

Placing Kosovo in the didaktik-curriculum continuum: A quantitative, international comparative perspective

Armend Tahirsylaj

Norwegian University of Science and Technology, Norway

This study examined the variation in four assessment practices, including mandatory standardised tests, non-mandatory standardised tests, teacher-developed tests, and teachers' judgmental ratings in Kosovo and six didaktik and five curriculum countries, and associations of the four assessment practices with students' science performance, using data from PISA 2015. The main objective of the study was to place Kosovo in the didaktik-curriculum continuum, while testing the differences in the key variables between the two traditions. Relying on mean comparison, Mann-Whitney two-sample rank-sum tests, and hierarchical linear modelling for data analysis, the findings suggest that Kosovo is placed towards the curriculum end of the continuum as a result of the shift of education policy towards curriculum tradition-based policies such as standardised testing over past 20 years. Overall, the findings reveal that theoretical claims about differences between didaktik and curriculum traditions regarding teacher autonomy still hold when tested empirically and quantitatively and using assessment practices as evidence. The hierarchical linear modelling results were underwhelming, as in most sample countries the four assessment practices were not statistically and significantly associated with students' science performance, with a few exceptions. The article ends with potential avenues for further research warranted by the results of the present study.

Introduction and purpose

Kosovo, a small landlocked country in the Western Balkans, has gone through an eventful history over the past 20 years. Developments in the education sector have been affected by the political, social, and cultural changes that accompanied Kosovo's transitions from being an occupied country in the 1990s, to a territory under the United Nations protectorate during 2000s, to an independent country since 2008. Kosovo became the youngest country in Europe when it declared independence from Serbia in 2008. Traditionally, Kosovo's education system has been under the influence of didaktik education tradition, albeit one that arrived to Kosovo as part of former Yugoslavia since the end of the Second World War. Also, the didaktik model dominant in Kosovo more closely followed didactical teacher authority-based approaches imported to former Yugoslavia via Russia's influence, a model often dubbed 'monitorial instruction' (Hopmann, 2015, "Encounters," par. 5), rather than Bildung-based didaktik of Western and/or Northern Europe tradition. Still, since the end of the most recent war in Kosovo in 1999, educational developments have put Kosovo under varied influences arriving through national and multi-national organisations in the form of direct or indirect technical assistance. Subsequently, after 20 years of education reforms, it is not clear what education tradition is more dominant in Kosovo, not least because of lacking comparable data internationally. However, this situation changed when Kosovo decided to participate in international large-scale assessments to compare the performance of its students with those of other countries in Europe and globally.

Kosovo's participation in the OECD *Programme for International Student Assessment* (PISA) 2015 was its first ever in any international large-scale assessment. Thus, with publication of PISA data in December 2016, Kosovo education data became available to address various research questions from a comparative perspective. To this end, it is important to pinpoint where Kosovo stands in the didaktik-curriculum education traditions continuum, an area of international comparative education studies that has gained traction among educational researchers since 1990s (e.g. Gudem & Hopmann, 1998; Westbury, Hopmann & Riquarts, 2000; Hudson, 2007; Biesta, 2011; Pantić & Wubbels, 2012; Pettersson et al., 2015; Tahirsylaj, Niebert & Duschl, 2015). While earlier studies have focused on theoretical, policy, and discourse analyses of similarities and differences between didaktik and curriculum traditions, a small, but growing, number of researchers are turning to quantitative studies to examine didaktik and curriculum traditions comparatively (Tahirsylaj, 2019; Wermke & Pröitz, 2019). In this regard, this study explored PISA 2015 data from the school survey to examine variation in assessment practices in Kosovo in comparison with a representative set of six didaktik countries - Austria, Germany, Denmark, Norway, Sweden and Finland; and five curriculum countries - United Kingdom, Ireland, Canada, USA, and New Zealand; following Tahirsylaj's (2019) criteria and rationale for categorisation of countries into didaktik and curriculum grouping.

The study addressed two main research questions: (1) What is the variation in assessment practices in Kosovo in comparison with didaktik and curriculum countries in the sample in PISA 2015? (2) How are assessment practices items associated with students' science performance in Kosovo and in the representative didaktik and curriculum countries in PISA 2015? Descriptive and inferential statistical approaches are applied to examine the questions, using a framework that distinguishes between the German (and to some extent Nordic) didaktik education tradition and Anglo-Saxon curriculum tradition. Addressing the research questions is aimed at meeting a number of research objectives.

The first major objective of the study is to place Kosovo in the didaktik-curriculum continuum by examining the variation in assessment practices across countries in the sample, since due to lacking data, there is very little, if any, research about Kosovo education context or student achievement studies applying quantitative methods. A second objective is to apply the didaktik-curriculum framework on addressing educational issues pertaining to the use of varied assessment practices and how they are associated with students' performance in comparative achievement studies. This is a new and innovative way of exploring education phenomena as the field has been dominated by looking at educational issues primarily from sociological and economic frameworks and theories. A third objective is to continue to test empirically the theoretical claims made about differences between didaktik and curriculum. Lastly, the study seeks to extend the dialogue over didaktik and curriculum education traditions and examine how they are becoming more similar or different as a result of global trends in education, when considering empirical data collected internationally.

Theoretical framework, literature review, and context

Didaktik and curriculum are the two most prevalent educational traditions that provide the framework for education systems in the Western world (Biesta, 2011; Tahirsylaj et al., 2015), also referred to as "... dominant modes of understanding of schooling ..." (Hopmann, 2015, "Introduction," para. 4). German *Didaktik* theory is central to curriculum, teaching and learning in Continental Europe generally and the German speaking world specifically, as well as in Nordic Europe, but is mostly unknown in the English speaking world (Hopmann, 2007; Westbury et al., 2000; Kansanen, 1995). Here the original German term *Didaktik* is used throughout to avoid the use of English term *didactics*, which is often ascribed negative connotations such as frontal teaching (Kansanen, 1995). In its original conceptualisation, "Didaktik is about how teaching can instigate learning, but learning that is a content-based student activity not as swallowing a sermon or a monologue or otherwise one-sided distribution of knowledge by a teacher" (Hopmann, 2007, p. 113).

Curriculum, on the other hand, is a widely applied tradition amongst many countries, primarily in the English-speaking world. Curriculum here refers to the prevailing curriculum model that has been in place in the U.S. since early 1900s, when the so-called social efficiency model of curriculum won the American education battle against humanistic-based models of curriculum (Kliebard, 2004). Didaktik and curriculum frameworks claim, amongst else, that there is higher level of teacher autonomy among teachers working in didaktik than those in curriculum countries (Westbury et al., 2000) and that curriculum tradition is more assessment-oriented than didaktik (Tahirsylaj et al., 2015). This issue of teacher autonomy is relevant for the present study as it assists in differentiating between teacher-based assessment practices such as teacher-developed tests and teacher judgmental ratings, and externally developed tests such as mandatory or non-mandatory standardised tests. The use of the former practices would indicate a higher level of autonomy, while the use of the latter would show a lower level of teacher autonomy since such choices reflect the decision-making of teachers within their classrooms, and external pressures they are influenced by.

Viewed from the didaktik/curriculum education traditions, it can be argued that Kosovo used to apply a 'downgraded' version of didaktik in the past, with the main emphasis on transmission of disciplinary knowledge. However, since the end of the war in 1999, international aid and education programs have pushed Kosovo to adopt national external standardised testing as an accountability measure, a practice employed and promoted through and from the curriculum tradition (Tahirsylaj, 2020; Tahirsylaj, 2018). In turn, the findings of the present study address a relevant issue as it seeks to place Kosovo on the didaktik/curriculum continuum based on quantitative data collected and made available through OECD's PISA study, which in turn reveals whether Kosovo teachers enjoy higher or lower level of autonomy in decision-making about assessment practices.

Literature review

Educational research covering educational issues in the Kosovo context is scarce, and investigating issues quantitatively almost inexistent. Therefore, the use of PISA data to address educational questions pertaining to Kosovo context fills an important gap in the literature. However, where literature exists, questions that have been addressed were related to the extent to which Kosovo students have opportunities to develop learning to learn competences and civic competences within the framework of latest competence-based curriculum reform (Tahirsylaj, 2020; Tahirsylaj, 2018). Next, Tahirsylaj and Wahlström (2019) have examined the role that education policy context plays in mastery of critical thinking competences; Saqipi (2019) analysed the role that context plays in challenging environments such as Kosovo in adopting far reaching educational reforms such as competence-based curriculum; while Shala and Grajcevcic (2018) have examined Kosovo students' low performance in PISA 2015 from a socio-economic perspective and socio-economic-related variables such as parental education.

In another study, Grajcevcic and Shala (2021) analysed Kosovo PISA 2015 data to examine the role of teacher characteristics, such as teacher availability, training and collaboration, classroom management, teacher responsibility and tendencies, and teaching style/pedagogical techniques, on students' performance, and found that teacher characteristics played a detrimental role in Kosovo students' performance, contrary to previously reported findings internationally that teacher characteristics are positively associated with students' performance. No prior studies have examined assessment practices in use, or their relationship with students' performance in PISA in the Kosovo context. The selection of assessment practices is a strategic decision to focus the study on assessment as a factor that differentiates curriculum and didaktik traditions as it has already been shown that curriculum is more assessment intensive compared to didaktik (Tahirsylaj et al., 2015).

Internationally, few studies have examined the role of assessment practices in students' performance in international large-scale assessments such as PISA or *Trends in International Mathematics and Science Study* (TIMSS). For example, Liang (2010) found only weak associations of teacher-developed tests as an assessment practice with PISA 2015 students' mathematics performance in Canada, the US, and Finland. Using TIMSS data and focusing on the US sample, Rodriguez (2004) found stronger significant associations of teacher classroom assessment practices with students' classroom performance. Other studies were mostly completed within specific national contexts focusing on the role of classroom assessment practices on students' motivation or their achievement goals (e.g. Alkharusi, 2008; Brookhart & Durkin, 2003; Brookhart, 2013). In the OECD's own volumes related to PISA 2015 students' performance, it is reported that teacher-developed tests and teachers' judgmental ratings are the most frequent assessment practices, while mandatory and non-mandatory standardised tests are used less, across OECD countries (OECD, 2016a). The contribution of the present study fills the gap in literature regarding the uses of assessment practices in Kosovo and in comparison to the reference countries included in the sample. Further, the study offers a more elaborate framing of the role

assessment practices play on students' science performance relying on established didaktik and curriculum education traditions.

An overview of education systems in Kosovo and reference countries

A number of major education reforms have been introduced in Kosovo in past 20 years, of which the 2011 competence-based curriculum was the most consequential. It shifted the curriculum approach from a traditional content-based one to the competence-based one, to better align local education with the reforms at the European Union level, particularly the 2006 recommendation of the European Commission on eight key competences for lifelong learning (MEST, 2011; OJEU, 2006). This major change also could be considered as a clear shift of Kosovo's education towards curriculum tradition ideology of efficiency, as the most dominant ideology within the curriculum tradition (Tahirsylaj, 2017; Schiro, 2013; Deng & Luke, 2008). Therefore, from the education and curriculum policy perspective, Kosovo can be categorised as a curriculum country as opposed to the previously dominant didaktik tradition, and this study explores this issue empirically and quantitatively. Further, other reforms have also shaped Kosovo's education system in recent years, including the introduction of standardised testing at the end of grade 5, grade 9, and grade 12, major teacher professional development programs, and restructuring of teacher education programs (Tahirsylaj & Wahlström, 2019; Tahirsylaj, 2018).

Overall, considering Kosovo's education context in comparison with all other 11 countries in the sample, there are some similarities and differences worth noting. First, there is a clear tendency across all countries to expand education towards early childhood education, even though in Kosovo there is still very low participation of students in any early childhood education programs (MEST, 2018). For example, in the 2017/2018 school year, there were about 5000 under 5 year-old students in preschool programs compared to 21,500 students in the mandatory pre-primary grade including 5-6 year-old age group (MEST, 2018). Next, almost all countries follow a similar structure of four or five years of elementary education, four/five years of lower secondary, and three or four years of upper secondary education, with the exception of Ireland, which divides its education into two phases only – primary schooling including ages 4-12 years, and post-primary schooling including 12-17/18 years. Regarding content covered in respective pre-university curricula, some differences between didaktik and curriculum countries are present, most notably regarding tracking of students into academic and vocational tracks of education, with curriculum countries providing a more comprehensive, less-vocational-based education, and didaktik countries varying between less tracking in Nordic countries and more vocational-based tracking in Austria and Germany – with Germany being a clear outlier with introduction of tracking as early as grade 5.

Kosovo follows the didaktik model at upper secondary education where there is about 50/50 percent division of students into academic and vocational tracks, but similar to curriculum countries, there is no tracking before grade 10 (start of upper secondary education). Regarding monitoring and evaluation of schools' performance, countries have legislative frameworks in place that provide guidelines for internal and external evaluation

of schools' and sometimes teachers' performance, as well as students' performance in standardised tests, and usually this supervisory and evaluative work is undertaken by national educational agencies focusing in pre-university education (e.g. *Skolverket* - National Agency for Education - in Sweden or Ministry of Education, Science and Technology in Kosovo) (Eurydice, 2021). In terms of PISA performance, all 11 reference countries have been performing and perform either at or above the OECD average of 500 score points.

Table 1 shows that only United States in mathematics (470), and Austria in reading (485) achieved a PISA 2015 average score that is statistically significantly below the OECD average in. Kosovo, in comparison, scored significantly below the OECD average in 2015 and well lower than any of the 11 reference countries, with scores of 347 in reading, 362 in mathematics, and 378 in science.

Table 1: Representative study sample of Kosovo, and didaktik and curriculum countries and their test scores in PISA 2015 on three achievement scales

Country	PISA 2015 achievement scale		
	Overall reading	Mathematics	Science
Finland (didak.)	526	511	531
Canada (curric.)	527	516	528
New Zealand (curric.)	509	495	513
Kosovo (curric.)	347	362	378
Norway (didak.)	513	502	498
United States (curric.)	497	470	496
Sweden (didak.)	500	494	493
Germany (didak.)	509	506	509
Ireland (curric.)	521	504	503
Denmark (didak.)	500	511	502
United Kingdom (curric.)	498	492	509
Austria (didak.)	485	497	495
Cell colour key		Statistically significantly above the OECD average	
		Not statistically significantly different from the OECD average	
		Statistically significantly below the OECD average	

Note: Didaktik countries in red; curriculum countries in blue.
Kosovo is placed under curriculum countries for practical reasons here.
Source: Adapted from OECD, PISA 2015 database.

Data and methods

This study used PISA 2015 data collected through student achievement tests, student background questionnaires and school questionnaires completed by the school principal. PISA tests 15-year old students' skills in three cognitive domains including science, mathematics, and reading, on a rotation basis, the focus of the 2015 study being on

science literacy (OECD, 2016b). Data from 12 countries are used, including Kosovo as the country of focus, and 11 reference countries, where six representing didaktik include Denmark, Finland, Norway, Sweden, Austria and Germany, and five representing curriculum include Canada, Ireland, New Zealand, UK and USA. Table 2 shows the sample sizes for schools and students in individual countries. In total, the data used come from 3517 schools and 92482 students.

Table 2: Representative study sample of countries, schools and students

Didaktik			Curriculum		
Country	No. schools	No. students	Country	No. schools	No. students
AUT	269	7007	KOS	224	4826
DEU	256	6504	CAN	759	20058
DNK	333	7161	GBR	550	14157
FIN	168	5882	IRL	167	5741
NOR	229	5456	NZL	183	4520
SWE	202	5458	USA	177	5712
Total	1457	37468	Total	2060	55014

Source: Adapted from OECD, PISA 2015 database.
 Note: Didaktik countries in red; curriculum countries in blue.
 Kosovo (KOS) is placed under curriculum countries for practical reasons here.

Four variables are used as measures of assessment practices, derived from the following question to which school principals responded: Generally, in your school, how often 15 year olds are assessed using: a. standardised tests (mandatory); b. standardised tests (non-mandatory); c. teacher developed tests; and d. teachers' judgmental ratings. The answers range was never; 1-2 times a year; 3-5 times a year; monthly; and more than once a month (recoded from 0 to 4). The definitions for the four measures adopted for the present study rely on those provided in the OECD documentation so that the conceptualisations remain closer to what was meant to be collected when OECD developed the data collection instruments. In this regard, standardised mandatory tests, including those mandated, for example, by national, state or district authorities, as well as standardised non-mandatory tests, including for example publicly or commercially available standardised test materials, are tests that are consistent in design, content, administration and scoring, and thus results can be compared across students and schools, while teacher-developed tests, and teachers' judgmental ratings are tests or assessment practices developed by teachers for use in specific classrooms and thus results cannot be compared across students or schools at national level (OECD, 2016b).

To address the first research question, (1) "What is the variation in assessment practices in Kosovo in comparison with didaktik and curriculum countries in the sample?", mean comparison is presented to examine whether the means of assessment practices in individual countries were different from one another. This descriptive procedure is helpful to test the hypothesis whether didaktik countries show higher use of teacher-based assessment practices than curriculum countries, as the respective framework suggests, and where Kosovo is placed in that continuum. The question was broken down to four individual variables and respective dataset from PISA 2015.

For the second question, (2) "How are assessment practices associated with students' science performance in Kosovo and in the representative didaktik and curriculum countries?", linear random intercept models are fitted for each of the 12 countries in the sample, with level-1 covariates including economic, social and cultural status (ESCS), gender, age, grade, immigration status, and level-2 covariate related school type, where school is the second level in the hierarchical linear model (HLM). HLM is employed to examine the association of assessment practice items in particular with students' science performance in PISA 2015 in Kosovo and 11 other countries in the sample to capture the nature of nested data applied in the PISA dataset (Raudenbush & Bryk, 2002).

To develop the HLM models, first an unconditional model was run for each country using the dependent variable, i.e. students' science performance, which is captured applying equation (1):

$$science_{ij} = \beta_{0j} + e_{ij} \quad (1)$$

Each school's intercept, β_{0j} , is then set equal to a grand mean, γ_{00} , and a random error u_{0j} .

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

where j represents schools and i represents students with a given country. Substituting equation (2) into (1) produces

$$science_{ij} = \gamma_{00} + u_{0j} + e_{ij} \quad (3)$$

where: β_{0j} is mean science achievement for school j

γ_{00} is grand mean for science achievement

$\text{Var}(e_{ij}) = \theta$ is within school variance in science achievement

$\text{Var}(u_{0j}) = \tau_{00}$ is between school variance in science achievement

This model explains whether there is variation in students' standardised science scores across j schools for the given country. From here, a linear random-intercept model with covariates was set up. This model is an example of a linear mixed effects model that splits the total residual or error into two error components. It starts with a multiple-regression model, as follows:

$$Science\ scores_{ij} = \beta_1 + \beta_2 x_{2ij} + \dots + \beta_p x_{pij} + \zeta_{ij} \quad (4)$$

Here β_1 is the constant for the model, while $\beta_2 x_{2ij}$ to $\beta_p x_{pij}$ represent covariates included in the given model. ζ_{ij} is the total residual that is split into two error components, equation (5):

$$\zeta_{ij} = u_j + e_{ij} \quad (5)$$

where u_j is a school specific error component representing the combined effects of omitted school characteristics or unobserved heterogeneity. It is a random intercept or the level-2 residual that remains constant across students, while level-1 residual e_{ij} is a student-

specific error component, which varies across students i as well as schools j . Substituting ξ_{ij} into the multiple-linear regression model (4), we obtain the linear random-intercept model with covariates, equation 6:

$$\text{Science scores}_{ij} = \beta_1 + \beta_2 x_{2ij} + \dots + \beta_p x_{pij} + u_j + e_{ij} \quad (6)$$

$\beta_2 x_{2ij}$ to $\beta_p x_{pij}$ represent the covariates included in the model, and they vary depending on how many covariates are included in a specific model. The final model focuses on four level-2 covariates representing assessment practice items, and it also includes one school-level covariate of school type (public vs. private) as well as a number of student level-1 covariates including students' socio-economic status (SES), gender (girl), age, grade, immigration status, and controlling for dummy missing variables. Mean substitution was used to address missing data in HLM models. The same full HLM model is then run for each of the twelve countries in the study. The study relies entirely on secondary data analyses and does not create new scales of any sort, and as such relies on variables already on the PISA datasets; therefore, validity and reliability issues corresponding to assessment practices and science test scores for example are dealt with in PISA documentation and technical reports (e.g. OECD, 2016b).

Results

Overall, the findings do not show a clear-cut difference in means of assessment practices across countries, while assessment practice items are not strongly associated with students' science performance in many of the countries in the sample. The detailed results are provided below, first related to the first research question on the use of assessment practices across countries in the sample, and then related to associations of assessment practices to students' science performance in PISA 2015.

Means of assessment practices across sample countries

Figure 1 shows the means of use of standardised tests for each country in the sample, a variable coded on a scale from 0 to 4, where 0 indicates no use of standardised tests and 4 indicates the use of standardised tests more than once a month.

As Figure 1 shows, didaktik countries in red report less use of standardised tests, especially Germany with 0.44. The exception is Sweden, which reports the highest use of standardised tests at 1.56, indicating the students in Sweden on average take these tests at least once a year or more. Curriculum countries on average report higher use of standardised tests than didaktik countries. Kosovo, as shown in the graph, is located on the curriculum end of the continuum on this item with a score of 1.25. A Mann-Whitney two-sample rank-sum test was used to compare curriculum and didaktik countries as two groups. The overall analysis resulted with curriculum countries having a higher sum of ranks compared to the expected rank sums under the null hypothesis than didaktik countries and there was a significant difference of $z = -34.03$ with $p < 0.001$. As a result, taken together curriculum countries have a higher mean on the use of standardised tests than didaktik countries. The higher use of standardised tests indicates a higher influence

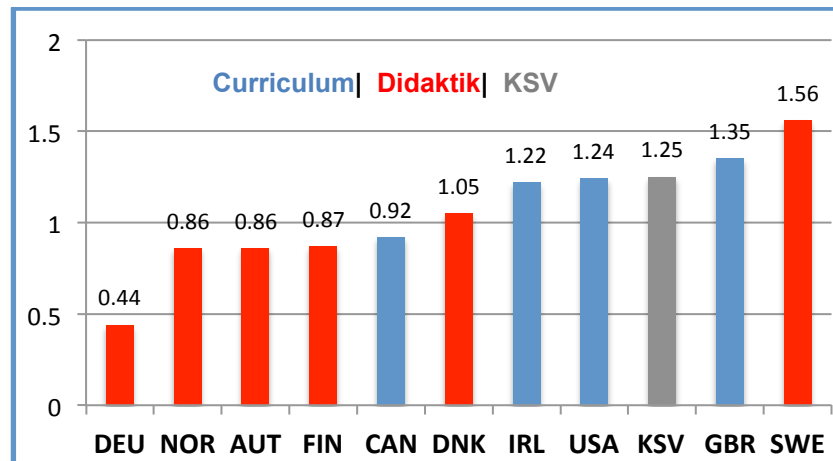


Figure 1: Means of use of mandatory standardised tests per country
 0=Never; 1=1-2 a year; 2=3-5 a year; 3=monthly; 4=more than once a month.

of external stakeholders on the use of assessment practice within the classroom, which in turn reduces teachers' autonomy to make decisions on their own. The results here show that teachers in didaktik countries, on average, are less influenced by externally mandated standardised tests than curriculum countries.

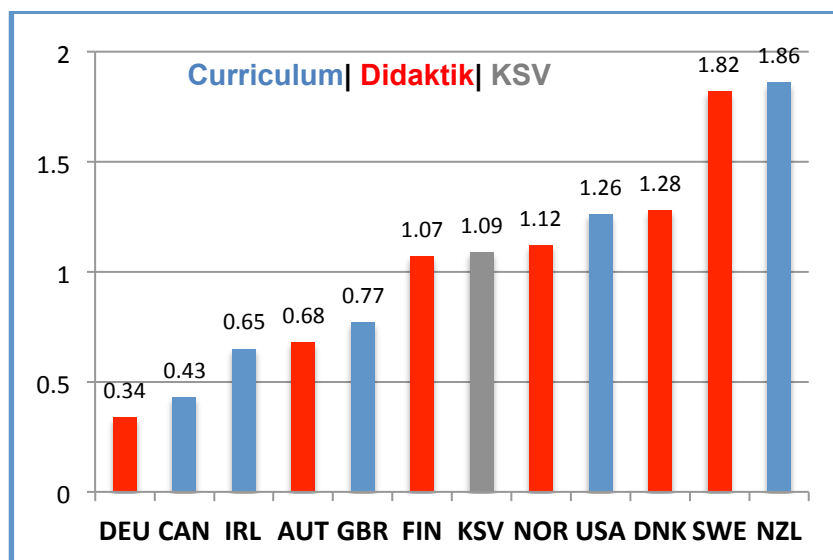


Figure 2: Means of non-mandatory standardised tests per country
 0=Never; 1=1-2 a year; 2=3-5 a year; 3=monthly; 4=more than once a month.

Figure 2 shows the means of use of non-mandatory standardised tests by each country in the sample. Countries of the two groups do not follow a pattern on this variable, with New Zealand and Sweden as outliers with the highest means in the use of non-mandatory

standardised tests, while Kosovo is in the middle of continuum. The Mann-Whitney two-sample rank-sum test comparing didaktik and curriculum countries showed a higher mean of use of these tests by curriculum than didaktik countries and there was a significant difference of $z = 57.81$ with $p < 0.001$.

Figure 3 shows the means of use of teacher-developed tests per country. On average, the didaktik countries are grouped in the middle of the continuum, Canada and USA are outliers with the highest use of this assessment practice with 3.67 and 3.63 respectively, and interestingly, Kosovo shows the lowest use of these tests with 2.22 among all countries in the sample. Indeed, with exception of Kosovo, Denmark, and Ireland, all other countries have a mean close to 3 or higher, meaning teachers in those countries are reported to use tests developed by them on a monthly basis during the school year. The results from Mann-Whitney rank-sum test was indicated that curriculum countries in total had a higher sum of ranks than the expected sum of ranks, while for didaktik countries the sum of ranks was lower than expected and there was a significant statistical difference of $z = -61.66$, with $p < 0.001$. Here, teachers in curriculum countries, on average, are reported to be more in control of the decision-making about the use of an assessment practice developed by them. However, if teacher-developed tests represent a form of preparation for students to take standardised tests rather than for teachers' classroom use for within-classroom summative purposes then, indirectly, a higher mean on this variable might indicate teachers making decisions about assessment practices under the external pressure of standardised tests.

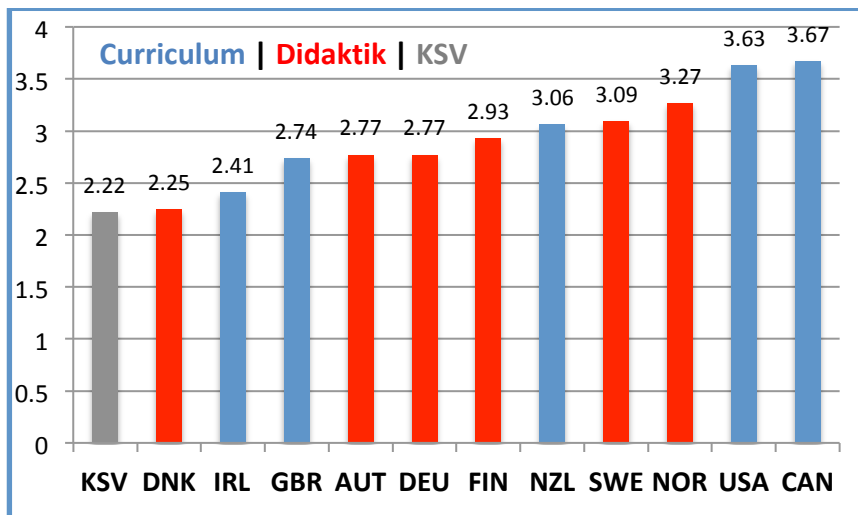


Figure 3: Means of use of teacher-developed tests per country
 0=Never, 1=1-2 a year, 2=3-5 a year, 3=monthly, 4=more than once a month

Lastly, the results for the fourth assessment practice of teachers' judgmental ratings are shown in Figure 4. Here, a clear separation between curriculum and didaktik countries is observed with all didaktik countries grouped in the higher end of the continuum, and all

curriculum countries grouped on the lower end. Kosovo, with a score of 2.48 is grouped in the curriculum end of the continuum. The Mann-Whitney rank-sum test confirmed the statistical difference of $z = 94.94$ at $p < 0.001$ between the curriculum and didaktik countries, with didaktik countries having higher sums of ranks than expected while curriculum countries' sums of ranks were lower than expected.

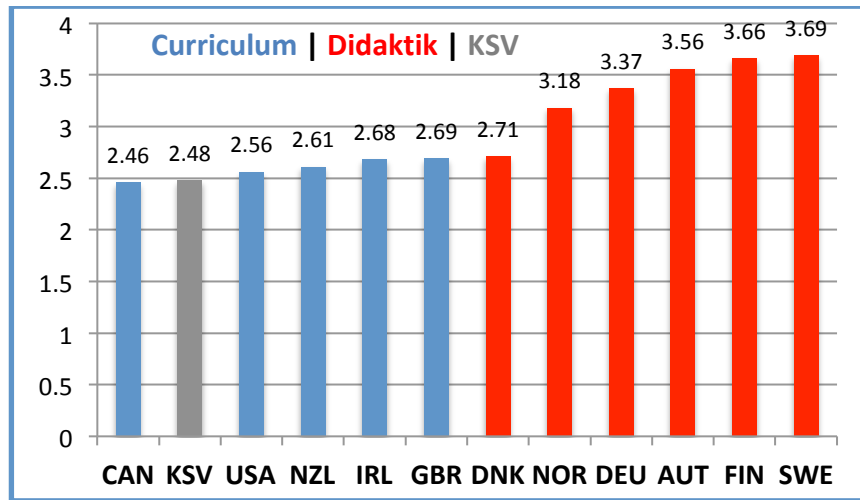


Figure 4: Means of use of teachers' judgmental ratings per country
0=Never; 1=1-2 a year; 2=3-5 a year; 3=monthly; 4=more than once a month.

Table 3: HLM results showing associations of model variables with science performance in PISA 2015 - Kosovo and curriculum countries

	Kosovo	Canada	United Kingdom	Ireland	New Zealand	United States
Mandatory tests	n.s.	n.s.	-10.24	n.s.	no data	n.s.
Non-mandatory tests	n.s.	n.s.	-7.62	-13.02	n.s.	n.s.
Teacher-developed	n.s.	n.s.	n.s.	n.s.	n.s.	13.31
Judgmental ratings	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
SES	8.35	24.42	22.02	31.41	33.09	16.53
Gender	4.21	-5.21	n.s.	-10.86	-8.86	-11.53
Age	-21.59	n.s.	n.s.	n.s.	n.s.	-20.66
Grade	-21.12	39.69	29.76	8.77	27.32	40.49
First immigration	n.s.	n.s.	-22.91	n.s.	-11.35	-18.81
Second immigration	-23.66	n.s.	n.s.	n.s.	n.s.	n.s.
Public school	-37.88	-27.19	-47.79	-15.07	-58.87	n.s.

Notes: Only results significant at $p < 0.05$, $p < 0.01$ and $p < 0.001$ shown. If **bold**, significant at $p < 0.001$. n.s.=not significant. Gender here represents 'girls' and boys are the reference group.

On average, teachers' judgmental ratings is the most used assessment practice across the countries in the sample, and interestingly this variable shows the clearest division between the didaktik and curriculum countries, with Kosovo being placed among curriculum

countries. Next, maximum likelihood estimates from HLM models are provided in Tables 3 and 4 for curriculum and didaktik countries respectively, with Kosovo results presented in Table 3.

Table 4: HLM results showing associations of model variables with science performance in PISA 2015 - Didaktik countries

	Austria	Germany	Denmark	Finland	Norway	Sweden
Mandatory tests	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Non-mandatory tests	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Teacher-developed	n.s.	n.s.	n.s.	n.s.	10.84	9.84
Judgmental ratings	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
SES	9.11	9.96	23.25	30.08	29.25	28.52
Gender	-24.42	-24.07	-11.74	14.68	n.s.	n.s.
Age	n.s.	-18.57	n.s.	n.s.	17.04	n.s.
Grade	40.39	38.85	44.86	36.98	45.91	67.99
First immigration	-35.01	-37.99	-49.48	-72.51	-38.98	-51.67
Second immigration	-37.04	-32.05	-47.03	-61.38	-33.85	-33.71
Public school	n.s.	-58.69	n.s.	-29.07	n.s.	n.s.

Notes: Only results significant at $p < 0.05$, $p < 0.01$ and $p < 0.001$ shown. If **bold** significant at $p < 0.001$. n.s. indicates results were not significant. Gender here represents ‘girls’ and boys are the reference group.

Associations of assessment practices with students’ science performance in PISA 2015

The second question addressed in the study examined the associations of the four main variables of interest with students’ science performance in PISA 2015, controlling for seven student-related and school-related variables. To this end, the results provide evidence of whether the use of any of the assessment practices by teachers in the classroom settings across countries in the sample influenced students’ science performance in PISA 2015. Overall, across all countries, only in a few cases the assessment practices statistically and significantly affected science scores in PISA 2015. As Table 3 shows, mandatory standardised tests were statistically significant and negative only in UK, non-mandatory standardised tests were statistically significant and negative in UK and Ireland, teacher developed tests were statistically significant and positive in the US, and teachers’ judgmental ratings were not statistically significant in any of the curriculum countries. None of the assessment practices were statistically significant in Kosovo. The results show that based on the HLM model and controlling for the given variables, the use of assessment practices is either not significant or negative, and positive only in the case of teacher-developed tests in the US.

Regarding six level-1 covariates – SES, as expected, was significant and positive in all curriculum countries. Interestingly, in Kosovo the SES coefficient is the lowest at 8.35 showing that SES background did not largely affect students’ scores in Kosovo. Being a girl was significant and positive only in Kosovo (4.21), while significant and negative in Canada, Ireland, New Zealand, and the US, and not significant in the UK. Age was

significant and negative in Kosovo and the US, and not significant in other countries as shown in Table 3, while grade was significant and positive in all countries but Kosovo where it was significant and negative. The two immigration background variables were significant and negative in Kosovo, the UK, New Zealand, and the US, and not significant in others. The level-2 covariate, public school, was significant and negative in all countries but the US, indicating that students' in public schools in these countries are expected to perform worse in science than their peers attending privately-run schools, as per the model employed and covariates used.

As shown in Table 4, the assessment practices among didaktik countries were also overwhelmingly not significant, being significant and positive only in the case of the use of teacher-developed tests in Norway and Sweden, suggesting that the use of this specific assessment practice contributed to higher science performance in these two countries among 15-year olds that participated in PISA 2015.

Among the covariates included in the model, the coefficients related to immigration background and public school are most dramatic when compared to results of curriculum countries in Table 3. The coefficients of belonging to either first or second generation of immigrants are negatively, significantly, and alarmingly high across all six didaktik countries included in the sample, showing that 15-year olds with immigration background in all six countries are at a disadvantage when compared to native students. However, the role of public school on science performance was significant and negative only in Germany and Finland, while not significant in other countries, suggesting that students in public schools in Austria, Denmark, Norway and Sweden are not at a disadvantage when compared to the students attending private schools in these countries. In sum, results of the second research question do not suggest that assessment practices play a significant role in relation to students' science performance in PISA 2015, when controlling for the covariates in the models. Still, the results reveal interesting variations in the countries with significant results, both regarding assessment practices, and other covariates included in the HLM models, which will be discussed in the next section.

Discussion and conclusions

The theoretical and empirical differences between the didaktik and curriculum traditions have significant implications for educational research, policy making and educational practice in Kosovo, as well as in representative didaktik and curriculum countries. Educational research, particularly comparative international assessments, will have to pay more attention to the cross-country differences with regard to assessment practices and how they affect student achievement. Regarding policy-making, the two education traditions seem to be moving towards a similar educational trend with introduction of more accountability measures, which in turn means that educational practices are moving towards the turf of curriculum tradition which has utilised and promoted assessment-based accountability measures for a longer period of time. Still, these differences have potential to affect teacher education policy in particular, which in turn will affect educational practices in classroom settings in respective countries, with particular emphasis on the types of assessment practices that are used by teachers.

Returning to the first major objective of the study, i.e. placing Kosovo in the didaktik-curriculum continuum by examining the variation in assessment practices across countries in the sample, the study provides evidence from the first research question that Kosovo has shifted toward the curriculum end of the continuum over past 20 years, a trend that has already been observed from an education policy perspective. Specifically, in three (mandatory standardised tests, non-mandatory standardised tests, teachers' judgmental ratings) out of four assessment practices, Kosovo's means of use of assessment practices were similar to those of curriculum countries, and only in the case of teacher-developed tests the mean use was on the didaktik end of the continuum, as per the data used and models employed. Subsequently, the results suggest that Kosovo's shift in education policy over past 20 years towards curriculum's social efficiency ideas such as standardised testing are now present in the education practice in the form of assessment practices studied here. Therefore, policy-wise and practice-wise, Kosovo has moved towards the curriculum end of the didaktik-curriculum continuum.

From the perspective of the didaktik-curriculum framework, the findings suggest that theoretical claims regarding the differences in level of teacher autonomy between didaktik and curriculum countries still hold. Specifically, examining the four assessment practices, the results indicated that curriculum countries show higher use of both mandatory and non-mandatory standardised tests, which are externally developed tests that teachers have to use as a requirement of accountability policies in the given countries. On the other hand, regarding teacher-based classroom assessment practices of teacher-developed tests and teachers' judgmental ratings, curriculum countries had a higher mean of use on the former, and didaktik countries showed higher use of the latter. Therefore, only in one out of four assessment practices the empirical results are not in line with the theoretical claims. Overall, the findings suggest that teachers in didaktik countries are more in control of decision making about assessment practices they put in use within their classrooms, while still being under pressure of externally mandated standardised tests, but not at the same level as their counterparts in curriculum countries. Against these results then and considering the placement of Kosovo in the curriculum end of the didaktik/curriculum continuum, the empirical evidence suggests that Kosovo teachers' autonomy is under threat as they operate under the pressure of externally developed, standardised tests.

The second research question tested the relationship of assessment practices with students' science performance in PISA 2015, aiming to identify the effectiveness of any of the assessment practices in students' performance. The findings were underwhelming, as in most countries in the sample the use of assessment practices was not significant. However, in the few cases when the results were significant, some interesting results were observed. For example, the use of mandatory standardised tests was significantly and negatively associated with science performance in the UK, while the use of non-mandatory standardised tests was statistically significant and negative in the UK and Ireland, showing that these two forms of assessment practices hurt students' performance in the two countries. Therefore, policy-makers in both countries could review the use of these assessment practices. Only teacher-developed tests in the US were statistically significant and positively associated with science performance among curriculum countries. Further, among didaktik countries, the use of teacher-developed tests was

statistically significant and positive in Norway and Sweden. While teacher-developed tests are tests developed by teachers for their own classroom use, this issue needs further scrutiny as to whether it is used as opposed to, or in addition to and in preparation for, externally developed standardised tests. While not the focus of this study, the results of the second research question revealed striking negative influence of students' immigration background on their science performance in didaktik countries, and as such this issue also begs for further scrutiny in future studies.

In light of prior research, the findings of the present study align closer with those that were not able to identify strong relationships of assessment practices with students' performance (Liang, 2010), and indeed similarly to Liang (2010) the present study also identified positive association of teacher-developed tests with students' performance in the US, albeit Liang used mathematics as the dependent variable. However, in contrast to Liang (2010) who identified associations of teacher-developed tests with students' mathematics performance also in Canada and Finland, the present study didn't identify such associations when science scores were used as a dependent variable. To this end, it might be further explored whether some of the assessment practices on which PISA collects data affect different domains such as mathematics and science differently, as the findings of the present study seem to imply.

Limitations and further research

Considering the cross-sectional nature of the data used in the study, no causality is meant with the findings. However, the findings are applicable to the 12 sample countries that were examined more thoroughly here, as well as to schools within them. Still, a number of limitations need to be recognised. The study used secondary data, which are indeed representative of the countries as random sampling was applied in collecting them; however it was not possible to control what data were collected. The items used for descriptive and inferential analyses were derived from variables available on the PISA 2015 datasets. The analytical models employed were specifically focused on relationships of assessment practices with students' science performance, controlling for a number of variables that are routinely included in such models in related literature (see for example Tahirsylaj, 2019); however the variables used for predicting student's performance in PISA were selective and in no way exhaustive.

While the present study applied quantitative methods to place Kosovo within the didaktik-curriculum continuum, as well as based on the understanding and elaboration of Kosovo's policy context in past 20 years in the light of theoretical conceptions borrowed from didaktik and curriculum education traditions, it is but one way to pursue this line of research. To this end, further research relying on qualitative methodology can specifically focus on how any of the assessment practices is implemented within specific school contexts and countries, and also further explore whether other assessment practices are used in addition to or instead of the four that PISA questionnaires collected data on. This is particularly relevant for Kosovo's context, since in PISA data, school directors reported the use of non-mandatory standardised tests once a year on average; however no such non-mandatory standardised tests are administered in Kosovo. Lastly, a qualitative

investigation of Kosovo's pre-service and in-service teachers' understanding of didaktik-curriculum continuum and how they position themselves in it could contribute to more nuanced explanations of the identified shift from didaktik to curriculum as a dominant education tradition.

References

- Alkharusi, H. (2008). Effects of classroom assessment practices on students' achievement goals. *Educational Assessment*, 13(4), 243-266. <https://doi.org/10.1080/10627190802602509>
- Biesta, G. (2011). Disciplines and theory in the academic study of education: A comparative analysis of the Anglo-American and Continental construction of the field. *Pedagogy, Culture & Society*, 19(2), 175-192. <https://doi.org/10.1080/14681366.2011.582255>
- Brookhart, S. M. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, 20(1), 69-90. <https://doi.org/10.1080/0969594X.2012.703170>
- Brookhart, S. M. & Durkin, D. T. (2003). Classroom assessment, student motivation, and achievement in high school social studies classes. *Applied Measurement in Education*, 16(1), 27-54. https://doi.org/10.1207/S15324818AME1601_2
- Deng, Z. & Luke, A. (2008). Subject matter: Defining and theorizing school subjects. In F. M. Connelly, M. F. He & J. Phillion, J. (Eds.). *The SAGE handbook of curriculum and instruction*. (pp. 66-87). SAGE. <https://au.sagepub.com/en-gb/oc/e/the-sage-handbook-of-curriculum-and-instruction/book227122>
- Eurydice (2021). *Welcome to Eurydice*. European Commission. https://eacea.ec.europa.eu/national-policies/eurydice/home_en
- Gundem, B. B. & Hopmann, S. (Eds.) (1998). *Didaktik and/or curriculum: An international dialogue*. Peter Lang.
- Grajcevcic, A. & Shala, A. (2021). A review of Kosovo's 2015 PISA results: Analysing the impact of teacher characteristics in student achievement. *International Journal of Instruction*, 14(1), 489-506. <https://doi.org/10.29333/iji.2021.14129a>
- Hopmann, S. (2007). Restrained teaching: The common core of Didaktik. *European Educational Research Journal*, 6(2), 109-124. <https://doi.org/10.2304/eej.2007.6.2.109>
- Hopmann, S. (2015). 'Didaktik meets curriculum' revisited: Historical encounters, systematic experience, empirical limits. *Nordic Journal of Studies in Educational Policy*, 2015(1), article 27007. <https://doi.org/10.3402/nstep.v1.27007>
- Hudson, B. (2007). Comparing different traditions of teaching and learning: What can we learn about teaching and learning? *European Educational Research Journal*, 6(2), 135-146. <https://doi.org/10.2304/eej.2007.6.2.135>
- Kansanen, P. (1995). The *Deutsche Didaktik*. *Journal of Curriculum Studies*, 27(4), 347-352. <https://doi.org/10.1080/0022027950270401>
- Kliebard, H. M. (2004). *The struggle for the American curriculum, 1893-1958*. Routledge. <https://www.routledge.com/The-Struggle-for-the-American-Curriculum-1893-1958/Kliebard/p/book/9780415948913>
- Liang, X. (2010). Assessment use, self-efficacy and mathematics achievement: Comparative analysis of PISA 2003 data of Finland, Canada and the USA. *Evaluation & Research in Education*, 23(3), 213-229. <https://doi.org/10.1080/09500790.2010.490875>

- Ministry of Education, Science and Technology (MEST) (2018). Education statistics in Kosovo 2017-2018. <https://masht.rks-gov.net/uploads/2018/07/statistikat-e-arsimit-ne-kosove-2017-18.pdf>
- Ministry of Education, Science and Technology (MEST) (2011). *Curriculum framework for pre-university education in the Republic of Kosovo*. MEST. https://www.academia.edu/2912972/The_Kosovo_Curriculum_Framework
- OECD (2016a). *PISA 2015 results (Volume II): Policies and practices for successful schools*. OECD Publishing, Paris. <https://doi.org/10.1787/9789264267510-en>
- OECD (2016b). *PISA 2015 assessment and analytical framework: Science, reading, mathematics and financial literacy*. OECD Publishing: Paris. <https://doi.org/10.1787/9789264255425-en>
- Official Journal of the European Union (OJEU) (2006). Recommendation of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning (2006/962/EC). <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:394:0010:0018:en:PDF>
- Pantić, N. & Wubbels, T. (2012). Competence-based teacher education: A change from Didaktik to Curriculum culture? *Journal of Curriculum Studies*, 44(1), 61-87. <https://doi.org/10.1080/00220272.2011.620633>
- Pettersson, D., Prøitz, T. S., Román, H. & Wermke, W. (2015). Curriculum versus Didaktik revisited: Towards a transnational curriculum theory. *Nordic Journal of Studies in Educational Policy*, 1(1). <https://doi.org/10.3402/nstep.v1.27014>
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. SAGE. <https://au.sagepub.com/en-gb/oce/hierarchical-linear-models/book9230>
- Rodriguez, M. C. (2004). The role of classroom assessment in student performance on TIMSS. *Applied Measurement in Education*, 17(1), 1-24. https://doi.org/10.1207/s15324818ame1701_1
- Saqipi, B. (2019). Understanding the relation of policy discourse and re-conceptualising curriculum: A Kosovo perspective on a new meaning of context. *CEPS Journal*, 9(2), 33-52. <https://doi.org/10.26529/cepsj.559>
- Schiro, M. S. (2013). *Curriculum theory: Conflicting visions and enduring concerns* (2nd ed.) SAGE. <https://au.sagepub.com/en-gb/oce/curriculum-theory/book234465>
- Shala, A. & Grajcevcic, A. (2018). Kosovo's low performance in PISA 2015: An explanation from a socioeconomic perspective. *Educational Process: International Journal*, 7(1), 48-59. <https://doi.org/10.22521/edupij.2018.71.4>
- Tahirsylaj, A. (2020). What kind of citizens? Constructing 'Young Europeans' through loud borrowing in curriculum policy-making in Kosovo. *Comparative Education*, online first. <https://doi.org/10.1080/03050068.2020.1845066>
- Tahirsylaj, A. (2019). Teacher autonomy and responsibility variation and association with student performance in Didaktik and curriculum traditions. *Journal of Curriculum Studies*, 51(2), 162-184. <https://doi.org/10.1080/00220272.2018.1535667>
- Tahirsylaj, A. & Wahlström, N. (2019). Role of transnational and national education policies in realisation of critical thinking: The cases of Sweden and Kosovo. *The Curriculum Journal*, 30(4), 484-503. <https://doi.org/10.1080/09585176.2019.1615523>
- Tahirsylaj, A. (2018). Curriculum reform as a political statement in developing contexts: A discursive and non-affirmative approach. *Transnational Curriculum Inquiry*, 15(2), 38-49. <https://doi.org/10.14288/tci.v15i2.191089>

- Tahirsylaj, A. (2017). Curriculum field in the making: Influences that led to social efficiency as dominant curriculum ideology in progressive era in the U.S. *European Journal of Curriculum Studies*, 4(1), 618-628.
<http://pages.ie.uminho.pt/ejcs/index.php/ejcs/article/view/156>
- Tahirsylaj, A., Niebert, K. & Duschl, R. (2015). Curriculum and didaktik in 21st century: Still divergent or converging? *European Journal of Curriculum Studies*. 2(2), 262-281.
<http://pages.ie.uminho.pt/ejcs/index.php/ejcs/article/view/91>
- Wermke, W. & Proitz, T. S. (2019). Discussing the curriculum-Didaktik dichotomy and comparative conceptualisations of the teaching profession. *Education Inquiry*, 10(4), 300-327. <https://doi.org/10.1080/20004508.2019.1618677>
- Westbury, I., Hopmann, S. & Riquarts, K. (Eds.) (2000). *Teaching as a reflective practice: The German didaktik tradition*. Lawrence Erlbaum Associates.
<https://www.routledge.com/Teaching-As-A-Reflective-Practice-The-German-Didaktik-Tradition/Westbury-Hopmann-Riquarts/p/book/9781138983670>

Armend Tahirsylaj is an Associate Professor of Education at Norwegian University of Science and Technology (NTNU), Trondheim, Norway. His research focuses on curriculum theory and curriculum studies, didaktik/bildung, education policy, teacher education, international large-scale assessments, and international comparative education.
Email: armend.tahirsylaj@ntnu.no
<https://orcid.org/0000-0002-3902-0338>

Please cite as: Tahirsylaj, A. (2021). Placing Kosovo in the didaktik-curriculum continuum: A quantitative, international comparative perspective. *Issues in Educational Research*, 31(1), 255-273. <http://www.iier.org.au/iier31/tahirsylaj.pdf>