Mitigating Biases using an Additive Grade Point Model: Towards Trustworthy Curriculum Analytics Measures

Frederik Baucks[1] and Laurenz Wiskott[1]

Abstract: Curriculum Analytics (CA) tries to improve degree program quality and learning experience by studying curriculum structure and student data. In particular, descriptive data measures (e.g., correlation-based curriculum graphs) are essential to monitor whether the learning process proceeds as intended. Therefore, identifying confounders and resulting biases and mitigating them should be critical to ensure reliable and fair results. Still, CA approaches often use raw student data without considering the influence of possible confounders such as student performance, course difficulty, workload, and time, which can lead to biased results. In this paper, we use an additive grade model to estimate these confounders and verify the validity and reliability of the estimates. Further, we mitigate the estimated confounders and investigate their impact on the CA measures course-to-course correlation and order benefit. Using data from 574 Computer Science Bachelor students, we show that these measures are significantly confounded and mislead to biased interpretations.

Keywords: curriculum analytics, confounding, bias, mitigation, fairness

1 Introduction

Trustworthy results of measures used to describe and understand student activity in degree programs are essential to ensure fair and appropriate decisions for improving degree quality. Therefore, identifying student-dependent and course-dependent confounders and bias is central to guaranteeing equal opportunities for all students.

Curriculum Analytics (CA), as a sub-field of Learning Analytics (LA), aims to assess degree program quality and course relations using student data and program structure. CA methods associated with process mining and prediction use descriptive measures such as the correlation between course grades [Ba18; Ra21] and order benefit (OB) [Gu21] to describe course relations and build graphical curriculum representations. In order to measure relations, e.g., the content overlap of courses, using course grades, confounders [HR20] need to be identified and mitigated [WGD22]. Existing descriptive CA measures try to mitigate confounders, like student performance, by, e.g., normalizing student grades but do not quantify student and course-specific confounding that may bias the interpretation of the measured outcome. In addition, the normalization of grades often leads to a loss of information, e.g., correcting grades using the grade point average (GPA)

[1] Ruhr-Universität Bochum, Institut für Neuroinformatik, Universitätsstr. 150, 44801 Bochum, frederik.baucks@ini.rub.de; laurenz.wiskott@ini.rub.de

doi: 10.18420/delfi2023-12

[Oc16], which is known to be confounded by itself [HSS19]. The resulting confounded or even biased graphical representations of the curriculum are further used in process mining and prediction methods, leading to concept drift issues [BCR18] or violations of the IID assumption, respectively.

This paper addresses the open problem of quantifying and mitigating possible confounders in descriptive CA measures. We adopt a matrix factorization (MF) approach [RRS10] to model statistically independent confounders accounting for student performance, course difficulty, workload, and time. This methodology can quantify confounding in widely used CA measures such as course-to-course correlation and OB and mitigate it. Using course point grades of a Computer Science bachelor's degree program, we reveal that measure interpretations can be significantly and misleadingly biased, if confounders are not mitigated. When employing CA analyses, we reveal that our additive model is a valuable preprocessing tool. Our key contributions are **1.) Additive modeling as a CA methodology:** We modify MF to additively model different student and course-dependent confounders as statistically independent random variables, namely student performance, course difficulty, workload, and time. Further, we test the estimated confounders for validity and reliability. Using the independence of the modeled estimates, we can quantify their impact using variance and also mitigate them. **2.) Case study application:** Using data from CS students over nine years, we quantify the confounding in the grade data and the effect of mitigation on the widely used course-to-course correlation and the OB between courses. We observe that student performance and course difficulty confounders had the most impact, accounting for 54.8% of the variance, while workload and time accounted for less than 2%.

2 Related Work

Curriculum Analytics (CA) is a recognized branch of Learning Analytics (LA) [Gr16] and examines the structure of curricula to understand degree programs and improve their quality [Hi22]. One central tool of using curriculum structure is a representation of the curriculum, usually graphical, called a curriculum graph [Ra21]. These rely on descriptive measures (e.g., correlation), prerequisites, or process mining techniques (e.g., Bayesian networks). When no prerequisites are available, typically, curriculum representations use course grades to catch relations between courses, e.g., a prerequisite structure [Ba18]. Although grades are known to suffer from confounders [Oc16], sometimes called confounding bias [WGD22], these confounders are usually not addressed in the representations if the data set is to small for causal modelling. This paper focuses on grades used to calculate descriptive measures for building curriculum graphs. We tackle the open problem of optimally identifying, quantifying, and mitigating confounders to address distribution shifts in hindsight applications such as visualization, process mining (concept drift [BCR18]), or next-term grade prediction (IID assumption [BSW23]). Only some approaches, not necessarily used in CA, already consider potential confounders in grades and their induced biases, namely normalization, item response theory (IRT), MF, and covariate adjustment (CAD).

Normalization tries to remove the effects course difficulty and student performance from student grades in different contexts. The standard approaches here are to subtract the GPA as a measure of student performance, model the course difficulty as the distance between students' GPA and course outcomes [Oc16], or normalize grades course-wide to account for course difficulty [Gu21]. A significant downside to this is that the GPA is confounded by itself [Jo97] and that, therefore, the subtraction and normalization may lead to a considerable loss of information. IRT addresses this problem by modeling students' grades using student and course-dependent latent traits in a logistic model [BSW23]. In degree-wide applications, IRT is used to model students' GPAs independently of the courses attended [Jo97] or to model course difficulties independently of students attending [BSW23]. One limitation is that the latent traits shift to zero mean and unit variance, so the latent traits are not easy to interpret.

Like IRT, MF approaches model grades using latent variables and low dimensional estimates of the interaction between students and courses. MF is commonly used for recommender systems and next-term grade prediction [BBR21]. Most MF approaches incorporate a so-called student bias and course bias. In contrast to IRT, these biases are modeled additively and not with unit variance. This additivity allows us to fit more so-called biases while staying within the original scale of the grades, leaving grades and biases comparable. This approach is comparable with CAD, in which variables (e.g., confounded GPA) are fitted in a multivariate regression model to address their influence on course grades [WGD22]. In our approach, we are fitting latent variables using MF as approximately statistically independent summands, eliminating the confoundedness of the fitted variables in CAD approaches (e.g., GPA).

The grade prediction MF approach by Barrollet et al. [BBR21] is closest to our work. They tried to capture static student and course information in degree program data. The resulting one-dimensional MF model yielded a good fit using student and course bias estimates. In contrast, our paper employs a reduced MF approach, setting the lower dimensional representation to zero, fitting only bias factors, which we call *confounders*. Further, we re-purpose MF to describe exclusively historical course grade data, fitting student performance, course difficulty, workload, and time confounders. Our work mitigates these confounders and enables a bias-reduced perspective on CA measures.

3 Methodology

We assume a prerequisite-free curriculum consisting of courses offered several times over *different* semesters. We define these offerings as *course offerings* (CO) and write the word *course* for the time-aggregated COs.

3.1 Confounding and Bias

To reduce possible bias in interpreting CA measure results, it is essential first to define confounding. Assume, we are given grades in two COs, C_1 and C_2, and a third random variable G, where the grades in C_1 and C_2 depend on G. Suppose we calculate the correlation between the grades of COs C_1 and C_2, then variable G confounds the correlation as a measure. For example, G could be the student performance measured by the grade point average (GPA). That is, GPA (G) affects the measure correlation through the statistical dependence, e.g., we expect high GPA students to perform well in both COs C_1 and C_2.

To describe bias in that context, suppose we do not want to calculate the correlation between CO's C_1 and C_2 but rather estimate the courses' content overlap (OV) based on the grades. For example, an estimator for OV could again be the correlation of the grades of the two COs C_1 and C_2. That is, we approximate OV by correlation and denote the correlation as the estimate of OV. Then our result is impacted because the confounder G shifts our estimate (correlation) systematically. Therefore, it is, in particular, shifted away from the concept OV that we expect it to estimate. This shift, then, is called bias or biased interpretation. Overall, confounders can lead to bias if they systematically shift our estimate away from what we want to measure. So, it is highly relevant what concept we want to measure, and which estimate we use to measure it to distinguish between confounding and bias. The first step to counteract bias is to determine and eliminate confounders. Because then we can reduce the confounding effect on the grades of C_1 and C_2 and thereby get a more bias-free OV estimate. Therefore, confounders (e.g., student performance) and, in this context, as a source of bias, should be identified and eliminated. In theory, we can write any grade $g_{s,c}$ of a student s in CO c, as the sum of the confounder $b_{s,c}$ and the deconfounded grade $\xi_{s,c}$:

$$g_{s,c} = \xi_{s,c} + b_{s,c}. \tag{1}$$

To align with MF [RRS10], $\xi_{s,c}$ and $b_{s,c}$ are assumed to be statistically independent. If we know $g_{s,c}$, $b_{s,c}$, and $\xi_{s,c}$ for any student s and CO c, we can calculate the variances of each summand, which add up to the variance of $g_{s,c}$ due to the independence:

$$Var(g_{s,c}) = Var(\xi_{s,c}) + Var(b_{s,c}) \tag{2}$$

Further, we quantify the confounding in the grade data $g_{s,c}$ as the variance of the confounder $b_{s,c}$ relative to the variance of the overall grade $\xi_{s,c} + b_{s,c}$:

$$\frac{Var(b_{s,c})}{Var(g_{s,c})} = \frac{Var(b_{s,c})}{Var(\xi_{s,c}) + Var(b_{s,c})} \tag{3}$$

This equality becomes helpful in the following additive grade point model, where we approximate different confounders as independent summands, so that we can write the confounder $b_{s,c}$ as a sum of more specific confounders. Then, we can use the variances to quantify the confounding in the grades using Eq. 3.

3.2 Additive Grade Point Model

To model grades as the sum of independent summands, we use a *bias-only* approach motivated by MF, where we set the factorization to zero [KRB21]. The model approximates each grade $g_{s,c}$ as the sum of confounders, namely student performance b_s, CO difficulty b_c, workload b_w, and time b_t. Let Υ be the set of tuples of students and CO grades that exist in our data and μ_Υ the mean of all grades, and $\xi_{s,c}$ the model error, then we write:

$$g_{s,c} = \tilde{g}_{s,c} + \xi_{s,c} = (\mu_\Upsilon + b_s + b_c + b_w + b_t) + \xi_{s,c} \quad (4)$$

CA measures, such as the correlation between course grades, are only applied to historical data. Especially for COs held in the past, we do not expect to gain more grades in the future. Therefore, we estimate the confounders only in-sample since we know the exact error $\xi_{s,c}$ for each grade here. We declare $\mu_\Upsilon + \xi_{s,c}$ as the *deconfounded grade*. We optimize the grade estimates $\tilde{g}_{s,c}$ for all $(s,c) \in \Upsilon$ using the optimization problem

$$\min \textstyle\sum_{(s,c)\in\Upsilon} \left(g_{s,c} - \tilde{g}_{s,c}\right)^2 + \lambda(|b_s| + |b_c| + |b_w| + |b_t|), \quad (5)$$

and stochastic gradient descent for optimization with step size $\alpha = 0.005$ and regularization parameter $\lambda = 0.02$ since these led to convergence. Using the mean grade value μ_Υ as a positioning summand and the regularization in the optimization, we can identify confounders on the same scale as the confounded grade $g_{s,c}$. Using this identification, we can mitigate the confounders from $g_{s,c}$ by subtraction while staying on the same grade scale, so that the deconfounded grade $\mu_\Upsilon + \xi_{s,c}$ stays comparable to $g_{s,c}$.

Model Assumption: The central model assumption is the time-invariance of the confounders. For courses, we address the invariance by using COs. For workload and time, we use categorical variables. However, whether an invariant student confounder b_s is sufficient must be verified [BBR21]. Therefore, we establish reliability in our student confounder by employing a time-dependent split-half test described in Section 3.3. Here, we compare estimates of models fitted to student's first- and second-half grades.

Model Fit and Independence: To test the model fit, we compare the variances of the confounded grades $Var(g_{s,c})$ against the variances of the model estimates $Var(\tilde{g}_{s,c}) + Var(\xi_{s,c})$. Due to the variance argument made in Eq. 3, these should be equal if we achieve optimal model fit. We expect the variance of the fitted confounder estimates to be higher if they correlate. In addition, we investigate the statistical independence of each summand using Pearson correlation, which is a necessary criterion. All confounders should be uncorrelated to the deconfounded grade $\mu_\Upsilon + \xi_{s,c}$ and to each other.

3.3 Validity and Reliability Assessment

Concurrent Validity: We study concurrent validity by considering correlations between b_s vs. GPA and b_c vs. mean CO grade. In line with IRT research [BSW23], we expect high positive correlations indicating that the student performance confounder measures student performance and CO difficulty confounder measures CO's difficulty.

Internal Consistency Reliability: To test the stability of our model, we employ two split-half reliability tests [BSW23]. First, we split the training data into two randomly assigned disjoint data sets of equal size and calculate the Pearson correlation between the estimates of the two fitted models. Second, we employ a time-dependent split-half test for the time-invariance assumption (Section 3.2). We separate each student's grades in the data set into the earlier and later half. Then we compare the fitted confounders of the earlier and later half of the student's history using Pearson correlation. We expect a high correlation indicating a stable time-invariant student performance confounder.

3.4 Curriculum Analytics Measures

To visualize the impact of the estimated confounders, we calculate two established CA measures for our curriculum twice, where we aggregate the COs over time into single courses as it is common in the literature. First, we calculate each measure for all course combinations similar to the literature, and second, using the deconfounded estimates $\mu_Y + \xi_{s,c}$ as grades, leading to bias-reduced CA measure estimates of course relation.

Correlation: CA methods use correlation to estimate course content overlap or other course relations, e.g. [Ba18; Ra21], to construct graphical representations of the curriculum. We calculate Pearson correlation using minimum p-value and sample size according to the existing literature on CA. We use the standard threshold of $p< 0.05$ and calculate the correlations between grades of courses with a minimum size $N = 20$ [Ba18].

Order Benefit: Order benefit (OB) assesses the influence of the order of a pair of courses on the corresponding grades [Gu21]. For two courses, A, B let $S_{A\to B}$ be the subset of students who first took course A and then B in some COs. Then for all students $s \in S_{A\to B}$ exist grades $g_{s,A}(t_1)$ and $g_{s,B}(t_2)$ with student-dependent times t_1 and t_2 for courses A and B with $t_1 < t_2$. We denote the set of grades of course A of the group $S_{A\to B}$ as $G_A(S_{A\to B})$ and the corresponding ordered mean grade as $\mu_A(A \to B) \coloneqq \mu_{G_A(S_{A\to B})}$. Analogously for $\mu_B(A \to B)$, $\mu_A(A \to B)$, and $\mu_B(A \to B)$. Then, for two courses A, B, the order benefit $OB_{A\to B}$ from course A to course B is defined as:

$$OB_{A\to B} \coloneqq \mu_A(A \to B) + \mu_B(A \to B) - \mu_A(B \to A) - \mu_B(B \to A). \tag{8}$$

This measure is only significant if there is sufficient data. We set two thresholds to guarantee that the two order groups $S_{A\to B}$ and $S_{B\to A}$ are sufficiently large and comparable in size [Gu21]: 1.) $S_{A\to B}, S_{B\to A} > 50$, 2.) $\frac{\min(S_{A\to B}, S_{B\to A})}{\max(S_{A\to B}, S_{B\to A})} > 0.30$.

4 Experiments

Data Set: For our research, we utilized a data set containing exam scores from a CS bachelor's program at Ruhr-University Bochum in Germany. The data set includes data from 1098 students who took 19 compulsory courses between 2013 and 2022, including still enrolled, graduated or drop-out students. The grading scale for each exam is from 0 to 100, with a passing score of at least 50. Each CO grade was determined via a single written examination taken at the end of the semester. Before the data was obtained, anonymization was performed, which involved removing all demographic information and adding a uniform stochastic noise between -5 to 5 to each grade. To ensure a more stable parameter fit, we filtered the data to include only students with at least 10 non-zero grades and COs with at least 20 students. Further, we limited the grades to first try-exams to strengthen student confounders reliability. The resulting data set comprises 574 students and 127 COs. In addition to exam data, we have the CO‘s corresponding workload and recommended semester in which a CO should be examined. We model both workload and time as categorical variables. For workload, we measure each students mean compulsory CO workload over all semesters. The resulting distribution of all students' mean workloads is split into three equal-sized parts, where each part consists of the same amount of students‘ mean workloads. Then we label each student with {0, 1, 2} corresponding to the subpart of the distribution. We model time, similarly, as a variable in {0, 1, 2}. We assign the label 0 if the semester the student takes a CO is before the recommended time, 1 if taken as recommended, and 2 if taken later.

$Var(g_{s,c})$	$Var(\tilde{g}_{s,c}+\xi_{s,c})$	$Var(\xi_{s.c})$	$Var(b_s)$	$Var(b_c)$	$Var(b_w)$	$Var(b_t)$
715.248	737.831	303.787	318.746	100.504	14.498	0.295

Tab. 1: Variance model fit showing an error of 22.583 between grade point variance and fitted Variance, indicating a sufficient decorrelated fit. Student and CO confounders b_s, b_c account for most of the variance, whereas workload and time confounders show small variances.

4.1 Model Fit

Additive Model Fit: To evaluate the quality of the model fit, we calculate the difference of the variances of the grades in the data set and the respective estimates in Table 1. The difference of 22.583 is relatively small and indicates the approximated independence of the summands in Eq. 4. The variance of the student performance and CO difficulty confounders is the highest, accounting for 54.8% of the total variance. The workload confounder {0: -6.715; 1: 1.721; 2: 0.948} indicate poorer performance among students with a low mean workload. Categories 1 and 2 do not seem to have an impact, and the time confounder categories have negligible effects on grades (<1.000).

Independence of Summands: A statistically significant low Pearson correlation of less than 0.06 provides the necessary criterion that our modeled confounders are approximately independent. In Fig. 1, we see the correlations of all confounder estimates on the left side, and the scatter plot of b_s and the deconfounded grade $\mu_Y + \xi_{s,c}$ on the right side, showing the most structure of all variable combinations, which remains very weak.

	b_s	b_c	b_w	b_t	$\xi_{s,c}$
b_s	1	−0.024	0.055	−0.043	0.025
b_c		1	−0.053	−0.020	0.011
b_w			1	−0.007	0.005
b_t				1	0.001
$\xi_{s,c}$					1

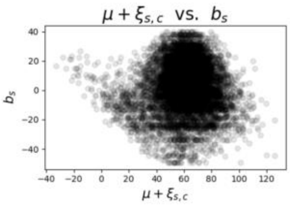

Fig. 1: [Left] Table showing Pearson correlations between the estimates of the additive model indicating very low correlations for all combinations. [Right] Scatter plot indicating statistically significant (p<0.001) absence of correlation between deconfounded grades $\mu_Y + \xi_{s,c}$ and student confounder estimate b_s as fitted by the additive model.

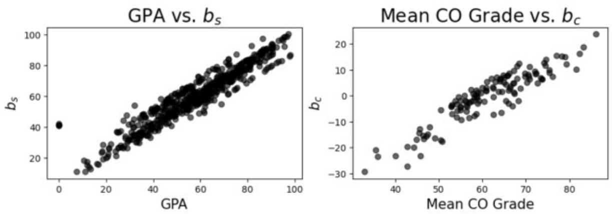

Fig. 2: [Left] Scatter plot indicating high correlation (r = 0.966, p < 0.001) between student's GPA and confounder b_s as fitted by the additive model. [Right] Scatter plot indicating high correlation (r = 0.928, p < 0.001) between mean CO grades and fitted CO confounder b_c.

4.2 Validity and Reliability Assessment

Concurrent Validity: In order to quantify the validity of the student performance and CO difficulty confounders, we show in Fig. 2 two scatter plots indicating very high statistically significant Pearson correlation (r > 0.900, p< 0.001) for both $b_s$vs. GPA and b_c vs. mean CO grades. These show us that the confounders we fitted do describe what we expect them to describe: student performance and CO difficulty.

Internal Consistency Reliability: Following the methodology, we first obtain Pearson correlations for the student performance and CO difficulty confounders using a random split-half test. The highly significant correlations ($p < 0.001$) for the student confounder b_s with 0.83 and the CO confounder b_c with 0.92 indicate a stable model fit. Second, the time-dependent split-half reliability test points to a similar correlation for b_s by 0.84, indicating that the student confounder estimates are stable for the first and second half of student grades. The high correlation suggests that a time-invariant student confounder is an adequate assumption in our setting.

4.3 Curriculum Analytics Measures

We calculate OBs and correlations for each course pair in the degree program. We visualize the pairs using graphs, where vertices represent courses and edges represent the value of a CA measure. The visualizations in Fig. 3 and Fig. 4 show graphs of the OBs and correlations, where the order of courses corresponds to the university's recommendation. We arrange the graph vertices from bottom to top in semester ascending order.

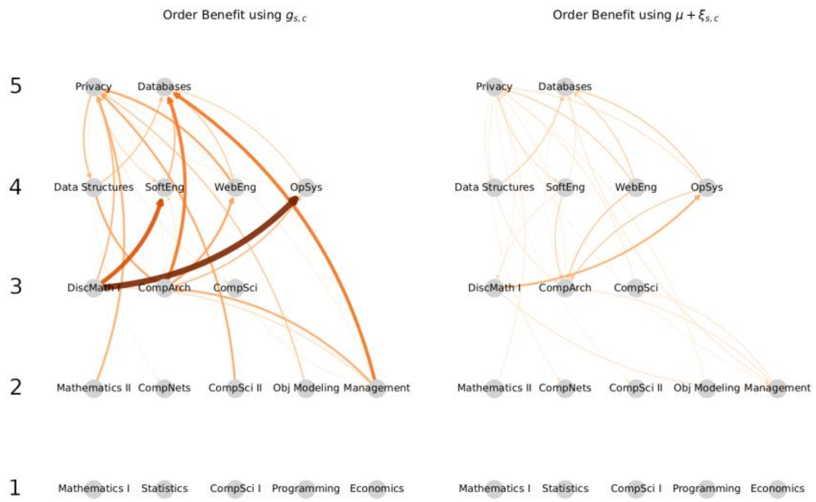

Fig 3: Curriculum graphs with colored OB edges indicating the beneficial direction of study. Color intensity and edge thickness corresponds to the value of the OB. [Left] We us the grades $g_{s,c}$. [Right] We use deconfounded grades $\mu_Y + \xi_{s,c}$.

Order Benefit: Since the OB is a directed and asymmetric measure, we visualize only positive edges in Fig. 3. The thickness and color intensity corresponds to the value of the OB, where the color scale is the same for both figures. The left side shows the OBs for the confounded grades $g_{s,c}$. Notably, there are no edges to first-semester courses, indicating that students violate the recommendation in first-semester courses not enough and, therefore, not meeting the thresholds. For existing edges, we observe that most point from bottom to top in the recommended direction of study. Compared to that on the right side, we can observe a different picture when using deconfounded grades. In general, edge weights appear less intense, and edges point far more often from top to bottom against the recommendation. The mitigation has an enormous impact on all edges of the graph, which often even change directions, indicating that the OB is largely confounded and biased as an estimate of beneficial course direction.

Correlation: Correlation is a symmetric undirected measure. Thus, we visualize undirected edges on the color scale [0, 1] using the absolute r-value. The graph in Fig. 4 on the left side, showing the correlation of the confounded grades, is characterized by high density. Almost every edge holds the conditions for sample size and p-value. The graph of the deconfounded grades (right) is much sparser. Only single edges stand the conditions. Similar to OB, mitigation dramatically impacts the results of the measure. Therefore, correlation is biased as an estimate of, e.g., course overlap. Interestingly, the edges now show expected insights for courses related strongly content wise, e.g., (Mathematics I/Mathematics II) and (CompSci I/CompSci II).

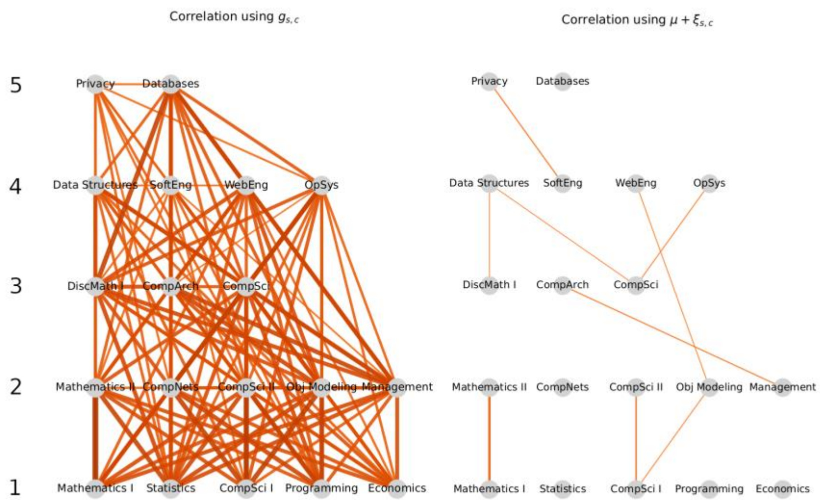

Fig. 4: Curriculum graphs with colored correlation edges indicating the correlation of grades pairs of courses. Color intensity and edge thickness corresponds to the absolute value of the correlation. [Left] We use the grades $g_{s,c}$. [Right] We use the deconfounded grades $\mu_Y + \xi_{s,c}$.

5 Discussion and Future Work

Our experiments show that the additive model in Eq. 4 can be used to estimate and mitigate course offering (CO) and student-specific confounders that induce bias in the context of graphical representations of the curriculum. We show that the amount to which Curriculum Analytics (CA) measures, such as the order benefit and correlation, are confounded can be substantially high. The confoundedness distorts the outcome completely and, ultimately, leads to severe bias.

We estimated reliable and valid confounder estimates relating to student performance, CO difficulty, workload, and time. Student and CO confounder estimates had the most impact, accounting for 54.8% of the variance, whereas categorical workload and time confounders account for less than 2%. Although the small variance of the workload confounder is in in line with research [PBY23], it should be interpreted carefully. Workload, in our setting, is limited to the compulsory courses of the degree program and given as a relative position to all students. The workload could be distorted by non-compulsory courses that are not included in the data, taken together with compulsory courses. Second, the workload assigned to courses by the university is not guaranteed to represent the courses well [MOC14]. Assessing non-compulsory course grades and the quality of workload assignments to courses will be important in future work.

CA-related process mining and prediction approaches use confounded graphical representations of the curriculum. They do not estimate confounders as statistically independent variables and even assume time-invariant COs as the concept drift issue in process mining [BCR18] or the IID assumption in prediction show. Our additive model can mitigate such limitations. However, one possible limitation of our model is the critical assumption of a time-invariant student confounder. We accounted for that in two ways. First, we limited the data set to first exam attempts. Second, we employed a time-dependent split-half reliability test that showed the stability of the student confounder. Future work can contain an extended model using a time-dependent student performance change factor. Then we can include the second exam tries on the student side.

This limitation does not occur for CO estimates. Similar to item response theory (IRT) [BSW23], an exciting line of future research is the investigation of model estimates for each CO to monitor CO difficulty independent of students attending. In contrast to IRT's unit variance transformation of confounders, our approach leaves them explainable. Compared to normalization and covariate adjustment approaches using student GPAs to mitigate confounders [Oc16], our approach yields optimized approximately statistically independent confounders. Conversely, mitigation using the possibly confounded GPA [Jo97] does not guarantee to catch confounders.

In conclusion, we hope that additive models for confounders and bias estimation become standard when employing CA measures, leading to reliable and fair measure results inducing bias-reduced interpretations.

Acknowledgements: This work is funded by the Ministry of Science (NRW, GER) via the project KI.edu:NRW.

Bibliography

[Ba18] Backenköhler, M.; Scherzinger, F.; Singla, A.; Wolf, V.: Data-Driven Approach Towards a Personalized Curriculum. In Proc. (Boyer, K; Yudelson, M.) of the 11th Conf. of Educational Data Mining, Raleigh, NC, USA, 246-251, 2018.

[BBR21] Barollet, T.; Bouchez Tichadou, F.; Rastello, F.: Do Common Educational Datasets Contain Static Information? A Statistical Study. In Proc. (Hsiao, I.; Sahebi, S.; Bouchet, F.; Vie, J.) of the 14th Int. Conf. of Educational Data Mining, Int. Data Mining Soc., Paris, France, 310-316, 2021.

[BCR18] Bogarın, A.; Cerezo, R.; Romero, C.: A survey on educational process mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8/1, 2018.

[BSW23] Baucks, F.; Schmucker, R.; Wiskott, L.: Tracing Changes in University Course Difficulty Using Item Response Theory. AI4EDU Workshop at AAAI 2023.

[Gr16] Greer, J.; Molinaro, M.; Ochoa, X.; McKay, T.: Learning analytics for curriculum and program quality improvement (PCLA 2016). In Proc. (Dawson, S.; Drachsler, H.; Penstein Rosé, C.) of the 6th Int. Conf. on Learning Analytics & Knowledge. Assoc. for Computing Machinery, New York, NY, USA, 494-495, 2016.

[Gu21] Gutenbrunner, T.; Leeds, D. D.; Ross, S.; Riad-Zaky, M.; Weiss, G. M.: Measuring the academic impact of course sequencing using student grade data. In Proc. (Hsiao, I.; Sahebi, S.; Bouchet, F.; Vie, J.) of the 14th Int. Conf. of Educational Data Mining, Int. Data Mining Society, Paris, France, 799-803, 2021.

[Hi22] Hilliger, I.; Aguirre, C.; Miranda, C.; Celis, S.; Pérez-Sanagustın, M.: Lessons learned from designing a curriculum analytics tool for improving student learning and program quality. Journal of computing in higher education, 633–657, 2022.

[HR20] Hernán, M. A.; Robins, J. M.: Causal Inference: What If, Chapman & Hall, 2020.

[HSS19] Hansen, J.; Sadler, P.; Sonnert, G.: Estimating high school GPA weighting parameters with a graded response model. Educational Measurement: Issues and Practice 38/1, 16–24, 2019.

[Jo97] Johnson, V. E.: An alternative to traditional GPA for evaluating student performance. Statistical Science 12/4, 251–278, 1997.

[KRB21] Koren, Y.; Rendle, S.; Bell, R.: Advances in collaborative filtering. Recommender systems handbook, 2021.

[MOC14] Méndez, G.; Ochoa, X.; Chiluiza, K.: Techniques for data-driven curriculum analysis. In Proc. (Pardo, A.; Teasley, S.) of the 4th Int. Conf. on Learning Analytics and Knowledge, Association for Computing Machinery, 148-157, 2014.

[Oc16] Ochoa, X.: Simple metrics for curricular analytics. In: CEUR Work. Proc., Bd. 1590, CEUR-WS, 20–26, 2016.

[PBY23] Pardos, Z.; Borchers, C.; and Yu, R.: "Credit hours is not enough: Explaining undergraduate perceptions of course workload using LMS records." The Internet and Higher Education 56 (2023): 100882.

[Ra21] Raji, M.; Duggan, J.; DeCotes, B.; Huang, J.; Zanden, B. V.: Modeling and Visualizing Student Flow. IEEE Transactions on Big Data 7/3, 510–523, 2021.

[RRS10] Ricci, F.; Rokach, L.; Shapira, B.: Introduction to recommender systems handbook. In: Recommender systems handbook. Springer, 1–35, 2010.

[WGD22] Weidlich, J.; Gašević, D.; Drachsler, H.: Causal Inference and Bias in Learning Analytics: A Primer on Pitfalls Using Directed Acyclic Graphs. Journal of Learning Analytics 9/3, 183–199, 2022.