Politische Vierteljahresschrift
https://doi.org/10.1007/s11615-025-00613-x

PVS Politische Vierteljahresschrift
German Political Science Quarterly

RESEARCH NOTE

Following Political Science Students Through Their Methods Training: Statistics Anxiety, Student Satisfaction, and Final Grades in the COVID Year 2021/22

Paul Vierus · Jonas Elis · Achim Goerres · Conrad Ziller · Jan Karem Höhne

Received: 23 April 2024 / Revised: 4 March 2025 / Accepted: 17 March 2025

Abstract Teaching empirical social research methods as a compulsory part of a curriculum involves several challenges. Students are often unaware of the relevance of methodological training for their political science education and its value as a transferable skill. In addition, some students are afraid of the mathematical components of their applied statistics training. These challenges can have a diminishing effect on student success. We examine three different perspectives of students' satisfaction with methods courses at a large political science department in a German university. We describe temporal changes in student satisfaction over the course of a complete term (6 months) and use a set of independent variables to explain the outcomes. To do this, we fielded a longitudinal survey in five in-person methods and statistics courses during 2021/2022 after the height of the COVID-19 pandemic in Germany. We demonstrate that statistics anxiety—the self-reported worries about getting lower grades, becoming nervous, or feeling helpless when solving tasks that focus on statistics—has a substantial negative effect both on student satisfaction and on their final grades. This clear pattern raises the question of how to optimally support stu-

Paul Vierus · Jonas Elis · ✉ Achim Goerres · Conrad Ziller
Department of Political Science, University of Duisburg-Essen, Duisburg, Germany
E-Mail: achim.goerres@uni-due.de

Paul Vierus
E-Mail: paul.vierus@uni-due.de

Jonas Elis
E-Mail: jonas.elis@uni-due.de

Conrad Ziller
E-Mail: conrad.ziller@uni-due.de

Jan Karem Höhne
German Centre for Higher Education Research and Science Studies (DZHW), Leibniz University Hannover, Hannover, Germany
E-Mail: hoehne@dzhw.eu

Published online: 24 April 2025

dents who exhibit high levels of negative emotions towards statistics. Our findings contribute to the understanding of course satisfaction in academic methodological training and can be used to improve the design of courses in order to significantly reduce failure and dropout rates.

Keywords Teaching · Longitudinal · Methodology · University · Germany · Student panel

Die Begleitung von Politikwissenschaftsstudierenden durch ihre Methodenausbildung: Statistikangst, Zufriedenheit der Studierenden und Abschlussnoten im COVID-Jahr 2021/22

Zusammenfassung Die Vermittlung von Methoden der empirischen Sozialforschung als obligatorischer Bestandteil der Curricula bringt mehrere Herausforderungen mit sich. Die Studierenden sind sich oft nicht über die Bedeutung der Methodenausbildung für ihre politikwissenschaftliche Ausbildung bewusst. Darüber hinaus haben einige Angst vor den mathematischen Komponenten in angewandter Statistik. Diese Herausforderungen können sich nachteilig auf den Erfolg der Studierenden auswirken. Unser Beitrag untersucht die Zufriedenheit der Studierenden mit den Methodenkursen an einem großen politikwissenschaftlichen Institut in Deutschland aus drei verschiedenen Perspektiven. Wir beschreiben die zeitlichen Veränderungen in der Zufriedenheit der Studierenden über den Verlauf eines kompletten Semesters und verwenden mehrere unabhängige Variablen, um die Messungen zu erklären. Dazu wurde eine Längsschnittdatenerhebung in fünf Methoden- und Statistikkursen durchgeführt, welche nach dem Höhepunkt der COVID-19-Pandemie in Deutschland in Präsenz stattfanden. Wir zeigen, dass Statistikangst, also die selbstberichtete Sorge, schlechtere Noten zu bekommen, nervös zu werden oder sich hilflos zu fühlen, wenn Aufgaben mit Statistikbezug gelöst werden, einen wesentlichen negativen Einfluss sowohl auf die Kurszufriedenheit als auch auf die Abschlussnoten hat. Dieses Muster wirft die Frage auf, wie Studierende mit einem hohen Maß an negativen Emotionen gegenüber Statistik optimal unterstützt werden können. Unsere Ergebnisse tragen zum Verständnis der Kurszufriedenheit in der akademischen Methodenausbildung bei und können dazu genutzt werden, die Gestaltung von Kursen zu verbessern, um das Scheitern in Prüfungen und die Abbrecherquote der Studierenden zu minimieren.

Schlüsselwörter Lehre · Längsschnitt · Methodik · Universität · Deutschland · Studierendenpanel

1 Introduction

Empirical methods training enables political science students to critically assess empirical research findings and to conduct their own research projects. It also endows them with a set of marketable skills for future jobs. Quantitative methods form a sizeable part of these empirical methods, which reflects the increased level of applied statistics in the social sciences (Włodzimierz 2012; Clogg 1992; Maravelakis 2019). Before having familiarised themselves with the details of the curriculum, many students-to-be are not aware that studying social sciences includes a substantial proportion of training in methods and statistics, a proportion that varies among universities. This lack of prior awareness might negatively affect students' satisfaction with their political science methods courses and jeopardise their studying success.

The didactics of political science methods are especially important for instructors, much more so than in other subfields of political science. Political science methods is a subject-specific amalgam of methods used in adjacent disciplines, such as anthropology, economics, history, psychology, and sociology. Its contents are, furthermore, built on the philosophy of science, logic, mathematics, and statistics. Whereas students are most motivated to learn methods hands-on when they apply them to political science topics, the fundamentals of the methods and the added value of knowing how to use them properly originate from and reach far beyond political science. Methods instructors thus need to provide courses that (1) provide an overview of the abstract fundamentals, often examined through written exams, and (2) give students the possibility of applying methods to political science research problems, often examined through written project reports.

Little is known about the in-class reception of political science methods training, especially about how the training evolves over the semester. This empirical research note answers the following two overarching research questions:

1. What drives student satisfaction across a semester of methods training in political science?
2. How important is student satisfaction for study success as measured by grades?

We will distinguish between structural factors that shape the student experience before they enter university—for instance, their parents′ educational background and the student′s final grades in high school—along with course characteristics such as attendance, bachelor′s or master′s level, or whether a course is mandatory, as well as individual characteristics such as optimism, procrastination, and self-efficacy. To motivate the relevance of student satisfaction, previous research suggests that students' perceived satisfaction with courses, as well as the individual learning process, is pivotal to their performance and appears to mediate how teaching quality translates into study success (Keri et al. 2021). Other studies, however, find hardly any correlation between student course satisfaction and student performance, as both have distinct roots: noncognitive factors versus cognitive factors (cf. Blanz 2014). Given the prominent role of student satisfaction ratings in course evaluations, and the attributed high value of these evaluations—for instance, in application packages for professorships—it is critically important to the political science profession to investigate the link between student satisfaction and performance, as well as the role of potential confounders.

To explore the ambiguous relationship between student satisfaction and performance in the context of methods training in the study of political science, we investigated the relationship between these two factors. The relationship can be attributed to (1) course-related characteristics, (2) personal attributes of the evaluator, and (3) other, rather arbitrary factors (e.g. gender or the migration background of students). To investigate this relationship, we fielded the longitudinal Pulse Survey in the Department of Political Science at the University of Duisburg-Essen with political science students in a single-major political science study programme. We repeatedly surveyed 219 students in five empirical methods and statistics courses across four points in time between October 2021 and February 2022 to "feel their pulse" during the difficult "COVID semester". These five courses differed in the amount of applied statistics, but all of them had at least some elements of it. Moving beyond previous studies on student satisfaction with methods curricula—studies that have focused on satisfaction with methods training as a predictor of self-assessed competencies in a cross-sectional research design (Auspurg et al. 2015)—we drew a more nuanced picture of the determinants of student satisfaction by including a range of course-related and student-related factors, and we collected data across a range of courses and over time. Moreover, we asked students for permission to merge their final course grades with their survey answers. This allowed us to investigate student satisfaction, together with several aspects of the courses they took, to assess their actual studying success.

The data collection was implemented during the height of the COVID-19 pandemic in the winter semester of 2021/2022. This period effect makes the findings particularly useful, as students were experiencing extraordinary levels of organisational and psychological strain. Over time, the dynamic of the pandemic created a volatile environment in which the methods and statistics training took place.

In sum, we demonstrate that the "statistics anxiety index"—the self-reported worries at the first survey about getting lower grades in statistics courses, becoming nervous, or feeling helpless when solving tasks with a focus on statistics—has a clear and substantive effect both on the unified satisfaction index and on the students' final grades. This clear pattern raises the question of how to optimally support students with high levels of negative emotions towards statistics.

Section 2 presents the theoretical framework, including a review of previous research and the expectations for our own analyses. Section 3 lays out the Pulse Survey data and our analytical strategy. Section 4 displays the empirical results, and Section 5 summarises our contributions and suggests some wider implications of intense student surveying.

2 Measuring Student Performance in Political Science

2.1 Previous Research

High-quality teaching in higher education is a desirable outcome, whether from the perspective of students, instructors, or employers. However, conceptualising, defining, and measuring high-quality teaching is a challenging task (Goerres et al. 2015; Lambach et al. 2017). Thinking of teaching situations as an increase in the knowledge base of students through knowledge transfer (Gow and Kember 1993), this task requires some evaluation criteria. While student performance can to some extent be mapped by grades, using average grades as indicators of teaching quality is problematic because it reflects unobserved characteristics of students, for instance, their motivation, intelligence, skills, and experience. Since students are typically not randomly assigned to courses—they select themselves for a course depending on specific criteria—average course assessment grades are likely to be confounded by the composition of the group of students. To mute such hard-to-observe confounding factors, a few approaches have measured the knowledge or competencies of students before and after the course and compared the students' progress over time (Wilson 2013). While standardised pre/post-testing would theoretically provide objective measurements, implementing such a system across numerous specialised courses, especially methods-related courses in which students typically possess little or no knowledge before the course starts, poses substantial practical challenges in terms of test development, validation, and administrative resources.

Given the limited feasibility of implementing knowledge-based tests, a typical approach to measuring teaching quality has been to rely on the students' subjective course evaluations (see Pineda and Steinhardt [2020] for an overview). This approach has various advantages, as it is time efficient and delivers comparable results across courses and over time. For example, studies show that student satisfaction, as a predictor variable, to some extent explains the perceived student performance, course grades, retention rates, and graduation rates (Keri et al. 2021; Kostagiolas et al. 2019). As a dependent variable, student satisfaction is explained by several academic and course-related factors, such as the quality of course instruction, advice, and class size (Tessema et al. 2012; Jamelske 2009), as well as by individual and psychological factors, such as expectations, self-esteem, and conscientiousness (Schaeper 2020). Consequently, some studies conceptualise student satisfaction as an intermediary variable that largely mediates aspects of teaching quality on student performance (Keri et al. 2021).[1]

Apart from studies that emphasise the role of satisfaction, other research finds that student satisfaction and performance are hardly correlated (Blanz 2014). At the same time, it is important to critically reflect on the validity of student perceptions in general and student satisfaction ratings more specifically. Several assumptions underlie the use of such perceptual measures: (1) it is assumed that students remember the course content correctly when they evaluate that content retrospectively, (2) it is assumed that students assess the course independently of individual characteristics (e.g. two students with the same experience of a course should come to a similar evaluation, regardless of their personal background), and (3) it is assumed that students take only course-related content into account rather than also considering arbitrary criteria (e.g. the perceived skin colour, attractiveness, age, or gender of the instructor). These three assumptions are not easy to defend.

[1] It might also be possible that performance causally affects student satisfaction, while most empirical studies model performance as a consequence of satisfaction. This nonetheless makes it even more important to theoretically conceptualise, and empirically measure, specific aspects of satisfaction that go beyond a global measure of student satisfaction.

Regarding the question of accurate recall (assumption 1), research shows that recall bias is common in survey research (Blome and Augustin 2015). The longer that events date back, and the more complex the chain of events has been, the greater the likelihood of measurement error due to memory bias (Manzoni et al. 2010). At the same time, evaluations of a one-term course typically only require reasonable amounts of memory and cognitive capacity. Moreover, course evaluations typically ask students how often they attended the course, which might provide some insight into the reliability of their responses. The second assumption calls for a careful control strategy in which the impact of individual characteristics is held constant so that systematic variations across courses do not hamper comparisons.

The third assumption—that only educational content should be relevant for evaluations made by students—is untenable. Decades of psychological research show that people evaluate others based on categories formed by socialisation and by experiential and societal processes (Rhodes and Baron 2019). While social categorisation represents a core psychological capacity that makes the social world accessible in an efficient and predictable way, it can also entail forms of stereotyping, prejudice, and discrimination. An example would be that teachers who are perceived as "outgroup" members receive extensively negative evaluations from students, not because of their teaching but because of their "otherness". Quasi-experimental studies on bias in student evaluations show that, for example, instructors who are female and persons of colour receive lower scores on student evaluations than do white men (Chávez and Mitchell 2020). Similarly, instructors who, on average, assign better grades to students or assign a low course workload receive better evaluations than those who do otherwise (Clayson et al. 2006; Marsh and Roche 2000). Students also give higher scores to more attractive instructors (Rosar and Klein 2009). To carry arbitrary reasoning to extremes, the availability of cookies during class has been shown to improve students' evaluations of teaching (Hessler et al. 2018). Taken together, the validity and usefulness of student satisfaction ratings as a meaningful correlate of teaching quality is far from unambiguous. However, instead of abandoning the whole concept of student satisfaction, we argue that it is all the more important to further investigate its determinants and consequences with appropriate research designs. After all, student satisfaction is positively linked to learning outcomes through motivation.

2.2 Our Expectations

To disentangle the multidimensional input that informs student satisfaction ratings from the context of higher education, we included survey questions that refer to (1) satisfaction with the organisation of the course, (2) satisfaction with the preparation of the instructor, (3) satisfaction with the learning progress, and (4) overall satisfaction with the course. Such a multidimensional approach is congruent with current conceptualisations of student satisfaction (Keri et al. 2021) and should facilitate the separation of foundations that underlie an assessment of these various factors (Blanz 2014). Our explanatory framework employs a multidimensional resource model of satisfaction that guides the selection of relevant predictor variables (Marsh 1980; Green et al. 2015). Specifically, this model states that student satisfaction only partially reflects teaching quality, as it also conveys the impact of other relevant factors that determine students' expectations, as well as perceived and actual learning success. We can summarise this in hypothesis 1: Factors other than course characteristics have measurable effects on satisfaction. This assertion relates to research that shows that factors largely unrelated to courses can also have a strong relationship with perceived satisfaction. These factors include personality traits, student motivation, and stress management (Keri et al. 2021; Cotton et al. 2002).

To assess the effects of alternative explanations, we included factors such as the students' educational background before studying, motivational aspects, statistics anxiety, procrastination, and self-efficacy, while also including measures of course characteristics. Incorporating these personal characteristics is also expected to mitigate concerns about biased assessments due to unobserved confounders.

Regarding the consequences of student satisfaction, we hypothesise—in line with previous research (Keri et al. 2021)—that student satisfaction with teaching is positively associated with studying success, measured as grades obtained in the final exams of the course (hypothesis 2). At the same time, we expect the empirical relationship between student satisfaction and studying success to be substantially absorbed once competing factors are taken into account (hypothesis 3). This contention is based on research that shows that student satisfaction and performance are hardly correlated because they are rooted in different processes (Blanz 2014). While student satisfaction is more systematically related to "noncognitive" factors, such as course-related factors, social competence, and personality traits, performance is largely based on "cognitive" factors, such as learning behaviour and previous grades (Blanz 2014: p. 282).

Regarding the specific attributes of the scope of this study, teaching methodology in political science involves a number of challenges that can have a particular impact on students' success in their studies and on how teachers cope with those challenges. Methods training is generally considered to be demanding, and psychological characteristics, especially statistics anxiety (Maloney and Beilock 2012), play a role here. We thus incorporated in our study indices of statistics anxiety, procrastination, and self-efficacy. This approach should map pitfalls specific to methods courses and help us to understand the role of time-constant characteristics and the resources of students, as well as the characteristics of the courses, in shaping student satisfaction and their academic success.

3 The Longitudinal Pulse Survey 2021/2022

3.1 Target Population and Field Period

The data originates from a four-wave panel survey called the Pulse Survey at the University of Duisburg-Essen in Germany. The University of Duisburg-Essen has major programmes in social science, both at the bachelor of arts and the master of arts levels, with students from diverse socioeconomic backgrounds, both with roots in Germany and from abroad. The target population of students consisted of all students from five courses with a political science methods focus at the Department of Political Science at the bachelor of arts and master of arts levels, both mandatory and elective and either lecture based or seminar based, that started in the winter semester 2021/22. All courses had at least some statistical components, either applied statistics or research designs that need statistical analysis during the implementation. That target population was about 450 students, with 198 who participated in the first survey wave. Due to panel dropouts, 98 students who took part in all four waves remained in the sample we used in the analyses. Considering the sample composition, no bias has been introduced by the panel dropouts (Online Appendix Table A.1). Online Appendix Table A.2 and Fig. A.1 show details of survey participation.

Students were surveyed four times between November 2021 and February 2022 with a field period of 7 days per wave. During this period, some restrictions due to the COVID-19 pandemic were still in place. For example, courses were offered in person on campus and were then moved to remote teaching in December 2021. On average, survey completion time was between 5 minutes (wave 1) and 3 minutes (waves 2–4). Student characteristics with no, or expectedly low, variation over time (e.g. psychological measures) were covered in wave 1 only. The remaining surveys included questions on course satisfaction and course attendance. All students received two reminders per wave by their course instructor.

For details on incentives, data handling, and ethics, see Note A.1 in the Online Appendix.

3.2 Variables

3.2.1 Dependent Variables

We used questions about student satisfaction with the methods courses taken as well as the final grade achieved by students in each course as dependent variables. The questions on course-specific satisfaction pertained to overall course satisfaction, satisfaction with the organisation of the course, satisfaction with the course instructor's performance, and satisfaction with the students' own learning progress. Each dimension of satisfaction was measured using a seven-point rating scale with labelled endpoints ranging from "very dissatisfied" to "very satisfied". We first calculated mean scores for each of the four satisfaction questions when a student was surveyed twice in different courses ($n = 20$; 10.1% of all students in wave 1). In a second step, we calculated the students' mean for each question over the four survey waves. This way, we transformed our longitudinal data structure into a cross-sectional one. In order to derive a unified satisfaction index from the four questions, we used principal component analysis (PCA) to reduce these dimensions of satisfaction to a single component (eigenvalue: 2.88, 72% explained variance).[2] We used the esti-

[2] We find that using an exploratory factor analysis (EFA; with promax rotation), as well as a confirmatory factor analysis (all standardised factor loadings > 0.6), produces factor scores that are highly correlated with the scores obtained from a PCA ($r_{PCA\_EFA} = 0.989$, $r_{PCA\_EFA} = 0.991$). Using factor scores from an EFA leads to virtually congruent regression results, as reported in this research note.

mated factor scores from this PCA as our first dependent variable. Results for the PCA estimation can be found in Online Appendix Table A.3.

For our second dependent variable, we used the final grade achieved by the students. Again, we calculated the average grade when a student was surveyed twice in different courses. We recoded the final grade—ranging from 1 “very good” to 5 “failed”. with intermediate steps of 0.3 and 0.7—in a way that higher values stand for better grades. Exam modes comprised either written exams on the lectures or academic term papers on the seminars. Students who did not take the exam or who failed the exam could make a second attempt several weeks later. In this case, we replaced the final grade with the outcome of the second attempt.

3.2.2 Independent Variables

All independent variables were assigned to one of four blocks: controls, study-related variables, student characteristics, and psychological attributes. The relevant sections from the Pulse Survey questionnaire are in the Online Appendix Table A.4. The coding and univariate distributions of all independent variables can be found in the Online Appendix Table A.5.

Block 1 (controls) includes two dummy control variables that indicate whether a course was part of a bachelor′s or master′s programme (1 = master′s programme) and whether it was seminar based or lecture based (1 = lecture based). Later, we used this block as a control in all models that explain the unified satisfaction index and final grade.

Block 2 (study-related variables) refers to variables of student behaviour during the courses and lectures. One question was on the self-reported percentage of homework done by students (100% if they reported to have done everything). The second variable is the self-reported frequency of course attendance, varying between 0 and 3 (0 = attended in none of the 3 weeks before a survey wave; 1 = attended in one of the 3 weeks before a survey wave; 2 = attended in two of the 3 weeks before a survey wave; 3 = all sessions attended). This question was asked only in waves 2–4 of the survey. In addition, we included the unified satisfaction index as a predictor variable in the models explaining the final grade. In order to enhance the comparison between the coefficients in our regression analyses, the three variables in this block were standardised with a minimum value of 0 and a maximum value of 1.

Block 3 (student characteristics) refers to time-constant characteristics of students. We included self-reported gender (female, male, and diverse) and recoded this variable as follows: male = 1 and female = 0. There were no reported cases of “diverse”. Age was surveyed in categories ranging from 1 to 5. Given the expected younger age distribution in the sample, students could classify their age in 3-year stages, which resulted in the following five categories: 18 years and younger, 19–21 years, 22–24 years, 25–27 years, 28 years and older. In addition, we asked students whether at least one parent had a German *Abitur* (at least one = 1, otherwise = 0). Students who either immigrated to Germany themselves or who had at least one parent who was not born in Germany were defined as being of immigrant origin (immigrant origin = 1, otherwise = 0). The last variable is the student’s own *Abitur* grade ranging from 1 ‘very good’ to 4 ‘passed’ with intermediate steps of 0.1. We recoded this variable so that higher values indicate a better *Abitur* grade and standardised it with a minimum value of 0 and a maximum value of 1. This also applies to the age variable.

Block 4 (psychological attributes) consists of three variables surveyed in the first wave of the panel (see Online Appendix Table A.4 for the wording of all questions and rating scales). The first variable represents an index from a statistics anxiety scale measured with four questions (Mang et al. 2018; Förster and Maur 2015) that were accompanied by a seven-point rating scale. Higher values represent higher levels of statistics anxiety. We transformed the four variables into a mean index by dividing respondents' additive score by the number of responses. We find that the mean index is internally consistent (Cronbach's $\alpha = 0.92$; eigenvalue: 3.25; 81% explained variance in a one-factor PCA). The second variable represents an index from a self-efficacy scale with three questions (Cronbach's $\alpha = 0.75$; eigenvalue: 2.02; 67% explained variance) using a five-point rating scale. Again, we built a mean index by dividing the additive score from the three questions by the number of responses. Higher values on the index express higher self-efficacy (Beierlein et al. 2014). The last variable of this block is a procrastination index based on eight questions in total (Cronbach's $\alpha = 0.89$; eigenvalue: 4.61; 57% explained variance) with each one using a four-point rating scale (Klingsieck and Fries 2018). We recoded all variables derived from these questions so that higher values indicate a higher tendency to procrastinate. The mean index was built by dividing respondents' additive score by the number of their responses. Finally, we standardised all variables running from a minimum of 0 to a maximum of 1.

3.3 Sample Composition

The sample reflects the diversity of the student population in terms of both university-related and sociostructural characteristics (see Online Appendix Table A.5 for the distributions of variables). The sample mainly consists of students aged between 19 and 27 years, 34% of whom were of immigrant origin and had an average *Abitur* grade of 2.7. Moreover, 51% of the students were female. In total, 40% of the students came from households in which the parents had no previous university experience. In addition, the sample shows a balanced representation of bachelor's and master's students, with 46% enrolled in a master's programme. The majority of students (83%) attended lectures, with the minority attending seminars.

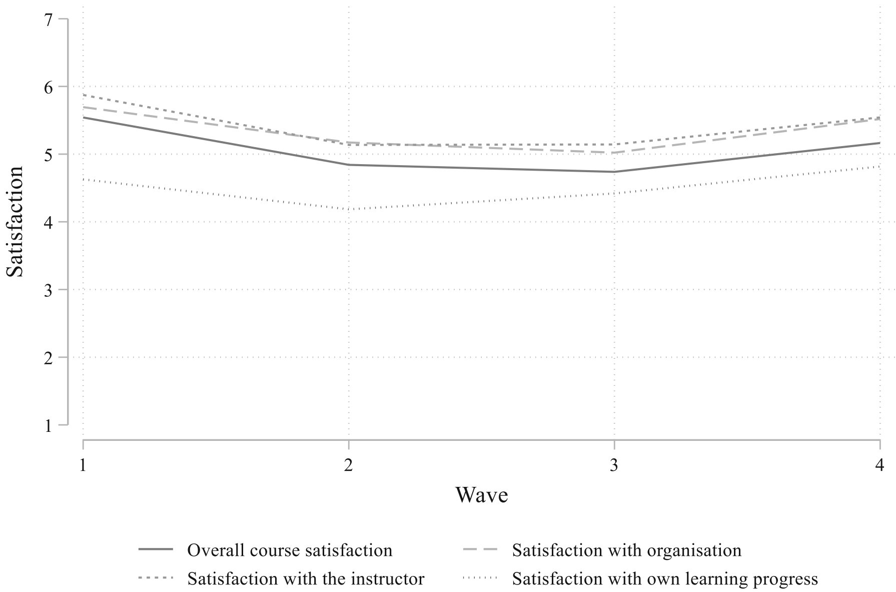
3.4 Analytical Strategy

To assess the explanatory factors for satisfaction and grades empirically, we analysed two dependent variables. These variables were (1) the unified satisfaction index, as the index consisting of all satisfaction measures extracted from a one-factor principal component analysis, and (2) the final grade achieved by students in each course.

All regression models have similar configurations. The first model setup examined the influence of study-related characteristics, including the amount of homework done and the frequency of attendance. In the second model, we analysed the influence of sociostructural variables such as gender, age, and immigrant origin. In the third model, we evaluated the relevance of psychological factors in the form of statistics anxiety and measures of procrastination and self-efficacy. The fourth model examined all predictor factors together. As the most restrictive model setup, the fifth model additionally included course dummy variables. We thus focused on within-course variance that accounts for unobserved heterogeneity from course and instructor characteristics, including differences in workload, exam forms, and the (perceived) difficulty of the course content and the course exam. This stepwise procedure allowed us to evaluate the association of each of our explanatory variable blocks with the dependent variables before combining them into a full model. Given the limited sample size, we could thus check whether the blocks of variables also showed consistent effects in combination.

In the first four model setups, we consistently controlled for the dummy variables that captured enrolment in a master's programme and attendance at a lecture, while in the fifth setup, the dummy variables for the courses were collinear with these variables and were thus not included. To improve interpretation, we used standardised variables with a minimum value of 0 and a maximum value of 1. Thus, all coefficients presented show the largest possible effects between students with the lowest and the highest empirically observable values.

We also performed regression diagnostics to test whether the assumptions of linear regression models were met. Indications of functional form deviations were observed. However, the use of transformed variables did not result in significant changes (results not shown). Additional analyses using bootstrapping procedures to deal with the small sample size and nonrandom sampling (Online Appendix Tables A.6 and A.7) do not show different results.



N = 89

Fig. 1 The development of means of student satisfaction variables over time

4 Empirical Analysis

4.1 Satisfaction Over Time

Figure 1 shows the means for each of the three dimensions of student satisfaction across methods courses per survey wave, as well as the means for overall course satisfaction as rated by the students. Higher values on the seven-point rating scale express higher satisfaction in each dimension. We used the full sample of the 89 students who took part in all four survey waves and who had no missing data on the relevant variables. We can see a slight U-shaped development over time, with high average satisfaction at the beginning as well as the end of the semester term, and with somewhat lower satisfaction values for the two measurement points in between. In comparison, satisfaction with students' own learning progress is lowest. At the same time, we see that all four satisfaction measures slightly converge across time.

To further investigate the variability of the satisfaction ratings over time, the development of "overall course satisfaction" with the methods courses is illustrated in a Sankey diagram (Fig. 2). We have divided the quasimetric scale, ranging from 1 to 7, into the following three categories: high, medium, and low overall satisfaction. The majority of students started the semester with high or medium overall satisfaction. During the course of the semester, we observed a decrease in highly satisfied students at the beginning of the semester as well as a sizeable increase from medium to high satisfaction at the end of the semester. For all four waves, the share of students with low overall satisfaction is small. Only a few students show a shift in overall satisfaction over more than one category between waves. In summary, students' overall satisfaction with the methods courses was relatively stable over the

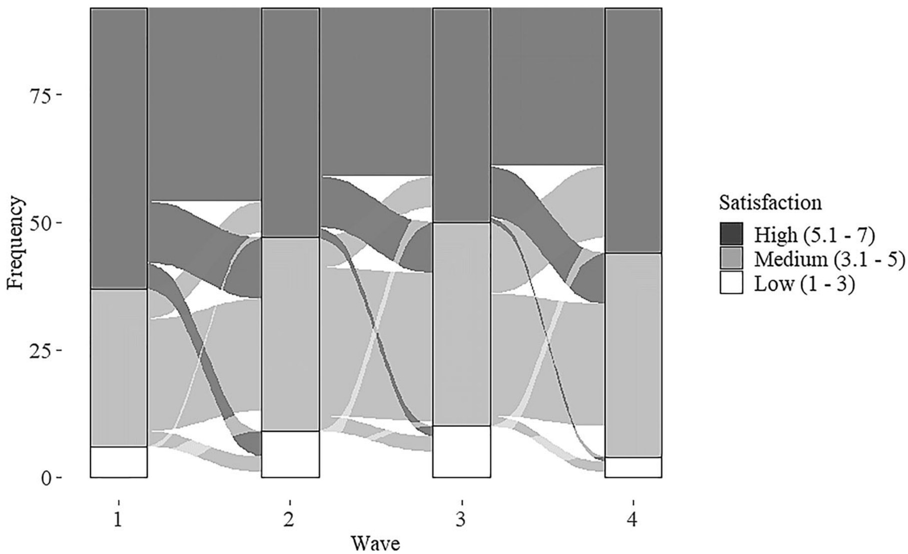


Fig. 2 Sankey diagram of overall course satisfaction

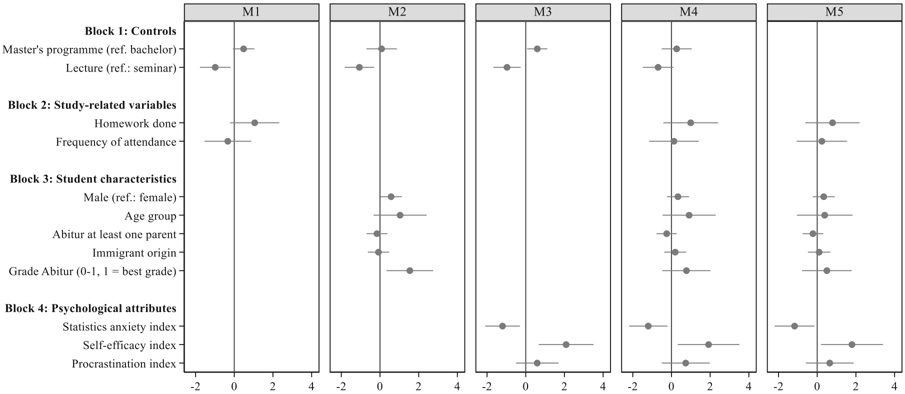


Fig. 3 Multiple regression results with the dependent variable unified satisfaction index (regression sample size $n = 89$). Circles represent coefficient estimates; bars refer to 90% confidence intervals

course, with only a few students moving between more than two adjacent categories (e.g. from high to low).

4.2 Explaining Levels of the Unified Satisfaction Index and the Final Grade

We begin our multiple regression analyses with the unified satisfaction index as dependent variable (a one-component score of all four satisfaction questions). The corresponding results are presented in Fig. 3 as coefficient plots displaying standardised coefficients with 90% confidence intervals. We decided to use 90% intervals because of the small sample size. Block 1 (controls) includes two control variables for models 1 to 4. Model 5 instead includes course dummy variables that control for various unobserved course characteristics, such as course composition, instructor effects, and the difficulty of the course content between the courses (see Online Appendix A.8 for the coefficients and standard errors).

Model 1 includes the control variables and study-related variables. The unified satisfaction index is shown to be lower for lectures than for seminars. The percentage of completed homework as related to the unified satisfaction index is not statistically significant. The relationship between frequency of attendance and the unified satisfaction index is negligible.

Model 2 examines the role of student characteristics. We find that being male and having achieved a better grade in the *Abitur* are both positively associated with higher levels of the unified satisfaction index. In contrast, age, having at least one parent with *Abitur*, and an immigrant origin are not related to the unified satisfaction index.

In the next step, we add the block of psychological attributes in model 3. The statistics anxiety index is negatively associated with the unified satisfaction index. For self-efficacy, we find a positively related coefficient. In contrast, the procrastination index shows no association at all.

Model 4 incorporates all variables from all blocks simultaneously. We find that the estimates of student characteristics are not related, whereas the psychological indices of statistics anxiety and self-efficacy are related to the unified satisfaction index. This illustrates that by holding various study and student characteristics constant, a lower level of the statistics anxiety index and a higher level of the self-efficacy index explain the unified satisfaction index.

Model 5 includes all predictor variables as well as the course dummy variables that control for various unobserved course characteristics, such as course composition, instructor effects, and the difficulty of the course content. If the unified satisfaction index strictly reflected course characteristics—as implied by the inherent logic of contemporary university course evaluation—then no predictor variables other than course characteristics should be relevant. However, this is not the case here. Statistics anxiety and self-efficacy are still relevant index variables even when controlling for course characteristics. Hence, the results provide evidence that is in line with hypothesis 1.

We now turn to those variables that influence student success in grades and present the corresponding regression results in Fig. 4 (see Online Appendix Table A.9 for the coefficients and standard errors). Model 6 includes the basic control variables and the study-related variables. Regarding the control variables, we find no systematic association between the variables of master's programme or lecture and final grade. Instead, we find a positive relationship between the unified satisfaction index and final grade. This means that those students who are more satisfied are more likely, on average, to get better final grades. Frequency of attendance is also positively related to course grades. These results provide evidence for hypothesis 2.

Model 7 examines student characteristics and shows that higher *Abitur* grades are positively related to course grades. In contrast, gender, age, parents' educational background, and immigrant origin have no relationship with academic performance.

In model 8, the psychological attributes are included. We find that the statistics anxiety index is negatively related to final grade. That is, students with a high level of discomfort with statistical tasks have a substantially lower grade in methods courses. For the psychological indicators of self-efficacy and procrastination, the coefficients tend to be positive, but they do not reach statistical significance.

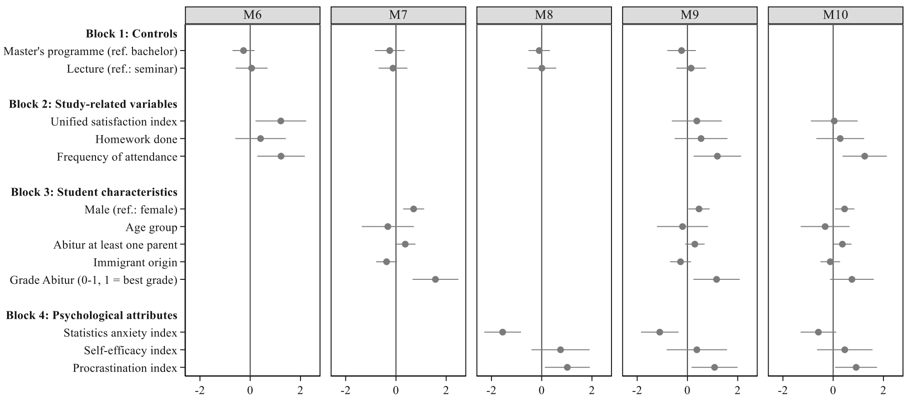


Fig. 4 Multiple regression results with the dependent variable final grade (regression sample size $n = 89$). Circles represent coefficient estimates; bars refer to 90% confidence intervals

Models 9 and 10 include all predictor variables, and in model 10 the course dummy variables are included. In both of these models, we find that frequency of attendance and the statistics anxiety index are significantly related to final grades. Moreover, the inclusion of these variables results in a drop in the coefficient estimate of the satisfaction measure. This pattern corroborates hypothesis 3. Specifically, the unified satisfaction index seems to have mediated student characteristics such as the statistics anxiety index. Once taken into account, the statistics anxiety index appears to be more indicative of the final grade than the unified satisfaction index.

5 Conclusions

We present three major results in explaining student satisfaction and course performance in political science courses. First of all, student satisfaction with the courses attended during the semester follows a slight U shape over time. Satisfaction initially decreases and then rises again almost to the starting level towards the end of the lecture period (wave 4). This trend of satisfaction is caused by only a few students moving between very high and very low levels of satisfaction between time points. This shows that there is some variation in satisfaction between students and over time.

Second, we see relatively few significant effects in two perspectives of the independent variables: characteristics of course participation and psychological as well as demographic characteristics of students. None of the course attributes has any precise effects on the unified satisfaction index. On the other hand, from the perspective of students' attributes, it is interesting that a higher level of their parents' education has a negative but statistically nonsignificant association with the unified satisfaction index. Most importantly, the statistics anxiety index shows a negative association with the unified satisfaction index. Also, it has a strong negative estimated effect on the unified satisfaction index and the student's satisfaction with their own learning process. Satisfaction therefore depends less on the course content and more on time-constant and individual characteristics.

Third, we analysed the effects of course and student characteristics on study success. The situation turns out to be similar to the final grade that students achieve for their course. Their final grade is primarily influenced by their own educational success in school (reported *Abitur* grade) and the statistics anxiety scale measured as an index (Mang et al. 2018; Förster and Maur 2015). The effects of maths anxiety—which can already have a negative impact on performance at school (Maloney and Beilock 2012)—also negatively influence satisfaction and grades at university. Another factor affecting study success is the students' frequency of attendance of the course: A higher frequency of attendance has a positive and substantial effect on their final grade. In this specific setting, where attendance is generally voluntary, there is clear potential for improving study performance by mobilising students into course participation throughout the lecture period. Finally, satisfaction with the student's own learning progress is associated with a better grade, while other dimensions of satisfaction show no precise effects.

The limitations of our study resemble those of other research projects that sample university students (Chávez and Mitchell 2020): The sample size is relatively small and is limited to one institution. However, it is worth noting that response rates throughout the longitudinal design are relatively high even before accounting for student dropout from the courses. It remains hard to assess the possible effects of the COVID-19 pandemic on study and response behaviour. Although we exercised a comprehensive control strategy—for instance, the inclusion of dummy variables for each course that absorbed differences in exams or differences in difficulty of course content—we need to emphasise that with the one-time measure of the final grade, it is not possible to map causal relationships between variables with our research design. Although randomised controlled trials are not easily feasible in the present context, panel studies with repeated measurement of satisfaction and performance outcomes are an important avenue for future research. Future research should therefore focus more closely on the specific effects of the subdimensions of student satisfaction. This would then go beyond the limitations of this research note.

For political science lecturers, our results are sobering. We show that student satisfaction—comprising a family of indicators—only partially predicts grades. A teaching culture based on measuring student satisfaction will thus fail to bring to light important patterns. The finding that students' statistics anxiety—as measured at the beginning of a course—maintains such a strong impact on both satisfaction and grades throughout the study period calls for a holistic approach to methods training. In this approach, teachers would collaborate with psychological specialists and would act as gatekeepers for strategies to deal with these negative emotions. While we are aware that a comprehensive implementation of these recommendations might be unrealistic, we nevertheless want, at least, to raise the awareness of instructors and students to these influential factors.

Supplementary Information The online version of this article (https://doi.org/10.1007/s11615-025-00613-x) contains supplementary material, which is available to authorized users.

Acknowledgements We thank Joshua Claaßen for implementing the online survey in Unipark survey software. The study was financed by seed money of the University of Duisburg-Essen to Prof. Achim Goerres (incentives and research assistance). Jan Karem Höhne was a postdoctoral researcher in the working group of Empirical Political Science, where most of the work was carried out.

Author Contribution Conceptualised research: AG; research design: all; literature research: PV; questionnaire design: CZ, JKH; pretesting: JKH; field management, data protection, and data linkage: JE; data management: PV; statistical analysis: CZ, PV; first draft: JE, PV, CZ, AG; final draft: all; revisions: AG (lead), all. PV and JE share equal first authorship in random sequence.

Funding Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest P. Vierus, J. Elis, A. Goerres, C. Ziller, and J. Karem Höhne declare that they have no competing interests.

References

Auspurg, Katrin, Sarah Brodhäcker, Lena Opitz, and Stefan Wender. 2015. *Nutzung, Zufriedenheit und Kompetenzerwerb. Evaluation der Methodenausbildung im "Starken Start"*. Starker Start ins Studium. Evaluationsbefragung des Methodenzentrums der Goethe-Universität Frankfurt/Main, Wintersemester 2014/15.

Beierlein, Constanze, Anastassiya Kovaleva, Christoph J. Kemper, and Beatrice Rammstedt. 2014. *Allgemeine Selbstwirksamkeit Kurzskala (ASKU)*. https://doi.org/10.6102/zis35.

Blanz, Mathias. 2014. How do study satisfaction and academic performance interrelate? An investigation with students of Social Work programs. *European Journal of Social Work* 17(2):281–292. https://doi.org/10.1080/13691457.2013.784190.

Blome, Christine, and Matthias Augustin. 2015. Measuring change in quality of life: Bias in prospective and retrospective evaluation. *Value in Health: the Journal of the International Society for Pharmacoeconomics and Outcomes Research* 18(1):110–115. https://doi.org/10.1016/j.jval.2014.10.007.

Chávez, Kerry, and Kristina M.W. Mitchell. 2020. Exploring bias in student evaluations: gender, race, and ethnicity. *Political Science & Politics* 53(2):270–274. https://doi.org/10.1017/s1049096519001744.

Clayson, Dennis E., Taggart F. Frost, and Mary Jane Sheffet. 2006. Grades and the student evaluation of instruction: a test of the reciprocity effect. *Academy of Management Learning & Education* 5(1):52–65. https://doi.org/10.5465/amle.2006.20388384.

Clogg, Clifford C. 1992. The Impact of sociological methodology on statistical methodology. *Statistical Science* 7(2):183–207. https://doi.org/10.1214/ss/1177011354.

Cotton, Sarah J., Maureen F. Dollard, and Jan de Jonge. 2002. Stress and student job design: satisfaction, well-being, and performance in university students. *International Journal of Stress Management* 9(3):147–162. https://doi.org/10.1023/A:1015515714410.

Förster, Manuel, and Andreas Maur. 2015. Angst im Fach Statistik und statistisches Selbstkonzept von Studenanfängerinnen und -anfängern in den Sozialwissenschaften: Eine Frage des Geschlechts und des sozio-kulturellen Hintergrunds? *Zeitschrift für Hochschulentwicklung* 10(4):67–90.

Goerres, Achim, Caroline Kärger, and Daniel Lambach. 2015. Aktives Lernen in der Massenveranstaltung: Flipped-Classroom-Lehre als Alternative zur klassischen Vorlesung in der Politikwissenschaft. *Zeitschrift für Politikwissenschaft* 25(1):135–152.

Gow, Lyn, and David Kember. 1993. Conceptions of teaching and their relationship to student learning. *British Journal of Educational Psychology* 63(1):20–23. https://doi.org/10.1111/j.2044-8279.1993.tb01039.x.

Green, Heather J., Michelle Hood, and David L. Neumann. 2015. Predictors of student satisfaction with university psychology courses: a review. *Psychology Learning & Teaching* 14(2):131–146. https://doi.org/10.1177/1475725715590959.

Hessler, Michael, Daniel M. Pöpping, Hanna Hollstein, Hendrik Ohlenburg, Philip H. Arnemann, Christina Massoth, Laura M. Seidel, Alexander Zarbock, and Manuel Wenk. 2018. Availability of cookies during an academic course session affects evaluation of teaching. *Medical Education* 52(10):1064–1072. https://doi.org/10.1111/medu.13627.

Jamelske, Eric. 2009. Measuring the impact of a university first-year experience program on student GPA and retention. *Higher Education* 57(3):373–391. https://doi.org/10.1007/s10734-008-9161-1.

Keržič, Damijana, Jogymol Kalariparampil Alex, Roxana Pamela Balbontín Alvarado, Denilson Silva Da Bezerra, Maria Cheraghi, Beata Dobrowolska, Adeniyi Francis Fagbamigbe, Moez Al Islam Ezzat Faris, Thais França, Belinka González-Fernández, Luz Maria Gonzalez-Robledo, Fany Inasius, Sujita Kumar Kar, Kornélia Lazányi, Florin Lazăr, Juan Daniel Machin-Mastromatteo, João Marôco, Bertil Pires Marques, Oliva Mejía-Rodríguez, Silvia Mariela Méndez Prado, Alpana Mishra, Cristina Mollica, Silvana Guadalupe Navarro Jiménez, Alka Obadić, Daniela Raccanello, Mamun Ur Rashid, Dejan Ravšelj, Nina Tomaževič, Chinaza Uleanya, Lan Umek, Giada Vicentini, Özlem Yorulmaz, Ana-Maria Zamfir and Aleksander Aristovnik. 2021. Academic student satisfaction and perceived performance in the e-learning environment during the COVID-19 pandemic: Evidence across ten countries. *PLOS ONE* 16 (10): https://doi.org/10.1371/journal.pone.0258807.

Klingsieck, Katrin B., and Stefan Fries. 2018. *Eine Kurzskala der General Procrastination Scale (GPS-K)*. https://doi.org/10.6102/zis255.

Kostagiolas, Petros, Charilaos Lavranos, and Nikolaos Korfiatis. 2019. Learning analytics: survey data for measuring the impact of study satisfaction on students' academic self-efficacy and performance. *Data in Brief* 25:104051. https://doi.org/10.1016/j.dib.2019.104051.

Lambach, Daniel, Caroline Kärger, and Achim Goerres. 2017. inverting the large lecture class: active learning in an introductory international relations course. *European Political Science* 16(4):553–569. https://doi.org/10.1057/s41304-016-0078-3.

Maloney, Erin A., and Sian L. Beilock. 2012. Math anxiety: who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences* 16(8):404–406. https://doi.org/10.1016/j.tics.2012.06.008.

Mang, Julia, Natalia Ustjanzew, Anja Schiepe-Tiska, Manfred Prenzel, Christine Sälzer, Katharina Müller, and Elisabeth Gonzaléz Rodríguez. 2018. *PISA 2012 Skalenhandbuch. Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.

Manzoni, Anna, Jeroen K. Vermunt, Ruud Luijkx, and Ruud Muffels. 2010. Memory bias in retrospectively collected employment careers: a model-based approach to correct for measurement error. *Sociological Methodology* 40(1):39–73. https://doi.org/10.1111/j.1467-9531.2010.01230.x.

Maravelakis, Petros. 2019. The use of statistics in social sciences. *Journal of Humanities and Applied Social Sciences* 1(2):87–97. https://doi.org/10.1108/JHASS-08-2019-0038.

Marsh, Herbert W. 1980. The influence of student, course, and instructor characteristics in evaluations of university teaching. *American Educational Research Journal* 17(2):219–237. https://doi.org/10.3102/00028312017002219.

Marsh, Herbert W., and Lawrence A. Roche. 2000. Effects of grading leniency and low workload on students' evaluations of teaching: popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology* 92(1):202–228. https://doi.org/10.1037/0022-0663.92.1.202.

Pineda, Pedro, and Isabel Steinhardt. 2020. The debate on student evaluations of teaching: global convergence confronts higher education traditions. *Teaching in Higher Education* https://doi.org/10.1080/13562517.2020.1863351.

Rhodes, Marjorie, and Andrew Baron. 2019. The development of social categorization. *Annual review of developmental psychology* 1:359–386. https://doi.org/10.1146/annurev-devpsych-121318-084824.

Rosar, Ulrich, and Markus Klein. 2009. Mein(schöner)Prof.de. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 61(4):621–645. https://doi.org/10.1007/s11577-009-0086-1.

Schaeper, Hildegard. 2020. The first year in higher education: the role of individual factors and the learning environment for academic integration. *Higher Education* 79(1):95–110. https://doi.org/10.1007/s10734-019-00398-0.

Tessema, Mussie T., Kathryn Ready, and W. Yu. 2012. Factors affecting college students' satisfaction with major curriculum: evidence from nine years of data. *International Journal of Humanities and Social Science* 2(2):34–44.

Włodzimierz, Okrasa. 2012. Statistics and sociology: the mutually-supportive development from the perspective of interdisciplinarization of social research. *Statistics in Transition* 13(2):365–386.