INSTITUTE for REPLICATION

No. 27
I4R DISCUSSION PAPER SERIES

Teacher Influence in Music Composition since 1450: A Replication of Borowiecki (2022)

Jori Korpershoek

Marco Musumeci

Renske A. Stans

Maddalena Totarelli

March 2023

INSTITUTE for REPLICATION

I4R DISCUSSION PAPER SERIES

I4R DP No. 27

Teacher Influence in Music Composition since 1450: A Replication of Borowiecki (2022)

Jori Korpershoek[1], Marco Musumeci[1], Renske A. Stans[1], Maddalena Totarelli[2]

[1]Erasmus University Rotterdam/The Netherlands
[2]University of Amsterdam/The Netherlands

MARCH 2023

E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research
Hohenzollernstraße 1-3
45128 Essen/Germany
www.i4replication.org

ISSN: 2752-1931

Teacher Influence in Music Composition since 1450: A Replication of Borowiecki (2022)

Jori Korpershoek[1], Marco Musumeci[1], Renske A. Stans[1], Maddalena Totarelli[2]

December 2022

Abstract

Borowiecki (2022) studies the influence of teachers on the style of their students in the domain of musical composition. The author finds that realized student-teacher pairs are on average 0.2-0.3 standard deviations more similar to unrealized, but possible, student-teacher pairs. In this report we provide the results of our replication of Borowiecki (2022). We direct our attention to the following tasks: 1) Replicating the outcome variables used in the paper, starting from the raw data, and generating alternative measures of similarity between students and teachers 2) Testing the validity of the random teacher-student pairing, a key assumption for the validity of the estimation strategy employed in the paper. We can replicate most of the outcome variables, but not all of them, due to incomplete raw data. Our alternative measures of similarity confirm the robustness of the original results. We find significantly different characteristics between paired and unpaired students, suggesting that matching between students and teachers does not occur randomly. However, controlling for these characteristics in the main regressions leads to quantitatively similar results to the ones reported in the original paper.

[1] Erasmus University Rotterdam, [2] University of Amsterdam

1 Introduction

Borowiecki (2022) studies how teachers and mentors in creative fields influence the style of their students. Building a novel data set on music composition over five centuries, the paper exploits teacher-student relationships to show that realized student-teacher pairs are on average 0.2-0.3 standard deviations more similar to unrealized, but possible, pairs. In addition, the author finds that the influence of teachers on students persists across generations and it is stronger for higher-quality teachers. Similarities between themes are computed as Jaccard indices and cosine similarities built from the notes, key signatures and tempo of the compositions. Data are retrieved from different sources, combining melodic themes (Barlow and Morgenstern, 1975, 1976) with biographical information on the composers (Grove, 2016; Pfitzinger, 2017).

The present report is centered around two tasks:

1) Reproducing the similarity indices used in the paper. This is of interest as the author does not provide the code needed to generate them from the raw data. We also construct alternative measures of similarity between compositions to check if the results presented in the paper are robust to the adoption of different similarity measures.

2) Checking the identifying assumption of random teacher-student pairs. Even though this is a key assumption for the validity of the estimation strategy, there is no formal test of the comparability of the chosen treatment and control group. We therefore create a balance table to compare paired and unpaired students and teachers. In addition, we perform a robustness check of the main results using matching as an alternative stastical method for comparing similiarity between paired and upaired teachers and students.

Starting from the raw data, we could reproduce the majority of the similarity indices used in the paper. However, this was not the case for the similarities based on tempo and key, as the raw data and documentation concerning the latter are not available. Our check of the random matching between teacher and students shows that there are significant differences between paired and unpaired students. However, including these characteristics as controls in the main specification leads to results that are quantitatively similar to the original ones.

2 Reproducibility

In this section we describe the issues we encountered when reproducing the main analysis using the data and codes provided by the author.

Reproducing the main outcome variables. The project files include both the raw data as well as the cleaned final version that is used to run the regression reported in the paper. However, missing data, documentation and code made it non-trivial to reproduce the main outcome variables (Jaccard indices and cosine similarities) used in the paper. We encountered three main issues while trying to reproduce the similarity indices starting from the raw data. First, there is no documentation describing the raw datasets (BM75 and BM76). Although some variable names are self-explanatory, this is not the case for all of them. For example, it appears that "value", "letter", and "number" are references to where in the original source a work can be found, but no clear explanation of the meaning of these variables is given in the documentation. Second, certain variables used in the main analysis cannot be found in the raw data, but only in the final cleaned dataset. This includes important variables such as the tempo and the key signature of a piece. Third, the code used to create the final dataset from the raw data is omitted from the project files. Checking whether we could replicate the main outcome variables from the raw data was one of our main focuses of this replication.

We wrote code in Python and R which imports the raw composition level data in the files BM75 and BM76 and then creates the final outcome variables as found in "composerpairs_sim_999" (the final dataset used to carry out the analysis). This includes the Jaccard index and cosine similarity for one until four-grams. Our replication was successful as it produced exactly the same similarity indices that the author provides in the cleaned final datafile. However, we could not reproduce the indices based on tempo and key signature as the raw data needed cannot be found in the files BM75 and BM76.

Like with the raw datafile, more documentation on how to reproduce the final dataset would have been useful. There are some steps where the researcher has to make arbitrary decisions and this is not described. For example, a few compositions included notes with typos (using "__" to denote a flat note rather than just "_"). We decided to drop these observations, even though

it would have been equally reasonable to fix these observations. Considering that we exactly replicated the results, this is what the original author did too. Similarly, in the final analysis, the outcome variable is standardized, but it is not described exactly what sample was used for this procedure. We standardized the similarity indices using the sample of all composer 1 - composer 2 possible pairs. We obtain similar standardized outcomes to the ones included in the final cleaned dataset, but not exactly the same. Regardless of the different standardization procedures, we find almost identically results to the ones reported in the paper when using our standardized outcome in the main regressions (Table 2 Section VI). A comparison between our results and the original ones is shown in Tables 1 and 2. Note that Table 1 includes specifications with key and time as outcome variables, while these are omitted from Table 2. However, Table 2 includes 1-grams as additional outcomes.

Table 1: Original specification

	Percent shared			Cosine similarity				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2-grams	3-grams	4-grams	2-grams	3-grams	4-grams	key	time
Connected	0.118**	0.283***	0.305***	0.0953**	0.178***	0.241***	0.169***	0.137***
	(0.0540)	(0.0922)	(0.111)	(0.0390)	(0.0639)	(0.0906)	(0.0572)	(0.0448)
N	23489	23489	23489	23489	23489	23489	23489	23489

Notes: The dependent variable is the **original** standardized coefficient that measures the similarity of a given pair of composers computed as Jaccard indices (columns 1-3, respectively) or as cosine similarities of N-grams, key and time signatures (columns 4-8, respectively). "Connected" is the main coefficient of interest and indicates realized teacher-student pairs. The reference group contains all the unrealized, but potentially possible, pairs of teacher-student. Controls include the natural logarithm of geographich distance and time distance, dummies for common birth country, time period and their interaction, common nationality, common descent and composer fixed effects. Clustered standard errors by candidate teacher are shown in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Replication main outcomes

	Percent shared				Cosine similarity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1-grams	2-grams	3-grams	4-grams	1-grams	2-grams	3-grams	4-grams
Connected	0.0528	0.121**	0.285***	0.306***	0.0302	0.0954**	0.178***	0.241***
	(0.0423)	(0.0544)	(0.0925)	(0.112)	(0.0205)	(0.0389)	(0.0639)	(0.0908)
N	23489	23489	23489	23489	23489	23489	23489	23489

Notes: The dependent variable is the standardized coefficient that we **replicated** in our analysis. It measures the similarity of a given pair of composers computed as Jaccard indices (columns 1-4 respectively) or as cosine similarities of N-grams (columns 5-8, respectively). "Connected" is the main coefficient of interest and indicates realized teacher-student pairs. The reference group contains all the unrealized, but potentially possible, pairs of teacher-student. Controls include the natural logarithm of geographich distance and time distance, dummies for common birth country, time period and their interaction, common nationality, common descent and composer fixed effects. Clustered standard errors by candidate teacher are shown in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Multicollinearity in baseline results. Due to multicollinearity, one of the control variables is not included in the main specification reported in Table 2 of the paper. In particular, the variable "common descent" is dropped when the variable "common country" is included. Hence, we repeated the main specification including "common descent" instead of "common country". The results can be found in Table 3. We find point estimates that are very similar to the original findings.

Table 3: Include common descent as control

	Percent shared			Cosine similarity				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2-grams	3-grams	4-grams	2-grams	3-grams	4-grams	key	time
Panel A: Original specification								
Connected	0.118**	0.283***	0.305***	0.0953**	0.178***	0.241***	0.169***	0.137***
	(0.0540)	(0.0922)	(0.111)	(0.0390)	(0.0639)	(0.0906)	(0.0572)	(0.0448)
N	23489	23489	23489	23489	23489	23489	23489	23489
Panel B: Include common descent								
Connected	0.114**	0.276***	0.292***	0.0900**	0.166***	0.229**	0.158***	0.142***
	(0.0534)	(0.0919)	(0.111)	(0.0384)	(0.0630)	(0.0898)	(0.0572)	(0.0452)
N	23489	23489	23489	23489	23489	23489	23489	23489

Notes: The dependent variable is a standardized coefficient that measures the similarity of a given pair of composers computed as Jaccard indices (columns 1-3, respectively) or as cosine similarities of N-grams, key and time signatures (columns 4-8, respectively). "Connected" is the main coefficient of interest and indicates realized teacher-student pairs. The reference group contains all the unrealized, but potentially possible, pairs of teacher-student. In Panel A controls include the natural logarithm of geographich distance and time distance, dummies for common birth country, time period and their interaction, common nationality, common descent and composer fixed effects. Panel B summarizes the results of the same specification but including "common descent" instead of "common country". Clustered standard errors by candidate teacher are shown in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Reproducing additional robustness checks. We could not find the code for the tables in appendix E.6, nor the variables used in these tables on class and wealth.

3 Replication

In this section we describe our replication analysis, which consists of two main exercises that investigate key aspects of the paper that we feel are underexposed. First, the paper directs little attention to checking the identifying assumption of random teacher-student pairs. In section

3.1 we investigate the validity of this assumption and accordingly propose alternative specifications. Second, we test the robustness of the main results to the adoption of alternative similarity measures that are not considered in the paper. These alternative outcome variables are described in section 3.2, together with the results obtained with them. Finally, section 3.3 describes some additional robustness checks.

3.1 Randomization assumption of matched pairs[1]

3.1.1 Balance tables and regressions

The main analysis relies on the assumption that the matching of students to teachers occurs randomly, conditional on time and geographical restrictions. Hence, there should not be any significant differences between matched teachers and students and unmatched teachers and students. Tables 4 and 5 test this assumption empirically for students and teachers, respectively, by looking at differences in pre-treatment characteristics. For both we find statistically significant differences between the realized and unrealized pairs. For students, we find differences in nationality, year of birth and age of meeting. For teachers, we find differences in nationality, occupation, quality indicators and year of birth.

Alternatively, we can see whether the dummy variable ”connected”, which indicates the realized pairs, is correlated with each of the pre-treatment characteristics. Tables 6 and 7 show the results of regressing the treatment dummy (i.e., ”connected”) on each of the pre-treatment characteristic. As before, we find significant effects pointing to non-randomness in who is matched and who is not.

[1]Notice that the similarity indices used in this part of the replication are the ones provided by the authors and not the ones that we created. We decided to use the original similarity indices because the ones we constructed would be available only at the end of (or after) the replication day.

Table 4: Balance table - students

	Unrealized pairs					Realized pairs					
	N	Min	Max	Mean	Stand. Dev.	N	Min	Max	Mean	Stand. Dev.	Diff
Born in Europe	23306	0.00	1.00	.81	0.39	183	0.00	1.00	0.84	0.37	0.031
French	23306	0.00	1.00	.18	0.39	183	0.00	1.00	0.34	0.47	0.157***
Italian	23306	0.00	1.00	.09	0.29	183	0.00	1.00	0.05	0.23	-0.035**
German	23306	0.00	1.00	.074	0.26	183	0.00	1.00	0.05	0.23	-0.019
English	23306	0.00	1.00	.11	0.31	183	0.00	1.00	0.10	0.30	-0.011
Austrian	23306	0.00	1.00	.056	0.23	183	0.00	1.00	0.02	0.13	-0.040***
American	23306	0.00	1.00	.19	0.39	183	0.00	1.00	0.14	0.34	-0.052**
Year of birth	23306	1450.00	1918.00	1867	41.04	183	1485.00	1918.00	1857.50	57.88	-9.039**
Age at meeting	9447	6.00	52.00	21	8.07	113	6.00	52.00	19.49	8.03	-1.269*

Notes: This table shows the difference in students' observable charecteristics by status, i.e. whether it is an unrealized or realized pair. The table reports the number of observations as well as some descriptive statistics of students' characteristics such as mean and standard deviation. The values displayed in the last column are the differences in the means across unrealized and realized pairs. Standard errors are robust. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Balance table - teachers

	Unrealized pairs					Realized pairs					
	N	Min	Max	Mean	Stand. Dev.	N	Min	Max	Mean	Stand. Dev.	Diff
Born in Europe	23306	0.00	1.00	.91	0.29	183	0.00	1.00	0.96	0.21	0.051***
French	23306	0.00	1.00	.25	0.43	183	0.00	1.00	0.42	0.50	0.167***
Italian	23306	0.00	1.00	.11	0.31	183	0.00	1.00	0.07	0.25	-0.041**
German	23306	0.00	1.00	.12	0.32	183	0.00	1.00	0.09	0.28	-0.028
Austrian	23306	0.00	1.00	.054	0.23	183	0.00	1.00	0.03	0.16	-0.027**
American	23306	0.00	1.00	.084	0.28	183	0.00	1.00	0.04	0.21	-0.041***
No other occupation	23306	0.00	1.00	.55	0.50	183	0.00	1.00	0.36	0.48	-0.191***
Murray index	23306	0.00	100.00	5.2	10.05	183	0.00	56.00	8.15	10.07	2.954***
Spotify followers	23175	0.00	816190.31	12738	50692.78	183	1.00	213233.67	11360.36	25412.13	-1,377.730
word count works	23306	0.00	46397.00	1990	3960.10	183	0.00	44714.00	3480.28	6464.42	1,490.205***
(max) citynumber	23306	1.00	23.00	6	3.90	183	1.00	23.00	5.84	4.25	-0.206
Number of BM works	23306	1.00	227.00	17	29.74	183	1.00	227.00	19.78	25.04	2.959
Year of birth	23306	1397.00	1913.00	1841	44.34	183	1450.00	1900.00	1831.43	58.63	-9.412**

Notes: This table shows the difference in teachers' observable charecteristics by status, i.e. whether it is an unrealized or realized pair. The table reports the number of observations as well as some descriptive statistics of teachers' characteristics such as mean and standard deviation. The values displayed in the last column are the differences in the means across unrealized and realized pairs. Standard errors are robust. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Balance regressions - teachers

	(1) EU	(2) FR	(3) IT	(4) G	(5) AU	(6) US	(7) Comp. only	(8) Quality I	(9) Quality II	(10) Quality III	(11) Words	(12) Cities	(13) Works
Connected	0.0301	0.0591	-0.0498**	-0.0298	-0.0177	-0.0625**	-0.225***	2.552**	-2105.2	1456.5*	0.159	3.253	1.628
	(0.0280)	(0.0694)	(0.0249)	(0.0315)	(0.0138)	(0.0261)	(0.0604)	(1.291)	(3878.3)	(872.2)	(0.510)	(2.623)	(5.089)
N	23489	23489	23489	23489	23489	23489	23489	23489	23358	23489	23489	23489	23489

Notes: This table shows the coefficients of a regression model in which the dummy variable "connected", which indicates the realized pairs, is regressed on each of the teachers' characteristics prior of matching with a student. The dependent variables in columns 1-6 are dummies indicating the nationality of the teacher. Columun 7 shows the coefficient for a regression model in which the dependent variable is a dummy variable indicating whether the teacher was only a composer. Columns 8-10 report the coefficients for the quality of the teacher, indicated by Murray index, Spotify followers and number of BM works, respectively. Finally, the dependent variable in columns 11-13 summarizes the number of words in each work, cities visited and number of works completed. Clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Balance regressions - students

	(1) EU	(2) FR	(3) IT	(4) G	(5) UK	(6) AU	(7) US	(8) Birth year	(9) Meeting age
Connected	-0.0683	0.0115	-0.0560***	-0.0373*	-0.0116	-0.0300***	-0.00203	-2.495***	-1.165
	(0.0439)	(0.0540)	(0.0201)	(0.0205)	(0.0365)	(0.0105)	(0.0399)	(0.535)	(0.889)
N	23489	23489	23489	23489	23489	23489	23489	23489	9560

Notes: This table shows the coefficients of a regression model in which the dummy variable "connected", which indicates the realized pairs, is regressed on each of the students' characteristics prior of matching with a teacher.The dependent variables in columns 1-7 are dummies indicating the nationality of the student. Columns 8-9 report the coefficients for birth year and age when the student met the teacher. Clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.1.2 Additional controls

The previous subsection shows that the assumption of randomly matched teacher-student pairs likely does not hold. However, as the main specification includes teacher fixed effects it is not a problem that teachers' pre-treatment characteristics differ as they are taken out. The same is not true for the student characteristics. Therefore, we repeat the original specification including the student characteristics that are found to significantly differ between realized and unrealized pairs. The results are shown in Table 8. Panel A reports the coefficients of the original specification used by the author. Panel B summarizes the results obtained including as additional controls to the previous regression model all the student characteristics that are found to be statistically different between realized and unrealized teacher-student pairs. The point estimates become slightly smaller when including the additional controls, however the interpretation stays the same.

3.1.3 Matching

In addition, we conducted a matching analysis to ensure the treated and control observations have comparable pre-treatment characteristics. Results are shown in Table 9. We conducted one-to-one matching, leading to a reduction in the sample size, which accounts only for 331 observations. Not surprisingly this lead to an increase in standard errors, and hence for most estimates to lose their statistical significance. Nonetheless, it is encouraging that the effect sizes stay all positive and most are larger than in the original specification.

Table 8: Addtional controls

	Percent shared			Cosine similarity				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2-grams	3-grams	4-grams	2-grams	3-grams	4-grams	key	time
Panel A: Original specification								
Connected	0.118**	0.283***	0.305***	0.0953**	0.178***	0.241***	0.169***	0.137***
	(0.0540)	(0.0922)	(0.111)	(0.0390)	(0.0639)	(0.0906)	(0.0572)	(0.0448)
N	23489	23489	23489	23489	23489	23489	23489	23489
Panel B: Additional controls								
Connected	0.103*	0.260***	0.288**	0.0879**	0.172***	0.245***	0.157***	0.122***
	(0.0533)	(0.0920)	(0.111)	(0.0399)	(0.0654)	(0.0912)	(0.0572)	(0.0444)
N	23489	23489	23489	23489	23489	23489	23489	23489

Notes: The dependent variable is a standardized coefficient that measures the similarity of a given pair of composers computed as a Jaccard indices (columns 1-3, respectively) or as cosine similarities of N-grams, key and time signatures (columns 4-8, respectively). "Connected" is the main coefficient of interest and indicates realized teacher-student pairs. The reference group contains all the unrealized, but potentially possible, pairs of teacher-student. Panel A reports the coefficient of the original specification. Controls include the natural logarithm of geographich distance and time distance, dummies for common birth country, time period and their interaction, common nationality, common descent and composer fixed effects. Panel B summarizes the results obtained including as additional controls to the previous regression model all the student characteristics that are found to be statistically different between realized and unrealized teacher-student pairs. Clustered standard errors by candidate teacher are shown in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Matching

	Percent shared			Cosine similarity				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2-grams	3-grams	4-grams	2-grams	3-grams	4-grams	key	time
Panel A: Original specification								
Connected	0.118**	0.283***	0.305***	0.0953**	0.178***	0.241***	0.169***	0.137***
	(0.0540)	(0.0922)	(0.111)	(0.0390)	(0.0639)	(0.0906)	(0.0572)	(0.0448)
N	23489	23489	23489	23489	23489	23489	23489	23489
Panel B: Matching specification								
Connected	0.109	0.453*	0.387	0.167	0.387*	0.627**	0.239	0.278**
	(0.201)	(0.272)	(0.274)	(0.164)	(0.226)	(0.278)	(0.250)	(0.114)
N	331	331	331	331	331	331	331	331

Notes: The dependent variable is a standardized coefficient that measures the similarity of a given pair of composers computed as a Jaccard indices (columns 1-3, respectively) or as cosine similarities of N-grams, key and time signatures (columns 4-8, respectively). "Connected" is the main coefficient of interest and indicates realized teacher-student pairs. The reference group contains all the unrealized, but potentially possible, pairs of teacher-student. Panel A reports the coefficient of the original specification. Controls include the natural logarithm of geographich distance and time distance, dummies for common birth country, time period and their interaction, common nationality, common descent and composer fixed effects. Panel B summarizes the results obtained using one-to-one matching. Clustered standard errors are shown in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.2 Checking robustness using alternative outcome variables

Using the code to replicate the main outcome variable, we also create some alternative outcome variables to check the robustness of the main results. Since the paper relies on similarity measures also used for text analysis, we check whether the results are driven by extremely common n-grams, or by very rare ones, like is often done in natural language processing. First, we create outcome variables where we drop the most common n-gram (which might be seen as equivalent to dropping stop words). Second, we drop rare n-grams, that we define as those with frequencies below the median. This radically reduces the dimensionality of the data, but at the expense of only few observed N-grams.

Table 10: Replication main outcomes

	Percent shared			Cosine similarity		
	(1)	(2)	(3)	(4)	(5)	(6)
	2-grams	3-grams	4-grams	2-grams	3-grams	4-grams
Panel A: N-grams above						
Connected	0.126**	0.290***	0.322***	0.0960**	0.178***	0.242***
	(0.0583)	(0.0948)	(0.117)	(0.0383)	(0.0635)	(0.0903)
N	23489	23489	23489	23489	23489	23489
Panel B: N-grams no max						
Connected	0.120**	0.291***	0.307***	0.108***	0.203***	0.227**
	(0.0548)	(0.0932)	(0.112)	(0.0382)	(0.0630)	(0.0890)
N	23489	23489	23489	23489	23489	23489

Notes: The dependent variable is a standardized coefficient that measures the similarity of a given pair of composers computed as a Jaccard indices (columns 1-3, respectively) or as cosine similarities (columns 4-6, respectively). "Connected" is the main coefficient of interest and indicates realized teacher-student pairs. The reference group contains all the unrealized, but potentially possible, pairs of teacher-student. Panel A summarizes the results obtained using similarities indices built by **dropping rare n-grams**. Whereas, Panel B reports the results using as outcome variable the similarities computed **without the most common n-grams**. Controls include the natural logarithm of geographich distance and time distance, dummies for common birth country, time period and their interaction, common nationality, common descent and composer fixed effects. Clustered standard errors are shown in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We find similar results to the main findings using these alternative specifications, as shown in Table 10. Panel A summarizes the results obtained using similarities indices built by dropping rare n-grams. Whereas, Panel B reports the results using as outcome variable the similar-

ities computed without the most common n-grams. These results show that the conclusions of the paper are not driven by the most common, or by very rare, n-grams.

Initially, we planned to construct other outcome measures, including the key and time signature in the cosine similarity. This was meant to build a richer measure of similarity, incorporating more information about a piece in a single measure. However, given that the key and time signature are not included in the raw data, this was not possible.

3.3 Additional robustness checks

3.3.1 Focus on 19th century

From Appendix Table B2 - from the original paper (Borowiecki, 2022, p. 1033) - it can be seen that the large majority of all observations is from the 19th century. Hence, it might be that the book that served as the basis for the data collection focused mainly on the 19th century and is much less accurate for the other centuries creating potential noise in the analysis. Therefore, we repeated the main specification with only observations from the 19th century. The results can be found in Table 11, and shows similar point estimates as the original results.

3.3.2 One-grams

One-grams are mentioned in the descriptive tables of the paper, however their similiraity measures are not used as an outcome variable in any of the regressions reported in the paper. We run the main specification for one-grams too and do not find significant results (see columns 1 and 5 of Table 2). This seems to show that single notes are not very informative about an author's style or that they do not represent the channel through which style is transmitted.

Table 11: Sample restricted to 19th century

	Percent shared			Cosine similarity				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2-grams	3-grams	4-grams	2-grams	3-grams	4-grams	key	time
Panel A: Original specification								
Connected	0.118**	0.283***	0.305***	0.0953**	0.178***	0.241***	0.169***	0.137***
	(0.0540)	(0.0922)	(0.111)	(0.0390)	(0.0639)	(0.0906)	(0.0572)	(0.0448)
N	23489	23489	23489	23489	23489	23489	23489	23489
Panel B: 19th century only								
Connected	0.107*	0.292***	0.293***	0.0901**	0.190***	0.257***	0.150**	0.138***
	(0.0608)	(0.0993)	(0.112)	(0.0426)	(0.0684)	(0.0947)	(0.0631)	(0.0473)
N	21162	21162	21162	21162	21162	21162	21162	21162

Notes: The dependent variable is a standardized coefficient that measures the similarity of a given pair of composers computed as a Jaccard indices (columns 1-3, respectively) or as cosine similarities of N-grams, key and time signatures (columns 4-8, respectively). "Connected" is the main coefficient of interest and indicates realized teacher-student pairs. The reference group contains all the unrealized, but potentially possible, pairs of teacher-student. Panel A reports the coefficient of the original specification. Controls include the natural logarithm of geographich distance and time distance, dummies for common birth country, time period and their interaction, common nationality, common descent and composer fixed effects. Panel B summarizes the results restricting the sample to observations from the 19th century. Clustered standard errors are shown in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4 Conclusion

In this report we focused on reproducing the similarity indices used in Borowiecki (2022) and on checking the validity of the randomness in student-teacher pairing.

The replication of the outcome variables based on the sequences of notes was successful. We obtained exactly the same values as reported by the author. On the contrary, we could not recreate the similarity indices based on tempo and key, as the raw data needed are not provided in the project file. In addition, we could not identify the sample used to standardize the similarities. In our analysis we based the standardization process on the sample of similarity indices computed between all pairs of composers included in the data. This approach produced standardized indices that are close to the original ones, but not exactly the same. However, these differences in standardization processes did not impact the results of the paper. More documentation would have been helpful to understand precisely how the indices are created.

Our tests of the random student-teacher pairing showed that paired and unpaired observations differ significantly in several pre-treatment characteristics. As the main specification

controls for teacher fixed effects, differences between teachers are not of concern. However, this is not the case for student characteristics. Therefore, we repeated the original specification including the student characteristics that are found to significantly differ between treatment and control. The results are quantitatively similar to the original ones. Finally, we implemented a one-to-one matching model that confirmed again the robustness of the original findings.

References

Barlow, Harold and Sam Morgenstern. 1975. *A Dictionary of Musical Themes*. New York: Crown.

———. 1976. *A Dictionary of Opera adn Song Themes*. New York: Crown.

Borowiecki, Karol Jan. 2022. "Good Reverberations? Teacher Influence in Music Composition since 1450." *Journal of Political Economy* 130 (4):991–1090.

Grove. 2016. "Oxford Music Online: Grove Music Online." URL https://www.oxfordmusiconline.com.

Pfitzinger, Scott. 2017. *Composer genealogies: A compendium of composers, their teachers, and their students*. Rowman & Littlefield.