

## Students' Perceptions of Teaching Quality: Validation of the teaCh Scale in a Sample of Croatian Upper Secondary School Students

Krešimir Jakšić and Irena Burić

University of Zadar, Department of Psychology, Zadar, Croatia

---

### Abstract

Teaching quality is an important construct related to students' affective and cognitive outcomes. However, it is usually measured with instruments that lack a clear theoretical background and/or whose psychometric properties have not been tested with appropriate statistical procedures such as multilevel modeling. Recently, Wisniewski et al. (2020) developed the teaCh scale, which measures the quality of teaching based on the students' perceptions of seven dimensions: Care, Control, Clarity, Consolidation, Conferment, Challenge, Captivation. The aim of the study was to further test the psychometric properties of this scale using a large sample of upper secondary school students in Croatia. The results of the study show that the seven factors have adequate predictive validity for task value, student self-efficacy, and final grade, as well as adequate convergent validity and reliability. However, there are strong intercorrelations between the dimensions. These results confirm that measuring teaching quality through student reports is challenging from a psychometric point of view, but that students' perceptions can still provide useful feedback for teachers.


*Keywords:* teaching quality, student ratings, upper secondary school students, validation

---

### Introduction

Teaching quality is an important construct related to student outcomes (Nilsen, et al., 2016), such as affective attitudes towards the subject (Leino et al., 2022), motivation (Scherer & Nilsen, 2016), and academic achievement (Nortvedt et al.,

---

Krešimir Jakšić  <https://orcid.org/0000-0003-2056-3604>

Irena Burić  <https://orcid.org/0000-0001-9182-968X>

#### *Acknowledgements*

This research was supported by the Croatian Science Foundation (Grant No. IP-04-2019-5472).

✉ Krešimir Jakšić, Department of Psychology, University of Zadar, Šime Vitasovića 1, 23000, Zadar, Croatia. E-mail: [kjaksic@unizd.hr](mailto:kjaksic@unizd.hr)

2016). One of the most common methods of measuring teaching quality is student evaluation, as students interact with teachers on a daily basis over an extended period of time (Senden et al., 2023). It is assumed that the aggregated students' ratings of teaching quality reflect their shared perceptions or the opinion of the class, which is considered a climate construct (Marsh et al., 2012). Since students are typically nested within classes (or teachers), teaching quality should be modeled as a climate construct using multilevel modeling to control for the interdependence of student ratings (Marsh et al., 2012). However, Wisniewski et al. (2020) note that the multilevel structure of student ratings is generally not considered when assessing the psychometric properties of instructional quality scales. Furthermore, the measures of teaching quality used in schools often lack grounded theory (Wisniewski et al., 2020).

Recently, Wisniewski et al. (2020) developed a new student-report scale to assess teaching quality (teaCh). This scale was developed based on clear theoretical assumptions from educational and cognitive psychology and underwent rigorous psychometric testing. The results of the validation study confirmed the expected seven-factor solution (i.e., Care, Control, Clarity, Consolidation, Conferment, Challenge, Captivation) through multilevel modeling. The model was invariant across subject groups (humanities/language arts, mathematics/science, and social sciences), school types (university preparatory high school, secondary intermediate school, vocational school), and grade levels (5–7, 8–10, 11–12), as well as across students and teachers.

In this paper, we will test the psychometric properties of the teaCh scale (Wisniewski et al., 2020) on a large sample (more than 14,000) of Croatian upper secondary school students to assess the suitability of the scale for measuring teaching quality in the national context.

## **Defining Teaching Quality**

Teaching quality can be considered as a generic construct that applies to different subjects, as a subject-specific construct, or as a global and specific construct (Senden et al., 2021). Regarding the generic approach to operationalizing teaching quality, most models define three dimensions of instruction: Classroom management, Student support, and Cognitive activation (Praetorius et al., 2018). Classroom management refers to the structure and organization of instruction as well as the management of student behavior; Student support means maintaining a positive learning climate, which includes teachers' caring behavior, a positive student-teacher relationship, and additional teacher engagement, such as providing feedback; Cognitive activation means facilitating students' deep processing of content (Dorfner et al., 2018).

Based on the generic approach to the teaching quality, Wisniewski et al. (2020) defined a broader model that defines teaching quality through seven dimensions:

Care, Control, Clarity, Consolidation, Conferment, Challenge, Captivation. The difference between this model and the 3-factor model of teaching quality is that the dimensions of student support and cognitive activation are further elaborated so that each of them includes three sub-dimensions: Captivation, Conferment, and Care as sub-dimensions of Student support as well as Challenge, Clarity and Consolidation as sub-dimensions of Cognitive activation (Wisniewski et al., 2020).

Care and Captivation are based on the assumptions of Taut and Rakoczy (2016) about the social (Care) and organizational (Captivation) aspects of the teacher-student relationship. Teacher care is important for both students and teachers. Students who perceive their teachers as caring report higher self-esteem, well-being, and school engagement, while their teachers report a higher sense of meaning at work (Lavy & Naama-Ghanayim, 2020). Captivation refers to the use of specific instructional features (e.g., interesting lectures) that help students develop self-efficacy (Wisniewski et al., 2020). Conferment is defined by teachers respecting students' perspectives, stimulating discussion, and asking students to contribute (Senden et al., 2021).

In terms of cognitive activation, challenging activities and tasks are related to students' enjoyment of learning, as students who perceive themselves as competent develop intrinsic motivation (Lazarides & Buchholz, 2019). Clarity refers to structural transparency and goal orientation (Wisniewski et al., 2020). Students' perception of the clarity of teaching is related to attitude towards the subject (Chen, 2023) and performance (Yagan, 2021). Consolidation refers to how teachers help students discuss the material and draw conclusions from it (Kersting et al., 2023). It can help to retain knowledge in the short term (Gerstner & Bogner, 2010).

## **Measuring Teaching Quality**

There are different ways (by observers, teachers, or students; Senden et al., 2021) to measure teaching quality, each with its own strengths and limitations. In this paper, the focus is on student self-reports as they are an efficient way to collect information from the whole class in a short period of time. The main assumption of this approach is that students can provide relevant feedback on the teacher's teaching quality as they spend a longer time together on a daily basis (Senden et al., 2023). Students' perceptions correlate with important outcomes such as their self-efficacy (Ruzek et al., 2022; Sánchez-Rosas & Esquivel, 2016), their academic achievement (Nortvedt et al., 2016; Ruzek et al., 2022), task value (Sánchez-Rosas & Esquivel, 2016), and affective attitudes towards the subject (Leino et al., 2022). In general, it can be said that teaching quality is relevant for both the cognitive and affective students' outcomes (Teig & Nilsen, 2022).

Measuring teaching quality using student reports requires appropriate data modeling. Teaching quality is considered a climate construct, i.e. a construct that reflects students' shared opinions about the same teacher (Marsh et al., 2012).

Therefore, it is necessary to separate individual-level variance from classroom-level variance, which is achieved through multilevel modeling (Morin et al., 2014). Class-level ratings reflect teaching quality, whereas individual deviations from the common opinion, while potentially interesting, do not reflect teaching quality *per se*, which is primarily a class-level construct (Marsh et al., 2012). Wisniewski et al. (2020) note that although the separation of individual and classroom variance is crucial for the correct modeling of students' ratings of teaching quality, it has often been ignored in previous research. Furthermore, most of the measures used do not have a clear theoretical background, which further complicates the comparison of results in different contexts (Wisniewski et al., 2020).

### **The Present Study**

Considering the limited number of studies measuring student-reported teaching quality with instruments that have a solid theoretical background, especially in the national context, the aim of the study was to investigate the psychometric properties (i.e. construct, convergent and criterion validity, reliability) of Wisniewski et al. (2020) teaching quality scale (teaCh) in the Croatian upper secondary school system. As the Croatian school system is still undergoing a profound reform process (e.g., transition to single-shift schools and all-day schools; European Commission, 2023), it is important to have valid instruments to track the impact of these reforms and enable teachers to continuously monitor their work. This scale was chosen because it has a clear theoretical background that allows for a comprehensive measurement of teaching quality based on the seven dimensions and because its solid psychometric properties have been demonstrated in previous research (Wisniewski et al., 2020).

In determining the psychometric properties of the teaCh scale, we focused on three research questions:

- Q1. What is the factor structure of the TeaCh scale in the sample of Croatian upper secondary students?
- Q2: How do the dimensions of the teaCh scale relate to the three generic dimensions of teaching quality (i.e., Student support, Cognitive activation, Classroom management) assessed with an alternative scale to demonstrate the convergent validity of the teaCh scale?
- Q3. Considering that teaching quality is associated with self-efficacy (Ruzek et al., 2022; Sánchez-Rosas & Esquivel, 2016), academic achievement (Nortvedt et al., 2016; Ruzek et al., 2022) and task value (Sánchez-Rosas & Esquivel, 2016), do we see the same pattern of results for the teaCh scale in the sample of Croatian upper secondary school students as evidence for the criterion validity of the teaCh scale?

We hypothesized:

- H1: Based on the results of the original validation of the TeaCh scale (Wisniewski et al., 2020), we expected that the seven-factor model (Care, Control, Clarity, Consolidation, Conferment, Challenge, Captivation) would best describe the data compared to alternative models (one-factor model, three-factor model, seven-factor model with one higher-order factor, and seven-factor model with three higher-order factors).
- H2: Based on the theoretical relationships between the seven dimensions of the teaCh scale and the three generic dimensions of teaching quality, we expected that teaCh scale will demonstrate convergent validity – Captivation, Conferment, and Care will be positively correlated with Student support, Challenge, Clarity, and Consolidation will be positively correlated with Cognitive activation, while Control will be positively correlated with Classroom management.
- H3: Considering the previous findings on the relationship between teaching quality, self-efficacy (Ruzek et al., 2022; Sánchez-Rosas & Esquivel, 2016), academic achievement (Nortvedt et al., 2016; Ruzek et al., 2022) and task value (Sánchez-Rosas & Esquivel, 2016), we expected that teaCh scale will demonstrate criterion validity – self-efficacy, academic performance, and task value will be positively correlated with the teaCh dimensions.

## Method

### Participants and Procedure

A total of 14,633 students (8800 female, 5825 male, 8 without a response) from 73 upper secondary schools and 972 classes took part in the survey. The average number of students per class was 15. The average age of the students was  $M = 16.50$ ,  $SD = 1.14$ . The majority of students attended grammar school ( $n = 8489$ ), followed by vocational school ( $n = 6144$ ). The data was collected in March 2022 using an online questionnaire. Students completed the questionnaire under the supervision of school psychologists. Students were randomly assigned to one of their teachers and asked to rate their teaching quality. Each student was assigned to only one teacher and each teacher to only one class. Students rated the teaching quality across a variety of subjects (humanities, social sciences, languages, arts, STEM, vocational subjects). Each school and class was given a unique password that was used to match the data. Prior to data collection, students were informed that their responses would be kept completely confidential, analyzed at the group level, and used for research purposes only. Participation in the study was voluntary and students were informed that they can withdraw from the study at any time. The study was approved by the Croatian Ministry of Science and Education, the Education and Teacher Training Agency, and the Ethics Committee of University of Zadar.

## Measures

### *Teaching Quality – teaCh (Wisniewski et al., 2020)*

The scale consists of 29 items grouped into 7 subscales: Care (3 items, “The teacher met me in a friendly and appreciative way”), Control (6 items, “During the lesson, clear rules were discernible, which the teacher set and enforced”), Conferment (4 items, “The teacher assessed my performance fairly”), Clarity (2 items, “The lesson had a clearly recognizable thread”), Challenge (2 items, “The tasks in the lesson were challenging for me”), Consolidation (4 items, “During the lesson, learning and practice phases alternated”), Captivation (6 items, “The content of the lesson was taught by the teacher in an interesting way”). Students rated their agreement with each statement on a scale from 1 (*completely disagree*) to 5 (*completely agree*). The intermediate values were also labeled (*disagree*, *neither agree nor disagree*, *agree*).

### *Teaching Quality (Burić & Frenzel, 2021; Burić & Kim, 2020)*

The general model of the teaching quality was measured with the items used in previous research on Croatian secondary school teachers (Burić & Frenzel, 2021; Burić & Kim, 2020). The classroom management scale consists of 4 items (“Our teacher makes sure that we pay attention”), Student support is measured with 5 items (“Our teacher is empathetic towards students”), while Cognitive activation is measured with 6 items (“Our teacher wants us to understand the material, not just to memorize it”). Students rated their agreement with each item on a 5-point scale from 1 (*strongly disagree*) to 5 (*strongly agree*). The internal reliability of the dimensions measured by Cronbach’s alpha was satisfactory:  $\alpha_{\text{classroom management}} = .86$ ,  $\alpha_{\text{student support}} = .92$ , and  $\alpha_{\text{cognitive activation}} = .92$ .

### *Motivational Beliefs (Pintrich et al., 1991)*

Motivational beliefs were measured with the items developed by Pintrich et al. (1991). Task value was measured with 6 items (“I think I will be able to use what I learn in this course in other courses”), while self-efficacy was measured with 8 items (“I believe I will receive an excellent grade in this class”). Participants rated their agreement with each item on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*). The internal reliability of the scales measured by Cronbach’s alpha was satisfactory:  $\alpha_{\text{self-efficacy}} = .95$ , and  $\alpha_{\text{task value}} = .93$ .

### *Academic Achievement*

Academic achievement was defined as the final school grade in the subject taught by the teacher being assessed.

## Data Analysis

To account for the hierarchical structure of the data (i.e., students nested within teachers), the factor structure of the Wisniewski et al. (2020) teaching quality scale was tested using a multilevel confirmatory factor analysis (MCFA). This model is a “doubly latent” model, as multiple indicators are used simultaneously to control for measurement error at the student level (L1), and student-level scores are used to control for sampling error as multiple indicators of the construct at the classroom level (L2) (Morin et al., 2014). In line with theory and previous empirical work (Praetorius et al., 2018; Wisniewski et al., 2020), six different models were tested: (1) one-factor model, (2) three-factor model, (3) seven-factor model, (4) seven-factor model with one higher-order factor, and (5) seven-factor model with three higher-order factors. In the one-factor model, we assume that all items reflect a common latent factor. The three-factor model represents a generic approach to teaching quality as defined by Praetorius et al. (2018), which is based on three generic dimensions: Classroom management, Student support, and Cognitive activation. The seven-factor model is the model proposed by Wisniewski et al. (2020) with seven dimensions: Care, Control, Clarity, Consolidation, Conferment, Challenge, Captivation. The seven-factor model with one and three higher-order factors is an extension of the seven-factor model. In the seven-factor model with one higher-order factor, we assume that the variance of the seven factors can be explained by a common factor. The seven-factor model with three higher-order factors assumes that the common variance can be explained by three higher-order factors that correspond to three generic dimensions of teaching quality: Classroom management (Control), Student support (Captivation, Conferment, Care), Cognitive activation (Challenge, Clarity, Consolidation).

Since teaching quality can be viewed as a climate construct that allows for aggregation of student ratings at the class level (L2) (Marsh, 2012), intraclass correlations (ICC1, ICC2) were calculated. ICC1 (also defined as ICC (1,1); Shrout & Fleiss, 1979) reflects how much variance in the items is attributable to the cluster variable (class membership in this case), and values above .06 (Chiu et al., 2015) are considered appropriate for the multilevel analyses. On the other hand, ICC2 (also defined as ICC [1, k]; Shrout & Fleiss, 1979) shows how reliable the estimates at class level (L2) are and can be interpreted similarly to Omega ( $\Omega$ ), which was also calculated as a reliability measure. A threshold of .70 is considered appropriate for both ICC2 and Omega (Morin et al., 2014). The full information maximum likelihood algorithm (FIML; Enders & Bandalos, 2001) was used to handle the missing data (the percentage of missing data ranged from 0.1% to 1.1% per variable).

To examine convergent validity, we calculated bivariate correlations between the Wisniewski et al.'s (2020) seven dimensions of teaching quality (i.e., care, control, conferment, clarity, challenge, consolidation, captivation) and three generic dimensions of teaching quality (i.e., classroom management, student support,

cognitive activation) measured with scales previously used on Croatian student samples (Burić & Frenzel, 2021; Burić & Kim, 2020). Finally, to examine the criterion validity of the Wisniewski et al. (2020) teaching quality scale, we calculated bivariate correlations between its seven dimensions and students' motivational beliefs (i.e., self-efficacy, task value) and academic achievement (the percentage of missing data for self-efficacy, task value, and academic achievement ranged from 0.5% to 3% per variable). The bivariate correlations were calculated at student and class level (L2) to account for the hierarchical structure of the data.

All analyses were conducted in Mplus 8.10 (Muthén & Muthén, 1998-2017) using the robust maximum likelihood estimator (MLR<sup>1</sup>). To examine the model fit for the MCFA, we used the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR). CFI and TLI values above .95 indicate a good fit, respectively, while RMSEA below .06 and SRMR below .08 indicate an adequate fit (Hu & Bentler, 1999).

## Results

### Descriptive Statistics

The descriptive statistics for the items can be found in Appendix 1. The item-level scores were approximately normally distributed and their skewness and kurtosis values were less than |2| (George & Mallery, 2010). The ICC1 values were above the threshold value of .06 (Chiu et al., 2015), justifying the use of multilevel analyses.

The descriptive data and the reliability coefficients (ICC2,  $\Omega$ )<sup>2</sup> for the seven factors are shown in Table 1. Students reported relatively high and similar estimates of teaching quality across all dimensions, demonstrating low dimension separability. Some of the absolute estimates were to some extent higher than in the research by Wisniewski et al. (2020).

The reliability of all factors at class level (L2), with the exception of Challenge, was satisfactory (ICC2,  $\Omega$  greater than .70; Morin et al., 2014). The lower reliability of the Challenge dimension can be explained by lower factor loading (.25) for one

<sup>1</sup> Given the discussion on the appropriateness of maximum likelihood estimation for ordinal data (Rhemtulla et al., 2012), we present the results for the diagonally weighted least squares estimation (WLSMV) in Appendix 3. The general conclusion is the same as for maximum likelihood estimation, although the model fit estimates are somewhat lower.

<sup>2</sup> The omega estimates are based on the unstandardized loadings and variances of the 7-factor model. Although the model with the higher order factor fits the data best, we still find it important to know the reliability of the individual factors. The general factor serves here primarily as evidence of the high intercorrelations between the factors.



item (i.e., “The tasks in the lesson were challenging for me”) and the fact that the scale consists of only two items. The factor loadings of all other items at the class level (L2) as well as the loadings on the global factor were satisfactory ( $> .70$ ; Cheung et al., 2024) and, in general, the factor loadings at the class level were greater than those at the student level (L1), as shown in Appendix 2.

**Table 1**

*Descriptive Data and Reliability Coefficients for the Seven Factors*

	<i>M</i>	<i>SD</i>	<i>SD<sub>w</sub></i>	<i>SD<sub>b</sub></i>	ICC1	ICC2	$\Omega_B$	$\Omega_W$
Care	3.85	.89	.74	.50	.32	.88	.97	.75
Control	3.79	.77	.67	.39	.25	.84	.97	.83
Conferment	3.91	.86	.74	.44	.26	.85	.99	.84
Clarity	3.88	.80	.69	.41	.26	.84	.98	.83
Consolidation	3.79	.82	.70	.42	.27	.85	.98	.82
Captivation	3.70	.82	.70	.44	.28	.86	.99	.86
Challenge	3.47	.80	.76	.23	.09	.57	.54	.48

*Note.* *SD<sub>b</sub>* – standard deviation at the class level (L2); *SD<sub>w</sub>* – standard deviation at the student level (L1);  $\Omega_B$  – reliability at the class level (L2);  $\Omega_W$  – reliability at the student level (L1).

## Factor Structure of the teaCh Scale

The fit indices of the tested models are given in Table 2. The estimation of the seven-factor model and the seven-factor model with three higher-order factors was followed by the warning of a non-positive definite PSI matrix. This could be due to a high dependency (correlations) between two or more variables (as shown in Table 3), resulting in at least one variable being expressed as a linear combination of the other variables. The seven-factor model with one higher-order factor resulted in a

**Table 2**

*Model Fit Statistics for the Maximum-Likelihood (MLR) Estimation*

	$\chi^2$ (df)	CFI	TLI	RMSEA	SRMR <sub>B</sub>	SRMR <sub>W</sub>	AIC	$\Delta$ AIC
1-factor model	22372.977 (754)	.892	.883	.044	.065	.039	905935.097	5432.563
3-factor model	20267.585 (749)	.902	.894	.042	.056	.038	903207.756	2705.222
7-factor model <sup>a</sup>	/	/	/	/	/	/	/	/
7-factor model with 1 higher order factor	18242.324 (741)	.912	.904	.040	.055	.036	900502.534	
7-factor model with 3 higher order factors <sup>a</sup>	/	/	/	/	/	/	/	

*Note.* <sup>a</sup> Estimation of the 7 – factor model and 7 – factor model with 3 higher order factors resulted in the warning about non-positive definite PSI matrix;  $\Delta$ AIC, in comparison to the 7 – factor model with 1 higher order factor.

warning-free estimate<sup>3</sup>, and had a better fit than the one-factor and three-factor model as described by  $\Delta AIC > 2$  (Fabozzi et al., 2014). Therefore, this model was accepted as the best possible of the models tested, albeit the fit itself is not entirely satisfactory. The selected model is shown in Figure 1.

**Table 3**

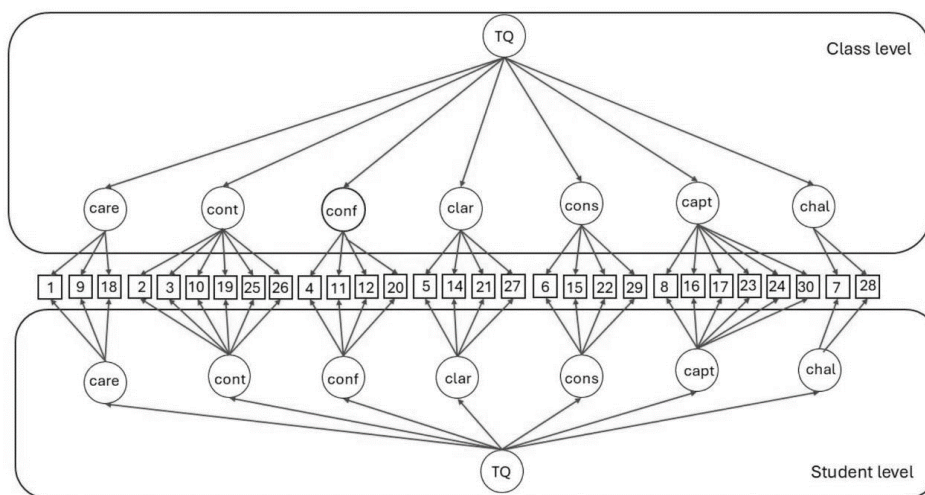
*Correlations Between Seven Factors*

	1.	2.	3.	4.	5.	6.	7.
1. Care	-	.72	.78	.74	.74	.76	.38
2. Control	.82	-	.74	.76	.75	.73	.42
3. Conferment	.94	.92	-	.76	.76	.74	.38
4. Clarity	.91	.95	.95	-	.82	.80	.43
5. Consolidation	.93	.91	.95	.96	-	.83	.44
6. Captivation	.94	.88	.93	.95	.96	-	.46
7. Challenge	.56	.63	.60	.63	.66	.62	-

*Notes.* All correlations are significant at  $p < .01$ ; Correlations at the student level (L1) are shown above diagonal, while correlations at the class level (L2) are shown below the diagonal.

**Figure 1**

*Model With the Best Fit – Seven-Factor Model With One Higher-Order Factor*



*Note.* TQ – teaching quality; cont – control; conf – conferment; clar – clarity; cons – consolidation; capt – captivation; chal – challenge.

<sup>3</sup> There was a small and nonsignificant negative residual correlation at the within level for Consolidation which was set to 0 (Hox, 2010).

## Convergent Validity

As shown in Table 4, there was a strong positive correlation between all seven factors of the Wisniewski et al. (2020) teaching quality scale and all three factors of the corresponding teaching quality scale (Burić & Frenzel, 2021; Burić & Kim, 2020) at L2. The correlations ranged from .58 (Challenge – Student support) to .97 (Care – Student support; Clarity – Cognitive activation) and were generally slightly lower for the Challenge factor. In terms of absolute values, the teaCh scale factors tended to show the strongest correlation with the corresponding dimension of the teaching quality scale (exceptions were the correlation between Conferment and Captivation with Student support and the correlation between Challenge and Cognitive activation). However, the correlations between the corresponding and non-corresponding dimensions were also high, which calls into question the separability of the teaCh’s factors.

In general, the correlations at the student level were in the same direction as those at the class level (L2) although they were less pronounced. It is important to note that the results at the student level (L1) do not necessarily reflect the teaching quality. Due to the nature of the scale, the ratings at this level primarily capture inter-individual differences between students in the same class in terms of perceived teaching quality for items relating to the individual student (e.g., “The teacher treated me in a friendly and appreciative way”). For items relating to the whole class (e.g., “During the lesson, clear rules were set and enforced by the teacher”), the variability in student responses is likely to reflect measurement error, as responses should be relatively consistent across all students (Marsh et al., 2012; Morin et al., 2014).

**Table 4**

*The Student Level (L1) and the Class Level (L2) Correlations Between Seven Factors and Criteria*

	Cognitive activation	Student support	Classroom management	Last grade	Self-efficacy	Task value
<i>Student level</i>						
Care	.73	.73	.66	.21	.50	.49
Control	.71	.64	.77	.15	.43	.46
Conferment	.72	.67	.67	.22	.48	.47
Clarity	.77	.67	.70	.19	.50	.52
Consolidation	.77	.69	.70	.20	.51	.52
Captivation	.77	.72	.71	.24	.58	.60
Challenge	.44	.40	.42	.13	.28	.31
<i>Class level</i>						
Care	.94	.97	.86	.48	.80	.74
Control	.91	.81	.96	.39	.71	.72
Conferment	.95	.91	.91	.47	.77	.75
Clarity	.97	.89	.94	.44	.78	.78
Consolidation	.95	.91	.92	.46	.78	.79
Captivation	.96	.93	.93	.52	.85	.83
Challenge	.66	.58	.66	.23	.39	.51

*Note.* All the correlations are significant at  $p < .01$ .

### Criterion Validity

In terms of criterion validity, as shown in Table 4, all seven factors at the class level (L2) were positively related to academic achievement, with correlations ranging from .23 (Challenge) to .52 (Captivation). In addition, all seven factors were positively correlated with self-efficacy beliefs, with correlations ranging from .39 (Challenge) to .85 (Captivation). Finally, all seven factors were positively correlated with task value, with correlations ranging from .51 (Challenge) to .83 (Captivation). In general, the correlations with the criteria are slightly lower for Challenge compared to the other dimensions and generally lower for academic achievement (last grade) compared to motivational beliefs.

Correlations at the student level (L1) were generally in the same direction as those at the class level (L2), but were of a smaller magnitude.

### Discussion

The aim of the study was to test the psychometric properties of the Wisniewski et al.'s (2020) teaching quality scale (teaCh) in the context of Croatian upper secondary education. In general, the seven-factor structure proposed by Wisniewski et al. (2020) was not confirmed and H1 was rejected. The seven-factor model could not be estimated and the seven-factor model with a higher-order factor showed the best fit, indicating a strong correlation between the seven dimensions at L2, which calls into question the separability of the dimensions. This finding is confirmed by the generally strong correlations between all teaCh factors and the three generic dimensions of teaching quality measured by the alternative scale, which partially refutes H2. Finally, the psychometric properties of the *Challenge* factor are unsatisfactory.

Regarding the intercorrelations of the factors, it is important to note that in the original study by Wisniewski et al. (2020), the latent intercorrelations were also high, ranging from .40 - .91. This could be related to students' competence in evaluating teachers' work and the separability of the dimensions of teaching quality, which has been the subject of debate (van der Scheer et al., 2019). Fauth et al. (2014) have shown that there is a positive correlation between students' perception of teaching quality and teachers' popularity. According to them, this result at the student level (L1) can be explained by the emotional relationship between students and teachers, while at the classroom level, teachers with high teaching quality could be more popular. In this sense, high intercorrelations could represent the existence of the teacher popularity effect. However, it may not necessarily be the case that teachers are popular. This effect may occur due to the socially desirable response. It has been found that students tend to give socially desirable responses due to the emotional relationship with the teacher (Maulana & Helms-Lorenz, 2016), which could explain

why students rated the teaching quality relatively high, somewhat higher than in the original study. Furthermore, Wagner et al. (2013) explained high intercorrelations between five dimensions of teaching quality in their study (i.e., motivation, comprehensibility, student involvement, structure, classroom management) with certain methodological factors that can also be applied to the present study. Namely, it is possible that the items are too abstract for the students, i.e., they require the students to evaluate teacher behavior that is not directly observed by the students. Due to the uncertainty, students adjust their ratings to a more general impression (Roch et al., 2009).

According to Wagner et al. (2013), the wording of the items can also play a role. Some items on the scale refer to the teacher's behavior towards the students themselves, while others refer to the teacher's behavior towards the entire class, which can lead to different cognitive processes in the students when evaluating the statements. Despite the high intercorrelations between the factors, it is important to note that the factor loadings at the class level (L2) were high and larger than their counterparts at the student level (L1). This means that each factor still accounts for a significant proportion of individual variance (Wisniewski et al., 2020). This is also confirmed by the significant positive correlations with academic achievement (final grade) and motivational beliefs (self-efficacy, task value), which is consistent with H3. This is in line with previous findings on the relationship between teaching quality and student outcomes as well as theoretical views on their relationship (Lazarides & Buchholz, 2019; Nortvedt et al., 2016; Senden et al., 2021; Taut & Rakoczy, 2016; Wisniewski et al., 2020). In this context, it is important to consider an issue related to the content of the *Challenge* factor. The factor loading for one of the items of this factor (i.e., "The tasks in the lesson were challenging for me") was not satisfactorily high. Consequently, the factor had the lowest convergent and criterion validity. We suspect that this could be due to the wording of the item. It is possible that this item has a negative connotation for some students. That is, they interpreted the question as asking them whether teachers were trying to put them under extra pressure, rather than whether teachers were trying to improve their knowledge by giving them tasks that required additional effort. However, this is speculative and needs to be investigated further by developing new item(s) and testing their psychometric properties. Additionally, the factor has only two items, which limits its reliability even if both items were satisfactorily loaded.

## **Limitations and Future Directions**

When interpreting the results of the tested models of teaching quality, it is generally important to be aware of possible cultural differences. Although there is some evidence of cross-cultural stability of models of teaching quality (Scherer et al., 2016), there is also evidence of limited cross-national comparability (Bellens et al., 2019; Senden et al., 2023). Moreover, Senden et al. (2021) pointed out that different perspectives (e.g., teachers, students, observers) should be used to assess

teaching quality, as each of them provides a unique perspective. Therefore, student ratings should be one of the measures used to assess teaching quality (given their subjectivity, e.g., due to their relationship with the teacher or their interest in the subject), and multiple measures (perspectives) should be used over multiple time periods to evaluate and support teachers (Ferguson, 2012). In addition, all measures are based on self-report, which potentially inflates intercorrelations due to common-method variance (Podsakoff et al., 2003). Moreover, although the data was collected from a diverse and large sample of students, grammar school students were overrepresented in the sample. In Croatia, approximately 30% of students attend grammar school (Matković & Šabić, 2022), while 60% of the sample consisted of grammar school students. Finally, given that the sample consisted of students from grammar and vocational schools, it would be beneficial to test the measurement invariance between these groups of students to see if school type is a relevant factor in measuring teaching quality.

## **Conclusion**

Using a large sample of upper secondary school students from across the country, this study is among the few studies that tested the psychometric properties of the teaching quality scale using multilevel modeling. The results of the study are not fully consistent with the original results of the seven-factor model. High intercorrelations between the factors and insufficient psychometric properties of the *Challenge* factor require additional work on the scale content. However, the finding that the factors are meaningfully related to the relevant criteria such as student motivation and academic achievement means that the scale could be useful as one of the tools for teachers to obtain feedback from students. This information could still have a practical value.

## **Declaration of Competing Interest**

Both authors declare that they have no conflict of interests.

## **Data Availability Statement**

The data is available on request. The data will be published at the end of the embargo period specified in the project's data storage protocol.

## References

- Bellens, K., Van Damme, J., Van Den Noortgate, W., Wendt, H., & Nilsen, T. (2019). Teaching 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, Article 1. <https://doi.org/10.1186/s40536-019-0069-2>
- Burić, I., & Frenzel, A. C. (2021). Teacher emotional labour, instructional strategies, and students' academic engagement: A multilevel analysis. *Teachers and Teaching: Theory and Practice*, 27, 335–352. <https://doi.org/10.1080/13540602.2020.1740194>
- Burić, I. & Kim, L. E. (2020). Teacher self-efficacy, instructional quality, and student motivational beliefs: An analysis using multilevel structural equation modeling. *Learning and Instruction*, 66, Article 101302. <https://doi.org/10.1016/j.learninstruc.2019.101302>
- Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pacific Journal of Management*, 41, 745-783. <https://doi.org/10.1007/s10490-023-09871-y>
- Chiu, C.-Y., Chia, S. I., & Wan, W. W. N. (2015). Measures of cross-cultural values, personality and beliefs. In G. J. Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of personality and social psychological constructs* (pp. 621–651). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-386915-9.00022-X>
- Dorfner, T., Förtsch, C., & Neuhaus, B. J. (2018). Effects of three basic dimensions of teaching quality on students' situational interest in sixth-grade biology instruction. *Learning and Instruction*, 56, 42–53. <https://doi.org/10.1016/j.learninstruc.2018.03.001>
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 8, 430–457. [https://doi.org/10.1207/S15328007SEM0803\\_5](https://doi.org/10.1207/S15328007SEM0803_5)
- European Commission (2023). *Education and training monitor 2023: Croatia*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2766/775582>
- Fabozzi, F. J., Focardi, S. M., Rachev, S. T., Arshanapalli, B. G., & Hoechstetter, M. (2014). *The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications*. Wiley.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1–9. <https://doi.org/10.1016/j.learninstruc.2013.07.001>
- Ferguson, R. (2012). Can student surveys measure teaching quality? *The Phi Delta Kappan*, 94, 24–28. <https://doi.org/10.2307/41763671>
- George, D. & Mallery, M. (2010). *SPSS for windows step by step: A simple guide and reference, 17.0 update*. Pearson.
- Gerstner, S., & Bogner, F. X. (2010). Cognitive achievement and motivation in hands-on and teacher-centred science classes: Does an additional hands-on consolidation phase (concept mapping) optimise cognitive learning at work stations? *International Journal of Science Education*, 32(7), 849–870. <https://doi.org/10.1080/09500690902803604>

- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Routledge/Taylor & Francis Group.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. <https://doi.org/10.1080/10705519909540118>
- Kersting, M., Karlsen, S., Ødegaard, M., Olufsen, M., Kjærnsli, M., & Lunde, M. L. (2023). Studying the quality of inquiry-based teaching in science classrooms. A systematic video study of inquiry-based science teaching in primary and lower-secondary schools. *International Journal of Science Education*, 45(17), 1463–1484. <https://doi.org/10.1080/09500693.2023.2213386>
- Lavy, S., & Naama-Ghanayim, E. (2020). Why care about caring? Linking teachers' caring and sense of meaning at work with students' self-esteem, well-being, and school engagement. *Teaching and Teacher Education*, 91, Article 103046. <https://doi.org/10.1016/j.tate.2020.103046>
- Lazarides, R., & Buchholz, J. (2019). Student-perceived teaching quality: How is it related to different achievement emotions in mathematics classrooms? *Learning and Instruction*, 61, 45–59. <https://doi.org/10.1016/j.learninstruc.2019.01.001>
- Leino, K., Nissinen, K., & Sirén, M. (2022). Associations between teacher quality, teaching quality and student reading outcomes in Nordic PIRLS 2016 data. *Large-Scale Assessments in Education*, 10, Article 25. <https://doi.org/10.1186/s40536-022-00146-4>
- Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, 47, 106–124. <https://doi.org/10.1080/00461520.2012.670488>
- Matković, T. & Šabić, J. (2022, September). *Studija o stanju u području srednjoškolskog obrazovanja u svrhu povećanja obuhvata učenika gimnazijskim programima i smanjivanja broja suficitarnih programa u strukovnome obrazovanju* [Study on the situation in secondary education to increase the number of students in grammar schools and reduce the number of deficient programs in vocational education]. <https://mzom.gov.hr/UserDocsImages/dokumenti/Obrazovanje/NacionalniPlan2021-2027/Studija-srednje-obrazovanje.pdf>
- Maulana, R., & Helms-Lorenz, M. (2016). Observations and student perceptions of the quality of preservice teachers' teaching behaviour: Construct representation and predictive quality. *Learning Environments Research*, 19, 335–357. <https://doi.org/10.1007/s10984-016-9215-8>
- Morin, A. J. S., Marsh, H. W., Nagengast, B., & Scalas, L. F. (2014). Doubly latent multilevel analyses of classroom climate: An illustration. *Journal of Experimental Education*, 82, 143–167. <https://doi.org/10.1080/00220973.2013.769412>
- Muthén, L. K. & Muthén, B. O. (1998-2017). *Mplus user's guide (8<sup>th</sup> edition)*. Muthén & Muthén.
- Nilsen, T., Gustafsson, J. E., & Blömeke, S. (2016). Conceptual framework and methodology of this report. In T. Nilsen & J. E. Gustafsson (Eds.), *Teacher quality, teaching quality and student outcomes* (pp. 1–21). Springer. [https://doi.org/10.1007/978-3-319-41252-8\\_5](https://doi.org/10.1007/978-3-319-41252-8_5)



- Nortvedt, G. A., Gustafsson, J. E., & Lehre, A. C. W. (2016). The importance of teaching quality for the relation between achievement in reading and mathematics. In T. Nilsen & J. E. Gustafsson (Eds.), *Teacher quality, teaching quality and student outcomes* (pp. 97–113). Springer. [https://doi.org/10.1007/978-3-319-41252-8\\_5](https://doi.org/10.1007/978-3-319-41252-8_5)
- Pintrich, P., Smith, D., García, T., & McKeachie, W. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. Ann Arbor.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903. <https://psycnet.apa.org/doi/10.1037/0021-9010.88.5.879>
- Practorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM – Mathematics Education*, 50, 407–426. <https://doi.org/10.1007/s11858-018-0918-4>
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17, 354–373. <https://doi.org/10.1037/a0029315>
- 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>
- Sánchez-Rosas, J., & Esquivel, S. (2016). Instructional teaching quality, task value, self-efficacy, and boredom: A model of attention in class. *Revista de Psicología*, 25, 1–20. <http://dx.doi.org/10.5354/0719-0581.2017.44966>
- Senden, B., Nilsen, T., & Blömeke, S. (2021). Teaching quality: A review of conceptualizations, measurement approaches, and research findings. In M. Blikstad-Balas, K., Klette & M. Tengberg (Eds.), *Ways of analyzing teaching quality. Potentials and pitfalls* (pp. 140–172). Scandinavian University Press. <https://doi.org/10.18261/9788215045054-2021-05>
- Senden, B., Nilsen, T., & Teig, N. (2023). The validity of student ratings of teaching quality: Factorial structure, comparability, and the relation to achievement. *Studies in Educational Evaluation*, 78, Article 101274. <https://doi.org/10.1016/j.stueduc.2023.101274>
- Scherer, R., & Nilsen, T. (2016). The relations among school climate, teaching quality, and achievement motivation in mathematics. In Nilsen, T., Gustafsson, J. E. (Eds.), *Teacher quality, teaching quality and student outcomes* (pp. 51–81). Springer. [https://doi.org/10.1007/978-3-319-41252-8\\_5](https://doi.org/10.1007/978-3-319-41252-8_5)
- Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating individual students' perceptions of teaching quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology*, 7, Article 110. <https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00110>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>

- Taut, S., & Rakoczy, K. (2016). Observing teaching quality in the context of school evaluation. *Learning and Instruction*, 46, 45–60. <https://doi.org/10.1016/j.learninstruc.2016.08.003>
- Teig, N., & Nilsen, T. (2022). Profiles of instructional quality in primary and secondary education: Patterns, predictors, and relations to student achievement and motivation in science. *Studies in Educational Evaluation*, 74, Article 101170. <https://doi.org/10.1016/j.stueduc.2022.101170>
- van der Scheer, E. A., Bijlsma, H. J. E., & Glas, C. A. W. (2019). Validity and reliability of student perceptions of teaching quality in primary education. *School Effectiveness and School Improvement*, 30, 30–50. <https://doi.org/10.1080/09243453.2018.1539015>
- Wagner, W., Göllner, R., Helmke, A., Trautwein, U., & Lüdtke, O. (2013). Construct validity of student perceptions of teaching 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>
- Wisniewski, B., Zierer, K., Dresel, M., & Daumiller, M. (2020). Obtaining secondary students' perceptions of teaching quality: Two-level structure and measurement invariance. *Learning and Instruction*, 66, Article 101303. <https://doi.org/10.1016/j.learninstruc.2020.101303>

## **Učeničke percepcije kvalitete poučavanja: Validacija ljestvice teaCh na uzorku učenika hrvatskih srednjih škola**

### **Sažetak**

Kvaliteta poučavanja važan je konstrukt povezan s afektivnim i kognitivnim ishodima učenika. Međutim, obično se mjeri instrumentima kojima nedostaje jasna teorijska podloga i/ili čija psihometrijska svojstva nisu testirana odgovarajućim statističkim postupcima poput višerazinskoga modeliranja. Nedavno su Wisniewski i suradnici (2020) razvili ljestvicu teaCh koja mjeri kvalitetu poučavanja na temelju percepcije učenika o sedam dimenzija: brizi, kontroli, jasnoći, konsolidaciji, povjerenju, izazovu te zaokupljenosti. Cilj je istraživanja bio dodatno testirati psihometrijska svojstva te ljestvice koristeći velik uzorak učenika srednjih škola u Republici Hrvatskoj. Rezultati istraživanja pokazuju da sedam faktora ima odgovarajuću prediktivnu valjanost za vrijednost zadatka, samoučinkovitost učenika i završni razred, kao i odgovarajuću konvergentnu valjanost i pouzdanost. Međutim, postoje visoke međusobne korelacije među dimenzijama. Rezultati potvrđuju da je mjerenje kvalitete poučavanja pomoću učeničkih izvješća sa psihometrijskoga gledišta izazovno, ali da učeničke percepcije i dalje mogu pružiti korisne povratne informacije nastavnicima.

*Ključne riječi:* kvaliteta poučavanja, procjene učenika, učenici srednjih škola, validacija

Received: 10. 6. 2025.

### Descriptive Statistics of the Items

Original item	Croatian translation	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	ICC1
<i>Care</i>						
The teacher met me in a friendly and appreciative way.	Profesor/ica mi pristupa na prijateljski i uvažavajući način.	3.98	1.01	-0.99	0.65	.25
The teacher created an atmosphere free of fear.	Profesor/ica stvara atmosferu u kojoj nema straha.	3.83	1.10	-0.80	0.00	.31
The teacher was interested in whether I really learned something.	Profesora/icu zanima jesam li uistinu naučio/la nešto.	3.73	1.01	-0.66	0.12	.19
<i>Control</i>						
During the lesson, clear rules were discernible, which the teacher set and enforced.	Tijekom nastave profesor/ica postavlja i primjenjuje pravila koja su jasna i poznata.	3.99	0.94	-0.89	0.64	.22
The teacher did not waste time due to delays or idling.	Profesor/ica ne gubi vrijeme zbog kašnjenja i praznog hoda.	3.74	1.16	-0.75	-0.21	.14
The teacher provided a trouble-free working atmosphere.	Profesor/ica stvara atmosferu u kojoj nema ometanja.	3.77	0.99	-0.65	0.11	.18
The teacher had a good overview of what was happening in the class.	Profesor/ica ima dobar pregled onoga što se događa u razredu.	3.77	0.99	-0.67	0.12	.20
When students violated the rules, the teacher intervened quickly and consistently.	Kada učenici krše razredna pravila, profesor/ica reagira brzo i dosljedno.	3.83	0.95	-0.71	0.35	.13
The course of instruction was smooth.	Poučavanje se odvija bez prekidanja.	3.64	0.98	-0.48	-0.09	.16

Original item	Croatian translation	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	ICC1
<i>Conferment</i>						
The teacher assessed my performance fairly.	Profesor/ica pravedno ocjenjuje moje uratke (radove, ispite).	4.06	0.99	-1.05	0.76	.21
The teacher gave me helpful feedback on my performance.	Profesor/ica mi daje korisnu povratnu informaciju o mojoj izvedbi (uratku).	3.87	1.00	-0.81	0.34	.20
The teacher was fair and unbiased towards me and my classmates.	Profesor/ica je pravedan i nepristran prema meni i ostalim učenicima u razredu.	3.88	1.06	-0.87	0.29	.18
The teacher gave me meaningful feedback on my contributions.	Profesor/ica mi daje smislenu povratnu informaciju o mom radu.	3.83	0.98	-0.74	0.28	.21
<i>Clarity</i>						
The lesson had a clearly recognizable thread.	Nastavni sat ima jasno prepoznatljiv tijek.	3.95	0.96	-0.88	0.57	.20
The teacher showed me what the new content is related to.	Profesor/ica mi pokazuje sa čime je povezano novo gradivo.	3.89	0.93	-0.76	0.49	.19
The teacher showed me what I could use the new content for.	Profesor/ica mi pokazuje za što sve mogu koristiti novo gradivo.	3.73	1.00	-0.62	0.03	.20
The teacher has tied in content that was already known to me.	Profesor/ica povezuje novo gradivo s onime što mi je poznato otprije.	3.94	0.91	-0.83	0.68	.18
<i>Consolidation</i>						
During the lesson, learning and practice phases alternated.	Tijekom nastavnog sata učenje i vježba se izmjenjuju.	3.86	0.97	-0.75	0.27	.20
During the lesson, the teacher showed me exactly how I could solve certain tasks.	Tijekom nastavnog sata profesor/ica mi točno pokazuje kako mogu riješiti određene zadatke.	3.83	0.98	-0.71	0.21	.21
I had enough time to concentrate on the content of the lesson.	Imam dovoljno vremena da se koncentriram na sadržaj nastavnog sata.	3.74	0.99	-0.66	0.12	.19

Jakšić, K., Burić, I.:  
*Validation of the TeaCh Scale*

<b>Original item</b>	<b>Croatian translation</b>	<b><i>M</i></b>	<b><i>SD</i></b>	<b>Skewness</b>	<b>Kurtosis</b>	<b>ICC1</b>
During the lesson there were plenty of opportunities to practice the new content.	Tijekom nastavnog sata postoji mnogo prilika za uvježbavanje novog gradiva.	3.72	0.96	-0.57	0.08	.20
<i>Captivation</i>						
The content of the lesson was taught by the teacher in an interesting way.	Profesor/ica poučava gradivo na zanimljiv način.	3.69	1.11	-0.63	-0.24	.29
The course of the lesson was varied.	Tijek nastavnog sata je dinamičan i promjenjiv.	3.53	1.03	-0.36	-0.31	.17
I was able to see personal learning progress through the lessons.	Mogu vidjeti svoj osobni napredak u učenju kroz nastavne sate.	3.63	1.05	-0.53	-0.21	.20
The requirement level in the lesson was appropriate for me.	Zahtjevi pojedinog nastavnog sata su prikladni za mene.	3.82	0.94	-0.70	0.39	.18
The learning pace in the class was appropriate for me.	Tempo učenja u razredu je prikladan za mene.	3.85	0.97	-0.76	0.36	.19
During the lesson I was able to apply strategies that are also useful for other problems/topics/areas.	Tijekom nastavnog sata mogu primjenjivati strategije koje su korisne i u drugim područjima i temama.	3.66	0.99	-0.53	-0.02	.18
<i>Challenge</i>						
The tasks in the lesson were challenging for me.	Profesor/ica mi zadaje zadatke koji su izazovni.	3.47	0.98	-0.27	-0.18	.09
The teacher had high expectations of me.	Profesor/ica ima visoka očekivanja u vezi mog uspjeha.	3.47	1.00	-0.25	-0.21	.07

**Factor Loadings**

<b>Item</b>	<b><math>\Lambda_w</math></b>	<b><math>\Lambda_b</math></b>
<i>Care</i>		
The teacher met me in a friendly and appreciative way	.68	.99
The teacher created an atmosphere free of fear	.68	.94
The teacher was interested in whether I really learned something	.77	.96
<i>Control</i>		
During the lesson, clear rules were discernible, which the teacher set and enforced	.71	.98
The teacher did not waste time due to delays or idling	.48	.80
The teacher provided a trouble-free working atmosphere	.74	.95
The teacher had a good overview of what was happening in the class	.77	.96
When students violated the rules, the teacher intervened quickly and consistently	.70	.91
The course of instruction was smooth	.68	.91
<i>Conferment</i>		
The teacher assessed my performance fairly	.70	.94
The teacher gave me helpful feedback on my performance	.81	.99
The teacher was fair and unbiased towards me and my classmates	.68	.96
The teacher gave me meaningful feedback on my contributions	.83	.99
<i>Clarity</i>		
The lesson had a clearly recognizable thread	.67	.92
The teacher showed me what the new content is related to	.77	.99
The teacher showed me what I could use the new content for	.78	.99
The teacher has tied in content that was already known to me	.76	.98
<i>Consolidation</i>		
During the lesson, learning and practice phases alternated	.67	.91
During the lesson, the teacher showed me exactly how I could solve certain tasks	.77	.98
I had enough time to concentrate on the content of the lesson	.73	.97
During the lesson there were plenty of opportunities to practice the new content	.73	.97
<i>Captivation</i>		
The content of the lesson was taught by the teacher in an interesting way	.76	.96
The course of the lesson was varied	.62	.89
I was able to see personal learning progress through the lessons	.76	.98
The requirement level in the lesson was appropriate for me	.73	.98

Jakšić, K., Burić, I.:  
*Validation of the TeaCh Scale*

<b>Item</b>	<b><math>\Lambda_w</math></b>	<b><math>\Lambda_b</math></b>
The learning pace in the class was appropriate for me	.70	.96
During the lesson I was able to apply strategies that are also useful for other problems/topics/areas	.73	.97
<i>Challenge</i>		
The tasks in the lesson were challenging for me	.47	.25
The teacher had high expectations of me	.65	.95
<i>TQ</i>		
Care	.98	.97
Control	.93	.96
Conferment	.94	.98
Clarity	.98	.99
Consolidation	1.00	.99
Captivation	.97	.98
Challenge	.71	.81

*Note.*  $\Lambda_w$  – factor loading at the student level (L1);  $\Lambda_b$  – factor loading at the class level (L2).

**Model Fit Statistics for the Diagonally Weighted  
Least Squares Estimation (WLSMV)**

	$\chi^2$ (df)	CFI	TLI	RMSEA	SRMR <sub>B</sub>	SRMR <sub>W</sub>
1-factor model	57986.705 (754)	.844	.832	.072	.041	.038
3-factor model <sup>a</sup>	53287.305 (749)	.857	.845	.069	.039	.035
7-factor model <sup>b</sup>	/	/	/	/	/	/
7-factor model with 1 higher order factor <sup>c</sup>	48270.954 (742)	.871	.859	.066	.037	.035
7-factor model with 3 higher order factors <sup>d</sup>	/	/	/	/	/	/

*Note.* <sup>a</sup> Due to the small negative residual variance, variance of one item (second) at the class level (L2) was set to 0. <sup>b</sup> Estimation of the model resulted with the warning about non-positive definite PSI matrix; <sup>c</sup> Due to the small negative residual variance, variance of one item (second) at the class level (L2) and one factor (consolidation) at the student level (L1) was set to 0. <sup>d</sup> Model was not identified.