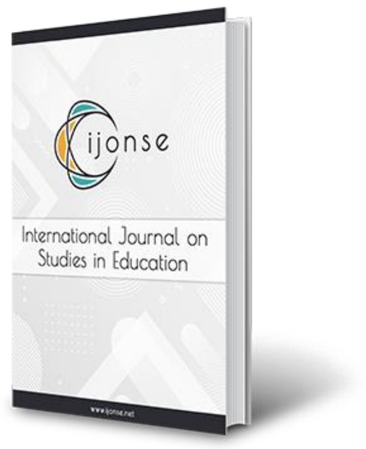

www.ijonse.net

Improving Student Performance with ILS Based Student-Teacher Pairings

Devon Zillmer
United States Military Academy, U.S.A.

Thomas Mussmann
United States Military Academy, U.S.A.

To cite this article:

Zillmer, D., & Mussmann, T. (2023). Improving student performance with ILS based student-teacher pairings. *International Journal on Studies in Education (IJonSE), 5*(3), 317-331. https://doi.org/10.46328/ijonse.133

International Journal on Studies in Education

2023, Vol. 5, No. 3, 317-331 https://doi.org/10.46328/ijonse.133

Improving Student Performance with ILS Based Student-Teacher Pairings

Devon Zillmer, Thomas Mussmann

Article Info

Article History

Received:
16 December 2022

Accepted:
19 March 2023

Keywords

Mathematics education
Classroom assignment
Learning styles
Index of learning styles
Meshing hypothesis

Abstract

Students are often assigned to instructors by student schedule or request. We present findings that assigning students to instructors based on learning style similarity can improve learning. We assume that instructors teach consistent with their learning style, and use the Felder and Silverman Index of Learning Styles (ILS) to describe learning style as a vector in $\mathbb{R}^4$. We examine the effect on grade difference a as a function of the learning style "distance" between students and their instructors . χ^2 tests show that the students (N=300) who were more similar (nearer) to their instructor performed better (p-value: 0.000453), and linear regression indicates that student performance could improve by up to 2% by being assigned to a more similar instructor. Our study shows that, with our three novel approaches (standardized learning style measurement, differenced grades, and the consideration of different norms), our data indicate a mild support for using the idea behind the meshing hypothesis to provide recommendations for student-teacher pairing.

Introduction

Many colleges have small classrooms that are part of a large course. A common example is recitations taught by a Teaching Assistant as part of a larger first-year calculus course or Fundamentals of Engineering classes that aim for smaller sections. We seek to find a way to improve the process of assigning students to instructors. While this process is constrained by scheduling requirements, there are often several potential instructors for a given student, and we seek to find a way to improve student performance by selecting an appropriate instructor. As instructors at the United States Military Academy (USMA, or West Point) students take a broad course load regardless of their major, and because of this are assigned to instructors rather than selecting their preferred schedule. The university also requires all classes have at most 18 students, which increases the number of required instructors and thus potential instructors for each student. If we could improve student performance simply by assigning students to different instructors, we could improve student learning with no additional cost to West Point.

We considered a common theory in the learning styles world, called the meshing hypothesis. Simply put, if students learn similarly to how teachers teach, we should see an improvement in student performance. While an initial literature review caused us to hesitate before further investigation (due to the consensus against the meshing hypothesis), a closer look at individual studies' data left us unconvinced. Our initial investigation was to consider a small change to how we assign students to instructors within our course, but our results indicated a significant correlation. Our data indicated that student performance improved when a student was assigned to a more similar instructor. However, the increase in student grades was modest, around 1-2% of overall course grade improvement. We were able to achieve these results by proposing three novel ideas. First, we measured method of instruction and learning style using the same standard; second, we controlled for many confounding factors by measuring grade difference rather than a single grade; finally, we use the mathematical idea of distance in a learning style space to attempt to predict this grade difference.

Background

A great deal of thoughts and research has gone into assessing the truth and utility of learning styles over the last few decades, including broad and deep examinations of a wide variety of frameworks (Coffield, 2004, and Pashler, 2009). Many authors posit that learning styles are a myth (Riener, 2010), or that the “learning style hypothesis” is all but “debunked” (Cuevas, 2015). These articles draw on published research, but often drawing conclusions outside the scope of the original research. Some examples are research claiming to debunk all learning styles, but only arguing against one specific learning style (Sankey, 2011), or which rely on an overly narrow interpretation of a framework (Kappe, 2009). These types of methodological handicaps limit the persuasiveness and generalizability of the findings.

We identified two major challenges in attempting to study the meshing hypothesis present in the literature. First, varying metrics between classifying content or teaching and in classifying a student’s learning style. In many studies, the learning style of the student is measured by a quiz or survey, whereas class content or teaching style is measured by a researcher’s decision (e.g., “this lesson is visually delivered, as compared to this lesson which is verbal”). A teacher can attempt to teach the same subject towards two very different learning styles and measure the difference in student performance, but the subjectivity of content classification can make it challenging to determine which technique is closer to a student’s learning style. This is exacerbated by the reality that many learning styles fail to account for the strength of a preference of the learner, and so the expected difference between a strongly verbal learner and a weakly verbal learner are lost.

Another component to the challenge in classifying content is a consideration of the material itself. Some subjects are best presented visually (such as graphing), while others (such as algebraic manipulation) require more verbal instruction. Thus, one can teach a lesson on graphing with more verbal emphasis, but that lesson may still be more visual than a lesson in algebraic manipulation lesson with more visual emphasis. This is difficult to deal with in an absolute sense, so is described subjectively. For example, one might consider how a lesson is usually taught or how the researcher would teach the class and then determine whether the taught lesson was more or less verbal. This level of subjectiveness reduces the utility of the content classification.

A second factor is controlling for noise. There are many factors, including study skills, mathematical background, or available time, that one would expect to have a greater impact on grades than learning style similarities. To examine the efficacy of learning styles, we must have a large enough sample size or control for these confounding variables.

To measure learning style, this paper uses the Felder and Silverman Index of Learning Styles (ILS) (Felder, 2002), which requires one to take a survey with 44 either/or questions and returns a vector of integers in four dimensions which indicate the learning style of the respondent. Previous analysis of the ILS highlights how these dimensions are roughly orthogonal (Zwanenberg, 2010), which indicates it may form an initial "basis" for considering a "learning space." See Figure 1 for a visualization of the four-dimensional response possibilities.

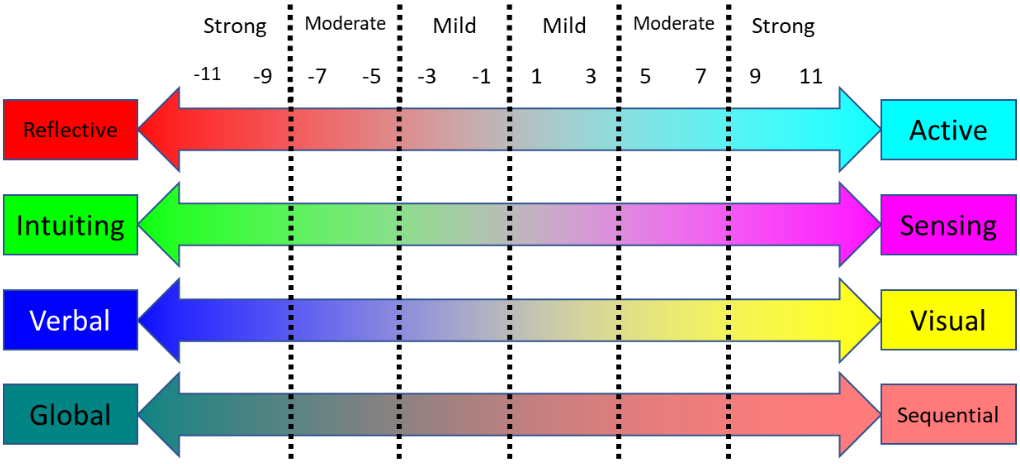

Figure 1. Visualization of ILS Learning Styles

This paper attempts to provide data relying on clearer assumptions behind a modest data set, and quantitative analysis of that data. In the following sections, we discuss our assumptions, data collection, analysis, and then address a slight modification of Pashler's et al.'s framework (Pashler, 2009) to highlight how our data indicate the potential benefits of matching students to more similar instructors. The essence of Pashler's framework was to dissociate superior teaching from consideration of the effect of learning styles, and included graphics to help assess if research findings were conflating the two.

Materials and Methods

In this section, we discuss the underlying assumptions of this project, how the data were collected, how we considered the notion of higher-dimensional distance, and the techniques used to analyse the data.

Assumptions

No study is methodologically perfect, including ones examining the impact of learning styles on student performance. This is unavoidable, but in some cases can limit the persuasiveness of the paper. Examples of this include: instruction and testing that is extremely close together (Choi, 2009), where only high-performing students are examined (Allcock, 2010), where only select "interventions" of different content are inserted (Hsieh, 2011), or examining performance based on a single data point (as discussed in Cuevas, 2015). This study was designed to remedy many of these methodological weaknesses.

The foundational assumption we make is that instructors teach in a way that makes sense to them. Every instructor in this study was given the freedom to prepare all lessons as they saw fit. The consistency between instructors was created by specifying the use specific chapters of a textbook, giving the same major graded events (instructors had 10% of available points to use for quizzes or other instructor specific events), specifying course objectives, and weekly meetings among instructors. From this freedom to design lessons, we assume that an instructor prepares material for class and teaches class consistent with their learning style. All students in this study were given the same book and resources to support the class, so one of the major sources of variation between students was their instructor.

A related key methodological weakness that differed from much of the previous research was that for our research, content was not simply classified as one learning style or another. The ILS presents learning styles in different dimensions and with different strengths, and that seems a suitable way to describe how content can be delivered. As an example of this principle: in place of classifying media or a block of instruction as either "active" or "reflective," we instead assume that the instructor teaches in accord with their learning style (presenting content in a way that "makes sense" to the instructor). This might be received better by a student with a more similar learning style and understood less well by a student with a more different learning style. Thus, we will consider how similar the learning style of the instructor is to the students, or in other words, the "learning style difference." This facilitates our first novel technique: we can now use the same measuring standard We will explore several different distance metrics later in this paper.

The critical connection between these two assumptions is that we now attempt to assess improvement in student performance (assuming course grades roughly correspond to the amount a student learned) as a function of the learning style distance. This clearly formulates our research question: does aligning student-instructor learning styles enhance learning?

Data

The greatest challenge in studying the interaction between learning style and student performance is the amount of noise. We expect many factors contribute to a student grade in the Advanced Math Program (AMP) at West Point. Some confounding factors are: prior math studies, proclivity for math, study skills, and competing interests such as clubs or sports. While it may be possible to try to consider these factors and control for them, we instead took a different approach uniquely possible at [university].

West Point prides itself on its core curriculum, which requires every student to broaden themselves with required courses. As an example, a student who majors in history must still take three semesters of math regardless of previous mathematical achievement, and eight other engineering and technology courses. Within this core curriculum, every student takes two math classes during their first year. In 2022, AMP brought in around 325 Students who take two courses: Mathematical Modelling with Ordinary Differential Equations (MA153) and Advanced Multivariate and Vector Calculus (MA 255). Additionally, West Point also caps the classroom size at 20 students, leading to 18 sections, taught by 8-9 instructors for each course. After taking MA153 in the fall, the students are "scrambled" into different sections for the spring semester.

With this unique West Point curriculum design, instead of trying to account for individual confounding factors, we simply compare a student's second-semester grades to their first-semester grades. This key step facilitates our second novel technique: measuring a difference of grades earned by the same student. This assumes that the student's study skills, mathematical ability, and competing requirements remained approximately constant, or increased at about the same rate as their classmates. This assumption controls for some of the confounding factors and allows us to achieve statistically significant results from our limited number of students.

In all cases of our analysis, we are focused on examining the change in student grade from the first semester to the second (subtracting the second grade from the first), and considering each student's performance as a percentage score distance from the average grade (i.e., a student who scores the mean on all exams across both semesters has a test grade of 0). This is designed to help control for factors intrinsic to each student such as study skills and mathematical proclivity.

Our data consists of 300 students who took the ILS survey and had grades for both MA153 and MA255. We considered three components of grades. First, test grades which included three midterms and one final per semester that were collectively graded by all the instructors. Second, "non-test" grades which consisted of 3-4 projects, homework, quizzes, and participation points. Each instructor graded their own non-test assignments. Finally, the total grade is a combination of test and non-test grades for the student final course grade. All grades are recorded as a proportion of total possible points (e.g., 0.87). All instructors for both MA153 (9 instructors) and MA255 (8 instructors) also took the ILS survey. There were 5 instructors who taught both courses.

Distance

To allow for thorough examination of the data, we examined several different types of distances from each student to their instructor across both semesters. We consider two different types of student-instructor learning style distances. First, the "raw" distance between two survey results is the number of questions in the ILS survey answered differently (returning an integer ranging from 0 to 44).

Second, we consider each person's learning style as a vector in $\mathbb{R}^4$ and use the p-norm definition to describe the distance between two individuals (returning a positive real number). Recall, the general form of the p-norm of some vector v, with n dimensions and elements $a_1, a_2, \dots, a_n$ is

$$\|\boldsymbol{v}\|_p = (|a_1|^p + |a_2|^p + \cdots + |a_n|^p)^{\frac{1}{p}}.$$

Where p is a real number greater than or equal to 1. We also recall that $\|v\|_\infty = \lim_{p\to\infty}\|v\|_p = \max_{i\in[1,n]}|a_i|$ is referred to as the infinity norm.

To provide an example of different types of norms, we consider two hypothetical students and their instructor, in only the active-reflection (AR) dimension and visual-verbal (VV) dimension. See Figure 2 for a visualization of this example. If our hypothetical instructor has AR score 5, VV score 4, and we consider their student-teacher distance with Student 1, who has AR score 1 and VV score 4. Because these two are only different in the AR dimension, all their norm distances are the same: 4. If we consider a second student with AR score of 1 and VV score 1, we begin to see different student-instructor distances by norm. The 1-norm distance is the sum of the difference in each dimension, and so with our instructor and Student 2 yields a distance of 7. The 2-norm is the square root of the sum of the squared distance in each dimension, and so in this case is distance 5. The infinity norm distance is the largest distance in any one dimension, and is thus 4.

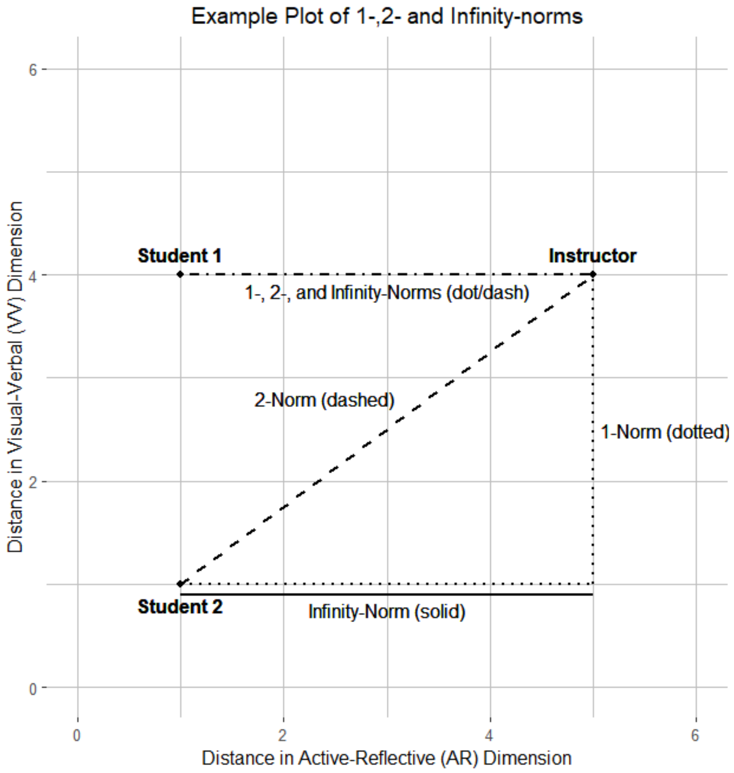

Figure 2. Visualization of 1-, 2-, and Infinity Norms

As two examples of the utility of these norms, the 1-norm "weights" each point in each dimension equally (i.e., a student-instructor difference of 10 could mean only slight differences across four dimensions or a huge difference in one dimension), whereas the infinity-norm only considers the magnitude of the largest distance in one dimension (i.e., small differences are neglected, only the single largest is important). These ideas form the core of our third novel technique: exploration of varying distance norms. Each of the different norms thus places a different emphasis on the relative importance of big or small changes across multiple dimensions and can inform what is significant when considering instructor-student difference. For this paper, we highlight only a few significant norms, focusing on the raw distance, 1-, 2-, and infinity-norm models.

Data Analysis

To assess if the learning style distance was associated with a change in performance, we considered two different statistical perspective: χ^2 tests and linear regression. In both cases, we begin with the assumption that there should not be any correlation between leaning style distance and grades.

To visualize our analysis, we created a single data point for each student: the x-position is the change in learning style difference (subtracting the second semester student-instructor distance from the first semester student-instructor distance), and the y-coordinate the change in grade across semesters (subtracting the second semester grade from the first). See Figure 3 for a depiction of the data considering the 2-norm.

As examples, we take two hypothetical data points. A data point in the first quadrant represents a student whose second instructor was “more similar” to the student than the first, and who performed worse (compared to the average performance in a class) in the second course than in the first. This student is in the first quadrant, and their performance is evidence refuting the meshing hypothesis, because their grade was worse when their instructor was more similar. A data point in the fourth quadrant is one whose second instructor was more similar to the student than the first instructor, but the student performed better in the second course than the first. This student’s performance supports the meshing hypothesis.

χ^2 Tests

For each χ^2 test, we first select the type of distance (e.g., 1-norm student-instructor distance change), and then count the number of observations in each of the four quadrants. We begin our analysis assuming that there is no correlation. Formally, the null hypothesis is that learning style distance is uncorrelated to performance in a class, which implies that each of the four quadrants seen in Figure 2 should have an equal number of observations. We note this is a non-parametric test which makes no assumptions about the distribution of the data (i.e., the data do not have to be normally distributed for this to remain true; we merely assume no correlation between learning style distance and performance).

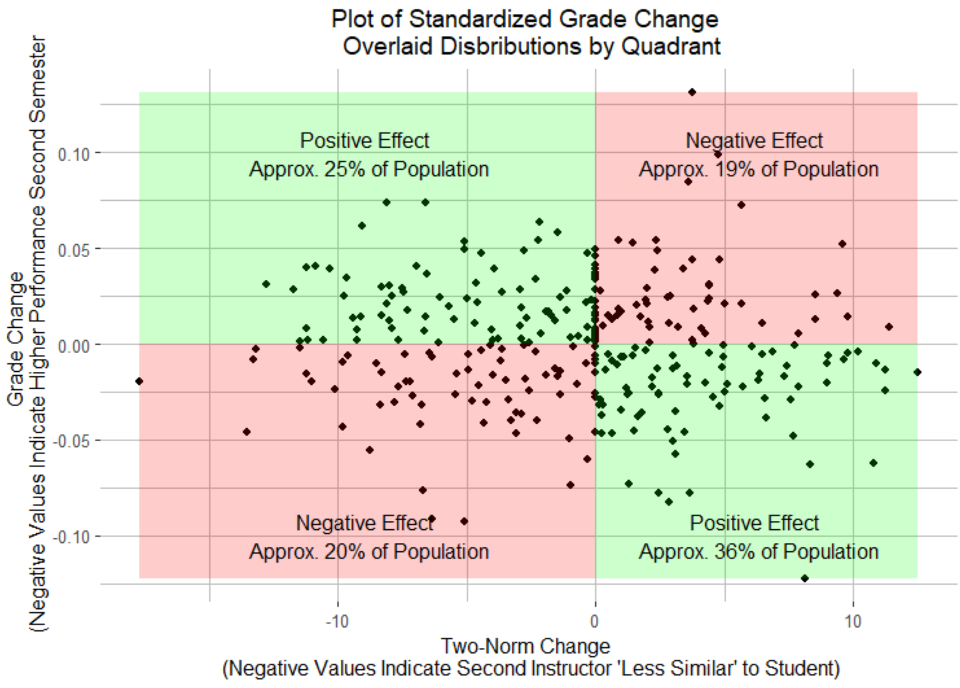

Figure 3. Visual Depiction of Data

A few situations could result in a student data point lying exactly on one of the dividing lines between categories. Two reasons are apparent: a student had the same instructor both semesters, or the instructor-student distance for a given norm remained zero as they changed teachers. We also note that a student with the same instructor both semesters does not help us demonstrate if "getting a more similar instructor" is correlated with higher test scores. For these reasons, these data were discarded (e.g., for the 1-norm, this reduced the by n=52 to a new sample size of n=248) . Another possible reason for being "on the edge" of a region is scoring the precise same distance from the course average; this should happen infrequently (i.e., a student scoring precisely 3% above the course average for both semesters), and so any such data was also not considered for this test.

Linear Regression

To examine more closely any connection indicated by the χ^2 tests, we examined several linear regression models to determine how an instructor being "more different" from the student in various ways was significant. We first considered linear regression models of the change in student grade across semesters as a function of instructor-student difference, using each of the distances discussed previously, considering all three types of grades. We then examined all possible subsets of the dimensions across each of the types of distances, continuing to examine the relationship with each of the grade types.

While single-semester grades as a function of instructor-student difference could be examined, the possible confounding variables mean that we will not focus on these results. Our study's methodology is an attempt to remedy this. To test the effect of changing the student-instructor distance, we only consider in student performance across semesters (controlling for "better" instructors and demonstrating the change in performance in a college math class over time, rather than just a "snapshot" data point). This also helps accounts for individual student bias (i.e., a consistently high-performing student will perform well across both semesters).

Results

χ^2Tests

The basic χ^2 test returned two significant results (beyond the p=.01) level: the raw instructor-student difference returned a p-value of .0004528, and the infinity-norm returned a p-value of .009032. Both of these indicate that the raw and infinity-norm instructor-student difference indicate sufficient evidence to reject our null hypothesis, and give evidence that there is some statistically significant correlation between instructor-student learning style difference and efficacy of learning, as measured by course grade. Both provide evidence in favour of the meshing hypothesis, as it is highly unlikely that this correlation is due to chance (see Appendix 1 for the full table of χ^2 test results.) (see Figure 2 for a visual depiction of the approximate distributions the raw distance overall grade χ^2 test results.).

Linear Regression

The only significant results for linear regression with all of the distance data considered are the raw student-instructor distance changes. The total grade was significant the p=0.01 level, returning a p-value of 0.00143 with a coefficient of -0.00130, and a total model adjusted R^2 of 0.0120 (see Appendix 2 for the summary of all linear regression models.).

The following were the most effective models, as measured by taking the models with the least likelihood of its results due to random chance, using fewer than four dimensions of the ILS as inputs (i.e., removing dimensions before conducting the same analysis). All possible subsets were considered, but only the major results are presented here.

Between the three categories of grades, the consistent best-performing models were all in the Non-Test Grade category (see Appendix 3 for the highlights of the results in subset analysis.). Table 1 highlights the most significant models by count of dimensions. Note that the three-dimension model is still a single factor model, we are using the distance in three dimensions to predict grade, not conducting multiple regression.

Table 1. Most Significant Subsets

Subset	4-dimension	3-dimension	2-dimension	1-dimension
Dimensions	AR,IS,VV,SG	AR, VV, SG	AR, VV	AR
P-Value	0.1342	0.0665	0.0313	.0803
$\boldsymbol{R^2}$	0.0075	0.0113	0.0155	0.0102
Norm	Infty-norm	Infty-norm	Infty-norm	N/A

Of note, the most statistically significant linear regression model of all possible subsets of our data occurred in the R^2 space of AR,VV. When the IS dimension is removed, we observe a significant improvement in the model's p-value, and when the SG dimension is also removed, we observe a similarly significant model improvement. This means that the infinity norm of AR and VV dimensions is the best predictor of grade improvement, and that the added dimensions of IS and SG increase the noise in the data. In other words, the IS and SG dimensions are occasionally the largest dimension (recorded as the infinity norm), but the model loses predictive power when we add this additional information. Unfortunately, when we remove the VV dimension, the p-value worsens indicating that the infinity norm of the AR, VV dimensions is the best predictor of student grade improvement.

Statistical Significance

As mentioned previously, Pashler et al proposed a standard of evidence which would indicate that learning styles were causing the effect. The methodological design of this paper's study, while different than what Pashler et al considered due to our allowance for learning style distance, lies solidly in the "Acceptable Evidence" standards, as we consider only those individual students whose performance increased or decreased. The consequence of the two-semester sequence, with its near-random reassignment of students, washes out the potential effects highlighted in by their "Unacceptable Evidence" categories.

As demonstrated by the χ^2 tests, there is likely a correlation between the change in learning style distance and the performance of a student as compared to the average performance in our classes. This is contrary to some researchers' assertions that the "meshing hypothesis" is all but "debunked" (Cuevas, 2015). The strength of the χ^2 tests compared to the relative weakness of the simple linear regression models indicates that the simple distances proposed in this paper may not adequately capture the interaction but do indicate that a relationship exists.

Practical Significance

To consider the practical effect measured in our analysis, we consider the strongest of the full linear regression models, the coefficient of -0.00130 on the raw difference metric (see Appendix 2, Raw Cumulative Model). For a hypothetical student that shifts from a "more distant instructor" distance of 26 to a "more similar instructor" distance of 8 (or from roughly the average of the 3rd quartile of instructor student distances to the average in the 1st quartile), our model indicates an improvement in course score of approximately 2%. Put a different way, a more similar instructor could be the difference between a B and a B+.

If we consider the most significant subset of the infinity-norm, using the active-reflective and visual-verbal dimensions to model cumulative grade, the slope coefficient is -.00146. Taking the effect of another hypothetical student shifting from a more different instructor of distance 18 to a more similar instructor at a distance of 4 (again from the middle of the 3rd quartile of instructor-student distances to the 1st quartile), our model indicates a course score improvement of just under 2%.

These were the most statistically significant of the models considered, and they demonstrate a mild shift in course performance without accounting for any other factors. So, while the results are statistically significant, our data indicate that learning style similarity with instructor, on its own, only accounts for up to a 2-3% change in a student's performance in a class. As previously discussed, this model fails to capture all the factors that affect student performance, but it does provide a no-cost method of improving student learning by assigning students to similar instructors.

Conclusions

Our study shows that, with our three novel approaches (standardized learning style measurement, differenced grades, and the consideration of different norms), our data indicate a mild support for using the idea behind the meshing hypothesis to provide recommendations for student-teacher pairing. The statistical significance of the χ^2 test and the consistent "negative slope" of the linear regression models show a relationship between learning style distance and student grades. While further data analysis or more data would provide greater fidelity, every statistically significant model examined in this study had a negative slope, showing the strength of our three techniques in measuring the meshing hypotheses. The data we offer seems to show that the most predictive norms are the raw norm and the infinity norm. Our data indicates that despite the noise we observe from hundreds of students' varying situations in the stress of life as a student at West Point, there is a correlation between more similar instructor-student learning styles and improved student performance. While we do not expect learning style distance to be the dominant feature in a student's ability to learn math, these results do demonstrate that it matters.

Our study also addresses a challenge in quantifying instructor-, student-, and content-specific "learning styles." The idea that an instructor teaches in a way that makes sense to them (i.e., consistent with their learning style) is a simplifying assumption that has proved to be useful. The challenge with this assumption is that it does not provide a way for an instructor to better reach students in a class that are not like the instructor. We hope it encourages instructors to deliberately broaden their perspectives to mitigate this potential weakness. We also hope this study serves as a tool that can help larger courses to direct their students in a way to help the students perform best (i.e., guide students to the instructor with which they will learn the best).

Further Study

To further study this concept researchers at another institution could attempt to replicate using techniques comparable to our novel approach, perhaps considering different courses or examining performance across different courses. Additionally, there is more data analysis we could do with our collected data. We briefly analysed other distance metrics in the p-norm family, but did not consider any of the many other norms in the math world. We also considered multiple regression, but abandoned it after limited results and given how few independent variables were present in the model.

Acknowledgements

The authors would like to thank the Department of Mathematical Science at West Point for its culture promoting asking questions and encouraging its instructors to lean how they teach best. We would also like to thank our peers and students for bearing with us as we offer ideas and request suggestions as this research developed. Most important, each author owes his wife a debt of gratitude for putting up with us.

References

Allcock, S.J. and Hulme, J.A. (2010). Learning styles in the classroom: educational benefit or planning exercise? *Psychology Teaching Review, 16*(2), 67–79.

Choi, I., Lee, S.J., & Kang, J. (2009). Implementing a case-based e-learning environment in a lecture-oriented anesthesiology class: Do learning styles matter in complex problem solving over time? *British Journal of Educational Technology, 40*(5), 933–947.

Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). *Should we be using learning styles? What research has to say to practice.* Retrieved from http://hdl.voced.edu.au/10707/64981

Cuevas, J. (2015). Is learning styles-based instruction effective? A comprehensive analysis of recent research on learning styles. *Theory and Research in Education, 13*(3), 308-333.

Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. *Engineering education, 78*(7), 674-681.

Hsieh SW, Jang YR, Hwang GJ, et al. (2011). Effects of teaching and learning styles on students' reflection levels for ubiquitous learning. *Computers & Education, 57*(1), 1194–1201.

Kappe, F.R., Boekholt, L., den Rooyen, C., et al. (2009). A predictive validity study of the Learning Style Questionnaire (LSQ) using multiple, specific learning criteria. *Learning and Individual Differences, 19*(4), 464–467.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological science in the public interest, 9*(3), 105-119.

Riener, C., & Willingham, D. (2010). The myth of learning styles. *Change: The magazine of higher learning, 42*(5), 32-35.

Sankey, M.D., Birch, D., & Gardiner, M.W. (2011). The impact of multiple representations of content using multimedia on learning outcomes across learning styles and modal preferences. *International Journal of Education and Development using Information and Communication Technology, 7*(3), 18–35.

Van Zwanenberg, N., Wilkinson, L. J., & Anderson, A. (2000). Felder and Silverman's Index of Learning Styles and Honey and Mumford's Learning Styles Questionnaire: how do they compare and do they predict academic performance? *Educational Psychology, 20*(3), 365-380.

Author Information

Devon Zillmer
https://orcid.org/0000-0003-2327-9529
United States Military Academy
MADN-MTH 646 Swift Road West Point, NY 10996
U.S.A.
Contact e-mail: *devon.zillmer@westpoint.edu*

Thomas Mussmann
https://orcid.org/0000-0002-1644-6635
United States Military Academy
MADN-MTH 646 Swift Road West Point, NY 10996
U.S.A.
Contact email: *thomas.mussmann@westpoint.edu*

Appendix 1. χ^2 Test p-values

Complete table of p-values for basic χ^2 investigation. The complete table of counts of the binned data were omitted due to size constraints; for a complete table of counts of results, please contact the authors.

Grade Type	Cumulative	Test	Non-Test
Raw Distance	.0004528	.00469	.1861
1-norm	.2024	.3818	.6348
2-norm	.07047	.2088	.7775
Infinity Norm	.009032	.1322	.1046

Appendix 2. Linear Regression Estimates and p-values

Below is the table showing the major linear regression results of the analysis of the norms. Again due to size constraints, please contact the authors for a similar examination of the by-semester scores (below are the change in grades across semesters, as a function in the change in student-instructor distances).

		Cumulative		Test		Non-Test	
		Value	p-val	Value	p-val	Value	p-val
Raw	Intercept	.000618	.742	.000716	.788	.000489	.839
	Slope	-.00130	.00143	-.00150	.00906	-.00103	.0479
	Adj. R^2	.0120		.0194		.00976	
1-norm	Intercept	-.000264	.890	-.0359	<2e-16	.00697	.00423
	Slope	-.000208	.262	-.000226	.387	-.000200	.395
	Adj. R^2	.000879		-.000832		-.000913	
2-norm	Intercept	-.000368	.847	-.0360	<2e-16	.00686	.00496
	Slope	-.000468	.167	-.000493	.303	-.000471	.273
	Adj. R^2	.00306		.000211		.000681	
Infinity-norm	Intercept	-.00400	.834	-.0360	<2e-16	.00669	.00601
	Slope	-.000595	.134	-.000493	.303	-.000800	.112
	Adj. R^2	.00418		.000211		.00513	

Appendix 3. Most Significant Subset Analysis

Following are the most significant models for each of the grade types, among all possible subsets of the four dimensions of the LSI. Please contact the authors for the full results of all possible dimensional subsets.

		Cumulative	**Test**	**Non-Test**
1-norm	Slope Value	-.00452	-.000433	-.000798
	P-value	.139	.184	.0390
	R^2	.00736	.00591	.0142
	Factors	AR, VV	AR, SG	SI, SG
2-norm	Slope Value	-.000803	-.000532	-.00129
	P-value	.0517	.240	.0136
	R^2	.0126	.00463	.0203
	Factors	AR, VV	AR, SI, SG	AR, VV
Infinity-Norm	Slope Value	-.000977	-.000564	-.00146
	P-value	.0313	.228	.0111
	R^2	.0155	.00487	.0214
	Factors	AR, VV	SI, SG	AR, VV
Single Factor	Slope Value	-.000672	-.000544	-.000875
	P-value	.0803	.316	.0725
	R^2	.0102	.00338	.0108
	Factors	AR	AR	AR