Validation of Scales for Measuring Factors of Teaching Quality from the Dynamic Model of Educational Effectiveness[*]

Bojana Bodroža[1], Jelena Teodorović[2], and Smiljana Jošić[3]

[1]Department of Psychology, Faculty of Philosophy, University of Novi Sad, Serbia
[2]Faculty of Education, University of Kragujevac, Serbia
[3]Institute for Educational Research, Belgrade, Serbia

Large-scale educational effectiveness research requires valid student questionnaires to assess teaching practices. This research validated eight scales for measuring teaching factors from the Dynamic Model of Educational Effectiveness (DMEE). Parallel versions of scales for measuring teaching factors in mathematics and biology were constructed and validated in two studies. In the first study, an exploratory factor analysis was conducted on data from 683 students. In the second study, the structure was cross-validated via a confirmatory factor analysis (CFA) on a sample of 5,476 students. The multi-group CFA resulted in an acceptable metric invarience for all scales, indicating that the scales have comparable factor loadings. However, unsatisfactory scalar invariance suggested that the scales could not be used to compare teachers of different subjects. Testing alternative structural relations between the teaching factors did not confirm that the data fit the DMEE model adequately, although the fit parameters were better than for the alternative theoretical models. For mathematics, the external validation of the scales showed that the scales correlated with job satisfaction, external control, and teacher self-efficacy reported by the teachers. The scales are reliable and valid and could be applied to different school subjects.

Keywords: teaching quality, scale validation, dynamic model of educational effectiveness

Highlights:

- Scales for measuring teaching factors from the Dynamic Model of Educational Effectiveness were developed.
- The scales had a stable factor structure and invariant item loadings across two subjects, mathematics and biology.
- The scales correlated with teachers' job satisfaction, teacher self-efficacy, and teachers' beliefs about students' external control in mathematics.
- The scales could be used for measuring teaching practices in various school subjects.
- Teachers of different school subjects should not be compared via these instruments.

Aside from individual and family factors, the quality of teaching is the most important determinant of student achievement (Baumert et al., 2010; Brophy & Good, 1986; Creemers & Kyriakides, 2008; Fauth et al., 2014; Hattie, 2009; Klieme, 2012; Sanders & Rivers, 1996; Scheerens, 2000, 2016; Wright et al., 1997). There is a consensus in the literature on the relevant factors of teaching quality, although their nomenclature and operationalization somewhat vary. According to Klieme (2012), teaching quality consists of three factors: structure/classroom management, supportive climate, and challenge/ cognitive activation. Muijs et al. (2014) and Kington et al. (2009) offer more specific and numerous aspects of teaching quality, such as structuring classes and materials, providing feedback to students, and proactively maintaining discipline.

One of the widely recognized theoretic models through which teaching quality has been conceptualized is the *Dynamic Model of Educational Effectiveness* (DMEE; Creemers & Kyriakides, 2008). The model represents an amalgam of theory and empirical research on the factors that affect students' learning and achievement. It has been empirically tested in several studies (Creemers & Kyriakides, 2015; Kyriakides et al., 2020). According to the DMEE, the determinants of student achievement come from different hierarchical levels – student, teacher, school, and system levels. Some of the factors at the student level are socio-economic status, motivation, high expectations of students, intellectual abilities, and perseverance. School-level factors (e.g., cooperation between teachers in school), and system-level factors (e.g., quality national curriculum) do not impact students directly, but enable lower-level factors to work well. For example, improving cooperation between teachers may lead to improvement of individual teacher's ability to better engage students in the classroom, which will impact student achievement. Within the education system, teaching factors affect students most.

The eight teaching factors included in the DMEE describe the following aspects of teacher work (Creemers & Kyriakides, 2008): *Orientation* refers to

emphasizing the relevance and purpose of teaching content and activities in the context of students' knowledge, everyday application, and scientific knowledge; *Modelling* relates to the development of students' strategies to solve difficult problems and evaluate and organize their own learning; *Application* implies exercising the taught content and applying it in different situations; *Questioning* refers to actively engaging students by asking them various questions, seeking argumentation, and so on; *Assessment* encompasses identifying students' needs and offering constructive feedback, as well as correcting teacher's own work; *Structuring* refers to active and clear teaching and the creation of well-structured and organized lessons, for example, by presenting the outline and repeating the most important points at the end of the class as well as positioning lesson content within the wider context of students' knowledge; *Classroom as a Learning Environment* refers to creating a positive and supportive climate in the classroom; *Management of Time* refers to the teacher's use of classroom management procedures to maximize students' time on task and to create an effective learning-oriented classroom without distractions.

The teacher level of Creemers and Kyriakides' DMEE (2008) corresponds well to another widely accepted model of educational effectiveness – the already mentioned Klieme's model of three overarching factors of teaching quality (2012). Klieme's model represents a theoretical basis for student questionnaires used in PISA international testing (OECD, 2019). According to Klieme (2012), the DMEE's Management of Time belongs to the Classroom Management part of the Structure / Classroom Management factor, while Structuring, Orientation, and Assessment belong to the Structure part of the Structure/Classroom Management factor. DMEE's Classroom as a Learning Environment corresponds to the Supportive Climate factor, while the DMEE's Modelling, Application, and Questioning correspond to the Challenge / Cognitive activation factor.

In order to examine the teaching quality and its effects, it is necessary to adequately assess the teaching practices. This is commonly done by surveying students or teachers. Having a valid and reliable instrument that assesses *all* relevant teaching practices that impact student achievement and that can be applied to various school subjects is invaluable to researchers but especially to practitioners and policy makers. Around two thirds of European countries require schools to undertake self-evaluation (European Commission/EACEA/Eurydice, 2015). Together with external evaluation, self-evaluation is seen as an important instrument for school development (European Commission, 2018). However, it is unrealistic to expect all schools to be thoroughly familiar with the latest literature on quality teaching practices and to possess the expertise to develop high-quality questionnaires to assess them. Furthermore, policymakers can make important decisions in the areas of professional development of teachers and accountability mechanisms based on the results of educational effectiveness studies that can use such questionnaires. That is why researchers should provide valid and reliable instruments that are both theory- and evidence-based, as well as comprehensive and practical.

The literature has indicated that the most common problems with surveying include socially desirable answers of teachers and incompetence of younger students to evaluate teaching (Nielsen & Gustafsson, 2016), along with teachers' popularity with younger students, which positively biases students' assessments (Fauth et al., 2014). However, even though individual responses of students can be unreliable, classroom aggregates of their responses are valid and reliable (Kane & Cantrell, 2010, 2012; Nielsen & Gustafsson, 2016). Younger students – third graders – can also give valid and reliable assessments of teaching quality (Fauth et al., 2014). A combination of observations, students' assessments, and previously determined teacher's value added[1] has proved to best predict teacher effectiveness. Since each measure has its pros and cons, combining them gives the best results (Kane & Cantrell, 2010, 2012).

This research focusing on scales for measuring teaching quality represents a part of a large-scale study. In the larger study, eighth-grade students in Serbia were surveyed and their responses were aggregated to the teacher level in order to adequately assess the teaching practices of their mathematics and biology teachers. The goal was to evaluate the quality of teaching in schools in Serbia and then identify its effects on student achievement on the final exam and student interest in these subjects. The findings on the effects of teaching practices on student achievement and interest in mathematics and biology are presented in another paper (Teodorović et al., 2021).

The aim of this research was to validate eight scales for measuring teaching factors from the DMEE. For this purpose, two studies were conducted. In the first study, a preliminary version of the eight scales was explored. The second study cross-validated scales' structure, tested their theoretical correspondence with the DMEE, and examined their external validity by exploring their relations with relevant teachers' psychological variables. Since the aim was to create scales that would be generic in nature and applicable to different school subjects, in this study, we validated scales referring to two subjects – mathematics and biology[2].

Study 1

The objectives of this study were the initial construction of scales for measuring the teaching factors from the DMEE and the exploration of their latent structure.

---

1 Teacher's value added has been determined by adjusting their students' achievement gains for student characteristics such as prior performance and demographics (Kane & Cantrell, 2010, 2012).

2 These subjects were selected because they were the only two subjects that were taught in all four grades of lower secondary education (ISCED 2) in Serbia and that had adequate measures on both TIMSS 2011 international testing of 4th graders in Serbia and on the final exams of 8th graders in Serbia in 2015, which were the requirements for our larger study. Having two achievement measures for the same group of students – one when they were in the 4th grade and another when they were in the 8th grade – was necessary for the larger study, in which we tried to capture the effect of four years of accumulated teaching practices on student achievement.

Methodology

Sample and Procedure

The sample included a total of 683 seventh-grade students from 16 primary schools (out of 20 that were contacted) in nine cities in Serbia (Belgrade, Ub, Kragujevac, Kruševac, Sokobanja, Novi Sad, Sombor, Sopot, and Jagodina). In 13 schools, two classes were sampled per school, and in three schools, one class was sampled per school. The sample of schools and classes was convenient. Within each class, one half of the students filled out a questionnaire assessing the teaching of mathematics ($n = 346$, 50.7%) and the other half assessed the teaching of biology ($n = 337$, 49.3%). Students completed the questionnaires during one school lesson in October 2014. In this study, we did not gather any socio-demographic information from respondents.

At the time when this study was performed (in 2014), there were no Institutional Review Boards (IRB) and their approval was not obligatory, nor was such an approval requested by the project funders and schools. The whole research project (including Study 1 and Study 2) was approved by the Ministry of Education of the Republic of Serbia and the study was carried out with the support and approval of school authorities. The project was implemented in accordance with the Law on Personal Data Protection and the best practices at the time.

Instruments

In the process of constructing the preliminary versions of the scales, the existing instruments for the assessment of the DMEE were examined (Creemers et al., 2012) and some items were modified to fit student age or the specific context of the school environment in Serbia. Then, the existing items from the PISA study (OECD, 2013) and the DaQS database (Datenbank zur Qualität von Schule, n.d.) were analyzed and the scales/items that corresponded to the theoretical content of factors from the DMEE were added to the item pool. Finally, the researchers constructed new items to better cover the theoretical content of the teaching factors. This process resulted in a total of 70 items distributed across eight scales. A smaller number of constructed items were reverse coded. Students assessed the frequency of a certain practice during class on a 4-point Likert scale (1 = *never*, 4 = *always or almost always*).

Statistical Analyses

An exploratory factor analysis (EFA) was conducted to investigate the structure of each individual scale. We applied the principal components method. To determine the optimal number of factors, we consulted two criteria: the Gutmann-Kaiser criterion and a parallel analysis with the 95th percentile of randomly generated eigenvalues. The criteria for item retention were a factor loading $\geq .50$ and no cross-loading. The factors were rotated via the Promax rotation. The EFA was applied iteratively until all items met the criteria. The EFA was carried out separately for the scales referring to mathematics and biology and only the items that fulfilled the described criteria for both subjects were kept. After the final versions of the scales were determined, we applied an EFA with the same criteria to the whole item pool to investigate whether the latent structure would correspond to the theoretical factors from the DMEE. For the final solution of eight scales, we calculated descriptive statistics, internal consistency (Cronbach's alpha), and mutual intercorrelation (Pearson correlation coefficients). All analyses were carried out in SPSS 20.0 software.

Results

Exploratory Factor Analysis

The Kaiser-Meyer-Olkin coefficient of sampling adequacy indicated that it was justified to apply an EFA on all scales (see Supplementary material 1,

Appendix A). The EFA resulted in five unidimensional scales: Orientation (59% of the variance was explained by the first principal component in mathematics and 55% in biology), Modelling (59% in mathematics and 60% in biology), Application (59% in mathematics and 57% in biology), Questioning (62% in mathematics and 54% in biology), and Assessment (56% in mathematics and 53% in biology). Detailed results of the EFA with the structure and the factor loadings for all scales are presented in Supplementary material 1, Appendix A (https://osf.io/acfpu/).

For the Structuring scale, Guttman-Kaiser resulted in two factors for both subjects, while the parallel analysis recommended a one-factor structure for mathematics and a two-factor structure for biology. Since we aimed to create a scale with the same structure for both subjects, we opted for the two-factor solution[3]. The two-component structure explained 59% of the variance in mathematics and 56% in biology. The first component referred to the teaching practices aimed at connecting the material with different lessons, different subjects or students' out-of-class knowledge. This component was named Connecting. The second component described clear and understandable teaching and setting clear learning goals. The component was named Clarity. The correlation of the extracted components was .59 for mathematics and .43 for biology.

Initially, a two-component structure was also obtained for the Classroom as a Learning Environment scale, explaining 51% of the variance in mathematics and 49% in biology. The first component described the relationship between the teacher and the students, with the teacher actively providing help and encouragement, while the second component described students' mutual relationships (see Supplementary material 2, Appendix A, Table 14 for the initial version of the EFA; https://osf.io/umcg5/). The extracted components did not correlate with each other ($r = .01$ for mathematics and $r = .16$ for biology), suggesting that these contents do not belong to the same conceptual space. Since the second component did not relate directly to the teacher's actions, we decided to exclude these items from the scale. The EFA was re-applied and the repeated analyses resulted in a unidimensional structure for both subjects. The first component explained 54% of the variance of items in mathematics and 43% in biology.

For the Management of Time scale, the Guttman-Kaiser criterion indicated it was optimal to keep three factors, while parallel analyses suggested two factors for mathematics and three for biology. Since our aim was to obtain correspondent structures for both subjects, we leaned on the Guttman-Kaiser criterion (see the other solution in the Supplementary material 2, Appendix A, Table 16). The three-dimensional structure explained 67% of the variance in mathematics and 64% in biology. Three components corresponded to the three aspects of this construct: Loss of time, Classroom disorder, and teacher's Classroom management. The three components' correlations varied in absolute

---

3   Interested readers can see the *initial* factor solution in Supplementary material 2, Appendix A. In this document, we presented only the initial EFA structure of the scales whose final structure was changed in the iterative process of EFAs, while the final EFA solution of all scales is presented in Supplementary material 1, Appendix A.

values from .27 to .47 for mathematics and from .40 to .50 for biology (in the theoretically expected direction).

Table 1 shows the number of items before and after the EFA of all scales, as well as the internal consistency (Cronbach's alpha) of the final version of the scales, which ranged from good to excellent. Only the Loss of time subscale from the Management of Time scale had suboptimal internal consistency when applied to biology.

Table 1
Number of items in the preliminary and final versions of the scales and their internal consistency

Scale	Initial number of items	Final number of items	Cronbach's α (mathematics)	Cronbach's α (biology)
Orientation	5	5	.82	.80
Modelling	7	7	.88	.89
Application	5	5	.83	.80
Questioning	7	6	.88	.83
Assessment	8	6	.84	.82
Structuring	14	9	.86	.81
Connecting		5	.79	.75
Clarity		4	.81	.76
Classroom as a Learning Environment	12	9	.89	.84
Management of Time	12	12	.87	.87
Classroom Disorder		4	.84	.86
Loss of Time		4	.77	.68
Management of Classroom		4	.83	.84

Table 2
Intercorrelations of teaching factors

	1	2	3	4	5	6	7	8
1. Orientation		.72**	.62**	.65**	.68**	.74**	.60**	.43**
2. Modelling	.66**		.75**	.80**	.76**	.77**	.69**	.45**
3. Application	.60**	.73**		.80**	.65**	.63**	.70**	.49**
4. Questioning	.59**	.63**	.67**		.72**	.75**	.78**	.51**
5. Assessment	.57**	.77**	.65**	.61**		.72**	.65**	.46**
6. Structuring	.75**	.64**	.57**	.64**	.61**		.71**	.50**
7. Classroom as a Learning Environment	.62**	.60**	.57**	.64**	.58**	.64**		.58**
8. Management of Time	.47**	.39**	.39**	.42**	.36**	.45**	.46**	

Note. Intercorrelations for mathematics are above the diagonal and for biology below.
**$p \leq .01$

Table 2 shows the intercorrelations of teaching factors. All correlations were moderate to high and positive. To determine whether the scales' items would group in line with the theoretical model of the DMEE, we applied a joint EFA on the pool of 59 items from the final versions of the scales. The Guttman-Kaiser criterion recommended eight components for mathematics and eleven

components for biology, while the parallel analysis suggested that the optimal number of components was three for mathematics and four for biology (see Supplementary material 1, Appendix B for a detailed presentation of the results). Since the Guttman-Kaiser criterion tends to overestimate the number of factors, we opted for the factor solution suggested by the parallel analysis[4]. Neither the number of components nor their structure corresponded to those from the DMEE. A simplified presentation of the structure of these factors is shown in Table 3. For mathematics, the first component predominantly gathered indicators of Orientation, Assessment, Modeling, and Structuring-Connecting; the second component gathered indicators of Classroom as a Learning Environment, Management of Time – Loss of time, Structuring-Clarity, Application, and Questioning; while the third component included two subscales of Management of Time – Classroom disorder and Classroom Management. For biology, the first component was saturated with indicators of Modeling, Assessment, and Application; the second component was saturated with indicators of Classroom as a Learning Environment, Questioning, and Structuring-Clarity; the third included items from the three subscales of Management of time; and the fourth gathered Structuring-Connecting and Orientation. The results of this study will be discussed together with the results of the second study in the General Discussion section.

Table 3
A simplified presentation of the results of joined EFA performed on the final versions of scales for measuring teaching factors from the DMEE

Factor	Mathematics	Biology
1st component	Assessment	Assessment
	Modeling	Modeling
	Structuring – Connecting	
	Orientation	
2nd component	Classroom as a Learning Environment	Classroom as a Learning Environment
	Structuring – Clarity	Structuring – Clarity
	Questioning	Questioning
	Application	
	Management of Time – Loss of Time	
3rd component	Management of Time – Classroom Disorder	Management of Time – Classroom Disorder
	Management of Time – Classroom Management	Management of Time – Classroom Management
		Management of Time – Loss of Time
4th component		Structuring – Connecting
		Orientation

---

4   We also analyzed the eight-factor solution for mathematics and the eleven-factor solution for biology. Even in the eight-factor solution obtained for mathematics, which had the theoretically expected number of factors, the items did not group in teaching factors proposed by the DMEE. The same was true for the eleven factors obtained for biology.

Study 2

In the second study, we aimed to: a) cross-validate individual scale structures on a new, larger sample of students, b) test which theoretical model best represents the overall structure and relations of the scales, and c) externally validate the final versions of the scales by examining their relations with specific teacher psychological variables – job satisfaction, teacher self-efficacy, and external control – which are associated with teaching practices (Gkolia, Belias, & Koustelios, 2014; Klusmann et al., 2008; Kunter et al., 2013; Rissanen et al., 2018; Rose & Medway, 1981).

Methodology

Sample and Procedure

The sample included students from 125 schools in Serbia[5]. In 115 schools, two 8th grade classes were sampled per school, while one 8th grade class was chosen from each of the remaining 10 schools, resulting in 240 classes attended by a total of 5,476 eighth-graders. In April 2015, one half of the students within each class evaluated the teaching of mathematics and the other half assessed the teaching of biology. In 33 classes, there were fewer than 20 students, so all those students assessed only the teaching of mathematics. Therefore, 2.895 students (53.4%) assessed mathematics classes and 2.527 (46.6%) assessed biology classes. The sample included 48.2% of males and 51.2% of females, while 37 participants (0.7%) did not indicate their gender. The average age was 14.54 years (SD = 0.33, range 13.16–17.66).

A sample of teachers was also included in the study. Out of 2.401 teachers of various subjects in the large study (20 from each school with certain dropout), there were 164 mathematics teachers (69.5% female; average work experience 14.1 years (SD = 10.48) and 135 biology teachers (85.9% female; average work experience 18.4 years (SD = 9.67) who could be paired with their students who filled out the questionnaires on teaching. An external validation of scales for measuring teaching factors was performed on this sample of teachers.

As mentioned in Study 1, ethical approval was not acquired for this study since there were no institutional IRBs at the time when it was performed (in 2014 and 2015) and their approval was not obligatory. Since we gathered more data about participants in this study, in accordance with the Law on Personal Data Protection and the best practices in force at the time when the project was implemented, all personal information about students, such as questions regarding house possessions (not reported in this paper), was gathered from students' parents who were informed about their children's participation in the study. Teachers who participated in the study were informed about project aims and consented to participate. Children, parents, and teachers participated anonymously and their data were paired through the system of codes.

Instruments

In Study 2, students filled out questionnaires for measuring teaching factors from the DMEE that resulted from the EFA of Study 1. Three teacher psychological variables were selected in order to determine the external validity of the scales.

---

5   The results reported in this paper are a part of a larger project, which utilized the nationally representative sample of 156 schools from the TIMSS 2011 study. Out of 156 schools, 27 schools were excluded because of the small number of students, which made subsamples from these schools inadequate for the design of the planned project. Another four schools refused to participate in the study. More information about this project and the sample is presented in the paper by Teodorović et al. (2021).

Job satisfaction scale (TIMSS, 2009) contains six items with a four-point Likert response scale and has a satisfactory internal consistency ($\alpha = .76$).

External control scale (Skaalvik & Skaalvik, 2007) implicitly measures teachers' beliefs about their influence on students' academic success. The scale contains five items with a four-point Likert response scale and has a satisfactory internal consistency ($\alpha = .72$). A higher score on the scale indicates the teacher's belief that students' success is primarily determined by their abilities and family environment, and not the teacher.

Self-efficacy scale (Tschannen-Moran & Woolfolk Hoy, 2001) measures teachers' perception of their performance in a variety of teaching activities. The scale has three subscales: efficacy in instructional strategies, efficacy in classroom management, and efficacy in engaging students. Each subscale is operationalized by four items with a five-point Likert response scale. The reliability of each subscale is good (from .75 to .89), with $\alpha = .89$ for the entire scale.

Statistical Analyses

A confirmatory factor analysis (CFA) was conducted to test the scale models derived as a result of the EFA. The CFA was applied separately to the data obtained for mathematics and biology. The following criteria were applied: a) corrected chi-square test, which is more tolerant to sample size than a simple chi-square and whose value should be under 5 (Mueller, 1996); b) root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR) with values < .05 indicating excellent fit and values < .08 indicating acceptable fit (Byrne, 2010); c) comparative fit index (CFI) and normed index of fit (NFI) with values > .90 indicating acceptable fit and values > .95 indicating excellent fit (Kline, 2005). If the model parameters were below acceptable levels, modification indices were analyzed and up to two corrections were made (Byrne, 2010; Kline, 2005).

Since the aim of this study was to test the stability of the theoretical model independently of the subject, a multi-group CFA was applied to each scale by examining the configural, metric, and scalar invariance, with the subject (mathematics/biology) as the grouping factor. The configural invariance implies that items have significant loadings on associated dimensions. The metric invariance assumes the invariance of item loadings across compared groups, while the scalar invariance tests the invariance of intercepts. Comparisons between the invariance models are performed by calculating the differences of their fit parameters (ΔCFI, ΔRMSEA, and ΔSRMR), where the differences should not be greater than .01 (values > .01 are considered significant at the $p < .01$ level; Chen, 2007). The lack of statistically significant differences between the models indicates that the more constrained model fits the data equally well.

Additionally, we used Tucker's congruence coefficient to verify the congruence between factor structures for mathematics and biology (Harman, 1976; Tucker, 1951). This statistic is used in large datasets because the standard CFA is sensitive to sample size and thus can easily dismiss the assumption of equal factor structures across groups (Lorenzo-Seva & Ferrando, 2003). The values between .98–1.00 imply excellent, .92–.98 good, .82–.92 marginal, .68–.82 weak, and < .68 poor congruence (MacCallum et al, 1999).

After the final versions of the eight scales were established, their internal consistency (Cronbach's alpha) and descriptive data were calculated.

In order to investigate whether the DMEE theoretical model best fits the data, we conducted a CFA (with the same criteria defined above) and contrasted it with the structure that corresponds to Klieme's latent-factor model (Klieme, 2012) and the structure that corresponds to the one-factor model. The DMEE model was defined by the eight factors based on the constructed scales that were mutually correlated (model 1). In Klieme's model, the items from DMEE scales were grouped so that all items from Modelling, Application, and Questioning

loaded on the Challenge / Cognitive activation factor; Classroom as a Learning Environment corresponded to the Supportive Climate factor, items from scales Structuring, Orientation, and Assessment loaded on the Structure factor; and Management of Time corresponded to the Classroom management[6] factor (model 2). In the one-factor model, all items loaded on one factor (model 3).

Finally, in order to examine the external validity of the scales, student evaluations of the eight teaching factors were aggregated at the teacher level by calculating the average value of students' responses. These values represented teachers' scores on the teaching factors from the DMEE. They were then correlated using the Pearson correlation with the variables obtained from teachers' self-assessment[7].

All analyses were performed in SPSS 20.0 and Amos 21.0 software.

Results

Confirmatory Factor Analysis

The majority of the scales had adequate fit parameters in the CFA (Table 4), with the exception of the corrected chi-square. Fit parameters were below the satisfactory values for Orientation, Application, and Assessment scales, so modifications were introduced. In each of the three scales, the errors between the two items were correlated (see Figure 1) and an acceptable fit was achieved. These modifications suggest that the pairs of items for which the error correlations were introduced share similarities beyond the main subject of measurement of the scale – in this case, both items refer to stressing out the importance of the teaching content or working through the most common students' mistakes. The parameters of these modified versions of the model are shown in Table 4. Except for these modifications, for all scales, the structure obtained in the EFA in Study 1 was confirmed in the CFA on the data from Study 2.

---

6 Although Klieme's model theoretically includes 3 teaching factors, he sometimes divides his factor Structuring / Classroom management into two - Structuring and Classroom management. In the EFA performed on all items from the eight scales constructed in Study 1, Management of time emerged as an individual component. Therefore, we decided to lean on Klieme's model with divided Structure and Classroom Management, which might better correspond to the data from our study and could be a more appropriate alternative to the DMEE.

7 We first planned to examine the unique contribution of individual teaching factors from the DMEE to these external criteria, by conducting regression analyses with teacher psychological variables as criterions and teaching factors as predictors. However, multicollinearity diagnostics indicated that collinearity was a problem (the highest VIF was 12.11 for mathematics and 11.24 for biology and values higher than 10 are usually taken as critical). Therefore, we decided to only carry out the Pearson correlations between these sets of variables.

Table 4
CFA fit parameters

Factor	Subject	χ^2	df	χ^2/df	CFI	RMSEA	RMSEA CI 90%		SRMR	NFI	NNFI
Orientation	M	67.47	5	16,868	.985	.074	.058	.089	.021	.984	.963
	B	73.40	5	14.68	.981	.073	.059	.088	.027	.980	.943
Modelling	M	183.99	14	13.14	.977	.065	.057	.073	.025	.975	.965
	B	148.00	14	10.57	.984	.061	.052	.070	.023	.982	.967
Application	M	58.23	4	14.59	.987	.068	.054	.085	.020	.986	.968
	B	102.39	4	25.60	.977	.071	.055	.082	.035	.977	.915
Questioning	M	110.83	9	12.31	.985	.063	.052	.073	.021	.984	.975
	B	92.76	9	10.37	.984	.060	.049	.071	.033	.983	.964
Assessment	M	103.00	7	14.71	.985	.069	.057	.081	.023	.984	.967
	B	103.94	7	14.85	.984	.073	.061	.086	.024	.983	.953
Structuring	M	471.12	26	18.12	.951	.077	.071	.083	.021	.948	.910
	B	708.22	26	27.24	.911	.080	.075	.086	.034	.901	.883
Classroom as a Learning Environment	M	473.20	27	17.53	.956	.076	.070	.081	.035	.953	.938
	B	823.05	27	30.48	.995	.073	.069	.078	.028	.954	.923
Management of Time	M	477.86	51	9.37	.973	.054	.049	.058	.031	.970	.961
	B	331.29	51	6.50	.978	.046	.041	.051	.029	.975	.961

Note. M = mathematics; B = biology.

The multi-group CFA was performed next and its results are shown in Supplementary material 1, Appendix C (https://osf.io/acfpu/). The values of the corrected chi-square test were unsatisfactory for all scales, but this is not uncommon when dealing with particularly large samples. As for other fit parameters, for all scales, both the configural and metric invariance had satisfactory CFI, RMSEA, and SRMR. The differences between them were not statistically significant, indicating invariance of factor loadings. For Management of time, Classroom as a Learning Environment, and somewhat Structuring (except for the CFI), the model assuming scalar invariance was also acceptable and not significantly different from the model with metric invariance, suggesting equal intercepts in the two subjects. For all other scales, the fit indices of the scalar invariance models were not acceptable, indicating different intercepts when these scales were used for different subjects. Tucker's congruence coefficient, however, showed excellent congruence for all scales, as its values were all .99. The structure of all eight tested models is presented in Figure 1.

In summary, the fit indices (except for the corrected chi-square) for all scales suggest that the factor structures of the scales are comparable for mathematics and biology. Specifically, item loadings on latent factors were invariant, but the intercepts generally differed. This means that it is justified to use these scales to compare teachers of the same subject, but not teachers of different subjects.
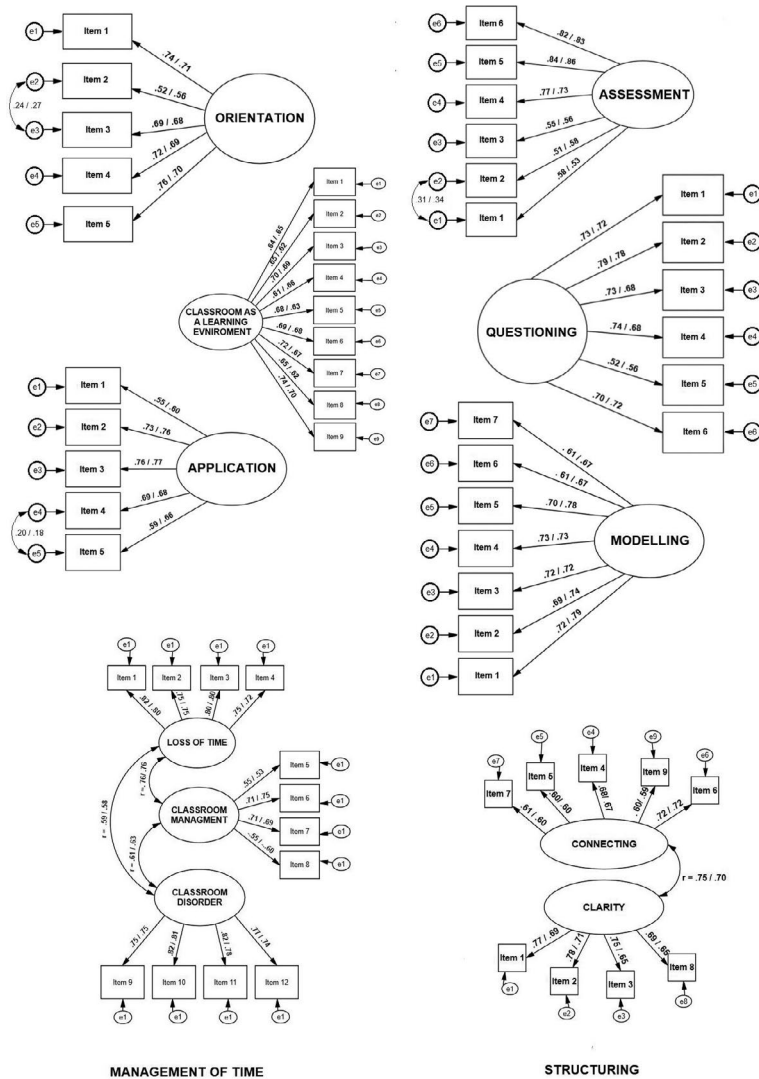
Descriptive Statistics and Reliability of Scales

After determining the final structures of the scales, scale scores were calculated as an average of responses to the belonging items. Standardized

skewness and kurtosis, as well as the Kolmogorov-Smirnov test of normality of distribution indicated that all variables leaned towards positive scores, i.e., that students' evaluations of teaching factors were positively biased (Table 5). The internal consistency of the scales (Cronbach's alpha) varied from good to excellent.

Descriptive statistics for teacher psychological variables (job satisfaction, external control, and teacher self-efficacy) are presented in Table 6.

Figure 1
Structure of scales for mathematics and biology

Note. Factor loadings and correlations for mathematics are shown before the slash and those for biology are shown after the slash.

Table 5

Descriptive statistics and reliability of scales

Scale	Mathematics								
	Min	Max	M	SD	α	Stand. Skewness	Stand. Kurtosis	K-S test (p)	KMO
Orientation	1.00	4.00	2.77	.72	.83	-8.12	-5.88	< .001	.83
Modelling	1.00	4.00	2.86	.67	.86	-9.25	-3.68	< .001	.90
Application	1.00	4.00	2.98	.66	.80	-11.38	-0.93	< .001	.82
Questioning	1.00	4.00	2.98	.68	.86	-12.23	-2.64	< .001	.89
Assessment	1.00	4.00	2.74	.72	.82	-6.02	-6.77	< .001	.82
Structuring	1.00	4.00	2.84	.60	.79	-10.41	-1.11	< .001	.89
Classroom as a Learning Environment	1.00	4.00	3.06	.63	.84	-15.08	-0.08	< .001	.93
Management of Time	1.00	4.00	3.00	.60	.89	-12.10	-1.61	< .001	.92

	Biology								
	Min	Max	M	SD	α	Stand. Skewness	Stand. Kurtosis	K-S test (p)	KMO
Orientation	1.00	4.00	2.98	.66	.81	-10.58	-3.07	< .001	.83
Modelling	1.00	4.00	2.61	.75	.89	-4.95	-7.16	< .001	.92
Application	1.00	4.00	2.56	.73	.83	-3.37	-6.17	< .001	.83
Questioning	1.00	4.00	2.98	.68	.85	-12.05	-0.74	< .001	.88
Assessment	1.00	4.00	2.51	.76	.84	-2.09	-8.45	< .001	.83
Structuring	1.00	4.00	2.90	.57	.79	-9.99	-0.12	< .001	.86
Classroom as a Learning Environment	1.00	4.00	3.07	.59	.82	-14.17	1.37	< .001	.92
Management of Time	1.00	4.00	3.00	.58	.89	-10.47	-2.60	< .001	.92

Legend: K-S test = Kolmogorov-Smirnov normality test.

Note. The values in columns titled stand. skewness and stand. kurtosis displayed in the table are calculated by dividing skewness and excess kurtosis values by their standard error. We will refer to these as standardized skewness and standardized kurtosis respectively. Values in excess of +-1.96 or +-2.56 (i.e. above 1.96 or 2.56 or below -1.96 or -2.56) would mean that the sample value of skewness/kurtosis fall outside the 95% or 99% confidence interval formed around the values of 0, which indicates a mesokurtic, symmetric distribution.

Table 6

Descriptive statistics for teacher psychological variables

Variable	Min	Max	M	SD	Fischer coef. of skewness	Fischer coef. of kurtosis
Job Satisfaction	9	24	19.53	3.29	-2.05	-1.64
External Control	5	19	11.98	2.69	-1.47	0.20
Teacher Self-Efficacy	29	60	49.45	5.44	-2.84	1.12
Teacher Efficacy in Instructional Strategies	10	20	17.59	1.76	-4.26	1.98
Teaching Efficacy in Classroom Management	8	20	15.47	2.27	0.45	-1.42
Teacher Efficiency in Student Engagement	7	20	16.38	2.59	-5.58	2.98

Note. The values in columns titled Fischer coef. of skewness and Fischer coef. of kurtosis displayed in the table are calculated by dividing skewness and excess kurtosis values by their standard error. Values in excess of +-1.96 or +-2.56 (i.e. above 1.96 or 2.56 or below -1.96 or -2.56) would mean that the sample value of skewness/kurtosis fall outside the 95% or 99% confidence interval formed around the values of 0 skewness/kurtosis, which indicate a mesokurtic, symmetric distribution.

The results of the CFA for different theoretical models – the eight-factor model from the DMEE (model 1), Klieme's latent factor model (model 2), and the one-factor model (model 3) – showed that only model 1 had satisfactory RMSEA and SRMR (Supplementary material 1, Appendix D). However, none of them had satisfactory $\chi2/df$, CFI, NFI, and NNFI, suggesting that none of these theoretical models described the data optimally. Additionally, we contrasted these models against each other and the results showed that model 1 is significantly better than models 2 and 3 (comparison of model 2 and model 1 for mathematics $\Delta\chi2(22) = 2826.93$, $p < .001$ and for biology $\Delta\chi2(22) = 3063.74$, $p < .001$; comparison of model 3 and model 1 for mathematics $\Delta\chi2(28) = 12394.37$, $p < .001$ and for biology $\Delta\chi2(28) = 12230.85$, $p < .001$; comparison of model 3 and model 2 for mathematics $\Delta\chi2(6) = 9567.44$, $p < .001$ and for biology $\Delta\chi2(22) = 9167.11$, $p < .001$).

External Validity of Scales

In order to examine the external validity of scales, aggregated teacher scores on the teaching factors were correlated with the measures of teachers' job satisfaction, external control, and their self-efficacy. The results are presented in Table 7.

Table 7
Correlations of teaching factors and teacher psychological variables

Scale	JS	EC	TSE	TE-IS	TE-CM	TE-SE
	Mathematics					
Orientation	.16*	-.13	.10	-.05	.02	.22**
Modelling	.21**	-.13	.09	-.04	.03	.18*
Application	.21**	-.14	.11	-.03	.06	. 20*
Questioning	.18*	-.18*	.06	-.08	.02	.16*
Assessment	.19*	-.06	.07	-.03	.04	.13
Structuring	.15*	-.14	.05	-.08	-.01	.16*
Classroom as a Learning Environment	.23**	-.24**	.11	.01	.03	.20*
Management of Time	.26**	-.18*	.22**	.08	.09	.33**
	Biology					
Orientation	.03	-.04	.08	.02	.01	.14
Modelling	.00	-.09	.08	.06	-.03	.15
Application	-.02	-.10	-.01	.01	-.12	.08
Questioning	-.02	-.08	.00	-.01	-.09	.09
Assessment	-.05	-.06	-.06	.02	-.00	.12
Structuring	-.02	-.05	.03	.01	-.07	.12
Classroom as a Learning Environment	.03	-.06	.07	.04	-.08	.15
Management of Time	.01	-.08	.11	.02	.02	.22**

Note. JS = Job Satisfaction; EC = External Control; TSE = Teacher Self-Efficacy (total score); TE-IS = Teacher Efficacy in Instructional Strategies; TE-CM = Teaching Efficacy in Classroom Management; TE– SE – Teacher Efficiency in Student Engagement.
$N_{math} = 164$, $N_{bio} = 135$.
* Correlation is significant at the level of $p < .05$ (bidirectional).
** Correlation is significant at the level of $p < .01$ (bidirectional).

General Discussion

The results of exploratory and confirmatory factor analyses showed that each individual scale for the measurement of teaching factors of the DMEE had a clear and theoretically adequate structure and good internal consistency. Six scales – Orientation, Modelling, Application, Questioning, Assessment, and Classroom as a Learning Environment – had a unidimensional structure, while two scales – Structuring and Management of Time – had a multi-dimensional structure. Two aspects of the Structuring scale – Clarity and Connecting – were interrelated and formed a single scale with good internal consistency. Out of the three subscales of the Management of Time scale – Loss of time, Classroom disorder, and Classroom Management – the first two positively correlated with each other and negatively correlated with the third, which is in line with theoretical expectations. The overall scale had good internal consistency. It should be noted that all reverse coded items were eliminated from the scales due to unsatisfactory psychometric characteristics and that the final version of the scales contains only positively formulated items. The scales in Serbian and English are publicly available on the OSF page of the project (https://osf.io/q85zu/).

The intercorrelations of the scales were moderate to high. While these correlations may seem higher than would be desirable in terms of discriminant validity of the scales, one should bear in mind two things. First, research has shown that quality teaching practices go hand in hand, i.e., that a good teacher does many things well, which explains high correlations (Muijs & Reynolds, 2000; Teodorović, 2011). Actually, one of the main assumptions of the Dynamic Model of Educational Effectiveness is that teacher factors are interrelated (Kyriakides et al., 2009). Secondly, student assessment of teaching practices may somewhat suffer from the halo effect, i.e., it is possible that student responses were partly influenced by the general impressions of the teacher. The fact that students' assessments of teaching factors were more distributed towards positive values on all scales may suggest that students were slightly biased toward their own teachers or that teachers on average exhibited quality teaching. However, this does not mean that student assessment does not have its place in research, as it has been established that students can give reliable and valid assessments of teaching practice even in younger grades (Fauth et al., 2014; Kyriakides et al., 2014). After all, a teacher's self-assessment measures are also influenced by personal biases. However, high intercorrelations of the scales may suggest that at least some of the teacher factors from DMEE are not separate theoretical constructs.

Taking into account mutual correlations of the scales, we decided to perform a joint EFA on all items from final versions of the eight scales and the results did not converge to the DMEE or to other alternative theoretical models such as Klieme's model (Klieme, 2012). Analyses revealed three general factors for mathematics and four for biology, meaning that the structure was not correspondent for the two subjects. There were, however, certain similarities

between the two obtained solutions. Management of Time converged into one factor, although for mathematics it did not include the Loss of time subscale. Classroom as a Learning Environment went along with Structuring-Clarity in both subjects, although it was accompanied by other theoretical concepts in mathematics. The factor that gathered Orientation, Structuring-Connecting, Modeling, and Assessment in mathematics, separated into two in biology – one that joined Orientation and Structuring-Connecting and the other that joined Modeling and Assessment. From the perspective of the DMEE, the most important conclusion could be that two aspects that theoretically belong to Structuring (Connecting and Clarity) seem to be conceptually different and that they are closer to other aspects of teaching practices than to each other. To conclude, although it seems that certain aspects of teaching are more likely to go together in the classroom, it seems that relations between different theoretical aspects of teaching depend, at least to some extent, on the subject that is taught. Fit of the scales to the theoretical model, as well as discriminant validity of the scales, would likely be improved with selection of items that are highly specific for each teaching factor. However, this may narrow the meaning of the teaching factors and limit their usefulness in predicting student achievement.

While our analyses did not show sufficient discriminant validity of the scales, this is not the reason to abandon either the Dynamic Model of Educational Effectiveness or the instrument. The structure of the theoretical model has been validated across and within several countries using CFA and Structural Equation Modeling – SEM (Belgium/Flanders, Cyprus, Germany, Greece, Ireland and Slovenia; Kyriakides et al., 2014), albeit with a smaller number of items than in our study (28 vs. 59, as different items had to be removed from the original 49-item questionnaire in the study) and with two identified second-order factors (quality of teaching practices and quantity of teaching practices). More than 20 studies conducted in different countries have provided empirical support for the model (for a review of these studies see Kyriakides et al., 2021). The DMEE has even been used in teachers' professional development after which student achievement results improved (Antoniou et al., 2011). Finally, although the confirmatory analyses in Study 2 did not show that data satisfactorily fit the theoretical structure from the DMEE, that structure was still somewhat better than the structure proposed by Klieme's model or the model that proposes that the whole variance in students' estimates of teaching may be due to the halo effect. Nevertheless, the results of our study indicate that either these new scales, the theoretical constructs from DMEE, or both could be further refined. While our instrument showed greater overlap of teaching practices than is desirable, the scales still reliably and validly measure important aspects of teaching quality that are based on one of the more utilized theoretical and empirical models of educational effectiveness in the world – the DMEE. Ideally, data from student questionnaires should be supplemented with classroom observation ratings in order to improve the reliability and validity of the model (Kyriakides et al., 2014).

The results of the multi-group CFA indicate that, when applied to different subjects, the scales measure the same constructs in a structurally identical manner. However, it is not justified to compare teachers of different subjects because the estimates may be affected by factors that are not the subject of measurement (e.g., various types of biases; Fisher & Karl, 2019; Xu & Tracey, 2017). Fisher & Karl (2019) claim that not reaching scalar invariance when the scale is applied to different groups should not be a problem as long as researchers keep in mind this limitation. On the other hand, the fact that all eight scales showed identical structures for mathematics and biology suggests that it is also appropriate to use these scales for other subjects. The very fact that the items describe general and not subject-specific aspects of teaching quality is in line with this conclusion.

Testing the relations between the teaching factors reported by students and teacher psychological variables reported by teachers showed that, in mathematics, the majority of teaching factors were associated with teachers' job satisfaction and self-efficacy in terms of engagement of students, which is consistent with previous literature (Gkolia et al., 2014; Klusmann et al., 2008; Kunter et al., 2013). Additionally, external control, i.e., the belief that the teacher does not have a major impact on student achievement, was partly related to lower teaching quality, particularly with poor time management, weaker ability to create a supportive learning environment, and inadequate questioning skills. It is interesting to note that these are the aspects of teaching that require good interpersonal and communication skills, which is consistent with the results of other research where teachers' psychological variables were more related to classroom management and supportive relationship with students than with instructional abilities related to cognitive activation of students (Klusmann et al., 2008; Kunter et al., 2013).

In contrast to mathematics, job satisfaction, external control, and self-efficacy had less "spillover" onto teaching practices in biology. Only self-efficacy in engaging students was associated with a better ability to manage time. These results suggest that it may be more demanding to bring students closer to the material and understanding of mathematics than biology, so teachers' satisfaction with their own profession and belief that they do their job well matter more in teaching mathematics than biology. Similarly, teachers who are prone to shifting responsibility from themselves for their students' achievement have an easy excuse to put in less effort to advance their own practice. These findings resonate with research that identifies mathematics as a subject that requires and engages logical reasoning more than other subjects (e.g., Gómez-Veiga et al., 2018).

When applied to biology, the teaching factors measured by the scales developed in this study did not correlate with variables chosen for external validation. However, it should be stressed that they successfully predicted students' interest in biology (Teodorović et al., 2021). Thus, the lack of statistically significant correlations between these variables could probably be due to subject specificities rather than inadequate external validity. It should also be noted that the scores on teaching factors and teacher psychological variables

came from different sources (students vs. teachers), which is most likely one of the reasons why correlations did not reach higher levels. Overall, we can conclude that the relations of teachers' job satisfaction, their self-efficacy. and external control with the scales for measuring teaching factors from the DMEE – which were statistically significant and in the expected direction for mathematics – indicate adequate external validity of these scales.

Conclusion

This paper presents the validation of scales for measuring teaching factors from the Dynamic Model of Educational Effectiveness. The scales have largely proven to be reliable and valid measures of teaching quality as described in the model and can be used to assess teaching in school self-evaluation, external evaluation or educational research, although they should not be used to compare teaching of different subjects. Although the scales in these two studies were used for the assessment of mathematics and biology teaching, the aim of their construction was to apply them to other subjects as well. Additional validation of the scales on a subject from social or humanistic sciences is recommended.

References

Antoniou, P., Kyriakides, L., & Creemers, B. P. M. (2011). Investigating the Effectiveness of a Dynamic Integrated Approach to Teacher Professional Development. *Center for Educational Policy Studies Journal, 1*(1), 13–41. https://doi.org/10.26529/cepsj.439

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y. M. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal, 47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching* (pp. 340–370). Macmillan.

Byrne, B. M. (2010). *Multivariate applications series. Structural equation modeling with AMOS: Basic concepts, applications, and programming (2nd ed.).* Routledge/Taylor & Francis Group.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Creemers, B. P. M, & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. Routledge.

Creemers, B. P. M., Kyriakides, L., & Panayiotou, A. (2012). *Designing evidence-based strategies and actions to promote quality in education. Handbook of the research project "Establishing a knowledge base for quality in education: testing a dynamic theory for education"*. University of Cyprus.

Creemers, B., & Kyriakides, L. (2015). Developing, testing, and using theoretical models for promoting quality in education. *School Effectiveness and School Improvement*, *26*(1), 102–119. https://doi.org/10.1080/09243453.2013.869233

Data bank for the quality of the school. [Databank for School Quality]. (nd) Retrieved June 20, 2014 from https://daqs.fachportal-paedagogik.de/constructlist/

European Commission. (2018). *Quality assurance for school development: Guiding principles for policy development on quality assurance in school education*. Directorate-General for Education, Youth, Sport and Culture.

European Commission/EACEA/Eurydice. (2015). *Assuring Quality in Education: Policies and Approaches to School Evaluation in Europe*. Eurydice Report. Luxembourg: Publications Office of the European Union.

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

Fischer, R., & Karl, J. (2019). A Primer to (Cross-Cultural) Multi-Group Invariance Testing Possibilities in R. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01507

Gkolia, A., Belias, D., & Koustelios, A. (2014). Teacher's job satisfaction and self-efficacy: A review. *European Scientific Journal, ESJ*, *10*(22). https://doi.org/10.19044/esj.2014.v10n22p%p

Gómez-Veiga, I., Vila Chaves, J. O., Duque, G., & García Madruga, J. A. (2018). A New Look to a Classic Issue: Reasoning and Academic Achievement at Secondary School. *Frontiers in Psychology, 9*(400). https://doi.org/10.3389/fpsyg.2018.00400

Harman, H. H. (1976). *Modern factor analysis (3rd ed.)*. Chicago: A University of Chicago Press.

Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyzes Relating to Achievement*. Routledge.

Kane, T., & Cantrell, S. (2012). *Gathering feedback for teaching. Combining high-quality observations with student surveys and achievement gains*. MET Project Research Paper, Bill & Melinda Gates Foundation, Seattle, WA. http://files.eric.ed.gov/fulltext/ED540960.pdf

Kane, T., & Cantrell, S. (2010). *Learning about teaching: Initial findings from the measures of effective teaching project*. MET Project Research Paper, Bill & Melinda Gates Foundation, Seattle, WA. https://docs.gatesfoundation.org/Documents/preliminary-findingsresearch-paper.pdf

Kington, A., Day, C., Sammons, P., Regan, E., & Brown, E. (2009, September 2–5). *Effective classroom practice – A mixed method study of influences and outcomes*. The British Educational Research Association Annual Conference, University of Manchester, UK.

Klieme, E. (2012, August 23). *Qualities and effects of teaching: Integrating findings across subjects and cultures*. EARLI Sig Educational Effectiveness Conference, Zurich, Switzerland.

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

Klusmann, U., Kunter, M., Trautwein, U., Ludtke, O., & Baumert, J. (2008). Teachers' Occupational Well-Being and Quality of Instruction: The Important Role of Self-Regulatory Patterns. *Journal of Educational Psychology*, *100*(3), 702–715. https://doi.org/10.1037/0022-0663.100.3.702

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

Kyriakides, L., Creemers, B. P. M., & Antoniou, P. (2009). Teacher Behaviour and Student Outcomes: Suggestions for Research on Teacher Training and Professional Development. *Teaching and Teacher Education 25*(1), 12–23. https://doi.org/10.1016/j.tate.2008.06.001

Kyriakides, L., Creemers, B. P. M., Panayiotou, A., & Charalambous, E. (2020). *Quality and Equity in Education: Revisiting Theory and Research on Educational Effectiveness and Improvement*. Routledge.

Kyriakides, L., Creemers, B. P. M., Panayiotou, A., Vanlaar, G., Pfeifer, M., Cankar, G., & McMahon, L. (2014). Using student ratings to measure quality of teaching in six European countries. *European Journal of Teacher Education*, *37*(2), 125–143. https://doi.org/10.1080/02619768.2014.882311

Lorenzo-Seva, U., & Ferrando, P. J. (2003). IMINCE: An unrestricted factor-analysis-based program for assessing measurement invariance. *Behavior Research Methods, Instruments & Computers*, *35*(2), 318–321. https://doi.org/10.3758/bf03202558

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99. https://doi.org/10.1037/1082-989x.4.1.84

Mueller R. O. (1996). Confirmatory Factor Analysis. In: *Basic Principles of Structural Equation Modeling*. Springer Texts in Statistics. Springer. https://doi.org/10.1007/978-1-4612-3974-1_2

Muijs, D., & Reynolds, D. (2000). School effectiveness and teacher effectiveness in mathematics: Some preliminary findings from the evaluation of the Mathematics Enhancement Program (Primary). *School Effectiveness and School Improvement, 11(3),* 273–303. https://doi.org/10.1076/0924–3453 (200009)11:3;1-g;ft273

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B. P. M., Timperley, H., & Earl, L. (2014). State of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*(2), 231–256. https://doi.org/10.1080/09243453.2014.885451

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks*. Timss & Pirls.

Nilsen, T., & Gustafsson, J. E. (2016). *Teacher quality, instructional quality and student outcomes: relationships across countries, cohorts and time*. Springer Nature.

OECD. (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. OECD Publishing.

OECD. (2019). *PISA 2018 Assessment and Analytical Framework.* OECD Publishing. https://doi.org/10.1787/b25efab8-en

Rissanen, I., Kuusisto, E., Hanhimäki, E., & Tirri, K. (2018). Teachers' Implicit Meaning Systems and Their Implications for Pedagogical Thinking and Practice: A Case Study from Finland. *Scandinavian Journal of Educational Research, 62*(4), 487–500. https://doi.org/10.1080/00313831.2016.1258667

Rose, J. S., & Medway, F. J. (1981). Measurement of Teachers' Beliefs in Their Control over Student Outcome. *The Journal of Educational Research, 74*(3), 185–190. https://doi.org/10.1080/00220671.1981.10885308

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. University of Tennessee Value-Added Research and Assessment Center.

Scheerens, J. (2000). *Improving school effectiveness* (Fundamentals of Educational Planning No. 68). Paris, France: UNESCO/International Institute for Educational Planning.

Scheerens, J. (2016). *Educational effectiveness and ineffectiveness: a critical review of the knowledge base*. Springer.

Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of Teacher Self-Efficacy and Relations with Strain Factors, Perceived Collective Teacher Efficacy, and Teacher Burnout. *Journal of Educational Psychology*, *99*, 611–625. https://doi.org/10.1037/0022-0663.99.3.611

Teodorović, J. (2011). Classroom and school factors related to student achievement: What works for students? *School Effectiveness and School Improvement, 22*(2), 215–236. https://doi.org/10.1080/09243453.2011.575650

Teodorović, J., Milin, V., Bodroža, B., Đerić, I., Vujačić, M., Jakšić, I., Stanković, D., Cankar, G., Charalambous, C., Van Damme, J., & Kyriakides, L. (2021). Testing the dynamic model of educational effectiveness: The impact of teacher factors on interest and achievement in mathematics and biology in Serbia. *School Effectiveness and School Improvement*. Advance online publication. https://doi.org/10.1080/09243453.2021.1942076

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing and elusive construct. *Teaching and Teacher Education, 17*, 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

Tucker, L. R. (1951). A method for synthesis of factor analysis studies. *Personnel Research Section Report No.984.* Department of the Army. https://doi.org/10.21236/ad0047524

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57–67. https://doi.org/10.1023/A:1007999204543

Xu, H., & Tracey, T. J. G. (2017). Use of Multi-Group Confirmatory Factor Analysis in Examining Measurement Invariance in Counseling Psychology Research. *The European Journal of Counseling Psychology, 6*(1), 75–82. https://doi.org/10.5964/ejcop.v5i2.120

Validacija skala za merenje faktora kvaliteta nastave iz Dinamičkog modela obrazovne efektivnosti

Bojana Bodroža[1], Jelena Teodorović[2] i Smiljana Jošić[3]

[1]Odsek za psihologiju, Filozofski fakultet, Univerzitet u Novom Sadu, Srbija
[2]Fakultet pedagoških nauka, Univerzitet u Kragujevcu, Jagodina, Srbija
[3]Institut za pedagoška istraživanja, Beograd, Srbija

Obimna istraživanja obrazovne efektivnosti zahtevaju validne učeničke upitnike za procenu nastavne prakse. U ovom istraživanju je validirano osam skala za merenje nastavnih faktora iz Dinamičkog modela obrazovne efektivnosti (eng. the Dynamic Model of Educational Effectiveness; DMEE). Konstruisane su paralelne forme skala za merenje faktora nastave matematike i biologije i validirane u dve studije. U prvoj studiji je primenjena eksplorativna faktorska analiza na podacima iz uzorka koji su činila 683 učenika. U drugoj studiji njihova struktura je kros-validirana konfirmativnom faktorskom analizom na uzorku od 5.476 učenika. Rezultati višegrupne konfirmativne faktorske analize su potvrdili postojanje metričke invarijantnosti svih skala što ukazuje na to da su njihova faktorska zasićenja veoma slična. Međutim, skalarna invarijantnost modela nije potvrđena što ukazuje na to da se skale ne mogu koristiti za poređenje nastavnika različitih predmeta. Testiranje alternativnih strukturnih relacija između nastavnih faktora nije ukazalo da se podaci adekvatno uklapaju u DMEE model iako su indeksi slaganja bili bolji nego za alternativne teorijske modele. Eksterna validnost za matematiku je pokazala da skale koreliraju sa zadovoljstvom poslom, eksternom kontrolom i samoefikasnošču nastavnika procenjenom od strane samih nastavnika. Skale su pouzdane i validne i mogu se primenjivati na različite školske predmete.

Ključne reči: kvalitet nastave, validacija skala, dinamički model obrazovne efektivnosti