

Article

Learning Outcomes in an Introductory Sociology Course: The Role of Learning Approach, Socio-Demographic Characteristics, Group and Teacher Effects

Péter Miskolczi ^{1,*}  and Márton Rakovics ²¹ Budapest Business School, Corvinus University of Budapest, 1149 Budapest, Hungary² Eötvös Loránd Science University, Corvinus University of Budapest, 1117 Budapest, Hungary; marton.rakovics@tatk.elte.hu

* Correspondence: miskolczi.peter@uni-bge.hu; Tel.: +36-30-445-0229

Received: 14 December 2017; Accepted: 7 January 2018; Published: 10 January 2018

Abstract: The article discusses a study on learning outcomes in an introductory Sociology course in a higher education setting. A quantitative model similar to education production functions is hypothesized. Independent variables include socio-demographic and psychological ones, as well as others pertaining to the learning environment. Learning is measured with the binary categories of “deep” and “surface” outcomes, which are qualitatively different, and are based on the comparison of mind maps drawn by students at the beginning and end of their first semester. Questionnaire data was collected from 264 first-year students of a Social Science Faculty. Significant effects on learning outcome are found for family background, major area of study, and teacher’s gender.

Keywords: teaching sociology; scholarship of teaching and learning; diagrammatic elicitation; deep learning; surface learning; education production function; group effects

1. Introduction

The goal of university education is to provide students with the skills and knowledge that help them become competitive in a rapidly changing labor market, and also in their quest for a fulfilling, meaningful life. The present study is investigating the outcomes of an introductory Sociology course. We believe that sociology can provide students with a conceptual toolbox that helps them get a hold on the rapidly changing social world, and also enhances their ability for reflective and critical thinking—important skills in the 21st-century labor market [1].

Sociology is the study of human activities—and as such, it studies education among many other phenomena. The exercise becomes truly reflexive when as sociologists, we investigate our own educational activities, among them, the effectiveness of our own courses. The main motivation behind the present study was to gain insight into how effective our introductory course is—what do students learn and how well do they understand what they learn? We believe that in order to achieve our goal, namely the ability to think reflexively and critically of the social world, students have to master, besides concepts, the ability of relating them to each other and their pre-existing knowledge, notions, and opinions—if needed, questioning and de-constructing previous beliefs, convictions and linkages in the process. What we are interested in are the socio-demographic, individual and group-level characteristics that might influence the success of the course. These considerations are reflected in the theoretical foundations and methodology of our research.

The study was carried out at the Faculty of Social Sciences at Corvinus University of Budapest, involving students majoring in four different areas (Sociology, Political Science, Media and Communication Studies, International Relations). Out of 304 first-year students registered for the

course, 264 ended up in the sample. We measured “Deep” and “Surface” learning outcomes as dependent variables in a logistic regression model. The most important results are the following: (higher) parental education is not in a clear relationship with learning outcome, and certainly cannot be said to be positively related to it; students who had female seminar instructors were deeper learners; students of political science had a higher chance of becoming surface learners; and university entry score (an indicator of previous achievement and, perhaps, ability) did not have an effect. The aim here is not to make broad general claims as to what predicts educational success, nevertheless, we believe that our findings are useful additions to the scholarship of teaching and learning, and we might not be the only ones who can utilize them in further research or teaching practice.

1.1. The Qualitative Conception of Learning

Educational practice often makes the impression that the point of studying is to amass a large body of knowledge, and that, indeed, knowledge is something that is about quantity: experts are those who know a lot of details of a given field. Elementary and secondary education, along with standardized tests, in which there is only one correct answer, reinforce and entrench this quantitative conception of knowledge [2] (pp. 23–24). However, the stated goal of university professors was different already half a century ago: when asked, the things they named most important were critical thinking, creativity, and mastering of both the technicalities and ways of thinking of a given field. However, academic practice at the time did not seek to reflect on the role of teaching and testing methods in achieving these goals [3] (pp. 4–6).

Besides the quantitative conception, qualitative approaches to learning have also long existed. They are strongly related to the tradition of cognitive psychology. Their most important feature is that they emphasize aspects of learning other than merely storing information, such as making connections between parts of the material, constructing a personal interpretation, and an overall mental model [2] (p. 27). An early but still influential framework in this qualitative approach is the Bloom Taxonomy, an ordinal system of 6 categories (starting from “Knowledge” meaning memorization, going up to “Evaluation”, the ability to judge the value of newly acquired information) [4,5].

While Bloom’s taxonomy was an a priori classification, the one used in the present study is the empirically grounded framework of Marton and Saljö [6,7]. Measuring student responses to a reading exercise in a university context, they described the following two categories of learning outcomes:

- Deep learning. The student wants to grasp the essence of the material. Distilling rules and mechanisms is more important than remembering particularities of the given example.
- Surface learning. The student concentrates on the details of the material. Constructing a personal meaning or interpretation is not a priority (it does not happen).

Marton and Saljö believed that differences in learning outcome can be due to differences in prior knowledge and ability, but placed much more emphasis on the approach to the task and the process of learning. Thereby they distinguished deep and surface approaches as well: those with the former strive for a general understanding of a given problem, and those with the latter go only for memorization of details [7] (pp. 39–43). Learning approach has been empirically linked to the end result in several studies [8]. Additionally, educational research has highlighted that (preferably intrinsic) motivation is also a very important contributor to the learning process and its outcome [7] (pp. 53–54).

Furthermore, we would like to mention another important psychological aspect of knowledge, namely that it is relational: it is made up at least as much of the connections between pieces of information as of the pieces themselves. This is why meaningful, good learning depends largely on whether the student can form links between parts of the material, and, more importantly, between their prior knowledge and the new material. This “forming of links” is a highly personal process, and also depends on whether the material holds any personal meaning for the student [3] (pp. 10–11).

1.2. Economic and Sociological Approaches to Learning

While learning is undoubtedly a personal process, taking place within one's brain, education has rightly been studied from economic and sociological viewpoints as well. Quantitative models in the latter vein can broadly be called education production functions [9]. Naturally, both the theoretical framework and empirical evidence of such studies will differ from the one briefly presented above. They are not typically concerned with the qualitative conception of knowledge [10] (pp. 1150–1154), and the majority of them focus on primary and secondary education, and less on the university level.

Typically, the following inputs are considered in education production functions:

- family background of the student (family size, socio-demographic characteristics),
- peers or other students (socio-demographic characteristics thereof),
- school effects (class sizes, facilities),
- teacher effects (education level, experience, sex, race),
- educational expenditure,
- abilities of the student.

With the addition of the caveat that these also interact with each other, and that the educational process is cumulative, meaning that past “treatment” (schooling) has a lasting effect [10] (p. 1155). While Hanushek's review of 147 empirical estimations came to the conclusion that results of such studies often point to different directions [10] (pp. 1159–1162), it should also be kept in mind that education production function models can suffer from specification problems as well [11].

However, the goal of the present study is not to construct a “generally applicable” education production function. The learning environment and outcomes of one particular introductory sociology course are investigated. Conclusions will first and foremost be drawn for the authors' own teaching practice. Therefore, the specification of a model and function without over-reaching ambitions is defensible. Next, we discuss empirical evidence on the factors to be included in the model.

Class size traditionally received ample attention in education research. Data from the Tennessee STAR experiment suggests that smaller class sizes are beneficial for the chances of entering higher education [12]. Using STAR data, Nye et al. came to the conclusion that there were also considerable teacher effects in elementary education—a variance in outcomes between classes within the same school that can be explained neither by the educational attainment nor by the experience of the teacher, and therefore has to come from unobserved or unobservable teaching abilities and traits [13].

Basow found compelling evidence of teacher gender effects in a higher education setting [14]. The study shows that the evaluation of teachers is conditional on both students' gender and teachers' gender, and in some cases, the interaction effect of the two is significant. Female teachers receive lower ratings overall in general and especially from male students. For the present study, this would suggest that students who value the instructor less will be less motivated to perform well. Consequently, students in groups with female instructors would be less likely to become deep learners. Compared to these results, a reverse association between teacher gender and student performance was found in a high school study by Duffy et al. [15]. They found evidence for the hypothesis that female teachers provide more feedback to students (mostly to male students), and this in turn facilitates better student performance. If these result are transferable to our higher education setting, then we can expect to see better learning outcomes for students in classes with female instructors.

The inclusion of family background as an explanatory variable in a sociologically oriented research is seemingly self-evident, since socioeconomic position is a primary influence on life chances in general. Ermisch and Francesconi—in an economically oriented study—found very strong association between parental education and child educational attainment [16]. Davis-Kean focused on beliefs, expectations and behaviors of higher educated parents that create an atmosphere that helps the child to academic success [17]. Buchmann and DiPrete also link family background to the fact that female students have become generally better performers in higher education, one possible mechanism being that better educated parents favor the education of both sons and daughters, generating higher “returns”

for the latter [18]. Finally, it should be added that “waning coefficients” have been described in the literature with regards to the advantage that family background provides as children move forward in the education system, suggesting that at higher levels, family matters less and less. However, this view has been contested by Holm and Jaeger who argue that previous findings were largely due to selection and that the family effect is largely constant through all levels of education [19].

To return briefly to the influence of student gender on educational success, Duckworth and Seligman reiterate the fact that women outperform men in university grades in most colleges and most subjects [20]. However, they find that one key trait leading to this result is their better self-discipline over an extended study period. In short and occasional “achievement tests” and “aptitude tests” (such as one measuring IQ) women’s advantage is less and less apparent. Then, Vermunt explicitly tested the influence of gender on different learning styles, for different majors in higher education, controlling for a number of possible confounding factors. The only substantial difference between genders was in the preference for cooperative learning. Male students were more individualistic in their learning style, while female students were more social [21]. Overall, these results indicate that gender is correlated with important determinants of educational success, although the existing evidence does not include qualitative measurements of knowledge.

The role of peers on individual achievement has received considerable attention in educational research, too. Epple and Romano provide an exhaustive summary of peer-based theoretical models, most of which concentrate on the distribution of ability in student groups—something we might call “composition effects”—and their consequences on the information available to and on the effort made by the individual [22] (pp. 1055–1069). They also discuss several empirical studies, most of which identified significant peer effects, either among randomly selected roommates or whole school classes (which possibly result from, among others, the alteration of preferences and habits of the student, from the average level of ability in the group, or from the propensity of some individuals for disruptive behavior) [22] (pp. 1112–1156).

Another way to interpret “group effects” is to look not at the composition, i.e., the characteristics of individuals making up a group, but at veritable group level attributes such as norms that guide the individual’s approach to education. An example for such a phenomenon is “the burden of «acting white»” described by Fordham and Ogbu—whereby black kids are held back from reaching academic success by a pressure from the group not to “betray” their social identity [23]. This debated concept was reassessed by Horvat and Lewis who found that this purported burden does not play a central role in the academic life of black students, although they did not dismiss it entirely. Instead, they reported that black students have a repertoire of “managing their academic success” depending on the particular peer group they are in [24]. Furthermore, group norms among students have received ample attention in the context of bullying and drinking behavior [25,26], and absenteeism in a workplace setting has been found to be linked to referent group norms [27]—we theorize that similar mechanisms operate with regard to studying as well.

1.3. Previous Studies in the Context of Teaching Sociology

A number of important quantitative studies have been done previously in the particular context of teaching Sociology [28–32]. These studies—differing from the one presented herein—used quantitative indicators such as course grade to measure learning outcomes. Otherwise, their design was similar to the present one, involving a pretest at the beginning of the course then comparing it to learning outcomes at the end. These studies have found that GPA scores were positively associated both with pretest scores and final learning outcomes, and found no effect for gender or previous high-school sociology studies [28,29]. Students with more college years did better both on pretests and final outcomes [28,32]. While it was found that a higher socio-economic background was associated with higher pretest scores [28], it was not associated with gains made during the course [29]; and having studied a foreign language was correlated with higher gains [29].

Notably, Neuman's study also compared two models of learning: "accumulated advantage" and "interest motivation". The former predicts that those with a good socio-economic and/or academic background—something we might call the social/economic element in the context of the present study—will make higher gains from a course; the latter that interest in the subject matter compensates for disadvantages in other areas—which is clearly related to the cognitive psychological concept of study approach. Eventually, the accumulated advantage model received more support in Neuman's work, not so much in its socio-economic dimension but in the one of academic background [29].

In summary of the literature on learning and educational success, let us reiterate the most important points here: namely that (1) learning can and should be conceptualized in a qualitative way; (2) a student's approach will influence the end result of the studying process; (3) there is a place for socio-demographic variables in our research; (4) group effects, either in the sense of size, composition or norms, can be consequential for learning outcomes; (5) teacher characteristics might also be influential.

1.4. Context of the Research

The context of the research is the following. Every year, the Social Science Faculty at the Corvinus University of Budapest accepts roughly 280 new students. They are admitted to 4 different major areas of study: Sociology, Political Science, Media and Communication, and International Studies. During their very first semester, they all take part as one year-group in the Foundations of Sociology course. There is a 90-min lecture held every week in a big hall; and there is one 90-min seminar block, held in groups of sizes up to 40, where participation is compulsory. Seminars are held by PhD students of the faculty. The program of these seminars is largely the same, following a standardized script (meaning, in the context of this study, that all students receive roughly the same "treatment"). Seminars offer the opportunity of discussion and reflection for students.

Evaluation is done on the basis of two multiple-choice tests, written at the middle and end of the semester (40 points each), and four "quick tests" written during seminars, which are mainly designed to ensure that students read the assigned texts. In addition, students can opt to do extra-curricular project work that requires autonomous efforts of reading and summarizing sociological sources, and linking them to real-life phenomena. Individual project works are rewarded with a maximum of 15 points, projects done in groups (thought to require more work by each member) are worth 20 points individually for all members.

1.5. Research Question and Hypotheses

The main aim of this research was to find out what influences the learning outcome of students in this introductory Sociology course, backed by the curiosity to simply gain a better understanding, and by the readiness to make adjustments to teaching practices based on the findings. On the basis of the literature reviewed above, the goal was to produce a synthesis of cognitive psychological and economic and sociological approaches to learning ("educational outcome"). Knowledge was conceptualized in a qualitative way, and the assumed education production function includes many of the inputs mentioned beforehand:

- student ability,
- learning approach,
- learning process,
- student ambition,
- gender of student,
- family background,
- group effects,
- teacher effect.

Taking theoretical foundations and previous findings into account, the following hypotheses were formulated.

Hypotheses 1. *Learning approach will have an effect on learning outcomes: the “deeper” the student’s approach, the “deeper” their outcome will be.*

Hypotheses 2. *Extra effort during the learning process, namely, doing extra-curricular project work, will deepen learning outcomes, since the point of the task is immersion of sociological material.*

Hypotheses 3. *Female students will be deeper learners than males.*

Hypotheses 4. *Students coming from higher educated families will be deeper learners than those whose parents are less educated.*

Anecdotal evidence from seminar leaders in the Fall 2014 and Fall 2015 semesters indicated that students majoring in Political Science approach our subject with a certain aloofness. We try to be careful about such impressions, as they might very well result from perception or confirmation bias, and this piece of research provided an opportunity to check if this was indeed the case. We took major area of study to be an instance of “group effect” or group identity, and therefore made the final hypothesis that

Hypotheses 5. *Political Science students will end up with “shallower” learning outcomes by virtue of not caring as much about Sociology.*

Finally, we would like to note that the literature was not conclusive concerning the effect of teacher gender, therefore we did not form a hypothesis on that.

2. Materials and Methods

The research was carried out during the Fall semester of 2015. Students filled short paper-based questionnaires during the seminars of the first (T1) and the last week (T2). The questionnaires consisted mostly of closed-ended items, but students also had to draw a mind map in both waves (see below). Each questionnaire received an individual code assigned by the seminar instructor. Questionnaire data was entered into a digital database by second-year student members of the research team who generally did not know first-year respondents and only saw the coded sheets, having no way to identify them. Some data that we did not ask for in the questionnaires themselves (such as seminar group, participation in project work, multiple-choice test scores) were registered in the database by seminar instructors. Student codes were then removed from the final database on which the analysis was done, making identification of individual respondents impossible. Out of 304 first-year students at the Faculty of Social Sciences who registered to the course, a sample of 264 (87%) provided information in both waves.

2.1. Specification of the Model

The data had to correspond to the theoretical foundations and the model outlined above. Therefore, each of the inputs and the output had to be represented by a variable backed with empirical data. Some of the concepts were easily operationalized, some less so. Apart from gender of student, the way of measuring each concept is briefly outlined below.

2.1.1. Student Ability

The so-called “university entry score” was used as a proxy for student ability. Admittedly, this is far from being a perfect solution, since true “ability” might be unmeasurable, or would require a lengthy process of testing on its own. Therefore “ability” might be rephrased as “preparedness for academic pursuits”, which this entry score might indeed indicate reliably, and is “brought” by students with themselves (i.e., not acquired during their university studies), which in turn is the point for its inclusion in the model. The predictive validity of entry scores on ensuing collegiate performance—at least for the first academic year—is well established [33,34], but in general it is very hard to tell what

the association is between student performance in higher education and the cognitive ability that fosters a deeper learning outcome.

In the Hungarian education system, all students finishing secondary school write a state sanctioned “final exam”. Depending on their school grades in the final 2 years, and on the results of these final exams, as well as on some extra credentials (most importantly, language exams), a “university entry score” is calculated for each student. Roughly a month after the centralized final exams, universities announce their thresholds for accepting students. Corvinus University of Budapest usually attracts students of excellent credentials, making average entry scores well above 400 in a system where 400 is the “normally” attainable level and some extra effort is required in order to gain another 100 points at maximum.

2.1.2. Learning Approach

The questionnaires in both waves (T1 and T2) included items pertaining to the learning approach of students. There exist multiple questionnaires for “learning style inventories”, but none of them are very short. In order to save time, 4 items from Biggs’s “Revised Study Process Questionnaire” were included [35]. The original questionnaire was designed to measure 2 main dimensions: deep and surface, and 4 subscales: deep/surface motive and deep/surface strategy. We were only interested in the 2 main scales, and included items 1, 3, 5 and 19 from the original questionnaire. Two of these were worded in ways that agreement with the statement (measured with a 5-grade Likert scale) indicated a deep approach to studies, while agreement with the other 2 indicated a surface approach. We posed these questions “free of context”, that is, we did not ask students to think of the Sociology course when indicating their approach to studying. Author instructions to the original questionnaire indicate that while students’ answers and emerging “approach profiles” cannot be taken as stable personality traits, they can serve to “describe how individuals differ within a given teaching context” [35] (p. 5)—which coincides with our aim.

The answers to these questions were condensed into a single variable that represented the learning approach of students. Answers to the questions measuring a deep approach were scored from 1 (never or only rarely true of me) to 5 (always or almost always true of me), those to the ones measuring a surface approach from −1 (never or only rarely) to −5 (almost/always). Eventually, the numbers corresponding to answers were simply added. Thus, the combined learning approach variable can range from −8 (a very surface approach) to +8 (a very deep approach).

It is well established in the literature that during their university studies, students generally slide towards a shallower approach than they begin with [35] (p. 6), [36] (p. 786). In our analysis, we assumed that the approach scores measured in T1, at the very beginning of the semester, reflected students’ intentions. However, the scores in T2 should reflect their “true” approach better: after having experienced how they pursue their university studies, they should evaluate themselves more correctly. Therefore, in the analysis, we included scores obtained from the T2 wave to represent learning approach.

2.1.3. Learning Process

As a proxy of learning process (effort), participation in project work was used. This ended up being a categorical variable with 3 possible values: did not do a project; participated in a group project; did an individual project.

2.1.4. Family Background

This variable was constructed from two basic items: highest educational attainment of mother, and of father. The eventual ordinal variable had 5 possible values:

- neither parent finished high school;
- one parent finished high school (the other did not);
- both parents have high school degrees at most;

- one parent has a university degree (the other does not);
- both parents have university degrees.

2.1.5. Group Effects

Group effects were interpreted in two different ways. First, we hypothesized that major area of study is a source of (group) identity for students, shaping their approach towards the Sociology class, therefore we included major as one of our variables (it was provided by seminar instructors).

Besides, the seminar group that students were part of could also be of consequence. During the Fall semester of 2015, there were 9 seminar groups. We did not assume that these had the strength to develop “group identities”, because they existed only for 90 min during the week, whereas students belonging to the same major spend much more of their time together in specialized classes. Still, the groups do put students in a certain “context”, which should be grasped in analysis. Therefore, we included these 9 groups not as an independent variable, but as a basis of clustering (see below).

2.1.6. Student Ambition

There was no explicit measure for student ambition or motivation available (although, notably, it might overlap with learning approach), but there was one useful proxy related to seminar groups. Out of the aforementioned 9 groups, 6 were very close to the size limit of 40 (the smallest of them was 37). The remaining 3 groups were below the size of 30. We firmly believe that small groups were results of selection: they were held in unattractive times of the week (two on Friday and one at a very late hour on Monday). Students who are generally serious and ambitious about their studies will typically register for their classes earlier, filling up the “good” slots, while those who care a little less about their academic careers will end up in the “fringe” (but also smaller) groups. Therefore, eventually this group size does not matter to us in the “classical” sense of class size, but as an indicator of students being ambitious in scheduling (members of large groups: yes, of small groups: no). Moreover, this variable is actually not only an indicator of individual ambition—the selection process described above should lead to group compositions of similarly ambitious students (even creating a possibility of either virtuous or vicious circles in the motivational climate of these groups).

2.1.7. Teacher Effect

Apart from the two Friday groups, each seminar had a different teacher. Thus, by adding the “particular person of teacher” as an independent variable to the analysis, estimation of effects would not have been possible, because of perfect multicollinearity or perfect separation. However, in order to capture some of the teacher effects that result from unobservable characteristics of instructors, gender of teacher was included as an explanatory variable in the model. Out of 9 groups, 6 had female teachers (164 total students in sample), and the remaining 3 had males (100 students in sample). All 3 of the “unfavorable” time slots were led by female seminar instructors.

2.1.8. Learning Outcome

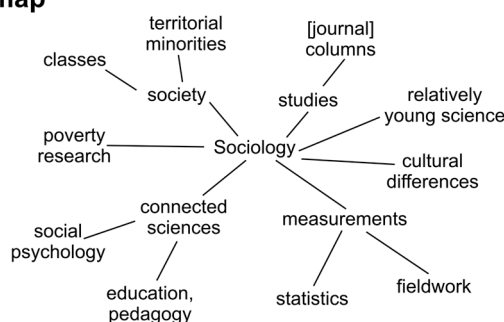
Finally, we come to the operationalization of the dependent variable. As mentioned earlier, the Sociology course under investigation here was evaluated with multiple-choice (MC) tests. How do such tests relate to the qualitative conception of learning outlined above? The literature is not unequivocal in this matter. On the one hand, Simkin and Kuechler, based on previous research, write that “MC items can reliably measure the same knowledge levels as CR items for the first four of Bloom’s levels” (CR meaning constructed response such as essays where deep understanding and analytical skills really show) [5] (p. 83); and Er et al. concur by stating that “well-constructed [multiple choice questions] can be used to assess higher-order cognitive skills such as interpretation, synthesis and application” [37] (p. 9). On the other hand, Simkin and Kuechler also draw attention to the fact that while in MC tests it is inevitable that the student’s thinking must converge to a single “good”

answer, true and able application of knowledge requires divergent production [5] (pp. 82–86). In our introductory course, we found it hard to construct MC questions that could meaningfully measure application of sociological knowledge, and as such, could go beyond the second level (comprehension) of the Bloom taxonomy—which is not deep enough to indicate “deep learning”.

Nevertheless, MC tests might still seem useful as a proxy for measuring deep or surface learning outcomes. However, Simkin and Kuechler’s review of the literature indicates that while in many studies, MC test results correlated with CR measures of knowledge, this finding is not ubiquitous, and MC tests are possibly an unreliable measure of deep understanding [5] (pp. 78–79).

Therefore, we had to find another way of measuring qualitative learning outcomes, and opted to use mind maps as a data source. Mind maps are generally well-known diagrams that are organized around a central concept. Starting from this center, new concepts are connected with lines, basically just recording a chain of association. Concepts can form chains and branches, and can also be connected by cross-links (see examples on Figures 1 and 2). The diagram allows so much freedom that besides words, small signs and pictures can also be added, parts can be of any color, lines of any width the creator likes [38] (pp. 21–24).

T1 mind map



T2 mind map

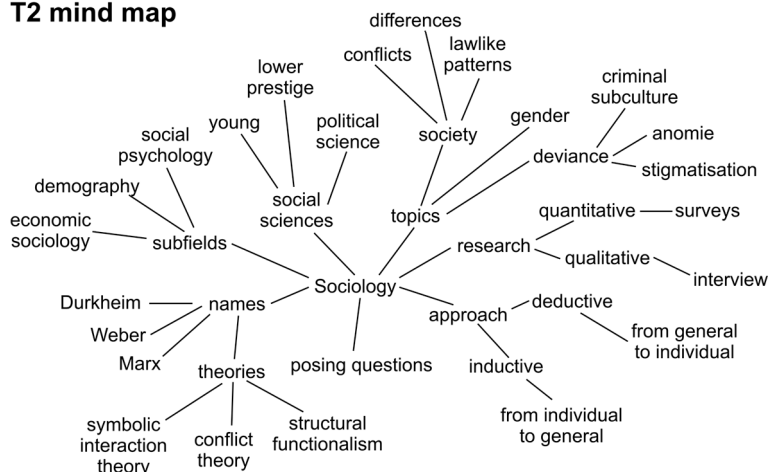
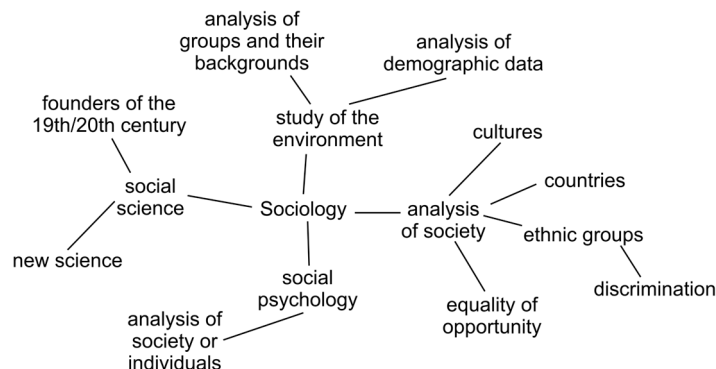


Figure 1. Example of a “deep learning” outcome. T2 shows a broader, more hierarchized structure, with some elements of T1 enriched with new knowledge acquired during the semester.

The end result is a diagram that reflects the highly personal ways of thinking and understanding of the author. Besides, concepts, i.e., pieces of information as “nodes”, are just as important as the links between them—and both these aspects reflect very well the considerations of cognitive learning theory, namely that the student has to be able to form a connection between parts, and especially between their existing knowledge and new information, based on a personal understanding of the material [38] (pp. 25–27, 79–80). Opposed to MC questions where the correct answer only has to be “recognized”,

mind maps have to be constructed without cues and should therefore indicate whether a student merely “memorized” the material or “understood” it as a whole, organized unit.

T1 mind map



T2 mind map

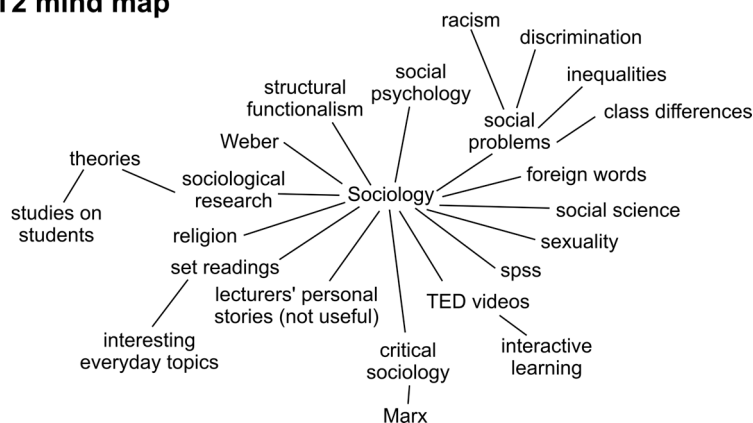


Figure 2. Example of a “surface learning” outcome. Even though T2 contains much more information, the structure barely goes beyond one level, many aspects are half-baked.

Diagrammatic elicitation, that is, the use of diagrams in data gathering, has many advantages [39]. Notably, diagrams help the respondent express what might be difficult to articulate in writing or speech, the influence of the researcher on responses given in this form is minimal, and a unique data structure is created, which opens up the possibility of both qualitative and quantitative analysis [38,40].

In both T1 and T2, our respondents drew a mind map around the central concept of “Sociology” (this was the only instruction). Thus, the T1 mind map represents their initial knowledge and the T2 one their knowledge at the end—recalled freely and registered expressively, showing connections between key parts without having to adopt the subtleties of a social scientific language. By comparing these two diagrams, we sorted each to one category of the qualitative learning outcome variable: deep or surface learning. Our guidelines for this categorization resemble closely those of Hay [41], who measured learning outcomes with a similar type of diagram (concept map). The basis of categorization was always the T2 diagram. The hallmarks of the two learning outcomes are as follows.

Deep learning:

- the T2 mind map refers to several (more than half) of the topics discussed during the semester, to each with more than one concept,
- the map is hierarchized: concepts are linked not only to the central node (Sociology) but are on several levels in chains,

- compared with the T1 map, good progress is made and/or new concepts learned during the semester are connected to pre-existing knowledge,
 - if the T2 mind map was very full and convincing on its own, no linking to concepts presented in T1 was required for a “deep” categorization.

Surface learning:

- the T2 mind map refers to less than half of the semester’s topics,
- the structure of the map is simple: contains links mostly only to the central concept,
- no previous knowledge is brought back in T2, or no links are formed between that and newly learned information,
- the T2 mind map contains elements that are mostly irrelevant to “Sociology” itself (overly “free” associations, recalling of personal experiences of the first university semester, etc.),
- the T2 mind map shows no progress compared with T1.

Based on their mind maps, each student was given a “Deep” or “Surface” verdict by two members of the research team separately. In case of disagreement, a third member assigned the final category. Figures 1 and 2 contain illustrative examples of mind maps that ended up receiving Deep and Surface labels.

2.1.9. Summary of the Model

In the preceding sections, variables were presented one by one. Figure 3 provides a visual summary of the model, where each variable is shown as part of a theoretical concept specified above.

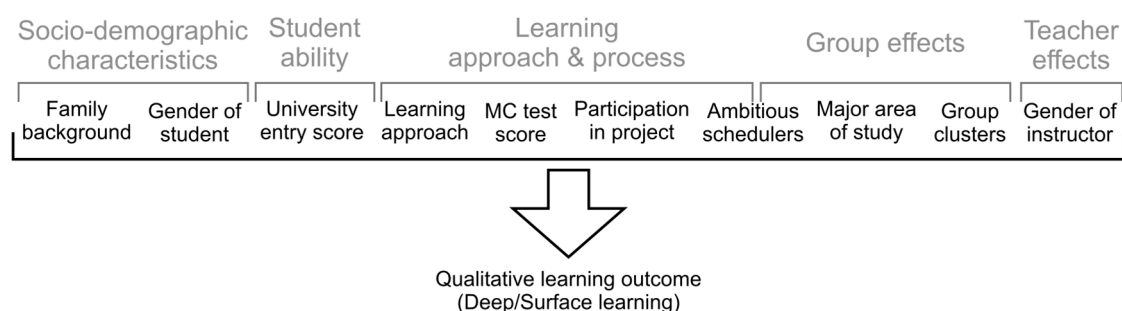


Figure 3. Variables of the research (in black). Corresponding concepts are indicated in gray.

What justifies the inclusion of multiple choice test score as an independent variable? While it was fully expected to be correlated with qualitative learning outcome (in accordance with the relevant literature), we wanted to control for what the MC score measured and see whether the effects of other variables still remained significant for deep learning.

2.2. Methods of Analysis

Since the aim was to investigate the effects of the measured factors influencing individual qualitative learning outcomes, three appropriate statistical model variants were investigated that could quantify these relationships.

It is natural to conceive the Deep and Surface qualitative learning as realizations of a binary random variable, for which the measured values were determined through expert classification, as described above. This meant that we were facing a supervised statistical learning problem concerned with inferring the correct learning outcome classification for individuals, given their measurements for some chosen attributes (listed in the model specification). There are a number of models that were designed to tackle such classification problems [42], but we had additional requirements from our

candidate models. In these settings, there is a trade-off between predictive performance of models, and model interpretability [43] (pp. 24–26). In our case, a model with high interpretability was a priority, since we wanted to be able to understand how different factors influenced the learning outcome, even at the expense of predictive performance, considering the fact that there may have been several influential omitted factors (i.e., unmeasurable in our study) that would hinder performance of less interpretable models. In connection with the latter consideration, we were looking for a classifier that derives its results from an underlying modelling of the probability of cases being in each group. Again, statements that pertain to changes in class probabilities with respect to changes in attribute values are preferred for a classifier from an interpretation standpoint, as opposed to one that solely relies on the confusion matrix [43] (pp. 127–168).

Based on our requirements and the above reasoning, we have chosen the binary logistic regression model to investigate the influence of measured attributes on the probability of attaining a “Deep learning” outcome class compared to “Surface learning”. Specifically, logistic regression implicitly models class probabilities through the logarithm of odds of the two levels of the dependent variable with a linear combination of the explanatory variables. Regression coefficients can thus be easily interpreted as the change in log odds, or more naturally the odds ratio if exponentiated. A regression coefficient (or log odds) of 0 would mean that a unit change in the explanatory variable would not affect the odds of “Deep learning” versus “Surface learning”. This “zero effect” is obviously an odds ratio of 1. The most important characteristic of logistic regression models is that effects are linear and additive only on the log scale of odds, and not on the linear scale of odds ratios where the actual probabilities of outcomes are measured.

Since we have gathered data from 9 seminar groups, we had to take into consideration this clustered structure of our sample. Three model variants were selected, all accounting for this clustering, but with a different approach, with increasing complexity in structure [44]. Explanatory variables were the same set, as motivated by theory, for all models.

In the simplest variant, we used a robust standard error estimation of the regression coefficients, which resulted in increased coefficient estimate confidence intervals, as standard errors increase due to a smaller effective sample size if sample elements are not independent of each other. The other two variants—the fixed effects and the random effects versions—explicitly modelled the presence of clusters (i.e., seminar groups) in the dataset to separate within cluster and between cluster effects. The fixed effects binary logistic regression is appropriate for estimating within cluster coefficients of the individual level explanatory variables, while the random effect model also allows for the estimation of group level effects at the expense of using an inconsistent estimator if group level effects are correlated with the independent variables [45]. Conceptually, fixed effects would mean that groups in the sample (with their respective mean outcomes) are treated as a given population of groups, which could be argued for, since all groups in the studied year-group were measured, while random effects would model groups as random draws from larger population of groups, which makes sense if we generalize from our surveyed year-group to at least other year-groups close in time.

2.2.1. Structure of Models

Let us define $\Pr(y = \text{Deep learner}) = p(X)$ as the probability of an individual achieving a “Deep learning” outcome (y) with some fixed values on the explanatory variables X . In its most simple form, the binary logistic regression model is the following:

$$\ln\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta X$$

$$p(X) = \frac{e^{(\beta_0 + \beta X)}}{1 + e^{(\beta_0 + \beta X)}}$$

where β_0 is the intercept, and β is the regression coefficient vector for the matrix of explanatory variables X .

For the fixed and random effects models, indices are needed to denote individuals and groups, since each group can have its own (fixed) intercept. The outcome of the i -th individual in the j -th group is y_{ji} . The fixed effect model takes the following form:

$$\Pr(y_{ji} = \text{Deep learner}) = p_{ji}(X) = \frac{e^{(\beta_{j0} + \beta X)}}{1 + e^{(\beta_{j0} + \beta X)}}$$

where β_{j0} is the intercept of the j -th group. Interpretation of odds ratios is the same as for simple logistic regression, since these are independent of group intercepts. This is not true however for odds, which complicates interpretation. Note that for X the indices were omitted to simplify notation.

The random effects model differs from the fixed effects in the fact that the group level intercepts are thought to be realizations of a normally distributed random variable. Without a group level explanatory variable, the random intercept model is the following:

$$\Pr(y_{ji} = \text{Deep learner}) = p_{ji}(X) = \frac{e^{(\beta_{00} + \beta X + u_j)}}{1 + e^{(\beta_{00} + \beta X + u_j)}}$$

where β_{00} is the grand mean, and u_j -s are modelled as i.i.d. draws from $N(0, \sigma^2)$.

If we were to assign explanatory variables to the group effect, that would mean substituting the 0 expected value of u_j -s with their own regression equation: $u_j \sim N(\gamma Z, \sigma^2)$, where the γ coefficient vector affects how a group mean deviates from the grand mean for given values of group level explanatory variables Z . Again, the only parameters that have an easy direct interpretation are the regression coefficients—for either the individual or the group level—as differences in log odds, and their exponentiated values as odds ratios.

2.2.2. Model Selection

Selecting the best model from the three considered variants is not a straightforward procedure, since both theoretical and statistical considerations may motivate a choice.

From a theoretical standpoint, choosing a simple logistic regression would mean that we were only concerned with grouped sampling, to adjust for the smaller effective sample size in the standard error estimation. Choosing the fixed effects model would mean estimating differences in group means (mean log odds), but considering these differences to be given as if predefined by us, while in the case of the random effects model, these means would also be modelled as random variables, known only when measured from the data. Since we wanted to make statements that generalize to larger population of groups from other year-groups (of at least neighboring years), and we were interested in certain group level effects, we have opted for the random effects binary logistic regression model.

The choice made based on theoretical considerations was not contradicted by statistical evidence. The likelihoods of the comparable models were very close, the Durbin-Wu-Hausman test for the comparison of the coefficients of the fixed effects and random effects models was inconclusive [46]. Lastly, even if we could have shown that model results are statistically different, it would not have mattered from an interpretation standpoint, since the regression coefficients (see Table 1) were qualitatively the same for all models.

2.2.3. Missing Values

The only type of missing values in the sample constituted of the missing T2 mind map measurements for 40 students. This does not introduce added bias to the analysis, since our models used these as dependent variables.

Table 1. Regression coefficients for the three models: simple-, fixed effects-, and random effects logistic regression (LR).

Variable	Level	Log Odds Ratio		
		Simple LR	Fixed Effects LR	Random Effects LR
Major	Sociology		reference	
	Political Science	−1.933 **	−1.649 *	−1.879 **
	Media and Communication	−0.816	−0.951	−0.843
	International Relations	−0.605	−0.529	−0.591
Gender	Male		reference	
	Female	0.335 *	0.428	0.361
Project work	Non		reference	
	Group	−0.772	−0.681 ′	−0.761 ′
	Individual	−0.004	−0.473	−0.088
Family background	neither parent finished high school		reference	
	one parent finished high school (the other did not)	−0.613	−0.731	−0.653
	both parents have high school degrees at most	−2.272 *	−2.389 *	−2.314 *
	one parent has a university degree (the other does not)	−1.183	−1.246	−1.202
	both parents have university degrees	−1.673	−1.777 *	−1.712 *
Gender of instructor	Male		reference	
	Female	1.095 *		1.083 **
Entry score		−0.002	−0.004	−0.002
Learning approach		0.102 ′	0.130 *	0.108 ′
Ambitious schedulers	No		reference	
	Yes	1.665 **	0.056 *	1.639 **
Test score		0.055 *		0.055 *
Intercept		−3.086		−2.959
Group level std. dev.				0.249

′ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$.

3. Results

Results below are presented in two forms. First, a quick quantitative overview of the sample is given, which offers a comparison of students of the four major areas, and offers some interesting insights in itself. Second, the main result is shown: coefficients obtained from the logistic regression model.

3.1. Quantitative Overview of the Sample

Our sample consisted of 264 students. Table 2 shows an overview of some key variables of the research.

Some patterns are conspicuous. One is that female students are a majority in all major areas of study, except for Political Science. University entry scores are high across the board, but International Studies and Political Science students stand out, and basically the same can be seen in MC test scores. Project participation is highest among Sociology students.

Table 2. Quantitative overview of the sample.

Major	Gender ¹		University Entry Score ²	Learning Approach Score ²	MC Test Score ²	Project Participation ¹		
	Male	Female				No	Group	Individual
Sociology	23 (37%)	39 (63%)	416.53 (26.15)	1.18 (2.59)	60.6 (10.73)	19 (31%)	22 (35%)	21 (34%)
Political Science	22 (63%)	13 (37%)	442.15 (15.92)	0.20 (2.78)	61.6 (6.38)	15 (43%)	8 (23%)	12 (34%)
Media and Communication	12 (18%)	54 (82%)	418.87 (28.96)	1.17 (2.81)	59.0 (8.41)	35 (53%)	11 (17%)	20 (30%)
International Studies	40 (40%)	61 (60%)	456.57 (19.87)	1.59 (2.74)	62.8 (8.70)	61 (60%)	16 (16%)	24 (24%)
Whole year-group	97 (37%)	167 (63%)	436.30 (29.59)	1.21 (2.75)	61.2 (8.99)	130 (49%)	57 (22%)	77 (29%)

¹ Row percentages are presented. ² The display format for these items is: Mean, then (standard deviation).

3.2. Model Results

The results for our chosen model variant are summarized in Table 3. For the qualitative learning outcome binary dependent variable, the reference category is “Surface learning”, so interpreting log odds ratios always has the form of “How much does the log odds of getting to Deep learning form Surface learning change when we move one unit from the reference level on the explanatory variable, keeping all other explanatory variables constant?”.

Table 3. Results for the random effects logistic regression model. Log Odds Ratios significant on a level of $p < 0.1$ are shown in bold.

Variable	Level	Log Odds Ratio	Std. Err.	z	$p > z $	95% Conf. Interval	
Major	Sociology	reference					
	Political Science	−1.879	0.666	−2.82	0.005	−3.18	−0.57
	Media and Communication	−0.843	0.477	−1.77	0.077	−1.78	0.09
	International Relations	−0.591	0.493	−1.20	0.231	−1.56	0.38
Gender	Male	reference					
	Female	0.361	0.347	1.04	0.298	−0.32	1.04
Project work	Non	reference					
	Group	−0.761	0.433	−1.76	0.079	−1.61	0.09
	Individual	−0.088	0.412	−0.21	0.832	−0.90	0.72
Family background	neither parent finished high school	reference					
	one parent finished high school (the other did not)	−0.653	0.954	−0.68	0.494	−2.52	1.22
	both parents have high school degrees at most	−2.314	0.921	−2.51	0.012	−4.12	−0.51
	one parent has a university degree (the other does not)	−1.202	0.842	−1.43	0.153	−2.85	0.45
	both parents have university degrees	−1.712	0.819	−2.09	0.037	−3.32	−0.11
Gender of instructor	Male	reference					
	Female	1.083	0.420	2.58	0.010	0.26	1.91
Entry score		−0.002	0.007	−0.36	0.717	−0.02	0.01
Learning approach		0.108	0.063	1.72	0.085	−0.01	0.23
Ambitious schedulers	No	reference					
	Yes	1.639	0.499	3.32	0.001	0.66	2.62
MC test score		0.055	0.022	3.28	0.013	0.01	0.10
Intercept		−2.959	2.899	−1.02	0.307	−8.64	2.72
Std. dev. within groups		0.249	0.318			0.02	3.05

Model Wald Chi-sq. = 34.89, with $p = 0.0026$.

First, we provide a brief textual overview of the regression output, then move on to connect the results to our hypotheses in the next section (Discussion).

We can state that *ceteris paribus* (a remark always applicable but omitted for following statements) Sociology major students are more likely to be Deep learners than either Political Science or Communications majors. Political Science majors have a substantially lower chance of being Deep learners than all other groups. Male and female students have statistically equal odds of being Deep learners. Project workers, or more precisely those who chose to do group project work, have a lower chance to be Deep learners than those who did not choose this. Those levels of family background with parents having the same education significantly decrease the odds of ending up with a Deep learning outcome. Differences in entry score are not significant, but differences in learning approach are. A deeper approach toward learning does increase the odds of achieving the outcome corresponding to that attitude. While this sounds like a tautology, it actually informs us about the validity of the results, since learning attitudes and learning outcomes were measured independently.

The group level explanatory variable of gender of seminar group instructor, and the partially group level “ambitious schedulers” variable both had significant effects on the dependent variable. Individuals in groups with a female instructor had higher odds of becoming Deep learners than those in groups of male instructors. It is important to note, again, that this is true even when controlling for all other factors. Ambitious schedulers performed better, since those investing effort into getting selected to a seminar with a preferable time-slot were likely more proactive students (who could also form more ambitious groups).

Multiple choice test scores also had significant effects, but we only included this variable in the model to control for performance accounted by this non-qualitative element of learning.

We have termed effects significant for a p -level less than 0.1, which can be justified by the fact that 0.05 is not a strict cut-off limit [47], and we were aware of the fact that a stricter controlling of Type-I error can suggest a false confidence in results, since for our limited sample size and large number of variables, Type-II error can be just as problematic.

The rho coefficient (the estimated share of the between-groups variance) is just 1.8%, and the Likelihood-ratio test for its difference from zero is non-significant ($p = 0.22$), which confirms that the single-level and the multi-level models give very similar results.

4. Discussion

Let us start the discussion of results in the order of our hypotheses. Hypothesis 1, namely that learning approach will act positively on the outcome, did not gain unqualified support. Although its effect is positive, indicating that a deeper self-professed learning approach does “lead” to a deeper outcome, it can only claim significance with the “relaxed” threshold of 0.1, and the effect size is moderate as well (a log odds ratio of 0.108). This might be due to the fact that learning approach was measured only with 4 items in our questionnaire, which might not produce a result reliable enough, or might fail to grasp the true variance within the sample.

Our second hypothesis, concerning the expected positive effects of project work, can be resoundingly rejected on the basis of the evidence. Actually, doing a group project had a significant and large negative effect on learning outcome (log OR of -0.761). Admittedly, this result left us baffled and we struggle to explain it. We can only hypothesize that those who opt to do work in groups do so out of a motivation to “socialize”, and not because they strive to do extra-curricular work. Also, the seemingly negative but absolutely not significant—basically *absent*—effect of the individual project on learning outcome is hard to accept from a pedagogical standpoint. We can speculate that those students who are confident in their abilities do not opt for the safety of extra points that a project can provide, but make their way to success without it, also ending up with deep learning outcomes in the process.

After controlling for several variables, female students are not significantly deeper learners than their male counterparts, contrary to our 3rd hypothesis. This is not overly surprising, because the previous empirical evidence on which we based the assumption did not relate to a qualitative conception of knowledge. Also, since the university attracts students of the highest ability, there should

be no big difference between them on the basis of gender. Female students are still a little “ahead”, which is in line with indications of their reputedly better self-discipline and also their possibly higher level of appreciation of female teachers.

Our hypothesis 4 on the effect of family background has also been practically refuted. Instead of providing an advantage in the “depth” of learning, each category in the ordinal variable seems to do damage compared to the baseline (neither parent finished high school). What is more, the negative effect of both parents having high school (log OR of -2.314), and both parents having a university degree (log OR of -1.712) are statistically significant as well. This suggests to us a peculiar pattern in which children of families where the “educational horizon” of parents is “symmetrical” perhaps do not put as much effort into deep learning as others do. The fact that children of relatively lower (or “asymmetrically”) educated parents perform better could be explained by an “aspiration” hypothesis: the asymmetry puts the importance of education into sharp relief for them. Overall, our results make the impression that among those students who reach university, family background no longer provides a considerable advantage. Generally, this is good news from a social mobility standpoint.

Hypothesis 5 pertained to Political Science students and is strongly supported by the data: compared to Sociology students, their likelihood of ending up with a deep learning outcome is significantly and substantively smaller (log OR of -1.879). We believe that their relative indifference to our subject might result from the fact that out of the “other three” majors, Political Science is the closest to Sociology, and there might be a sense of competition between the two fields (e.g., in scientific prestige). Communication and International Studies students might not feel such pressures. The fact that Sociology students are the “deepest” learners is not a surprise and also comes as a relief to us.

Furthermore, it is notable in the results that university entry score—the proxy of ability and/or prior achievement—practically has zero effect on qualitative learning outcome in our course. This might suggest that the state-sanctioned exam does not reflect the ability for good understanding; indeed, it is generally assumed that it still concentrates on lexical knowledge to a great extent. Multiple-choice test score, a measure we do know to be based on lexical knowledge, has a significant effect on deep learning (log OR of 0.055)—a result we expected and is in line with existing literature; but again, the point of including MC scores in the regression was to control for it while measuring the effect of all other variables.

Lastly, we should discuss the effects of seminars. For one, we are not surprised to see that “ambitious schedulers” are positively and significantly related to deep learning (log OR of 1.639)—to us, this justifies our assumptions related to it both as an individual level and as a composition effect indicator along the lines of our reasoning presented above: more ambitious students end up together, filling up preferable class hours. Once again, while this variable coincided with seminar group size, we believe it did not measure a class size effect. In order to gain true insight into the effect of class size within our teaching context, we would clearly have to do random group assignment.

The positive effect that female teachers had on deep learning (log OR of 1.083) suggest to us that there indeed were unmeasured (unmeasurable) teacher characteristics that resulted in different outcomes between the groups. We believe that teacher gender is a proxy of them; we are very cautious of accepting it at face value that teacher gender in itself would be the cause of this difference. Such an unmeasured effect—as in [15]—could be the higher level of student-teacher interaction that is characteristic of female teachers.

Our results describe one year-group at the university in question, 87% of which was included in the sample, making the sample almost complete. Personally, we are confident that our results are at least indicative of neighboring year-groups at the university. Some of the results are in line with what educational research in general, and our hypotheses in particular, suggested and foreshadowed. For example, teacher effects are apparent in the findings. Other results we found surprising. One of them is the fact that in spite of their otherwise obvious academic excellence, Political Science students end up being “shallower” learners in this course than students of other majors. Contrary to our

expectations, the extra-curricular project work does not seem to have a “deepening” effect on learning outcomes—something that could deserve further attention in future studies. Also curious was the pattern in which family background influenced learning outcomes—seemingly, parents with “both high school” or “both university” degrees are “detrimental” to deep learning. Otherwise, as teachers of Sociology we are happy that family background or prior achievement (i.e., entry score) do not seem to be barriers to understanding our subject, and that Sociology majors perform well generally.

Educational researchers and teachers of Sociology might find our study interesting because of some of its counter-intuitive findings, and because of the insight it provides into group norms or identities that major areas of study constitute, and are influential to openness towards our subject. We believe that its design—utilizing diagrammatic elicitation—is an easy but powerful tool of measuring qualitative learning outcomes. Mind maps used within the research contain valuable qualitative data as well, the analysis of which will be the next step of our research undertaking.

Acknowledgments: The authors received funding from the project titled “for innovative solutions to improve the efficiency of R&D activity in Zala county. Project identifier EFOP-3.6.1-16-2016-00012”. The authors would like to express their gratitude to the members of the research team: Gábor Király, Kinga Emese Kovács, Yvette Lovas, and Bence Pálóczi, who helped us with advice, data processing, and—in the case of Yvette—revision of deep and surface classifications.

Author Contributions: Péter Miskolczi provided the theoretical framework and organized and oversaw data collection. Márton Rakovics suggested the data analysis methods, and was responsible for their execution. Both authors have read and approved the final version of the article.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Arum, R.; Roksa, J. Limited Learning on College Campuses. *Society* **2011**, *48*, 203–207. [[CrossRef](#)]
2. Dalhgren, L.-O. The Quantitative Conception of Knowledge. In *The Experience of Learning: Implications for Teaching and Studying in Higher Education*, 3rd ed.; Marton, F., Hounsell, D., Entwistle, N., Eds.; University of Edinburgh, Centre for Teaching, Learning and Assessment: Edinburgh, UK, 2005; pp. 23–38.
3. Entwistle, N. Contrasting Perspectives on Learning. In *The Experience of Learning: Implications for Teaching and Studying in Higher Education*, 3rd ed.; Marton, F., Hounsell, D., Entwistle, N., Eds.; University of Edinburgh, Centre for Teaching, Learning and Assessment: Edinburgh, UK, 2005; pp. 3–22.
4. Bloom, B.S.; Engelhart, M.D.; Furst, E.J.; Hill, W.H.; Krathwohl, D.R. *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*; David MacKay: New York, NY, USA, 1956; Volume 19.
5. Simkin, M.G.; Kuechler, W.L. Multiple-Choice Tests and Student Understanding: What Is the Connection? *Decis. Sci.* **2005**, *3*, 73–98. [[CrossRef](#)]
6. Marton, F.; Säljö, R. On qualitative differences in learning: I—Outcome and process. *Br. J. Educ. Psychol.* **1976**, *46*, 4–11. [[CrossRef](#)]
7. Marton, F.; Säljö, R. Explaining Differences in Outcome. In *The Experience of Learning: Implications for Teaching and Studying in Higher Education*, 3rd ed.; Marton, F., Hounsell, D., Entwistle, N., Eds.; University of Edinburgh, Centre for Teaching, Learning and Assessment: Edinburgh, UK, 2005; pp. 39–58.
8. Scouller, K. The influence of assessment method on students’ learning approaches: Multiple choice question examination versus assignment essay. *High Educ.* **1998**, *35*, 453–472. [[CrossRef](#)]
9. Monk, D.H. The education production function: Its evolving role in policy analysis. *Educ. Eval. Policy Anal.* **1989**, *11*, 31–45. [[CrossRef](#)]
10. Hanushek, E.A. The economics of schooling: Production and efficiency in public schools. *J. Econ. Lit.* **1986**, *24*, 1141–1177.
11. Dewey, J.; Husted, T.A.; Kenny, L.W. The ineffectiveness of school inputs: A product of misspecification? *Econ. Educ. Rev.* **2000**, *19*, 27–45. [[CrossRef](#)]
12. Krueger, A.B.; Whitmore, D.M. The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *Econ. J.* **2001**, *111*, 1–28. [[CrossRef](#)]

13. Nye, B.; Konstantopoulos, S.; Hedges, L.V. How large are teacher effects? *Educ. Eval. Policy Anal.* **2004**, *26*, 237–257. [\[CrossRef\]](#)
14. Basow, S.A. Student evaluations of college professors: When gender matters. *J. Educ. Psychol.* **1995**, *87*, 656–665. [\[CrossRef\]](#)
15. Duffy, J.; Warren, K.; Walsh, M. Classroom interactions: Gender of teacher, gender of student, and classroom subject. *Sex Roles* **2001**, *45*, 579–593. [\[CrossRef\]](#)
16. Ermisch, J.; Francesconi, M. Family matters: Impacts of family background on educational attainments. *Economica* **2001**, *68*, 137–156. [\[CrossRef\]](#)
17. Davis-Kean, P.E. The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *J. Fam. Psychol.* **2005**, *19*, 294–304. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Buchmann, C.; DiPrete, T.A. The growing female advantage in college completion: The role of family background and academic achievement. *Am. Sociol. Rev.* **2006**, *71*, 515–541. [\[CrossRef\]](#)
19. Holm, A.; Jæger, M.M. *Selection Bias in Educational Transition Models: Theory and Empirical Evidence*; Social Policy and Welfare Services Working Paper; SFI—The Danish National Centre for Social Research: Copenhagen, Denmark, 2008; Volume 11.
20. Duckworth, A.L.; Seligman, M.E. Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *J. Educ. Psychol.* **2006**, *98*, 198–208. [\[CrossRef\]](#)
21. Vermunt, J.D. Relations between student learning patterns and personal and contextual factors and academic performance. *High Educ.* **2005**, *49*, 205–234. [\[CrossRef\]](#)
22. Epple, D.; Romano, R. Peer effects in education: A survey of the theory and evidence. *Handb. Soc. Econ.* **2011**, *1*, 1053–1163. [\[CrossRef\]](#)
23. Fordham, S.; Ogbu, J.U. Black students' school success: Coping with the "burden of 'acting white'". *Urban Rev.* **1986**, *18*, 176–206. [\[CrossRef\]](#)
24. Horvat, E.M.; Lewis, K.S. Reassessing the "burden of 'acting White'": The importance of peer groups in managing academic success. *Sociol. Educ.* **2003**, 265–280. [\[CrossRef\]](#)
25. Salmivalli, C.; Voeten, M. Connections between attitudes, group norms, and behaviour in bullying situations. *Int. J. Behav. Dev.* **2004**, *28*, 246–258. [\[CrossRef\]](#)
26. Perkins, H.W. Social norms and the prevention of alcohol misuse in collegiate contexts. *J. Stud. Alcohol Suppl.* **2002**, *14*, 164–172. [\[CrossRef\]](#)
27. Bamberger, P.; Biron, M. Group norms and excessive absenteeism: The role of peer referent others. *Organ. Behav. Hum. Decis. Process.* **2007**, *103*, 179–196. [\[CrossRef\]](#)
28. Szafran, R.F. What do introductory sociology students know and when do they know it? The results of pretesting students. *Teach. Sociol.* **1986**, *14*, 217–223. [\[CrossRef\]](#)
29. Neuman, W.L. Which students learn the most, and why? A replication and extension of the Szafran pretest study. *Teach. Sociol.* **1989**, 19–27. [\[CrossRef\]](#)
30. Dietz, T.L. Predictors of success in large enrollment introductory courses: An examination of the impact of learning communities and virtual learning resources on student success in an introductory level sociology course. *Teach. Sociol.* **2002**, *30*, 80–88. [\[CrossRef\]](#)
31. Howard, J.R. An examination of student learning in introductory sociology at a commuter campus. *Teach. Sociol.* **2004**, *33*, 195–205. [\[CrossRef\]](#)
32. Howard, J.R.; Novak, K.B.; Cline, K.M.; Scott, M.B. Another Nibble at the Core: Student Learning in a Thematically-focused Introductory Sociology Course. *Teach. Sociol.* **2014**, *42*, 177–186. [\[CrossRef\]](#)
33. Alon, S.; Tienda, M. Diversity, opportunity, and the shifting meritocracy in higher education. *Am. Sociol. Rev.* **2007**, *72*, 487–511. [\[CrossRef\]](#)
34. McWhorter, J.H. Eliminating the SAT Could Derail the Progress of Minority Students. *The Chronicle of Higher Education*, 9 March 2001, pp. B11–B12.
35. Biggs, J.; Kember, D.; Leung, D.Y. The revised two-factor study process questionnaire: R-SPQ-2F. *Br. J. Educ. Psychol.* **2001**, *71*, 133–149. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Mogashana, D.; Case, J.M.; Marshall, D. What do student learning inventories really measure? A critical analysis of students' responses to the Approaches to Learning and Studying Inventory. *Stud. High Educ.* **2012**, *37*, 783–792. [\[CrossRef\]](#)

37. Er, H.M.; Ramamurthy, S.; Pook, P.C. Can learning outcomes in cognitive domain be assessed effectively using multiple choice questions? A study in an undergraduate pharmacy curriculum. *Int. J. Sci. Med. Educ.* **2014**, *8*, 9–18.
38. Wheeldon, J.; Ahlberg, M.K. *Visualising Social Science Research*; SAGE Publications: London, UK, 2011.
39. Miskolczi, P.; Király, G.; Kovács, K.E.; Lovas, Y.; Pálóczi, B. Fogalmak mentén: Hallgatói gondolatterképek alkalmazása oktatási újítás vizsgálatára. *Prosperitas* **2016**, *3*, 111–132.
40. Umoquit, M.J.; Tso, P.; Burchett, H.D.; Dobrow, M.J. A multidisciplinary systematic review of the use of diagrams as a means of collecting data from research subjects: Application, benefits and recommendations. *BMC Med. Res. Methodol.* **2011**, *11*, 11–20. [[CrossRef](#)] [[PubMed](#)]
41. Hay, D.B. Using concept maps to measure deep, surface and non-learning outcomes. *Stud. High Educ.* **2007**, *32*, 39–57. [[CrossRef](#)]
42. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, 2nd ed.; Springer: New York, NY, USA, 2009; ISBN 9780387848570.
43. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: New York, NY, USA, 2013; Volume 6, ISBN 978-1461471370.
44. Rabe-Hesketh, S.; Skrondal, A. *Multilevel and Longitudinal Modeling Using Stata*; Stata Press: College Station, TX, USA, 2008.
45. Agresti, A. *Categorical Data Analysis*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2013; ISBN 9780470463635.
46. Gelman, A.; Hill, J. *Data Analysis Using Regression and Multilevel Hierarchical Models*, 1st ed.; Cambridge University Press: New York, NY, USA, 2007; ISBN 9780521867061.
47. Ziliak, S.T.; McCloskey, D.N. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*; University of Michigan Press: Ann Arbor, MI, USA, 2008; ISBN 9780472050079.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).