.	NA−	When I don't know how to do it yet, I still have to continue with the next exercise.	Als ik iets nog niet kan, moet ik toch al verder met de volgende opdracht.
13.	NA+	When I am already able do it, this teacher still helps me.	Als ik het al kan, komt deze docent mij toch helpen.
14.	NA−	When I do not yet understand the exercise, this teacher still makes it more difficult for me.	Als ik de opdracht nog niet snap, maakt deze docent het toch moeilijker voor mij.
15.	A−	When I know how to do it, I get a more difficult exercise.	Als ik het kan, krijg ik een iets moeilijkere opdracht.
16.	A−	When I understand an exercise well, this teacher makes it nice and challenging for me.	Als ik een opdracht goed begrijp, maakt deze docent het lekker moeilijk voor mij.
17.	NA+	This teacher asks me things that I already know.	Deze docent vraagt mij dingen die ik al lang weet.
18.	A+	When I get totally confused, this teacher helps me to find a solution.	Als ik helemaal in de war raak, dan helpt deze docent mij een oplossing te vinden.
19.	NA+	This teacher gives me an explanation, even though I already understand it.	Deze docent geeft mij uitleg ook al snap ik het al lang.
20.	NA−	This teacher makes it a bit more difficult for me, even though I do not understand it yet.	Deze docent maakt het ietsje moeilijker voor mij, ook al begrijp ik het nog niet.
21.	A+	When I really do not get it, this teacher helps me to find out what I need to write down.	Als ik er echt niet uit kom, helpt deze docent mij met wat ik op moet schrijven.

Note. A five point likert scale ranging from totally disagree (1) to totally agree (5) is used. A + refers to adaptive support with much regulation upon low student understanding, A − refers to adaptive support with little regulation upon high student understanding, NA + refers to non-adaptive support with much regulation upon high student understanding, and NA − refers to non-adaptive support with little regulation upon low student understanding.

Appendix B

Additional model results, factor correlations, and descriptive statistics for Study 1 and Study 2

Table B1. Factor correlations from final model Study 1.

Factor	A+	A−	NA+	NA−
A+	1	.566	−.155	−.968
A−	.328	1	−.115	−.506
NA+	−.079	.110	1	.230
NA−	−.752	−.095	.346	1

Note. Values above diagonal are teacher level correlations, values below the diagonal are student level correlations. A+: adaptive support with much regulation upon low understanding; A−: to adaptive support with little regulation upon high understanding; NA+: non-adaptive support with much regulation upon high understanding; NA−: non-adaptive support with little regulation upon low understanding.

Table B2. Standardized (and unstandardized) model results from final model Study 1.

		Student level		Teacher level		
Facet	Item	Factor loading	Residual variance	Factor loading	Residual variance	Intercept
A+	2	.688 (1.000)	.527 (.511)	.993 (1.000)	.014 (.002)	11.012 (4.002)
	5	.671 (.930)	.550 (.484)	.931 (.930)	.134 (.017)	11.574 (4.171)
	9	.647 (.927)	.581 (.548)	.944 (.927)	.108 (.014)	11.144 (3.949)
	11	.674 (.914)	.546 (.461)	1.000 (.914)	.001 (.000)	12.475 (4.117)
	22	.731 (1.082)	.466 (.469)	.962 (1.082)	.075 (.012)	9.760 (3.961)
	27	.595 (.933)	.646 (.729)	.891 (.933)	.206 (.029)	10.005 (3.782)
A−	1	.616 (1.000)	.620 (.816)	.872 (1.000)	.240 (.028)	7.860 (2.664)
	6	.707 (1.130)	.500 (.638)	.927 (1.130)	.141 (.018)	7.755 (2.794)
	13	.613 (1.023)	.625 (.870)	.889 (1.023)	.210 (.024)	8.660 (2.945)
	18	.781 (1.238)	.390 (.489)	1.000 (1.238)	.001 (.000)	7.596 (2.780)
	20	.682 (1.020)	.535 (.598)	.972 (1.020)	.056 (.005)	7.853 (2.437)
NA+	8	.669 (1.000)	.552 (.573)	.938 (1.000)	.120 (.013)	7.581 (2.509)
	15	.520 (.802)	.729 (.804)	.816 (.802)	.334 (.031)	7.646 (2.332)
	21	.458 (.698)	.790 (.850)	.849 (.698)	.280 (.018)	10.245 (2.616)
	23	.630 (.959)	.604 (.650)	.999 (.959)	.001 (.000)	9.065 (2.700)
NA−	7	.607 (1.000)	.632 (.735)	.995 (1.000)	.009 (.001)	6.481 (2.254)
	10	.673 (1.063)	.548 (.586)	.944 (1.063)	.109 (.017)	5.395 (2.104)
	12	.677 (1.037)	.542 (.544)	.957 (1.037)	.085 (.012)	5.242 (1.966)
	14	.486 (.840)	.763 (.975)	.999 (.840)	.001 (.000)	9.424 (2.744)
	16	.545 (.834)	.703 (.705)	.949 (.834)	.100 (.009)	6.512 (1.981)
	24	.526 (.791)	.724 (.701)	.878 (.791)	.229 (.022)	6.622 (2.066)

Note. A+: adaptive support with much regulation upon low understanding; A−: to adaptive support with little regulation upon high understanding; NA+: non-adaptive support with much regulation upon high understanding; NA − non-adaptive support with little regulation upon low understanding.

Table B3. Descriptive statistics of Study 2 data per QTSA facet and Questionnaires used for convergent and divergent validity on student level and teacher level.

Facet		N	M	SD	Min	Max
A+	Student	1156	0.00	.572	−2.49	1.81
	Teacher	55	4.16	.354	3.19	4.75
A−	Student	1154	0.00	.744	−2.63	2.53
	Teacher	55	2.82	.434	1.81	3.81
NA+	Student	1165	0.00	.647	−2.12	2.24
	Teacher	55	2.46	.235	1.96	3.22
NA−	Student	1164	0.00	.578	−1.57	2.39
	Teacher	55	2.01	.290	1.47	2.76
Pressurized Teaching Questionnaire	Student	1112	0.00	.818	−1.98	3.06
	Teacher	54	2.01	.412	1.20	2.98
Teachers' Adaptation of Instruction Questionnaire	Student	1117	0.00	.578	−1.81	1.70
	Teacher	54	2.67	.306	1.96	3.33
Diagnostic Competence Questionnaire	Student	1094	0.00	.700	−2.59	1.89
	Teacher	54	3.38	.433	2.34	4.23
Adaptive Intervention Questionnaire	Student	1098	0.00	.487	−2.03	1.57
	Teacher	54	2.99	.301	2.23	3.60
Teacher Autonomy Support Questionnaire	Student	1081	0.00	.504	−1.90	1.33
	Teacher	54	3.79	.270	3.25	4.53
Diagnostic Competence of Social-Emotional Well-Being Questionnaire	Student	1106	0.00	.769	−2.27	2.36
	Teacher	54	3.19	.471	2.10	4.16
Interpersonal warmth dimension (communion) (QTI *16 items*)	Student	484	0.00	.452	−2.17	1.42
	Teacher	25	0.93	.520	−0.17	1.87
Interpersonal warmth dimension (communion) (QTI *24 items*)	Student	583	0.00	.601	−2.68	1.77
	Teacher	29	1.25	.562	0.03	2.26
Interpersonal agency dimension (QTI *16 items*)	Student	484	0.00	.700	−3.20	1.91
	Teacher	25	0.47	.450	−0.75	0.99
Interpersonal agency dimension (QTI *24 items*)	Student	583	0.00	.423	−2.69	1.33
	Teacher	29	.47	.296	−0.21	1.21

Note. Negatively worded items were recoded. The interpersonal agency and teacher warmth dimensions of the QTI were computed by weighting each item separately for each of the interpersonal dimensions (cf. Mainhard et al., 2018). Student level descriptives are centered on the group mean (i.e., teacher-level); therefore the student-level mean is always zero.

Table B4. Standardized (and unstandardized) model results from final model Study 2.

		Student level		Teacher level		
Facet	Item	Factor loading	Residual variance	Factor loading	Residual variance	Intercept
A+	2	.678 (1.00)	.540 (.355)	1.000 (1.00)	.000 (.000)	12.303 (4.230)
	5	.672 (.990)	.548 (.360)	.960 (.990)	.079 (.010)	12.220 (4.336)
	9	.526 (.822)	.723 (.534)	.960 (.822)	.079 (.007)	13.821 (4.070)
	11	.693 (.973)	.519 (.310)	1.000 (.973)	.001 (.000)	12.659 (4.235)
	22	.717 (1.097)	.486 (.345)	.996 (1.097)	.008 (.001)	10.885 (4.122)
	27	.496 (.853)	.754 (.675)	.886 (.853)	.215 (.024)	11.677 (3.867)
A−	1	.584 (1.000)	.659 (.715)	.781 (1.000)	.390 (.062)	6.591 (2.637)
	6	.704 (1.231)	.505 (.572)	.964 (1.231)	.070 (.011)	7.270 (2.900)
	13	.570 (1.011)	.676 (.787)	.824 (1.011)	.032 (.047)	8.010 (3.071)
	18	.793 (1.321)	.371 (.380)	1.000 (1.321)	.001 (.000)	7.026 (2.900)
	20	.690 (1.124)	.524 (.514)	.972 (1.124)	.055 (.007)	7.014 (2.534)
NA+	8	.645 (1.000)	.585 (.559)	.822 (1.000)	.324 (.000)	10.811 (2.383)
	15	.575 (.845)	.669 (.595)	.593 (.845)	.648 (.045)	8.458 (2.225)
	21	.390 (.575)	.848 (.768)	.739 (.575)	.545 (.009)	17.938 (2.591)
	23	.573 (.871)	.672 (.685)	.998 (.871)	.004 (.000)	16.175 (2.670)
NA−	7	.525 (1.000)	.725 (.679)	.917 (1.000)	.159 (.013)	7.490 (2.148)
	10	.604 (1.044)	.636 (.491)	.999 (1.044)	.002 (.000)	6.855 (1.886)
	12	.604 (1.024)	.636 (.460)	.961 (1.024)	.076 (.000)	6.155 (1.698)
	14	.455 (.916)	.793 (.828)	.999 (.916)	.002 (.000)	10.330 (2.491)
	16	.573 (1.040)	.672 (.570)	.825 (1.040)	.319 (.036)	5.687 (1.884)
	24	.538 (.950)	.710 (.571)	.781 (.950)	.389 (.024)	6.319 (2.021)

Note. A+: adaptive support with much regulation upon low understanding; A − to adaptive support with little regulation upon high understanding; NA+: non-adaptive support with much regulation upon high understanding; NA−: non-adaptive support with little regulation upon low understanding.

Table B5. Factor correlations from final model Study 2.

Factor	A+	A−	NA+	NA−
A+	1	.622	−.604	−.953
A−	.231	1	−.203	−.591
NA+	−.121	.102	1	.675
NA−	−.696	.056	.485	1

Note. Values above diagonal: teacher level correlations; below the diagonal: student level correlations. A+: adaptive support with much regulation upon low understanding; A−: to adaptive support with little regulation upon high understanding; NA+: non-adaptive support with much regulation upon high understanding; NA − non-adaptive support with little regulation upon low understanding.