Empirical Paper

EUROPEAN JOURNAL OF Personality

European Journal of Personality
2024, Vol. 0(0) 1–22

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/08902070241301810
journals.sagepub.com/home/ejop
Sage

Students' perceptions of instructional quality and learning achievement in everyday life: Do personality traits matter?

Irma Talić[1], John F. Rauthmann[2], Karl-Heinz Renner[1], Jens Möller[3] and Christoph Niepel[4]

Abstract

Students' perceptions of instructional quality (SPIQ) are subjective and time-specific to some extent. Yet, they are mostly aggregated across students and assessed at one time point, neglecting student- and lesson-specific variance. The present study examined the role of students' personality traits in state SPIQ and their relation to perceived lesson-specific learning achievement (i.e., self-reported comprehension). Thereby, we distinguished between idiosyncratic and consensual (classroom) SPIQ. We assessed the three basic dimensions of instructional quality, teacher support, cognitive activation, and classroom management, as state perceptions of all students within classrooms in mathematics' instruction ($N_{observations}$ = 2681) across three weeks of 372 German secondary school students' (M_{age} = 15.3 years) daily life. Linear mixed effect models revealed (a) that students' agreeableness and negative emotionality were positively and negatively, respectively, related to state SPIQ, (b) particularly pronounced positive relations between teacher support and perceived learning achievement, which were (c) stronger for lower levels in agreeableness. Differences across idiosyncratic and consensual perceptions could hardly be detected. Thus, the present study shed light on personality traits' relations to SPIQ and within-student SPIQ–learning achievement associations, while demonstrating a new application for classroom-based state SPIQ that bridges the gap between intra- and intersubjective perceptions of instructional behavior.

Plain language summary

Typically, students' perceptions of teaching quality are summarized based on classes or schools to inform on their teachers' teaching quality. However, these perceptions are subjective to some degree: One student may generally perceive teaching quality higher than another student, who is more critical. The very same student may also differ in his or her perceptions depending on the time point, perceiving teaching quality lower in one lesson as compared to another lesson of the same teacher. To gain more insights into how perceptions of teaching quality differ across students and time, we assessed students repeatedly and examined multiple lesson-specific perceptions of teaching quality over a three-week period in school life. We explored whether students' personality traits played a role in these daily perceptions. Additionally, we assessed students' self-reported lesson comprehension to determine if, for instance, higher ratings of teaching quality were related to higher comprehension. 372 German secondary school students (mean age = 15.3 years) took part in our study. Our results indicated that, based on the five broad personality traits open-mindedness, conscientiousness, extraversion, agreeableness and negative emotionality, the latter two traits were related to perceptions of teaching quality: Higher levels of agreeableness (i.e., the tendency to agree with others and avoid conflict) were related to more favorable perceptions, while higher levels of negative emotionality (i.e., tendencies of stressful and anxious reactions) were related to less favorable perceptions of teaching quality. We found that particularly one aspect of teaching quality, namely perceived teacher support, was associated with students' lesson comprehension. However, we could hardly identify differences between subjective and classroom-based perceptions, such that future research should further investigate the conditions and consequences of subjectivity in perceptions of teaching quality.

Keywords

Students' perceptions of instructional quality, personality, experience sampling, idiosyncratic perceptions, linear mixed effects models

Received 22 January 2024; Revised 15 June 2024; accepted 17 August 2024

Teachers' instructional quality substantially impacts students' achievement (Kunter et al., 2013). Accordingly, students' perceptions of instructional quality (SPIQ) are implemented worldwide in educational large-scale assessments such as the Programme for International Student Assessment (PISA) to provide information on teaching effectiveness at the country level (OECD, 2014). Thus, major efforts have been directed toward assessing the

[1]University of the Bundeswehr Munich, Neubiberg, Germany
[2]Ludwig-Maximilian-University of Munich, Munich, Germany
[3]Kiel University, Kiel, Germany
[4]University of Luxembourg, Esch-sur-Alzette, Luxembourg

Corresponding author:
Irma Talić, Institute of Psychology, University of the Bundeswehr Munich, Werner-Heisenberg-Weg 39, Neubiberg 85579, Germany.
Email: irma.talic@unibw.de

validity of these perceptions (e.g., Fauth et al., 2014; Kunter & Baumert, 2007; Ruzek et al., 2022; Wagner et al., 2013; Wisniewski et al., 2020; Wisniewski et al., 2022). In the majority of studies on SPIQ (Praetorius et al., 2018), student perceptions are assessed at one time point, aggregated to the class- or school-level, and related to correlates of interest such as achievement (Lüdtke et al., 2009). Such aggregated data neglect (a) individual student characteristics and (b) lesson-specific dynamics, although SPIQ are influenced by both the student raters and time points (Feistauer & Richter, 2017; Wagner et al., 2016). Why different students perceive the same instructional quality differently and how these differing perceptions relate to student outcomes remain poorly understood so far. By employing a multi-rater experience sampling approach, the present study leverages methodological and content-specific advances compared with cross-sectional, aggregated data, with the goal of addressing individual student characteristics and lesson-specific within-person associations of state SPIQ. To do so, we used an experience sampling design that assessed state SPIQ of all students within a classroom attending the same lessons. Assessing perceptions of multiple individuals (i.e., students) in the same situations (i.e., lessons) on the same target (i.e., instructional quality) enables us to transfer insights from situation perception research (e.g., Rauthmann et al., 2015) to instructional quality research, conceptualizing perceptions of instructional quality in the classroom as situation perceptions. The SPIQ of multiple students, assessed within shared lessons, allow for bridging the gap between intrasubjective versus intersubjective perceptions of instructional quality by differentiating idiosyncratic from consensual perceptions. To shed light on relevant associations of consensual and idiosyncratic components of state SPIQ, we examined—for the first time—both the role of students' Big Five personality traits in relation to state SPIQ and the short-term relations between state SPIQ and students' lesson-specific perceived learning achievement (PLA) in the same lessons (i.e., their self-reported comprehension per lesson). To operationalize SPIQ, we used the popular and well-validated framework of the Three Basic Dimensions (TBDs) teacher support, cognitive activation, and classroom management (Klieme et al., 2001). Ultimately, the present study thus provides insight into the occurrence and relevance of individual SPIQ components. In doing so, we highlight the role of the *student* in SPIQ, and with this, combine the two disciplines of personality and educational psychology.

Background

Differentiating idiosyncratic and consensual components within state SPIQ

The TBDs describe SPIQ parsimoniously within the three dimensions of *teacher support* (i.e., providing support during the learning process by, e.g., being sensitive to student needs and avoiding achievement pressure), *cognitive activation* (i.e., encouraging students' thinking and metacognition by, e.g., offering challenging tasks), and *classroom management* (i.e., efficiently using classroom time as learning time by, e.g., imposing clear rules and dealing with disruptions in class effectively; Klieme et al., 2001). The TBDs framework is widely used and empirically well validated, although most of the validation studies used higher-level (e.g., classroom- or school-level) aggregated, habitual perceptions (see Praetorius et al., 2018, for an overview), whose results cannot be transferred to the level of the individual (Molenaar, 2004; Murayama et al., 2017). Trait SPIQ that are only assessed at one point in time (a) lack the examination of within-student (i.e., intraindividual) variability (in contrast to between-student or interindividual variability) and (b) represent self-reported perceptions whose degree of subjectiveness cannot be estimated without further sources of information (e.g., teachers' self-perceptions; Wisniewski et al., 2022). To overcome Point (a), the experience sampling method can be implemented where individuals repeatedly report on their momentary (i.e., state) experiences in multiple situations in daily life (Bolger & Laurenceau, 2013). If these state experiences are assessed for multiple individuals regarding the same situation, one can also address Point (b) by differentiating idiosyncratic perceptions (that reflect intrasubjectiveness) from overlapping, consensual perceptions (that reflect intersubjectiveness; Rauthmann & Sherman, 2019). Drawing upon the same state SPIQ data as in the present study, Talić et al. (2022) conducted a first validation study on the state perceptions of the TBDs within an experience sampling design. They showed a reliable differentiation of the TBDs from lesson to lesson and statistically significant relations to crucial trait outcomes of student motivation and achievement. State SPIQ in mathematics lessons were reported to entail up to 61 % within-student variance across three weeks that remained virtually the same after statistically controlling for shared lesson perceptions of all students in a specific lesson. These shared perceptions were not investigated in more detail.[1] To overcome this drawback, the present investigation builds upon situation perception research (Rauthmann et al., 2015). Using situation perception terminologies, students' perceptions of instructional quality that are commonly used (i.e., "raw" scores) are termed *experience*, and can be differentiated into the unique and idiosyncratic individual perception (i.e., *construal*) and the shared perception of all students within the classroom (i.e., *consensus*; Rauthmann et al., 2015).[2] The distinction of the three SPIQ components experience, construal, and consensus[3] enables a comparison of intrasubjective SPIQ versus intersubjective perceptions of instructional quality.

The role of students' Big Five personality traits in SPIQ

Being exposed to the same situational stimuli does not mean that all individuals form the same psychological representation of that situation (Rauthmann & Sherman, 2019). For instance, opportunity-and-use models assume that instruction is a co-constructive process of actions related to the subject matter where both teacher traits (e.g., subject knowledge, competence) and student traits (e.g., motivation) impact learning opportunities and their use, respectively (Vieluf et al., 2020). In other words, learning opportunities offered by the teacher do not implicitly guarantee their use by the students but rather depend on the students' perceptions and interpretations. It is highly plausible that student traits impact the *perception* of instructional behaviors (e.g., Feistauer & Richter, 2017; Wisniewski et al., 2022). Rauthmann and Sherman (2019) have discussed personality traits as possible correlates of differences in information processing and subsequent ratings. The Big Five personality traits open-mindedness, conscientiousness, extraversion, agreeableness, and negative emotionality capture major trait domains (John, 2021; McCrae & Costa, 1987). These personality traits have been shown to be related to individual differences in situational perceptions. For instance, higher negative emotionality was related to more perceptions of threat (Jonason & Sherman, 2020). In the educational context, the Big Five traits were mostly examined regarding academic achievement as student outcome where open-mindedness, conscientiousness, and agreeableness were statistically significantly related to academic achievement, with conscientiousness as the strongest predictor even after controlling for intelligence (e.g., Franzen et al., 2022; see also meta-analyses from Mammadov, 2022; Poropat, 2009). Regarding perceptions of instructional quality, *teachers'* traits—such as their self-efficacy—have been considered in some works to be relevant for SPIQ (Holzberger et al., 2013; Toropova et al., 2019). For example, Roloff et al. (2020) found that teachers' agreeableness positively predicted students' perceptions of teacher support. Although substantial variance in SPIQ has been shown to be attributable to the student raters (Feistauer & Richter, 2017; Ruzek et al., 2022; Wagner et al., 2016), there is a lack of studies on the role of *students'* personality traits in SPIQ. In a study on perceptions of online learning experiences, students' conscientiousness has been shown to predict positive evaluations positively and negative evaluations negatively, while agreeableness and open-mindedness additionally predicted the value of online courses positively (Keller & Karau, 2013). Yet, relations between students' Big Five personality traits and perceptions of instructional quality within the TBDs are poorly understood so far.

Lesson-specific relations between SPIQ and PLA

The relations between SPIQ and academic achievement as one of the most important educational outcomes are of utmost interest. Of the TBDs, particularly the dimensions of cognitive activation and classroom management have theoretical relations to achievement, whereas teacher support is assumed to be more closely related to student motivation (Praetorius et al., 2018). Cognitive activation enhances student achievement by stimulating higher-order thinking, ultimately resulting in the construction of deep and flexible knowledge (Hardy et al., 2006). Classroom management is related to student achievement by enhancing the time effectively spent on tasks (Seidel & Shavelson, 2007). Empirically, relations to achievement were demonstrated for cognitive activation (Klieme et al., 2001), classroom management (Fauth et al., 2014; Scherer et al., 2016), and teacher support (Fauth et al., 2014; for an overview see Praetorius et al., 2018). Again, these relations are based on between-person research designs, where interindividual variance (between students, classes, schools, or countries) is used and individual student deviations from mean perceptions are neglected. Thus, evidence for SPIQ–achievement relations is extended in the present study to the within-student level, assessing short-term, lesson-specific relations for the first time.

At the same time, the postulated SPIQ–achievement relations could not always be confirmed, and could be moderated by various aspects such as different perspectives (students, teachers, or observers), latent versus manifest modeling, or different operationalizations (Praetorius et al., 2018). To shed light on the role of students' personality traits not only on SPIQ themselves, but also on their possible role on SPIQ–achievement relations, we examine students' Big Five personality traits as moderators of the lesson-specific SPIQ–PLA relations exploratively for the first time. Such interactive situation perception × trait effects have been found for state behavior (Breil et al., 2019), while other authors have not identified such effects (Abrahams et al., 2021). Thus, we extend research on SPIQ–achievement relations by examining within-student associations. Considering students' personality traits in these associations can possibly reveal new insights to the relevance of personality traits for student achievement, as well as ultimately aid in fostering student achievement in individual lessons.

The present study

This experience sampling study addresses three distinct yet interrelated research questions that—to the best of our knowledge—have not been examined before. We focus on state measures of SPIQ and lesson-specific PLA and on trait measures of students' Big Five personality traits in German secondary school students in the domain of math instruction, where the majority of research on the TBDs has been conducted (Praetorius et al., 2018). Enabled by our multi-rater experience sampling design, we apply insights from situation perception research (Rauthmann & Sherman, 2019) to perceptions of instructional quality research with the goal of disentangling different components that are confounded within raw SPIQ scores. Specifically, in all examined relations, we differentiated between the SPIQ components of experience (i.e., raw SPIQ scores provided by the students), construal (i.e., the purely idiosyncratic portion within the SPIQ that one respective student does not share with their classmates), and consensus (i.e., the intersubjective, overlapping classroom perception; Rauthmann et al., 2015) for each of the TBDs of teacher support, cognitive activation, and classroom management. Although we focused on state SPIQ, we also provided intercorrelations between trait SPIQ and all study variables for the interested reader.

In terms of constructs and variables focused on, we used the TBDs framework (Klieme et al., 2001) for state SPIQ, lesson-specific PLA (i.e., self-reported lesson-specific comprehension; Niepel et al., 2022) as a subjective achievement indicator—both across three weeks of German secondary school students' daily life—as well as the Big Five framework (Costa & McCrae, 1992) for students' personality traits. Prior research has shown students' gender to be related to selfperceived math abilities (Niepel et al., 2019). Gender differences in personality traits have also been reported (Weisberg et al., 2011). Reasoning ability was related to personality traits (Sutin et al., 2022), while school grades were associated with SPIQ (Jaekel et al., 2021). To control for possible confounding effects, we therefore included students' gender, math grade, and reasoning ability as covariates, and show model results for models with and without covariates.

We derived three research questions (RQs). Due to the novelty of analyses (RQ1 and RQ3) or unknown applicability to within-person designs (RQ2), respectively, these three RQs are addressed in an explorative manner.

RQ1: How are students' personality traits associated with their state perceptions of instructional quality?

In perceptions of online learning experiences, students' open-mindedness, conscientiousness, and agreeableness were related to positive evaluations (Keller & Karau, 2013). However, it is unclear as to how these findings can be transferred to the TBDs in a classroom setting.

RQ2: How are students' state perceptions of instructional quality associated with their perceived learning achievement in the same lesson?

Relations between all three dimensions of instructional quality and student achievement have been found (Praetorius et al., 2018). These between-person-based results cannot be necessarily generalized to RQ2, as RQ2 addressed short-term, lesson-specific relations in a within-person design.

RQ3: Do students' personality traits moderate the relationship between state perceptions of instructional quality and perceived learning achievement in the same lesson?

Addressing possible situation perception × trait effects (e.g., Abrahams et al., 2021; Breil et al., 2019), we examined the role of students' Big Five traits in lesson-specific SPIQ–PLA relations exploratively. For instance, it could be that higher degrees of open-mindedness are related to a stronger relation between SPIQ and lesson-specific PLA.

Differentiating the SPIQ components (experience, construal, and consensus) in all RQs allowed for a more fine-grained consideration of their associations. For instance, it could be that only consensual SPIQ are related to higher lesson-specific PLA (i.e., the higher the individual student agrees with the classmates on their SPIQ), but not experienced SPIQ (i.e., the rating the student provides in a specific lesson whose degree of subjectiveness cannot be estimated).

Method

Procedure and participants

We used data from the larger intensive longitudinal "Dynamics of Academic Self-Concept in Everyday Life" (DynASCEL) project (Niepel et al., 2022), where a three-week experience sampling study was conducted. Prior to and following the experience sampling phase, respectively, a pre- and post-assessment was carried out in paper-and-pencil format that obtained exhaustive student trait variables (e.g., Big Five personality traits, SPIQ traits). Data from the project have been used in other manuscripts on different research questions (e.g., Talić et al., 2022, used the data on math state SPIQ to examine their factorial within- and between-student structure; Niepel et al., 2022, used the data on lesson-specific PLA to investigate reciprocal relations to academic self-concept; and Hausen et al., 2022, used the data on the Big Five personality traits in relation to mean level and within-person variability of general academic self-concept. A full list of all other manuscripts drawing on DynASCEL data can be found here: https://osf.io/e8jqm/). To address the present RQs, we focused on the experience sampling data on SPIQ and lesson-specific PLA in every math lesson and used trait data from the pre-assessment. Within the overarching DynASCEL project, we assessed a convenience sample of German secondary, academic-track schools, aiming at attaining a sample size of different classrooms that allowed for statistically controlling for variance a the highest level of analysis (Hox et al., 2018). To be able to differentiate the three SPIQ components, we further aimed at assessing all students within classes, which was predominantly the case. To capture day-to-day variability whilst considering the participant burden, we chose a three-week experience sampling period (see also Tsai et al., 2008 for a similar procedure), during which we aimed at assessing all math lessons. Specific power analyses targeted at the present exploratory analyses based on expected effect sizes were not conducted.

We sampled $N = 372$ German secondary school students attending the highest ability track (i.e., the German *Gymnasium*)[4] who participated in the experience sampling part of the study. Of these, $n = 308$ students attended the 9th and $n = 64$ students attended the 10th grade. Our sample consisted of 34.1 % boys (from $n = 301$ students with available gender information). Students reported a mean age of 15.3 years ($SD = 0.68$; range = 13.3–17.4 years; based on $n = 298$ students with available age information) and were nested in 18 classes in six schools from four German states (Rhineland-Palatinate, North Rhine-Westphalia, Baden-Wuerttemberg, and Mecklenburg-Western Pomerania). Data collection took place between January and July 2018. Based on these 18 classrooms, we assessed students' perceptions of 17 different math teachers' (58.8 % male) instructional quality (i.e., one teacher instructing math to two separate classes). On average, there were 20.6 students in a classroom ($SD = 4.65$; range = 10–27). Students generally remained in the same classes with their peers across school grades.

In the experience sampling phase, students completed e-diaries on smartphones where they responded to a short electronic questionnaire assessing their perceptions of a specific lesson on the application movisensXS (versions 1.3.0–1.3.4; movisens GmbH, Karlsruhe, Germany). The smartphones were given to the students for the duration of the study by the research team and were all identical models. We preprogrammed the smartphones such that the

Table 1. Descriptive Statistics.

	M	*SD*	ω	ICC	ω_{within}	$\omega_{between}$
State SPIQ experience						
Teacher support	3.11	1.37	-	.49	.84	.97
Cognitive activation	3.06	1.29	-	.44	.83	.97
Classroom management	3.41	1.37	-	.61	.76	.94
State SPIQ construal						
Teacher support	0.00	1.29	-	.47	-	-
Cognitive activation	0.00	1.26	-	.43	-	-
Classroom management	0.00	1.20	-	.53	-	-
State SPIQ consensus						
Teacher support	0.00	0.94	-	.49	-	-
Cognitive activation	0.00	0.82	-	.44	-	-
Classroom management	0.00	0.97	-	.62	-	-
Lesson-specific achievement						
Lesson-specific PLA	3.46	1.17	-	.40	.89	.95
SPIQ traits						
Teacher support	2.74	1.32	.93			
Cognitive activation	2.94	0.72	.81			
Classroom management	3.25	1.15	.90			
Personality traits						
Open-mindedness	2.24	0.63	.85	-	-	-
Conscientiousness	2.36	0.60	.89	-	-	-
Extraversion	2.37	0.61	.89	-	-	-
Agreeableness	2.65	0.57	.82	-	-	-
Negative emotionality	1.70	0.62	.89	-	-	-
Covariates						
Math grade	4.36	1.09	-	-	-	-
Reasoning ability	43.52	5.70	.77	-	-	-

Note. Response formats: SPIQ experience [0, 5]; lesson-specific PLA [0; 5]; SPIQ traits [0, 5]; personality traits [0, 4]; math grade [1, 6]; reasoning ability [0, 60].

experience sampling prompts were triggered three minutes prior to the regular ending of every math lesson during the three-week period according to class-specific timetables. The number of math lessons thus varied between classes per design (i.e., $M = 10.11$ math lessons; $SD = 3.39$; range = 3–16). In total, we obtained 2,681 valid responses (i.e., at least one out of nine items of interest answered; see *Measures* section below), representing a compliance rate of 70.81 %. Causes of missingness included absences from lessons (e.g., student illness), cancellation of classes, exams or similar events, and technical issues (e.g., empty smartphone batteries).

Students' participation in the study was voluntary. Single items and prompts were skippable. The students were ensured that their responses were confidential and their responses were not disclosed to their teachers in any way. Written parental consent was obtained for participating students and the local ethics review panel of the University of Luxembourg and all involved German federal education authorities approved of all procedures. This study was not preregistered.

Measures

State measures. Students responded to all state measures in the three-week experience sampling phase. These measures were all self-reported. Descriptive statistics and ω reliability coefficients can be found in Table 1.

State SPIQ. In the three-week experience sampling phase, state SPIQ in math instruction were assessed within the TBDs of teacher support, cognitive activation, and classroom management using the two-item state scales described by Talić et al. (2022). These were based on the PISA 2012 scales (Mang et al., 2018) and adapted for the use in intensive longitudinal designs. Example items are "*During this lesson, the teacher helped students with their learning"* (teacher support), "*During this lesson, the teacher gave problems that required us to think for an extended time*" (cognitive activation), and "*During this lesson, there was noise and disorder*" (classroom management; reverse scored). Items used a six-point Likert-type response scale ranging from 0 (*false*) to 5 (*true*), with higher ratings representing higher perceived instructional quality. Reliability, validity evidence concerning the factor structure and relations to trait SPIQ scales, school grades, and interest were provided, altogether suggesting the applicability of the state SPIQ scales in an experience sampling design (Talić et al., 2022).

Lesson-specific PLA. Lesson-specific PLA was assessed in each math lesson during the three-week experience sampling phase (i.e., in the same situations as state SPIQ). Three items that were shown to be applicable and meaningful in an experience sampling design (Niepel et al., 2022) were used to assess lesson-specific perceived comprehension and learning progress. Niepel et al. (2022) derived the items from previous research that implemented similar items in e-diaries to assess lesson-specific PLA (e.g., Peterson & Miller, 2004; Shernof et al., 2017). The item wordings were "*I was able to follow the last lesson well*," "*I understood a lot in the last lesson*," and "*I learned a lot in the last lesson.*" Items used a six-point Likert-type response scale ranging from 0 (*false*) to 5 (*true*), with higher scores indicating higher lesson-specific PLA.

Trait measures. Students responded to all trait measures in the pre-assessment (i.e., prior to the three-week experience sampling phase). These measures were all self-reported. Descriptive statistics and ω reliability coefficients can be found in Table 1.

Personality traits. We assessed students' Big Five personality traits open-mindedness, conscientiousness, extraversion, agreeableness, and negative emotionality using the German version (Danner et al., 2019) of the Big Five Inventory 2 (Soto & John, 2017).[5] Items used a five-point Likert-type scale ranging from 0 (*disagree completely*) to 4 (*agree completely*).

Trait SPIQ. Trait SPIQ in math instruction were assessed within the TBDs of teacher support, cognitive activation, and classroom management. We used the original full scales implemented in PISA 2012 (Mang et al., 2018), consisting of five (for teacher support and classroom management, respectively) and nine (for cognitive activation) items that assess general (habitual) perceived instructional quality that is not tied to specific lessons (i.e., aggregated perceptions). Example items are "*The teacher helps students with their learning"* (teacher support), "*The teacher gives problems that require us to think for an extended time*" (cognitive activation), and "*There is noise and disorder*" (classroom management; reverse scored). Items used a six-point Likert-type scale ranging from 0 (*never [in no lesson]*) to 5 (*always [in every lesson]*).

Report card math grade. We obtained students' self-reported math grade from their most recent report card. Prior research has shown that self-reported school grades serve as reliable achievement indicators in German student samples that do not contain systematic reporting biases (Dickhäuser & Plenter, 2005; Sparfeldt et al., 2008). School grades in Germany are assigned on a six-point numerical scale which we recoded into 1 (*insufficient*) to 6 (*very good*) such that higher scores represented better achievement.

Reasoning ability. Students' reasoning ability was assessed using the Intelligenz-Struktur-Test-Screening (IST-Screening; Liepmann et al., 2012), the short version of the well-established Intelligenz-Struktur-Test (IST; Amthauer, 1970; Liepmann et al., 2007). We used Version A of the available Versions A and B of the test. The IST-Screening measures students' reasoning ability in the three task areas of verbal analogies, number sequences, and figural matrices using 20 items each. We used the composite raw score across the three task areas as an indicator of general reasoning ability.

Statistical analyses

We followed recommendations by a recent experience sampling study that examined the three components experience, construal, and consensus in situation perceptions in an educational context (Abrahams et al., 2021). We conducted all analyses using the statistical software R (R Core Team, 2021) with the exception of two-level ω reliability coefficients which were calculated using the software Mplus 8.3 (Muthén & Muthén, 1998-2017). For fitting linear mixed effects models, we used the lme4 package with the optimizer bobyqa to improve convergence (Bates et al., 2015). We used the effectsize package (Ben-Shachar et al., 2020) to obtain standardized parameters and the confint() function to obtain bootstrapped 95 % confidence intervals.

To address our RQs, we first disentangled the three SPIQ components experience, construal, and consensus (see also Abrahams et al., 2021; Rauthmann et al., 2015). First, SPIQ *experience* is reflected by the raw individual SPIQ scores (as commonly used in previous SPIQ research). Second, we calculated lesson-specific SPIQ class means such that for each row in the dataset (i.e., for a specific student in a specific lesson), the respective lesson-specific student SPIQ mean was excluded. Thus, the individual SPIQ did not enter the class mean perception of instructional quality in the same lesson. By doing this, we ensured that the class mean entailed only variance from all other students to avoid an artificial overemphasis of student variance when relating lesson-specific student and class means. SPIQ *construal* was then obtained by extracting the standardized residual scores from regression analyses, where lesson-specific individual SPIQ (i.e., experience) were regressed on lesson-specific class-mean SPIQ. Residual variance of the individual SPIQ experience that was not explained by class SPIQ was considered idiosyncratic (i.e., SPIQ construal). Third, SPIQ *consensus* reflected consensual perceptions of *all* students within a class in a specific lesson. It was obtained by extracting factor scores from exploratory factor analyses with the default "oblimin" rotation on individual SPIQ experience and class SPIQ where one factor was specified to be extracted for each SPIQ dimension. Variance that was shared across individual and class SPIQ was considered as overlapping (i.e., SPIQ consensus). The above procedures were conducted for all three SPIQ dimensions (i.e., the three TBDs teacher support, cognitive activation, and classroom management).

The experience sampling phase produced data where measurement points (Level 1) were nested within students (Level 2) that were, in turn, nested within classes (Level 3). Clustering in 18 classes at Level 3 was controlled for by adding 17 dummy-coded class-based predictor variables in each model (Hox et al., 2018). Although it would be more desirable to use three-level models, we used this approach to control for between-classroom variance due to a limited sample size at Level 3. All dummy-coded classroom variables' fixed effects are shown in the OSM. To estimate the reliability of our implemented measures, we computed single-level (for traits) and two-level (i.e., within- and between-student for states) McDonald's ω coefficients (Geldhof et al., 2014). To estimate dependency in the data due to repeated measurements within persons, we calculated intraclass correlation coefficients (ICCs) for the state measures (Aarts et al., 2014). Due to model convergence issues when implementing random slopes, we conducted random intercept models. Predictors at Level 1 were centered within students, while predictors at Level 2 were centered at the grand mean (Enders & Tofighi, 2007). To control for gender, math grade, and reasoning ability, we conducted two sets of models for each RQ that exclude or include these covariates, respectively. Results are presented for both model sets. We report unstandardized fixed effects coefficients (*b*s) and their bootstrapped 95 % confidence intervals. To estimate the fixed effects' fit to the model, we calculated marginal multiple *R*s (R_m; Nakagawa & Schielzeth, 2013) and also derived standardized regression coefficients as a multilevel model effect size measure (Lorah, 2018). To adjust for multiple testing, we followed the procedure implemented by Abrahams et al. (2021) and used the more conservative level of $p < .001$ to test for statistical significance. Additionally, we also highlight findings at the level of $p < .05$ in the tables for interested readers, but we refrain from interpreting them due to the explorative nature of the analyses. Exact model specifications are described at the beginning of the respective *Results* section for enhanced clarity.

Results

Preliminary analyses

Prior to addressing our research questions, we examined descriptive statistics of all measures (see Table 1). For state measures, within-student [between-student] ω coefficients ranged from $\omega_{within} = .76$ to $\omega_{within} = .84$ [$\omega_{between} = .94$ to $\omega_{between} = .97$] for SPIQ experiences across the TBDs. Lesson-specific PLA showed a reliability of $\omega_{within} = .89$ and $\omega_{between} = .95$. For trait measures, coefficients ranged from ω = .81 to ω = .93 for trait SPIQ, and from ω = .85 to ω = .89 for personality traits.[6] ICC values ranged from .40 to .62 across the SPIQ components and lesson-specific PLA, indicating a substantial amount of within-student variance in these measures.[7]

Having disentangled the three SPIQ components experience, construal, and consensus for the TBDs, we preliminarily examined their intercorrelations at the within- and between-student level (see Table 2 and the corresponding exact *p*-values in Table S4 in the OSM). Correlations at the between-student level were higher than at the within-student level. Here, we focus on the within-student level, where in general the three components showed close to perfect correlations to one another across dimensions (e.g., teacher support experience and teacher support construal), ranging between $r = .88$ to $r = 1$. Thus, the three components experience, construal, and consensus showed an extensive overlap in all three state SPIQ dimensions. Teacher support components were moderately related to cognitive activation components (ranges of $r = .43$ to .45, all $ps < .001$), while classroom management was uncorrelated with either of the two.

We further calculated correlations between personality traits and SPIQ *traits* (i.e., habitual SPIQ not tied to specific lessons; see Table S1 and the corresponding exact *p*-values in Table S2 in the OSM). Teacher support was mostly unrelated to the Big Five personality traits, with the exception of negative emotionality ($r = -.13$, $p = .021$). Cognitive activation showed positive relations to open-mindedness ($r = .16$, $p = .007$), conscientiousness ($r = .21$, $p < .001$), extraversion ($r = .12$, $p = .031$), and agreeableness ($r = .19$, $p = .001$). Classroom management showed positive relations to conscientiousness ($r = .13$, $p = .028$) and extraversion ($r = .15$, $p = .011$).[8] Thus, trait SPIQ showed some (and mostly positive) relations to personality traits, with the dimension of cognitive activation showing most ties as compared to the other two SPIQ dimensions.

Students' personality traits as predictors of state SPIQ (RQ 1)

Before we addressed RQ1, we examined correlations between personality traits and state SPIQ experience, construal, and consensus for the TBDs (see Table 3 and the corresponding exact *p*-values in Table S6 in the OSM). First, we noted that relations of SPIQ components to personality traits were similar across all three components within dimensions (e.g., teacher support experience and construal were similarly related to open-mindedness).

Second, we detected some differences in correlations between personality traits and SPIQ traits (see *Preliminary Analyses*). The trait correlations revealed substantial relations of teacher support to negative emotionality only, of cognitive activation to each personality trait except for negative emotionality, and of classroom management to conscientiousness and extraversion only. In comparison, relations of state SPIQ components revealed that all teacher support components were only related to agreeableness (mean $r = .19$, $ps < .001$) and negative emotionality (mean $r = -.23$, $ps < .001$). Cognitive activation components were related to each personality trait except for extraversion (mean $r = .10$, *p*s ranged between $p = .035$ and $p = .045$ for open-mindedness; mean $r = .11$ and *p*s between $p = .034$ and $p = .040$ for conscientiousness; mean $r = .19$ and *p*s between $p = .002$ and $p = .010$ for agreeableness; mean $r = -.17$ with $p = .002$ for experience and consensus in negative emotionality, and $p < .001$ for construal in negative emotionality). All classroom management components were related to agreeableness (mean $r = .19$, $p < .001$ for experience and consensus, and $r = .15$, $p = .004$ for construal) and negative emotionality (mean $r = -.12$, and *p*s between $p = .003$ and $p = .028$), while the classroom management components of experience and consensus were additionally related to open-mindedness (mean $r = .11$, and $p = .046$ and $p = .029$, respectively) and to conscientiousness (mean $r = .16$, $ps = .002$). Thus, with regard to state SPIQ components, agreeableness and negative emotionality seemed to show the most pronounced relations to SPIQ components, whereas extraversion was unrelated to all state SPIQ components.

To address RQ1, we conducted linear mixed effect models with 17 dummy-coded variables controlling for classroom membership (shown in Table S7 in the OSM), personality traits (and covariates) as simultaneous predictors, and SPIQ components experience, construal, and consensus for the three dimensions teacher support, cognitive activation, and classroom management as outcome variables. Results are presented in Table 4 (see exact *p*-values in Table S8 in the OSM). For all

Table 2. Correlations between SPIQ components and lesson-specific PLA.

	SPIQ experience			SPIQ construal			SPIQ consensus			
	Teacher support	Cognitive activation	Classroom management	Teacher support	Cognitive activation	Classroom management	Teacher support	Cognitive activation	Classroom management	Lesson-specific PLA
SPIQ experience										
Teacher support	—	**.72**	**.16**	**.94**	**.67**	**.14**	**1**	**.72**	**.16**	**.62**
Cognitive activation	**.44**	—	.09	**.68**	**.97**	.05	**.72**	**1**	.09	**.47**
Classroom management	.01	−.02	—	**.13**	.06	**.86**	**.16**	.09	**1**	**.23**
SPIQ construal										
Teacher support	**.95**	**.43**	.02	—	**.70**	**.15**	**.94**	**.68**	**.13**	**.59**
Cognitive activation	**.43**	**.97**	−.02	**.45**	—	.05	**.67**	**.98**	.06	**.44**
Classroom management	.03	−.01	**.92**	.03	−.01	—	**.14**	.05	**.81**	**.19**
SPIQ consensus										
Teacher support	**1**	**.44**	.01	**.95**	**.43**	.03	—	**.72**	**.16**	**.62**
Cognitive activation	**.44**	**1**	−.02	**.43**	**.97**	−.01	**.44**	—	.09	**.47**
Classroom management	.00	−.03	**1**	.01	−.02	**.88**	.00	−.03	—	**.23**
Lesson-specific PLA	**.49**	**.27**	**.07**	**.45**	**.27**	**.07**	**.49**	**.27**	**.07**	—

Note. Correlations below the diagonal represent within-student correlations, and correlations above the diagonal represent between-student correlations. Correlation coefficients printed in **bold** are statistically significant at $p < .05$, and correlation coefficients printed in **bold and gray shading** are statistically significant at $p < .001$. Exact p-values can be found in Table S4.

teacher support components, negative emotionality emerged as the only statistically significant predictor at $p < .001$ in the model without covariates (mean $b = -0.36$, mean $\beta = -.19$). In other words, for every unit increase in negative emotionality, an average of 0.36 decrease in experienced, construed, and consensual teacher support would be expected. However, after including the covariates gender, math grade, and reasoning ability, this relation no longer reached statistical significance. Instead, in this model, agreeableness was now statistically significantly, positively related to all three teacher support components at $p < .001$ (mean $b = 0.34$, mean $\beta = .17$; $p < .001$). The models predicting experience, construal, and consensus of cognitive activation and classroom management displayed no statistically significant relation to personality traits at $p < .001$. Across models with teacher support as the outcome variable, the average model fit was $R_m = .38$, whereas for those with cognitive activation and classroom management it was $R_m = .35$ although the latter revealed no statistically significant fixed effects. Notably, though, the 17 dummy-coded classroom predictor variables partly produced statistically significant fixed effects that inflated R_m estimates.

State SPIQ as predictors of lesson-specific PLA (RQ 2)

Before we addressed RQ2, we calculated within- and between-student correlations between the three SPIQ components and lesson-specific PLA (see Table 2 and the corresponding exact p-values in Table S4 in the OSM). Again, correlations at the between-student level were higher than at the within-student level. We focus on the within-student level, where relations to lesson-specific PLA were descriptively strongest for teacher support (mean $r = .48$) and cognitive activation (mean $r = .27$) and lowest for classroom management (mean $r = .07$, all $ps < .001$), with an almost identical pattern across the components of experience, construal, and consensus.

To address RQ2, we conducted linear mixed effect models with 17 dummy-coded variables controlling for classroom membership (whose relations are shown in Table S9 in the OSM), three SPIQ dimensions per component (and covariates) as simultaneous predictors, and lesson-specific PLA as the outcome variable. Results can be found in Table 5. All components of all SPIQ dimensions were statistically significantly and positively related to lesson-specific PLA at $p < .001$ in the models without covariates. There were clear descriptive differences in effect sizes. Teacher support showed the largest association (mean $b = 0.48$, mean $\beta = .32$). The associations of cognitive activation and classroom management were of similar extent (mean $b = 0.08$, mean $\beta = .05$). In the models with covariates, these results remained virtually unchanged. Further, female gender was negatively related to lesson-specific PLA ($b = -0.36$, $\beta = -.16$). The math grade was positively related to lesson-specific PLA ($b = 0.27$, $\beta = .24$). Reasoning ability did not show any incremental relation to lesson-specific PLA above and beyond gender and the math grade. Including the covariates improved the model fit (mean $R_m = .44$ without covariates, and mean $R_m = .55$ with covariates). In summary, experienced, construed, and consensual teacher support were most related to PLA in the same lesson.

Students' personality traits as moderators of the association between state SPIQ and lesson-specific PLA (RQ 3)

Finally, addressing RQ3, we examined personality traits as possible moderators of the link between SPIQ components and lesson-specific PLA. We ran a set of preliminary models where we included all possible interaction terms between

SPIQ components and personality traits, of which we only used those interaction terms that were statistically significant at $p < .05$ for our final models (for a similar procedure, see Abrahams et al., 2021; Sherman et al., 2015). The elevated alpha level was chosen here for the preliminary models to facilitate the detection of interaction effects that are usually very small (Rauthmann, 2021) and thus might aid in generating new hypotheses. For interpreting moderation effects in the final models, however, we used the criterion of $p < .001$.

In the final models, we included 17 dummy-coded variables controlling for classroom membership (whose relations are shown in Table S10 in the OSM), three SPIQ dimensions per component (e.g., experience of teacher support, cognitive activation, and classroom management), the respective personality traits and interaction terms between state SPIQ components and personality traits that were statistically significant predictors in the preliminary models (and covariates) as simultaneous predictors, and lesson-specific PLA as the outcome variable. Results are displayed in Table 6. Relations between state SPIQ components and lesson-specific PLA and between covariates and lesson-specific PLA remained virtually the same as those reported for RQ2 and only showed marginal differences in effect sizes. We only found interactions of components of the state SPIQ dimension teacher support with the personality traits agreeableness and negative emotionality to be related to lesson-specific PLA. Regarding the preliminary models, of the 45 (3 dimensions per state SPIQ component * 5 personality traits * 3 state SPIQ components) possible interactions, only five interactions reached statistical significance at the $p < .05$ level and were included in the final models. In our test of the final models, only one of those interactions was statistically significant at the $p < .001$ level (see Table 6). Specifically, agreeableness moderated the relation between construed teacher support and lesson-specific PLA in the models with and without covariates (bs = −0.13, $\beta = -.06$ and $\beta = -.05$, respectively). In other words, the less agreeable students reported themselves to be, the stronger was the positive association between construed teacher support and lesson-specific PLA (see Figure 1). Concerning main effects, agreeableness showed a positive relation to lesson-specific PLA in the model using SPIQ construal and covariates ($b = 0.31$, $\beta = .15$; $p < .001$), while negative emotionality showed negative relations to lesson-specific PLA in the models using SPIQ experience and consensus (mean $b = -0.36$, mean $\beta = .19$; $p < .001$). The average model fit was $R_m = .52$ for the models without covariates and $R_m = .59$ for the models with covariates.

Discussion

The present study addressed "the perception problem" (Wisniewski et al., 2022) within instructional quality research—differences between perceptions across rating sources—from a different angle. Using an experience sampling design with repeatedly assessed multiple students' perceptions of the same lesson-specific instructional quality, we (a) examined state SPIQ and within-student relations to PLA and (b) disentangled idiosyncratic perceptions from consensual perceptions that are usually confounded within raw SPIQ scores while (c) shedding light on the role of students' personality traits in an educational context. Such analyses are not possible in traditional research designs that assess SPIQ at one point in time with an unclear target time frame and aggregate them to higher levels of analyses, thereby considering within-student variation merely as disturbance. We detected small to medium associations between students' personality traits of agreeableness and negative emotionality and state SPIQ. Within-student analyses revealed that the dimension of teacher support showed particularly strong positive relations to lesson-specific PLA. Additionally, this relation was more pronounced in less agreeable students. Clear differential relations across the three components of SPIQ experience, construal, and consensus could not be detected. Shifting the focus of instructional quality research to individual lessons, within-student relations, and student factors associated with both SPIQ and within-student relations of SPIQ essentially shifts the focus of instructional quality research to the student perceiver instead of merely the teachers' behavior (Bellens et al., 2019; Scherer et al., 2016). This ultimately casts a more differentiated picture on instructional quality, classroom interactions, and dynamics in specific lessons.

Experience, construal, and consensus could hardly be differentiated in state SPIQ

This study was the first one to differentiate the components of experience, construal, and consensus within SPIQ and the TBDs' perceptions. An initial examination of intercorrelations revealed large to perfect associations between the different components within the three dimensions of the TBDs framework. In other words, a higher experienced instructional quality (i.e., students' raw perceptions of instructional quality) was substantially related to higher construed (i.e., students' idiosyncratic perceptions) and higher consensual (i.e., students' agreement with classmates' perceptions) instructional quality. Consistent with this, in all examined relations to personality traits and lesson-specific PLA, the three components showed virtually the same results that only slightly differed in effect sizes in almost all examined relations. Given prior research that discussed the role of the student in SPIQ and the subjectiveness of SPIQ (e.g., Feistauer & Richter, 2017; Talić et al., 2022; Wagner et al., 2016; Wisniewski et al., 2022), this finding was rather surprising. Some overlap is inherent due to the fact that the components are confounded within one another (i.e., shared variance between individual experience and class mean experience yields consensus, and individual experience variance not explained by class mean experience yields construal). This overlap leads to partially artificially higher correlations and more similar relations to outcome criteria. One way to address this in future research could be applying designs that allow for a variance decomposition approach with the goal of identifying how much variance in SPIQ experience is attributable to SPIQ construal, and how much is attributable to SPIQ consensus. This would enable more insights into the structure and idiosyncracies of SPIQ.

Further, it is important to note that prior research investigated the TBDs as state SPIQ and identified substantial and meaningful within-student variation, where students reliably differentiated between the TBDs from lesson to lesson, even after controlling for shared lesson perceptions (Talić et al., 2022). On this sample-based approach, approximately 53 % of the variance in state SPIQ were

Table 3. Correlations between SPIQ components, lesson-specific PLA, covariates, and personality traits.

	SPIQ experience			SPIQ construal			SPIQ consensus			
	Teacher support	Cognitive activation	Classroom management	Teacher support	Cognitive activation	Classroom management	Teacher support	Cognitive activation	Classroom management	Lesson-specific PLA
Gender	**−.15**	−.09	.03	**−.19**	**−.12**	.00	**−.15**	−.09	.03	**−.23**
Math grade	**.24**	**.22**	**.11**	**.20**	**.18**	.04	**.23**	**.22**	**.11**	**.40**
Reasoning ability	.08	.06	.07	**.10**	.05	.04	.08	.06	.07	**.27**
Open-mindedness	.02	**.10**	**.10**	.02	**.11**	.04	.02	**.10**	**.11**	**.15**
Conscientiousness	.09	**.11**	**.16**	.07	**.11**	.10	.09	**.11**	**.16**	**.15**
Extraversion	.09	.05	.10	.07	.04	.09	.09	.05	.10	**.11**
Agreeableness	**.18**	**.13**	**.18**	**.21**	**.16**	**.15**	**.19**	**.13**	**.19**	**.18**
Negative emotionality	**−.22**	**−.16**	**−.11**	**−.25**	**−.19**	**−.15**	**−.22**	**−.16**	**−.10**	**−.31**

Note. Gender is coded with 0 = male; 1 = female.
Correlation coefficients printed in **bold** are statistically significant at $p < .05$, and correlation coefficients printed in **bold and gray shading** are statistically significant at $p < .001$. Exact p-values can be found in Table S6.

Table 4. Personality traits as predictors of SPIQ components (RQ1).

Predictors	SPIQ experience				SPIQ construal				SPIQ consensus			
	b [95% CI]	β [95% CI]	*t*	R_m	b [95% CI]	β [95% CI]	*t*	R_m	b [95% CI]	β [95% CI]	*t*	R_m
Outcome: Teacher support												
Model without covariates												
Open-mindedness	−0.09 [−0.27, 0.10]	−.04 [−.12, .04]	−0.93	.40	−0.09 [−0.28, 0.09]	−.05 [−.14, .05]	−0.96	.27	−0.06 [−0.18, 0.06]	−.04 [−.12, .04]	−0.93	.40
Conscientiousness	−0.11 [−0.29, 0.08]	−.05 [−.13, .04]	−1.08		−0.11 [−0.31, 0.07]	−.05 [−.15, .04]	−1.08		−0.07 [−0.19, 0.06]	−.05 [−.13, .04]	−1.08	
Extraversion	−0.02 [−0.19, 0.17]	−.01 [−.09, .07]	−0.21		−0.03 [−0.22, 0.16]	−.01 [−.10, .08]	−0.27		−0.01 [−0.13, 0.12]	−.01 [−.09, .07]	−0.22	
Agreeableness	**0.28** [0.05, 0.47]	**.12** [.03, .21]	2.71		**0.30** [0.08, 0.52]	**.14** [.04, .23]	2.79		**0.19** [0.04, 0.33]	**.12** [.03, .21]	2.72	
Negative emotionality	−**0.39** [−0.57, −0.19]	−**.18** [−.27, −.09]	−4.03		−**0.41** [−0.63, −0.21]	−**.20** [−.30, −.10]	−4.08		−**0.27** [−0.39, −0.13]	−**.18** [−.27, −.09]	−4.03	
Model with covariates												
Open-mindedness	−0.06 [−0.24, 0.15]	−.03 [−.11, .06]	−0.67	.45	−0.07 [−0.28, 0.12]	−.03 [−.13, .06]	−0.69	.34	−0.04 [−0.17, 0.09]	−.03 [−.11, .06]	−0.67	.44
Conscientiousness	−0.08 [−0.26, 0.11]	−.04 [−.13, .05]	−0.80		−0.09 [−0.31, 0.12]	−.04 [−.14, .06]	−0.84		−0.06 [−0.20, 0.10]	−.04 [−.13, .05]	−0.80	
Extraversion	0.09 [−0.09, 0.26]	.04 [−.04, .13]	0.99		0.09 [−0.09, 0.27]	.04 [−.05, .14]	0.90		0.06 [−0.08, 0.20]	.04 [−.04, .13]	0.99	
Agreeableness	**0.37** [0.16, 0.59]	**.16** [.07, .25]	3.49		**0.40** [0.20, 0.62]	**.18** [.08, .29]	3.58		**0.26** [0.13, 0.40]	**.16** [.07, .26]	3.49	
Negative emotionality	−0.21 [−0.43, 0.00]	−.10 [−.20, .00]	−1.92		−**0.23** [−0.47, −0.01]	−**.11** [−.22, .00]	−1.99		−0.14 [−0.30, 0.00]	−.10 [−.20, .00]	−1.92	
Gender	−**0.42** [−0.67, −0.16]	−**.15** [−.25, −.06]	−3.26		−**0.44** [−0.70, −0.17]	−**.17** [−.27, −.07]	−3.23		−**0.29** [−0.48, −0.12]	−**.15** [−.25, −.06]	−3.25	
Math grade	**0.16** [0.04, 0.29]	**.13** [.03, .22]	2.61		**0.17** [0.05, 0.29]	**.15** [.04, .25]	2.70		**0.11** [0.03, 0.20]	**.13** [.03, .22]	2.61	
Reasoning ability	0.00 [−0.02, 0.03]	.02 [−.08, .11]	0.36		0.00 [−0.02, 0.03]	.01 [−.09, .12]	0.23		0.00 [−0.01, 0.02]	.02 [−.08, .11]	0.36	
Outcome: Cognitive activation												
Model without covariates												
Open-mindedness	0.11 [−0.07, 0.29]	.05 [−.03, .14]	1.19	.30	0.11 [−0.07, 0.31]	.06 [−.03, .15]	1.21	.20	0.07 [−0.05, 0.17]	.05 [−.03, .14]	1.19	.29
Conscientiousness	−0.06 [−0.23, 0.15]	−.03 [−.12, .06]	−0.64		−0.05 [−0.25, 0.14]	−.03 [−.12, .07]	−0.54		−0.04 [−0.15, 0.09]	−.03 [−.12, .06]	−0.62	
Extraversion	−0.06 [−0.22, 0.12]	−.03 [−.11, .05]	−0.67		−0.06 [−0.23, 0.12]	−.03 [−.12, .06]	−0.71		−0.04 [−0.15, 0.07]	−.03 [−.11, .06]	−0.66	
Agreeableness	0.17 [−0.03, 0.37]	.08 [−.01, .17]	1.69		0.18 [−0.02, 0.38]	.08 [−.01, .18]	1.73		0.10 [−0.02, 0.23]	.08 [−.01, .17]	1.68	
Negative emotionality	−**0.26** [−0.45, −0.07]	−**.13** [−.22, −.04]	−2.79		−**0.27** [−0.45, −0.09]	−**.14** [−.23, −.04]	−2.86		−**0.16** [−0.28, −0.05]	−**.13** [−.22, −.04]	−2.79	
Model with covariates												
Open-mindedness	0.11 [−0.08, 0.30]	.05 [−.04, .14]	1.13	.32	0.11 [−0.10, 0.32]	−.03 [−.13, .06]	1.14	.23	0.07 [−0.06, 0.19]	−.03 [−.11, .06]	1.13	.32
Conscientiousness	−0.04 [−0.23, 0.16]	−.02 [−.11, .08]	−0.35		−0.03 [−0.23, 0.22]	−.04 [−.14, .06]	−0.28		−0.02 [−0.14, 0.10]	−.04 [−.13, .05]	−0.32	
Extraversion	0.01 [−0.17, 0.20]	.00 [−.09, .09]	0.09		0.01 [−0.18, 0.19]	.04 [−.05, .14]	0.01		0.01 [−0.12, 0.11]	.04 [−.04, .13]	0.10	
Agreeableness	0.20 [−0.02, 0.41]	.09 [−.01, .19]	1.83		0.21 [−0.02, 0.41]	.18 [.08, .29]	1.91		0.12 [−0.02, 0.25]	.16 [.07, .26]	1.83	
Negative emotionality	−0.16 [−0.38, 0.07]	−.08 [−.18, .03]	−1.45		−0.16 [−0.36, 0.07]	−.11 [−.22, .00]	−1.44		−0.10 [−0.24, 0.04]	−.10 [−.20, .00]	−1.44	
Gender	−0.25 [−0.51, −0.02]	−.10 [−.20, .00]	−1.94		−**0.27** [−0.54, −0.01]	−**.17** [−.27, −.07]	−2.06		−0.16 [−0.32, 0.00]	−.15 [−.25, −.06]	−1.97	
Math grade	0.10 [−0.01, 0.22]	.09 [−.01, .19]	1.71		0.11 [−0.02, 0.24]	.15 [.04, .25]	1.80		0.07 [−0.01, 0.15]	.13 [.03, .22]	1.70	
Reasoning ability	0.00 [−0.03, 0.02]	−.02 [−.11, .08]	−0.30		0.00 [−0.03, 0.02]	.01 [−.09, .12]	−0.39		0.00 [−0.02, 0.01]	.02 [−.08, .11]	−0.31	
Outcome: Classroom management												
Model without covariates												
Open-mindedness	0.02 [−0.16, 0.20]	.01 [−.08, .09]	0.16	.50	0.00 [−0.20, 0.17]	.00 [−.10, .10]	0.03	.19	0.01 [−0.13, 0.13]	.01 [−.07, .09]	0.19	.55
Conscientiousness	0.06 [−0.14, 0.23]	.03 [−.06, .12]	0.64		0.07 [−0.11, 0.27]	.04 [−.07, .14]	0.69		0.04 [−0.08, 0.17]	.03 [−.06, .11]	0.64	
Extraversion	0.03 [−0.14, 0.22]	.02 [−.07, .10]	0.38		0.04 [−0.12, 0.22]	.02 [−.07, .12]	0.44		0.02 [−0.10, 0.15]	.02 [−.06, .09]	0.38	
Agreeableness	0.11 [−0.07, 0.31]	.05 [−.04, .14]	1.12		0.11 [−0.10, 0.31]	.06 [−.05, .16]	1.08		0.07 [−0.07, 0.20]	.05 [−.04, .13]	1.12	
Negative emotionality	−0.10 [−0.29, 0.07]	−.05 [−.14, .04]	−1.10		−0.12 [−0.31, 0.10]	−.06 [−.16, .04]	−1.20		−0.07 [−0.19, 0.05]	−.05 [−.13, .04]	−1.08	

(continued)

Table 4. (continued)

Predictors	SPIQ experience				SPIQ construal				SPIQ consensus			
	b [95% CI]	β [95% CI]	*t*	R_m	b [95% CI]	β [95% CI]	*t*	R_m	b [95% CI]	β [95% CI]	*t*	R_m
Model with covariates												
Open-mindedness	0.07 [−0.12, 0.26]	.03 [−.06, .12]	0.68	.50	0.06 [−0.15, 0.24]	.03 [−.07, .13]	0.55	.19	0.05 [−0.08, 0.18]	.03 [−.05, .12]	0.71	.55
Conscientiousness	0.08 [−0.12, 0.32]	.04 [−.06, .13]	0.77		0.09 [−0.13, 0.29]	.05 [−.06, .16]	0.87		0.05 [−0.07, 0.20]	.04 [−.06, .13]	0.78	
Extraversion	0.03 [−0.18, 0.20]	.01 [−.08, .10]	0.31		0.04 [−0.15, 0.24]	.02 [−.08, .13]	0.45		0.02 [−0.11, 0.13]	.01 [−.07, .10]	0.30	
Agreeableness	0.08 [−0.14, 0.30]	.04 [−.06, .13]	0.76		0.08 [−0.14, 0.31]	.04 [−.07, .15]	0.75		0.06 [−0.07, 0.19]	.04 [−.06, .13]	0.76	
Negative emotionality	−0.09 [−0.31, 0.14]	−.04 [−.15, .06]	−0.83		−0.10 [−0.32, 0.13]	−.05 [−.17, .07]	−0.85		−0.06 [−0.22, 0.08]	−.04 [−.14, .06]	−0.81	
Gender	−0.05 [−0.32, 0.22]	−.02 [−.12, .08]	−0.36		−0.07 [−0.36, 0.20]	−.03 [−.14, .09]	−0.49		−0.03 [−0.20, 0.14]	−.02 [−.11, .08]	−0.37	
Math grade	−0.03 [−0.16, 0.09]	−.02 [−.12, .08]	−0.42		−0.03 [−0.16, 0.09]	−.03 [−.15, .09]	−0.48		−0.02 [−0.10, 0.05]	−.02 [−.12, .08]	−0.41	
Reasoning ability	0.00 [−0.02, 0.02]	.00 [−.10, .10]	−0.01		0.00 [−0.02, 0.03]	.01 [−.11, .12]	0.09		0.00 [−0.02, 0.01]	.00 [−.10, .09]	−0.06	

Note. Each model additionally contains 17 dummy-coded predictor variables indicating class membership to control for clustered data at Level 3. For brevity, these fixed effects are not displayed in this table, but shown in Table S7 in the OSM. *b* = unstandardized multilevel regression coefficient; β = standardized multilevel regression coefficient; R_m = marginal multiple R for generalized linear mixed effect models. Gender is coded with 0 = male; 1 = female. Personality traits and covariates were centered at the grand mean. Regression coefficients printed in **bold** are statistically significant at $p < .05$, and regression coefficients printed in **bold and gray shading** are statistically significant at $p < .001$. Exact *p*-values can be found in Table S8.

Table 5. SPIQ components as predictors of lesson-specific PLA (RQ2).

	Lesson-specific PLA				
Predictors	*b* [95% CI]	β [95% CI]	*t*	*p*	R_m
SPIQ experience					
Model without covariates					.45
Teacher support	**0.42** [0.39, 0.46]	**.33** [.30, .36]	22.65	<.001	
Cognitive activation	**0.07** [0.03, 0.10]	**.05** [.02, .08]	3.49	<.001	
Classroom management	**0.07** [0.04, 0.11]	**.05** [.02, .08]	3.85	<.001	
Model with covariates					.56
Teacher support	**0.46** [0.41, 0.50]	**.34** [.31, .37]	21.96	<.001	
Cognitive activation	**0.07** [0.02, 0.11]	**.05** [.02, .08]	3.30	<.001	
Classroom management	**0.09** [0.05, 0.14]	**.06** [.03, .09]	4.45	<.001	
Gender	−**0.39** [−0.55, −0.22]	−**.16** [−.24, −.09]	−4.28	<.001	
Math grade	**0.27** [0.18, 0.36]	**.24** [.16, .32]	5.68	<.001	
Reasoning ability	0.01 [−0.01, 0.03]	.07 [−-.02, .15]	1.56	.120	
SPIQ construal					
Model without covariates					.43
Teacher support	**0.40** [0.36, 0.44]	**.30** [.27, .32]	19.77	<.001	
Cognitive activation	**0.08** [0.04, 0.12]	**.06** [.03, .09]	4.04	<.001	
Classroom management	**0.07** [0.03, 0.11]	**.05** [.02, .07]	3.40	<.001	
Model with covariates					.54
Teacher support	**0.43** [0.38, 0.47]	**.30** [.27, .33]	18.90	<.001	
Cognitive activation	**0.08** [0.04, 0.12]	**.06** [.03, .09]	3.72	<.001	
Classroom management	**0.08** [0.03, 0.12]	**.05** [.02, .08]	3.52	<.001	
Gender	−**0.39** [−0.54, −0.19]	−**.16** [−.24, −.09]	−4.28	<.001	
Math grade	**0.27** [0.19, 0.36]	**.24** [.16, .32]	5.67	<.001	
Reasoning ability	0.01 [0.00, 0.03]	.07 [−.02, .15]	1.56	.120	
SPIQ consensus					
Model without covariates					.45
Teacher support	**0.62** [0.57, 0.67]	**.33** [.30, .36]	22.65	<.001	
Cognitive activation	**0.10** [0.04, 0.16]	**.05** [.02, .08]	3.49	<.001	
Classroom management	**0.11** [0.05, 0.16]	**.05** [.02, .08]	3.84	<.001	
Model with covariates					.56
Teacher support	**0.67** [0.60, 0.73]	**.34** [.31, .37]	21.96	<.001	
Cognitive activation	**0.11** [0.04, 0.17]	**.05** [.02, .08]	3.30	<.001	
Classroom management	**0.14** [0.08, 0.20]	**.06** [.04, .09]	4.49	<.001	
Gender	−**0.39** [−0.56, −0.20]	−**.16** [−.24, −.09]	−4.29	<.001	
Math grade	**0.27** [0.18, 0.35]	**.24** [.16, .32]	5.68	<.001	
Reasoning ability	0.01 [0.00, 0.03]	.07 [−.02, .15]	1.55	.123	

Note. Each model additionally contains 17 dummy-coded predictor variables indicating class membership to control for clustered data at Level 3. For brevity, these fixed effects are not displayed in this table, but shown in Table S9 in the OSM. *b* = unstandardized multilevel regression coefficient; β = standardized multilevel regression coefficient; R_m = marginal multiple R for generalized linear mixed effect models. Gender is coded with 0 = male; 1 = female. SPIQ components were centered within students. Covariates were centered at the grand mean and added as predictors of random intercepts. Regression coefficients printed in **bold** are statistically significant at $p < .05$, and regression coefficients printed in **bold and gray shading** are statistically significant at $p < .001$.

attributable to the within-student level, suggesting substantial fluctuations within students. The exact conditions of these fluctuations remained unclear (e.g., fluctuations due to idiosyncratic student characteristics, teacher states, lesson content, or interactions among them; Talić et al., 2022). The present study found no clear separation of lesson-specific idiosyncratic and consensual SPIQ components within SPIQ experience that could have explained such fluctuations. Perhaps, perceptions of instructional quality with regard to a certain teacher assimilate between students over time due to a social construction process that leads to a kind of stereotypical judgment of a teacher as good or bad also in terms of his or her instructional quality.

The question of how idiosyncratic SPIQ actually are remains. Generally, Rauthmann et al. (2015) noted that “most people perceive situations as most other people do,” leaving little remaining variance after extracting consensual perceptions. In the present study, it might be that variance in SPIQ construal and consensus was too limited to draw reliable conclusions on this question due to limited variance across lessons. Indeed, Talić et al. (2022) reported a maximum of 11 % of variance between lessons (in contrast to a maximum of 54 % of variance between students) on the same dataset. Future research might consider assessing a longer time frame to capture more variability across lessons or to compare multiple subjects that might change lesson content more frequently. It is also important to keep in mind that one focus of situation research is the examination of why certain people are in certain situations (Rauthmann, 2021). For instance, extraverted people might go to parties or get coffee with their friends because they enjoy the settings (Matz & Harari, 2021). In the present study, however, the situations that were assessed (i.e., lessons in math instruction across three weeks) were not created or deliberately chosen by the students, but constitute a forced environment. The examination of elective subjects might thus offer more insight into idiosyncrasies in SPIQ that go along with a more self-directed choice of attended lessons.

Notwithstanding, the lack of detecting differential relations across the three components might imply the question of the usefulness of differentiating these components within SPIQ. We assert that this differentiation is useful for the examination of SPIQ in shared lessons. First, the advanced insights gained by differentiating different components are theoretically informative. For instance, one can distinguish between higher levels of agreeableness related to higher ratings for teacher support possibly because more agreeable students tend to agree more with the posed item in the questionnaire (i.e., reflecting a relation of agreeableness to *experienced* teacher support) versus more agreeable students construing instructional behavior as more supportive above and beyond their classmates' perceptions (i.e., reflecting a relation of agreeableness to *construed* teacher support) versus more agreeable students' overlapping in their perceptions with their classmates (i.e., reflecting a relation of agreeableness to *consensual* teacher support). Drawing and testing such distinctions is not possible if the components are not separated.

Second, although result patterns were largely similar across components in our findings, there are still some noteworthy differences. For instance, we found a statistically significant moderator effect of personality traits on SPIQ–PLA relations only for teacher support construal and agreeableness, indicating that it is not the mere rating of instructional behavior as supportive that is related to a lower positive effect of perceived teacher support on lesson-specific PLA, but rather the idiosyncratic construal of more agreeable students. Hence, the differentiation in the SPIQ components yields more nuanced insights that offer the elaboration of further hypotheses to be addressed in future studies. In the present sample of German 9th and 10th grade students in math instruction, the TBDs' SPIQ components could hardly be distinguished, suggesting that SPIQ experience might generalize to SPIQ construal and consensus in similar contexts. However, it is needed to consider all components, SPIQ experience, construal, and consensus, and their implications to be able to draw conclusions on associations of the raw ratings, the idiosyncratic or consensual perceptions, respectively.

Students' agreeableness and negative emotionality were most closely linked to state SPIQ (RQ1)

From the Big Five personality traits, agreeableness and negative emotionality were most strongly tied to state SPIQ and particularly teacher support (RQ1). Agreeableness was positively related to all components of teacher support. Negative emotionality was negatively related to all components of teacher support, whereas these relations did not reach statistical significance at $p < .001$ after including the covariates (see a discussion on the role of covariates below). We could not detect any statistically significant relations between personality traits and the other dimensions of cognitive activation and classroom management. The dimension of teacher support captures more affective perceptions of instructional quality than the dimensions of cognitive activation and classroom management do (e.g., indicating teachers' sensitivity for student needs), focuses on the quality of interactions and relationships of agents in the classroom and is strongly linked to students' self-determination (Praetorius et al., 2018; Ryan & Deci, 2000). The current data indicate that students' personality traits are related to these more affective perceptions of instructional quality in contrast to task-focused (i.e., cognitive activation) or classroom-focused (i.e., classroom management) perceptions. Fittingly, a previous study has also shown that students' perceptions of teacher support were predicted by teachers' agreeableness (Roloff et al., 2020). The dimensions of cognitive activation and classroom management might be less tied to personality traits because of their less affective content and clearer indications (targeted at task specifics or the learning environment, respectively). The finding that agreeableness was related to higher perceived teacher support is in line with what the trait domain captures (e.g., being cooperative and trusting; Costa & McCrae, 1992) and its positive link to positive course evaluations (Keller & Karau, 2013). Further, negative emotionality was related to lower perceived teacher support (yet only without considering the covariates gender, math grade, and reasoning ability). In other words, students reporting a higher tendency of experiencing stress and anxiety tended to perceive the same instructional behavior as less supportive. This might indicate a higher need for supportive instructional behavior for those students to benefit from it in the classroom. The personality traits of open-mindedness, conscientiousness, and extraversion did not show any relations to SPIQ. In prior research, conscientiousness was found to be the most important Big Five trait in terms of student achievement (Mammadov, 2022). Indeed, conscientiousness showed a statistically significant positive relation to perceived lesson-specific learning achievement at $p < .05$ in the present study only when considering bivariate correlations, though this relation was smaller than those with the other traits.

It is important to note that the correlational pattern of trait SPIQ and personality traits differs in some parts from relations between state SPIQ and personality traits (see *Preliminary Analyses*). In general, relations between state SPIQ and personality traits were more numerous than between trait SPIQ and personality traits and differed slightly regarding the personality traits they correlated with. The present study focused on examinations at the level of individual lessons such that relations between SPIQ trait measures were only of secondary interest and only reported to inform interested readers. For further generation of hypotheses, trait SPIQ relations as well as relations for the 15 personality subfacets of the Big Five traits (Soto & John, 2017) are provided in Table S1 in the OSM.

Lesson-specific PLA showed the largest within-student associations to state teacher support (RQ2)

Lesson-specific PLA was predominantly positively related to all SPIQ components, yet with a descriptively larger

Table 6. Personality traits as moderators of the association between SPIQ components and lesson-specific PLA (RQ 3).

Predictors	Lesson-specific PLA				
	b [95% CI]	β [95% CI]	*t*	*p*	R_m
SPIQ experience					
Model without covariates					.54
Teacher support	**0.45** [0.41, 0.49]	**.34** [.31, .37]	22.53	<.001	
Cognitive activation	**0.07** [0.03, 0.11]	**.05** [.02, .08]	3.29	.001	
Classroom management	**0.08** [0.03, 0.12]	**.05** [.02, .08]	3.58	<.001	
Teacher support x agreeableness	−**0.08** [−0.14, −0.02]	−**.03** [−.06, −.01]	−2.74	.006	
Teacher support x negative emotionality	**0.06** [0.01, 0.12]	**.03** [.00, .05]	2.07	.038	
Agreeableness	0.11 [−0.04, 0.27]	.05 [−.02, .13]	1.34	.181	
Negative emotionality	−**0.45** [−0.59, −0.32]	−**.24** [−.31, −.16]	−6.10	<.001	
Model with covariates					.60
Teacher support	**0.46** [0.42, 0.50]	**.34** [.31, .37]	22.07	<.001	
Cognitive activation	**0.06** [0.01, 0.10]	**.04** [.01, .07]	2.73	.006	
Classroom management	**0.08** [0.04, 0.12]	**.05** [.03, .08]	3.85	<.001	
Teacher support x agreeableness	−**0.08** [−0.14, −0.03]	−**.04** [−.06, −.01]	−2.87	.004	
Teacher support x negative emotionality	0.06 [0.00, 0.11]	.03 [.00, .05]	1.90	.058	
Agreeableness	**0.21** [0.04, 0.37]	**.10** [.02, .18]	2.55	.011	
Negative emotionality	−**0.27** [−0.44, −0.12]	−**.14** [−.22, −.06]	−3.46	<.001	
Gender	−**0.37** [−0.56, −0.16]	−**.15** [−.23, −.07]	−3.80	<.001	
Math grade	**0.23** [0.14, 0.32]	**.21** [.13, .29]	4.95	<.001	
Reasoning ability	**0.02** [0.00, 0.04]	**.09** [.01, .17]	2.15	.033	
SPIQ construal					
Model without covariates					.47
Teacher support	**0.43** [0.37, 0.46]	**.30** [.27, .33]	19.54	<.001	
Cognitive activation	**0.08** [0.05, 0.13]	**.06** [.03, .09]	4.01	<.001	
Classroom management	**0.07** [0.02, 0.11]	**.04** [.02, .07]	3.04	.002	
Teacher support x agreeableness	−**0.13** [−0.19, −0.07]	−**.05** [−.08, −.03]	−4.34	<.001	
Agreeableness	**0.25** [0.10, 0.40]	**.12** [.04, .20]	3.07	.002	
Model with covariates					.57
Teacher support	**0.43** [0.39, 0.48]	**.30** [.27, .33]	19.05	<.001	
Cognitive activation	**0.07** [0.03, 0.12]	**.05** [.02, .09]	3.43	<.001	
Classroom management	**0.07** [0.02, 0.11]	**.04** [.01, .07]	3.00	.002	
Teacher support x agreeableness	−**0.13** [−0.19, −0.07]	−**.06** [−.08, −.03]	−4.43	<.001	
Agreeableness	**0.31** [0.16, 0.46]	**.15** [.08, .23]	3.94	<.001	
Gender	−**0.48** [−0.67, −0.29]	−**.20** [−.27, −.12]	−5.16	<.001	
Math grade	**0.25** [0.15, 0.34]	**.23** [.14, .31]	5.32	<.001	
Reasoning ability	0.02 [0.00, 0.03]	.08 [−.01, .16]	1.82	.070	
SPIQ consensus					
Model without covariates					.54
Teacher support	**0.66** [0.60, 0.72]	**.34** [.31, .37]	22.54	<.001	
Cognitive activation	**0.11** [0.04, 0.17]	**.05** [.02, .08]	3.29	.001	
Classroom management	**0.11** [0.05, 0.17]	**.05** [.02, .08]	3.63	<.001	
Teacher support x agreeableness	−**0.11** [−0.21, −0.03]	−**.03** [−.06, −.01]	−2.75	.006	
Teacher support x negative emotionality	**0.08** [0.01, 0.17]	**.03** [.00, .05]	2.08	.038	
Agreeableness	0.11 [−0.05, 0.25]	.05 [−.02, .13]	1.34	.183	
Negative emotionality	−**0.45** [−0.60, −0.32]	−**.24** [−.31, −.16]	−6.10	<.001	
Model with covariates					.60
Teacher support	**0.67** [0.61, 0.73]	**.34** [.31, .37]	22.07	<.001	
Cognitive activation	**0.09** [0.02, 0.16]	**.04** [.01, .07]	2.73	.006	
Classroom management	**0.12** [0.05, 0.19]	**.06** [.03, .08]	3.92	<.001	
Teacher support x agreeableness	−**0.12** [−0.20, −0.04]	−**.04** [−.06, −.01]	−2.87	.004	
Teacher support x negative emotionality	0.08 [−0.01, 0.16]	.03 [.00, .05]	1.90	.058	
Agreeableness	**0.21** [0.05, 0.38]	**.10** [.02, .18]	2.55	.011	
Negative emotionality	−**0.27** [−0.43, −0.11]	−**.14** [−.22, −.06]	−3.46	<.001	
Gender	−**0.37** [−0.58, −0.16]	−**.15** [−.23, −.07]	−3.81	<.001	
Math grade	**0.23** [0.13, 0.32]	**.21** [.12, .29]	4.94	<.001	
Reasoning ability	**0.02** [0.00, 0.03]	**.09** [.01, .17]	2.14	.034	

Note. Each model additionally contains 17 dummy-coded predictor variables indicating class membership to control for clustered data at Level 3. For brevity, these fixed effects are not displayed in this table, but shown in Table S10 in the OSM. *b* = unstandardized multilevel regression coefficient; β = standardized multilevel regression coefficient; R_m = marginal multiple R for generalized linear mixed effect models. Gender is coded with 0 = male; 1 = female. SPIQ components were centered within students. Covariates and personality traits were centered at the grand mean and added as predictors of random intercepts. Regression coefficients printed in **bold** are statistically significant at $p < .05$, and regression coefficients printed in **bold and gray shading** are statistically significant at $p < .001$.

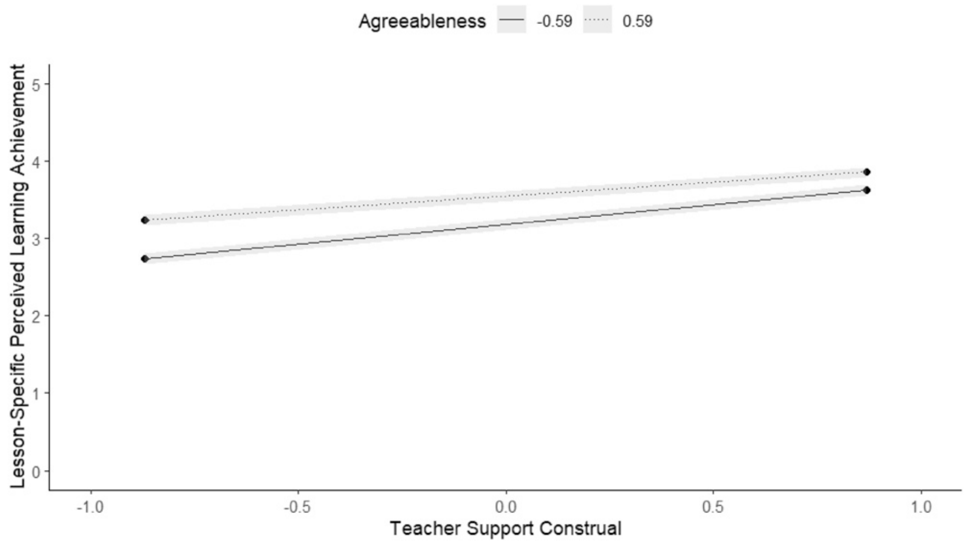

Figure 1. Agreeableness as moderator of the association between teacher support construal and lesson-specific perceived learning achievement.
Note. Teacher support construal was centered within persons and agreeableness was centered at the grand mean. For illustrative purposes, high agreeableness is reflected as + 1 *SD* from the grand mean (i.e., +0.59), and low agreeableness is reflected as—1 *SD* from the grand mean (i.e., −0.59).

effect size for teacher support than for cognitive activation and classroom management (RQ2). Given theoretical assumptions and empirical findings on the relation between the TBDs and achievement, these findings are rather unexpected although also relations between teacher support and achievement have been reported (Fauth et al., 2014). Based on between-person research designs, positive relations between cognitive activation and classroom management to student achievement are expected, while teacher support is more closely related to student motivation (Praetorius et al., 2018). The present study identified the unambiguously strongest relation between teacher support and a lesson-specific, subjective learning achievement indicator at the within-student level. First, this contrast between expected and actual results might reveal differential SPIQ–achievement relations at different levels of analyses due to using interindividual versus intraindividual variance (Molenaar, 2004; Murayama et al., 2017). Second, the strong relations between teacher support and lesson-specific PLA might also in part stem from a semantic item wording overlap, focusing on the support or progress, respectively, during the learning process. However, this overlap is likely not strong enough to exclusively explain their relations. Third, prior findings that used, for instance, standardized test scores to examine SPIQ–achievement relations can only vaguely be compared to our findings that are based on *perceived* learning achievement (i.e., not reflecting objective achievement). In contrast, the math grade showed substantial relations to teacher support and cognitive activation across all components (see Table S1), demonstrating some differential result patterns for students' lesson-specific PLA and their math grade. Specifically, while teacher support and PLA seemed to be positively related to each other within lessons, the supposed positive relation of cognitive activation was not likewise reflected in students' lesson-specific PLA.

Taken together, lesson-specific PLA has been shown to be suitable as a daily measure in experience sampling designs, thus maintaining high ecological validity (see Niepel et al., 2022; see also *Limitations and Future Research* below) and offering new insights into the dynamics of PLA in students' daily life within lessons. Within-student relations between state SPIQ components and lesson-specific PLA remained virtually the same after including covariates.

Students' agreeableness moderated the within-student relation between lesson-specific teacher support and PLA (RQ3)

Concerning the role of students' Big Five personality traits on lesson-specific relations between SPIQ and PLA, we only identified agreeableness as a statistically significant moderator with teacher support construal (RQ3), further underpinning the relevance of agreeableness in relation to teacher support. This moderation describes a stronger positive relation between construed teacher support and lesson-specific PLA for lower levels of agreeableness. Students reporting to be more agreeable might uniquely construe teacher support in a more lenient way (irrespective of actual teacher support). In fact, prior research showed that agreeableness was positively related to rating leniency, with more agreeable persons providing more favorable ratings even in light of poorer performance (Bernardin et al., 2000, 2009; Randall & Sharples, 2012; Yun et al., 2005). The effect sizes of the moderation effect were rather small, which is in line with previous research on interaction effects between personality traits and perceived situation characteristics on personality states (Rauthmann, 2021). Two moderator effects of negative emotionality on the relation experienced and consensual teacher support, respectively, and lesson-specific PLA reached statistical significance only at $p < .05$ and only without considering covariates. In conclusion, this study underscores the importance of students' agreeableness and negative emotionality for perceptions of instructional quality.

The roles of gender, math grade, and reasoning ability

Students' gender, math grade, and reasoning ability were entered in all models as potentially relevant covariates. The presentation of model results with and without covariates allows interested readers to scrutinize the covariates' effects. Here, we discuss some noteworthy findings that were not the central focus of the present study. The relations between negative emotionality and teacher support components did not reach statistical significance after including the covariates, indicating some confounding of these variables. While reasoning ability did not play a significant role above and beyond gender and math grade in any of the examined relations, gender and math grade were each related to teacher support, and seemed to be confounded with negative emotionality. Students with higher math grades perceived more teacher support. Female students perceived less teacher support as well as less learning achievement in the lesson, and they reported higher negative emotionality than male students. This clearly highlights the need to investigate gender-related dynamics in education. While girls and boys show similar math achievement levels in the early school years, gender differences eventually emerge—with boys scoring higher than girls—throughout the school career (Buchmann et al., 2008). In addition, prior studies reported lower self-reported math abilities in female students even if actual achievement levels were equivalent (Niepel et al., 2019; OECD, 2015). Understanding gender-related processes with the goal of approaching gender equality (UNESCO, 2017) should be focally pursued in future studies.

Limitations and future research

First, we focused on *students'* perceptions of instructional quality. However, *teachers'* self-perceptions of instructional quality and their lesson-to-lesson variation would also be of great interest. Even if consensus (as calculated here) captures something intersubjective among students, this is only an approximation and might still diverge from teachers' self-perceptions or independent observers' perceptions. Thus, future research should assess teachers' state perceptions to gain a more balanced picture of classroom dynamics. In such a design, consensus scores can be based on multiple rating sources (i.e., teachers' and students' perceptions; see Abrahams et al., 2021, for the computation of consensus between self- and other ratings).

Similarly, although we assessed students' state perceptions of instructional quality, the assessment of students' personality *states* (in addition to their personality traits) would have been beneficial to gain a more in-depth understanding on personality "in action" in different situations (Baumert et al., 2017). For instance, Whole Trait Theory posits that traits are enacted based on social-cognitive mechanisms in situations, thus yielding within-person variability across situations in trait manifestations (Fleeson & Jayawickreme, 2021). Future research examining the relevance of personality in different contexts should incorporate personality states besides personality traits to allow for a more holistic picture (see also Ching et al., 2014).

In a similar vein, the present study did not sample situational characteristics in form of a collection of different dimensions (e.g., Situational Eight Diamonds; Rauthmann et al., 2014). Rather, we conceptualized perceptions of instructional quality as situational characteristics. It is unclear, however, whether and how perceptions of instructional quality relate to frameworks of situational characteristics. Future research could assess both perceptions of instructional quality as well as situational perceptions to gain further insight into how these perceptions relate to each other, and with this, how students attribute the TBDs in terms or broader, more general characteristics.

To estimate the relevance of our observed effects, it is crucial to discuss effect sizes. For effects that were statistically significant at $p < .001$, we observed effect sizes that ranged between $\beta = |.05|$ and $\beta = |.33|$, with a mean of $\beta = |.17|$. In particularly small effects below .10, their statistical, content, or practical relevance seems questionable at first sight. Yet, it is important to keep in mind that even comparatively small effects can have a crucial impact when accumulating over time and at scale (Funder & Ozer, 2019; Götz et al., 2022; Matz et al., 2017; Rauthmann et al., 2015). The present study examined dynamics at the level of school lessons (i.e., 45-minute intervals), something experienced by students many thousands of times during their school career.

Further, it is important to note that our study is correlational and, therefore, cannot be used to infer causality (although we do refer to statistical predictions). For instance, it could also be that lesson-specific PLA causally influences the perceptions of instructional quality (e.g., "If I have understood the lesson well, the teacher must have been teaching good"). Similarly, our analyses might omit potentially relevant third variables such that a causal interpretation of our results might be biased (Wilms et al., 2021). To infer causality, experimental research designs including control groups and possible mediators are needed. Although the present study cannot ascertain causal pathways, it still implemented an intensive longitudinal design where all personality traits and covariates were assessed prior to the experience sampling phase, such that the direction of effects with traits predicting subsequent states seems more plausible than vice versa.

We used lesson-specific PLA targeted at the conceptual comprehension of the lesson content as a state achievement indicator. In between-person research designs, usually standardized test scores or school grades are used as achievement indicators (see Arens et al., 2017 for a balanced discussion on different achievement indicators), which are more objective than our perceived state achievement indicator. Yet, in an experience sampling design, the implementation of a standardized test in each lesson is hardly feasible. Further, the positive and substantial relation between math grade and our PLA indicator supports the latter's validity. In addition, a previous study has demonstrated the empirical distinction of students' lesson-specific PLA versus their perceived lesson-specific math abilities (math self-concept; Niepel et al., 2022), further suggesting lesson-specific PLA's validity in an experience sampling design. In line with our focus on individual perceptions within SPIQ, we thus use lesson-specific PLA as a subjective achievement indicator. Future research should, nevertheless, address the question of different indicators of student achievement and their respective implications in an experience sampling design.

Finally, we note that our findings are based on a sample of German secondary school students attending the 9th and 10th grades in schools of the highest ability track, and we only considered math instruction. To test the generalizability of our findings, students from other countries, ability tracks, age groups, and teaching subjects are needed.

Implications and conclusion

Perceptions of instructional quality are omnipresent in daily school life and have wide-ranging implications for both students in terms of their achievement and motivation, and for teachers in terms of evaluations of their teaching effectiveness (OECD, 2014). The present experience sampling study contributed to the understanding of such perceptions within the framework of Three Basic Dimensions (teacher support, cognitive activation, classroom management; Klieme et al., 2001) from the students' perspectives by considering students' personality traits and perceived learning gains in individual lessons. We disentangled idiosyncratic from consensual student perceptions that are confounded in the raw perceptions to uncover their respective relations with students' traits and states, which we found to be remarkably similar across the different components of experience, consensus, and construal. Students' personality traits, in particular agreeableness and negative emotionality, were related to perceptions of instructional quality, underscoring their relevance in daily school life and providing more insights into perception tendencies based on student characteristics. For teachers, possible implications might include shaping their teaching practices such that they can adapt to different student personality traits more purposefully by providing tailored teaching approaches. Such adaptation would require highly individualized instruction, which has long been discussed, particularly in the context of intelligent computer-assisted instruction (e.g., Fletcher, 1992). However, it could be helpful for teachers to understand that the perception of their teaching partially depends on certain personality traits of the students, such as agreeableness and negative emotionality. Teachers could thus interpret the SPIQ in the context of personality traits and deal with student feedback on their teaching in a reflective manner (Spooren et al., 2013). For students, a self-awareness of their own personality traits could be beneficial for their own well-being in and outside the classroom. For instance, students high in agreeableness could benefit from knowledge on their tendency to avoid conflict and its manifestations (e.g., needs possibly not being met) and implications (e.g., adapting communication styles) in daily life (see also Jelley, 2021, for a discussion on personality-feedback interventions in the workplace). For educational policy-makers and in teacher training, the implications could extend to more personalized school models. If different personality traits relate to the same instructional behavior differently, it is conceivable that different students require different instructional behavior to thrive. For instance, students' personality traits have been shown to be related to learning strategies (Komarraju et al., 2011; Zhang, 2003). Moreover, undergraduate students' learning strategies and personality traits were related to the perception of and performance in a gamified learning intervention (Buckley & Doyle, 2017), suggesting the potential of individualized learning opportunities. Implementing more personalized school models with more choice options for students according to their interests, needs, and abilities could account for different, individual experiences in school rather than pursuing a rigid one-size-fits-all approach.

Author Notes

The views and opinions expressed in this article are the authors' and do not reflect the official opinions or policies of the authors' host affiliations or any of the supporting institutions.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by a grant from the Luxembourg National Research Fund (FNR) to Christoph Niepel (C16/SC/11333571).

Open Science Statement

Data cannot be made available because of data protection concerns. Readers interested in the data can contact the first author. The R analysis code can be accessed at https://osf.io/e8jqm/.

Ethical statement

The local ethics review panel of the University of Luxembourg and all involved German federal education authorities approved of all procedures.

ORCID iD

Irma Talic https://orcid.org/0000-0001-5994-8371

Supplemental Material

Supplemental material for this article is available online.

Notes

1. For more information on the data used in the present study, please see the *Methods* section below.
2. Rauthmann et al. (2015) distinguish between liberal and conservative contact: the former labels consensus as the shared perception among all perspectives of a situation, and the latter labels consensus as a shared perception of only external perspectives. Throughout the remainder of this article, we use the term consensus in the sense of liberal contact.
3. Throughout this article, we refer to the methodological distinction of SPIQ into experience, construal, and consensus as SPIQ *components*, and to the content-specific distinction (i.e., the framework of TBDs) into teacher support, cognitive activation, and classroom management as SPIQ *dimensions*.

4. The German school system encompasses many different educational paths that partly differ between individual states. In general, all children attend (the same) primary schools. Based on their achievement level in primary school and their preferences, they are subsequently allocated to secondary schools, which are generally divided into three different ability tracks (i.e., the lower, intermediate, and highest ability track).
5. Note that these five broad domain traits can be distinguished into three facets each (i.e., aesthetic sensitivity, intellectual curiosity, and creative imagination for open-mindedness; organization, productiveness, and responsibility for conscientiousness; sociability, assertiveness, and energy level for extraversion; compassion, respectfulness, and trust for agreeableness; and anxiety, depression, and emotional volatility for negative emotionality). Due to the explorative nature of analyses and a magnitude of multiple comparisons, the present study does not examine relations of personality traits at the facet level. For interested readers, intercorrelations between personality traits at the facet level and all other examined variables are provided in Table S1 and corresponding exact *p*-values in Table S2 in the Online Supplementary Material (OSM).
6. Descriptive statistics for personality traits at the facet level can be found in Table S3 in the OSM.
7. These analyses do not provide entirely new results. Drawing on DynASCEL data, Talić et al. (2022) reported ω and ICC coefficients for state SPIQ, and Niepel et al. (2022) reported ω and ICC coefficients for perceived lesson-specific learning achievement, yet the latter while drawing on a slightly different sample size. To provide all relevant information, we report these coefficients here anew.
8. To test the robustness of these results, we additionally conducted correlations of SPIQ traits using only the two corresponding items from the longer trait scales that were implemented in the state scales (see Table S5 in the OSM). Correlations with the Big Five were almost identical to those of the long trait scales with the one exception that teacher support was completely unrelated to personality traits.

References

Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., & van der Sluis, S. (2014). A solution to dependency: Using multilevel analysis to accommodate nested data. *Nature Neuroscience*, *17*(4), 491–496. https://doi.org/10.1038/nn.3648

Abrahams, L., Rauthmann, J. F., & Fruyt, F. D. (2021). Person-situation dynamics in educational contexts: A self- and other-rated experience sampling study of teachers' states, traits, and situations. *European Journal of Personality*, *35*(4), 598–622. https://doi.org/10.1177/08902070211005621

Amthauer, R. (1970). *Intelligenz-Struktur-Test [Intelligence structure test]*. Hogrefe.

Arens, A. K., Marsh, H. W., Pekrun, R., Lichtenfeld, S., Murayama, K., & vom Hofe, R. (2017). Math self-concept, grades, and achievement test scores: Long-term reciprocal effects across five waves and three achievement tracks. *Journal of Educational Psychology*, *109*(5), 621–634. https://doi.org/10.1037/edu0000163

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., Borkenau, P., Costantini, G., Denissen, J. J. A., Fleeson, W., Grafton, B., Jayawickreme, E., Kurzius, E., MacLeod, C., Miller, L. C., Read, S. J., Roberts, B., Robinson, M. D., Wood, D., & Wrzus, C. (2017). Integrating personality structure, personality process, and personality development. *European Journal of Personality*, *31*(5), 503–528. https://doi.org/10.1002/per.2115

Bellens, K., Van Damme, J., Van Den Noortgate, W., Wendt, H., & Nilsen, T. (2019). Instructional quality: Catalyst or pitfall in educational systems' aim for high achievement and equity? An answer based on multilevel SEM analyses of TIMSS 2015 data in Flanders (Belgium), Germany, and Norway. *Large-Scale Assessments in Education*, *7*(1), 1–27. https://doi.org/10.1186/s40536-019-0069-2

Ben-Shachar, M., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815

Bernardin, H. J., Cooke, D. K., & Villanova, P. (2000). Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology*, *85*(2), 232–236. https://doi.org/10.1037/0021-9010.85.2.232

Bernardin, H. J., Tyler, C. L., & Villanova, P. (2009). Rating level and accuracy as a function of rater personality. *International Journal of Selection and Assessment*, *17*(3), 300–310. https://doi.org/10.1111/j.1468-2389.2009.00472.x

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford Press.

Breil, S. M., Geukes, K., Wilson, R. E., Nestler, S., Vazire, S., & Back, M. D. (2019). Zooming into real-life extraversion – How personality and situation shape sociability in social interactions. *Collabra: Psychology*, *5*(1), 7. https://doi.org/10.1525/collabra.170

Buchmann, C., DiPrete, T. A., & McDaniel, A. (2008). Gender inequalities in education. *Annual Review of Sociology*, *34*(1), 319–337. https://doi.org/10.1146/annurev.soc.34.040507.134719

Buckley, P., & Doyle, E. (2017). Individualising gamification: An investigation of the impact of learning styles and personality traits on the efficacy of gamification using a prediction market. *Computers & Education*, *106*, 43–55. https://doi.org/10.1016/j.compedu.2016.11.009

Ching, C. M., Church, A. T., Katigbak, M. S., Reyes, J. A. S., Tanaka-Matsumi, J., Takaoka, S., Zhang, H., Shen, J., Arias, R. M., Rincon, B. C., & Ortiz, F. A. (2014). The manifestation of traits in everyday behavior and affect: A five-culture study. *Journal of Research in Personality*, *48*(1), 1–16. https://doi.org/10.1016/j.jrp.2013.10.002

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, *4*(1), 5–13. https://doi.org/10.1037//1040-3590.4.1.5

Danner, D., Rammstedt, B., Bluemke, M., Lechner, C., Berres, S., Knopf, T., Soto, C. J., & John, O. P. (2019). Die deutsche Version des Big Five Inventory 2 (BFI-2). *Diagnostica*, *65*(3), 121–132. https://doi.org/10.1026/0012-1924/a000218

Dickhäuser, O., & Plenter, I. (2005). "Letztes Halbjahr stand ich zwei": Zur Akkuratheit selbst berichteter Noten [On the accuracy of self-reported school marks]. *Zeitschrift Für Pädagogische Psychologie [German Journal of Educational Psychology]*, *19*(4), 219–224. https://doi.org/10.1024/1010-0652.19.4.219

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

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), 1–9. https://doi.org/10.1016/j.learninstruc.2013.07.001

Feistauer, D., & Richter, T. (2017). How reliable are students' evaluations of teaching quality? A variance components approach. *Assessment & Evaluation in Higher Education*, *42*(8), 1263–1279. https://doi.org/10.1080/02602938.2016.1261083

Fleeson, W., & Jayawickreme, E. (2021). Whole traits: Revealing the social-cognitive mechanisms constituting personality's central variable. In B. Gawronski (Ed.), *Advances in experimental social psychology* (pp. 69–128). Elsevier Academic Press. https://doi.org/10.1016/bs.aesp.2020.11.002

Fletcher, J. D. (1992). *Individualized systems of instruction*. Institute for Defense Analyses.

Franzen, P., Arens, A. K., Greiff, S., van der Westhuizen, L., Fischbach, A., Wollschläger, R., & Niepel, C. (2022). Developing and validating a short-form questionnaire for the assessment of seven facets of conscientiousness in large-scale assessments. *Journal of Personality Assessment*, *104*(6), 759–773. https://doi.org/10.1080/00223891.2021.1998083

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

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

Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small effects: The indispensable foundation for a cumulative psychological science. *Perspectives on Psychological Science*, *17*(1), 205–215. https://doi.org/10.1177/1745691620984483

Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and sinking". *Journal of Educational Psychology*, *98*(2), 307–326. https://doi.org/10.1037/0022-0663.98.2.307

Hausen, J. E., Möller, J., Greiff, S., & Niepel, C. (2022). Students' personality and state academic self-concept: Predicting differences in mean level and within-person variability in everyday school life. *Journal of Educational Psychology*, *114*(6), 1394–1411. https://doi.org/10.1037/edu0000760

Holzberger, D., Philipp, A., & Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality: A longitudinal analysis. *Journal of Educational Psychology*, *105*(3), 774–786. https://doi.org/10.1037/a0032198

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge Taylor & Francis.

Jaekel, A.-K., Göllner, R., & Trautwein, U. (2021). How students' perceptions of teaching quality in one subject are impacted by the grades they receive in another subject: Dimensional comparisons in student evaluations of teaching quality. *Journal of Educational Psychology*, *113*(4), 770–783. https://doi.org/10.1037/edu0000488

Jelley, R. B. (2021). Using personality feedback for work-related development and performance improvement: A rapid evidence assessment. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement*, *53*(2), 175–186. https://doi.org/10.1037/cbs0000230

John, O. P. (2021). History, measurement, and conceptual elaboration of the Big-Five trait taxonomy: The paradigm matures. In O. P. John, & R. W. Robins (Eds.), *Handbook of personality: Theory and research* (pp. 35–82). The Guilford Press.

Jonason, P. K., & Sherman, R. A. (2020). Personality and the perception of situations: The Big Five and Dark Triad traits. *Personality and Individual Differences*, *163*, Article 110081. https://doi.org/10.1016/j.paid.2020.110081

Keller, H., & Karau, S. J. (2013). The importance of personality in students' perceptions of the online learning experience. *Computers in Human Behavior*, *29*(6), 2494–2500. https://doi.org/10.1016/j.chb.2013.06.007

Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: "Aufgabenkultur" und Unterrichtsgestaltung [Mathematics instruction at secondary level. Task culture and instructional design]. In Bundesministerium für Bildung und Forschung (Ed.), *TIMMS - Impulse für Schule und Unterricht* (pp. 43–57). Bundesministerium für Bildung und Forschung.

Komarraju, M., Karau, S. J., Schmeck, R. R., & Avdic, A. (2011). The Big Five personality traits, learning styles, and academic achievement. *Personality and Individual Differences*, *51*(4), 472–477. https://doi.org/10.1016/j.paid.2011.04.019

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

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, *105*(3), 805–820. https://doi.org/10.1037/a0032583

Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000R [Intelligence structure test 2000R]*. Hogrefe.

Liepmann, D., Beauducel, A., Brocke, B., & Nettelnstroth, W. (2012). *Intelligenz-Struktur-Test-Screening [Intelligence structure test-screening]*. Hogrefe.

Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-Scale Assessments in Education*, *6*(1), 8. https://doi.org/10.1186/s40536-018-0061-2

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

Mammadov, S. (2022). Big Five personality traits and academic performance: A meta-analysis. *Journal of Personality*, *90*(2), 222–255. https://doi.org/10.1111/jopy.12663

Mang, J., Ustjanzew, N., Schiepe-Tiska, A., Prenzel, M., Sälzer, C., Müller, K., & Gonzaléz Rodríguez, E. (2018). *PISA 2012 Skalenhandbuch: Dokumentation der Erhebungsinstrumente*. Waxmann.

Matz, S. C., Gladstone, J. J., & Stillwell, D. (2017). In a world of big data, small effects can still matter: A reply to Boyce, Daly, Hounkpatin, and Wood (2017).

Psychological Science, *28*(4), 547–550. https://doi.org/10.1177/0956797617697445

Matz, S. C., & Harari, G. M. (2021). Personality-place transactions: Mapping the relationships between Big Five personality traits, states, and daily places. *Journal of Personality and Social Psychology*, *120*(5), 1367–1385. https://doi.org/10.1037/pspp0000297

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*(1), 81–90. https://doi.org/10.1037//0022-3514.52.1.81

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, *2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Murayama, K., Goetz, T., Malmberg, L.-E., Pekrun, R., Tanaka, A., & Martin, A. J. (2017). Within-person analysis in educational psychology: Importance and illustration. In D. W. Putwain, & K. Smart (Eds.), *British Journal of Educational Psychology Monograph Series II: Psychological aspects of education - Current trends: The role of competence beliefs in teaching and learning* (pp. 71–87). Wiley.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Niepel, C., Marsh, H. W., Guo, J., Pekrun, R., & Möller, J. (2022). Revealing dynamic relations between mathematics self-concept and perceived achievement from lesson to lesson: An experience-sampling study. *Journal of Educational Psychology*, *114*(6), 1380–1393. https://doi.org/10.1037/edu0000716

Niepel, C., Stadler, M., & Greiff, S. (2019). Seeing is believing: Gender diversity in STEM is related to mathematics self-concept. *Journal of Educational Psychology*, *111*(6), 1119–1130. https://doi.org/10.1037/edu0000340

OECD. (2014). *PISA 2012: Technical report*. OECD Publishing.

OECD. (2015). *The ABC of gender equality in education: Aptitude, behaviour, confidence. PISA*. OECD. https://doi.org/10.1787/9789264229945-en

Peterson, S. E., & Miller, J. A. (2004). Quality of college students' experiences during cooperative learning. *Social Psychology of Education*, *7*(2), 161–183. https://doi.org/10.1023/B:SPOE.0000018522.39515.19

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*(2), 322–338. https://doi.org/10.1037/a0014996

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM*, *50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4

Randall, R., & Sharples, D. (2012). The impact of rater agreeableness and rating context on the evaluation of poor performance. *Journal of Occupational and Organizational Psychology*, *85*(1), 42–59. https://doi.org/10.1348/2044-8325.002002

Rauthmann, J. F. (2021). Capturing interactions, correlations, fits, and transactions: A person-environment relations model. In J. F. Rauthmann (Ed.), *The handbook of personality dynamics and processes* (pp. 427–522). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-813995-0.00018-2

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*(4), 677–718. https://doi.org/10.1037/a0037250

Rauthmann, J. F., & Sherman, R. (2019). Toward a research agenda for the study of situation perceptions: A variance componential framework. *Personality and Social Psychology Review*, *23*(3), 238–266. https://doi.org/10.1177/1088868318765600

Rauthmann, J. F., Sherman, R. A., Nave, C. S., & Funder, D. C. (2015). Personality-driven situation experience, contact, and construal: How people's personality traits predict characteristics of their situations in daily life. *Journal of Research in Personality*, *55*, 98–111. https://doi.org/10.1016/j.jrp.2015.02.003

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Roloff, J., Klusmann, U., Lüdtke, O., & Trautwein, U. (2020). The predictive validity of teachers' personality, cognitive and academic abilities at the end of high school on instructional quality in Germany: A longitudinal study. *AERA Open*, *6*(1), 1–17. https://doi.org/10.1177/2332858419897884

Ruzek, E., Aldrup, K., & Lüdtke, O. (2022). Assessing the effects of student perceptions of instructional quality: A cross-subject within-student design. *Contemporary Educational Psychology*, *70*, Article 102085. https://doi.org/10.1016/j.cedpsych.2022.102085

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

Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology*, *7*, 110. https://doi.org/10.3389/fpsyg.2016.00110

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*(4), 454–499. https://doi.org/10.3102/0034654307310317

Sherman, R. A., Rauthmann, J. F., Brown, N. A., Serfass, D. G., & Jones, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. *Journal of Personality and Social Psychology*, *109*(5), 872–888. https://doi.org/10.1037/pspp0000036

Shernof, D. J., Ruzek, E. A., Sannella, A. J., Schorr, R. Y., Sanchez-Wall, L., & Bressler, D. M. (2017). Student engagement as a general factor of classroom experience: Associations with student practices and educational outcomes in a university gateway course. *Frontiers in Psychology*, *8*, 994. https://doi.org/10.3389/fpsyg.2017.00994

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, *113*(1), 117–143. https://doi.org/10.1037/pspp0000096

Sparfeldt, J. R., Buch, S. R., Rost, D. H., & Lehmann, G. (2008). Akkuratesse selbstberichteter Zensuren [The accuracy of self-reported grades in school]. *Psychologie in Erziehung und Unterricht*, *55*(1), 68–75.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, *83*(4), 598–642. https://doi.org/10.3102/0034654313496870

Sutin, A. R., Stephan, Y., Luchetti, M., Strickhouser, J. E., Aschwanden, D., & Terracciano, A. (2022). The association between five factor model personality traits and verbal and numeric reasoning. *Aging, Neuropsychology, and Cognition*, *29*(2), 297–317. https://doi.org/10.1080/13825585.2021.1872481

Talić, I., Scherer, R., Marsh, H. W., Greiff, S., Möller, J., & Niepel, C. (2022). Uncovering everyday dynamics in students' perceptions of instructional quality with experience sampling. *Learning and Instruction*, *81*, Article 101594. https://doi.org/10.1016/j.learninstruc.2022.101594

Toropova, A., Johansson, S., & Myrberg, E. (2019). The role of teacher characteristics for student achievement in mathematics and student perceptions of instructional quality. *Education Inquiry*, *10*(4), 275–299. https://doi.org/10.1080/20004508.2019.1591844

Tsai, Y.-M., Kunter, M., Lüdtke, O., Trautwein, U., & Ryan, R. M. (2008). What makes lessons interesting? The role of situational and individual factors in three school subjects. *Journal of Educational Psychology*, *100*(2), 460–472. https://doi.org/10.1037/0022-0663.100.2.460

UNESCO. (2017). *Cracking the code: Girls' and women's education in science, tech-nology, engineering and mathematics (STEM)*. UNESCO.

Vieluf, S., Praetorius, A.-K., Rakoczy, K., Kleinknecht, M., & Pietsch, M. (2020). Angebots-Nutzungs-Modelle der Wirkweise des Unterrichts: Ein kritischer Vergleich verschiedener Modellvarianten. *Zeitschrift für Padagogik*, *66*(1), 63–80. https://doi.org/10.3262/ZPB2001063

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

Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology*, *108*(5), 705–721. https://doi.org/10.1037/edu0000075

Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the big five. *Frontiers in Psychology*, *2*, 178. https://doi.org/10.3389/fpsyg.2011.00178

Wilms, R., Mäthner, E., Winnen, L., & Lanwehr, R. (2021). Omitted variable bias: A threat to estimating causal relationships. *Methods in Psychology*, *5*, Article 100075. https://doi.org/10.1016/j.metip.2021.100075

Wisniewski, B., Röhl, S., & Fauth, B. (2022). The perception problem: A comparison of teachers' self-perceptions and students' perceptions of instructional quality. *Learning Environments Research*, *25*(3), 775–802. https://doi.org/10.1007/s10984-021-09397-4

Wisniewski, B., Zierer, K., Dresel, M., & Daumiller, M. (2020). Obtaining secondary students' perceptions of instructional quality: Two-level structure and measurement invariance. *Learning and Instruction*, *66*, Article 101303. https://doi.org/10.1016/j.learninstruc.2020.101303

Yun, G. J., Donahue, L. M., Dudley, N. M., & McFarland, L. A. (2005). Rater personality, rating format, and social context: Implications for performance appraisal ratings. *International Journal of Selection and Assessment*, *13*(2), 97–107. https://doi.org/10.1111/j.0965-075X.2005.00304.x

Zhang, L.-F. (2003). Does the big five predict learning approaches? *Personality and Individual Differences*, *34*(8), 1431–1446. https://doi.org/10.1016/S0191-8869(02)00125-3