

Influence of Gender and Other Factors to Final Grade

Ludmila Domeova¹, Andrea Jindrova¹ & Jiri Fejfar¹

¹ Faculty of Economics and Management, Czech University of Life Sciences Prague, Prague, Czech Republic

Correspondence: Ludmila Domeova, Department of System Engineering, Czech University of Life Sciences Prague, Kamycka 129, 16521 Prague, Czech Republic. Tel: 420-224-382-181. E-mail: domeova@pef.czu.cz

Received: October 11, 2014 Accepted: December 5, 2014 Online Published: March 29, 2015

doi:10.5539/ies.v8n4p95

URL: <http://dx.doi.org/10.5539/ies.v8n4p95>

Abstract

The study focused on the relations between the partial evaluation and the final grade. The investigation has been done on a group of 269 students of the Czech University of Life Sciences in Prague, in the course of Mathematical Methods, who have to go through a strictly defined evaluation scheme. The results of statistical analysis confirmed that there is no influence of gender and branch of study to the final grade. The analysis of the relations between the final grade and the partial evaluation during the semester and by the examination test showed that the final grade is mostly dependent on the oral exam, i.e. the face to face evaluation of the student is deciding for passing or failing the exam. The research proved that the partial evaluation including on-line testing does not replace the deciding role of the teacher.

Keywords: partial evaluation, evaluation scheme, final grade, statistical analysis

1. Introduction

In the mathematical subjects the learning and teaching is not easy as well as the assessment of the acquired knowledge. The final grade should assess the student's knowledge and practical skills in a particular course, and it indicates passing or failing the exam. To support the final evaluation several steps of partial evaluation based on different examination methods have been introduced. Each of these steps is time and work demanding so it is important to investigate whether and to which level these partial evaluations influence the final grade.

1.1 Factors Influencing the Results

The results of each course and the study itself depend on many factors as social status, high school experiences, gender, race, family background, motivation and personality (Moorgat, 1996 in Laghal, Sévigny, & Frenette, 2013). These factors are connected with the performance of students. Other influences are connected with the teachers' personality and there are also random and unmeasured factors such as student-teacher chemistry or random deviations in student health (Grant, 2007). The question is to which extent the final grade can be used as a measure of the transmission of knowledge in a course and how the partial evaluation corresponds with the final grade.

The measure of student's learning might be also the student's course evaluation. Beleche et al. (2012) found a statistically significant relationship between individual student learning and the course evaluation. The evaluation of the quality of the student (grades) is often connected with measurement of the quality of the instructors (course, faculty and teachers). There is a significant positive effect of relative expected grades on evaluation scores (Ewing, 2012; Grant, 2007; Krautmann & Sander, 1999). This fact often leads to grade inflation, i.e. the teacher gives out easier grades in order to get better evaluation from the students. While higher standards may lead more motivated students to increase effort, they may cause others to give up (Betts & Grogger, 2003). The evaluation in several steps can bring more objectivity especially if it is done by several methods and by several persons.

The result of the exam can be affected by the student's ability to cope with a particular evaluation method. The evaluation methods investigated in this study are: multiple choice questions (on-line tests), calculation of examples (both on-line and off-line tests), written theoretical test with open answers and oral exam. By Chamorro-Premuzic et al. (2005) the most preferred method was continuous assessment, followed by multiple-choice exams, followed by supervised final year dissertation, followed by group work, followed by essay type of examinations, and finally oral or viva voce exams. The student's preference for evaluation methods can be explained by personality; the gender causes the higher difference in preferences of practical work and

written exams (Laghal, Sévigny, & Frenette, 2013). The results of Furnham, Batey, and Martin (2011) show that the multiple choice questions are preferred by bright, less open candidates; oral exams are better for stable, low conscientious students with a deep learning style. Chamorro-Premuzic et al. (2005) investigated relations between personality traits (Big Five), and preference for particular assessment methods, e.g. neurotic students seem to have a tendency to dislike oral examinations, open individuals are also more comfortable when they do not have to perform on analytic, concise, multiple-choice exams.

The mathematical proficiency is important for economic growth. Improving the maths performance should be a key focus for the economical faculties. The investigated subject is mathematical, and such subjects is said to be a source of stereotypes that harm girls. The gender gap in academic achievement is an important issue to explore, not only because the gender achievement gap itself is an important aspect of educational inequality, but also because it is closely related to the gender wage gap in the labor market (Lai, 2010). Niederle and Vesterlund (2010) note that the differences in observed maths scores may not necessarily match gender differences in maths skills but may in part reflect how men and women respond differently to test-taking environments. Falch and Naper (2013) found that the boys are better at final exams. Greater weight on coursework elements in the evaluation improves the relative performance of girls. The belief that school teachers have a grading bias against female students was not proved by Lavy (2008), over more there were indications that part of the gender difference is due to discrimination of male students. These results from Israel were not confirmed in Sweden (Hinnerich et al., 2011). In the study of Norwegian students in their final year of compulsory education the girls outperform boys in all subjects (Falch & Naper, 2013). The same results were reported by Lai (2010) from China: significantly more boys dropped out of the regular public school system by the end of middle school, and the male students demonstrated significantly inferior performance in all semester tests with exception of physic. Female students were found to be significantly more agreeable and more conscientious.

1.2 Continuity with Previous Research

The authors' previous work (Jindrová et al., 2013) deals with the evaluation of some courses in the University of Life Sciences in Prague and the e-support of the students. The evaluation was based on collection of self estimations and personal attitudes of the students. The views of students were compared with their real results and views of teachers. The present study uses only the test results and the evaluations made by teachers.

1.3 Goal of Present Investigation

The students of the Czech University of Life Sciences in Prague, in the course of Mathematical Methods, have to go through several evaluation steps before they get to the final exam. The final grade is influenced by the results of the partial evaluation and possibly by many other factors. The study focused on the relations between the partial evaluation and the final grade.

1.4 Goal and Hypotheses

The goal of the statistical analysis was to discover the mutual relations among the investigated factors. The strength of relations was in view for the partial evaluations during the semester and the final examination grade.

For the dependencies on the gender and the branch of study following hypotheses were formulated:

H1: The gender has no influence to the final grade.

H2: The branch of study has no influence on the final grade.

2. Method

The study is based on study results of the students of the Czech University of Life Sciences in Prague, Faculty of Economics and Management, with the majors "Economics and Management" and "Business and Administration" in the course "Mathematical methods in Economic II", winter semester, school year 2013/2014.

2.1 Partial and Final Evaluation of Students

All students have to complete 11 on-line tests during the semester without the presence of the lecturer. It is possible to get 1100 points all together. The students with 660 points get credit and are allowed to proceed to the final exam. The students with more than 900 points received at maximum 20 bonus points (1 bonus point for each 10 points over 900) which are added to their written test results. These bonus points are not added in the case of repeating the exam.

The final exam consists of written and oral part. The written test (in the presence of a lecturer) has two parts: theoretical (40 points) and practical, i.e. calculation of examples (60 points). The candidates with less than 50 points do not move further to the oral exam and failed. Another condition is to get at least 20 points from the

practical part of the test. In the further search for the correlations we skipped the students who did not get to the oral exam. The oral exam consists of asking two questions. The final grade is given by the lecturer after the oral exam and should summarize both the results of the written test and the oral exam.

2.2 Participant

The following analysis deals with the study results of the 278 students of second year, and of two major branches of study, which were recorded by 2 teachers–examiners. The data matrix was constructed after check up and simple analysis. It contained results of 269 participants. The factors were qualitative (branch of study, gender) and quantitative (all results evaluated by points). The set of responders consisted of 164 females, 105 males, 106 with major “Economics and Management”, and 163 with major “Business and Administration”.

2.3 Sampling Procedures

The students were chosen randomly by the authors of the study from the total number of 962 students who got the credit. For each final examination term (15 terms) was chosen a group of 10-20 students for the study. The number was dependent on the total number of students who had come to the particular term (20% of total number, rounded to natural numbers). The investigated group of students was taken from a random position of the alphabetical list of all the students for the term.

The students did not have any influence on being included into the study and the data do not include any self evaluation.

2.4 Collecting Data

The examiner (one of the authors of the study) first collected the data on the partial evaluation and bonus points from students’ information system. After that gave points for the practical part of the test. If the student got more than 20 points the teacher marked the theoretical part, added the bonus points, calculated the total result of the test, and entered all the points to the data matrix. If the result of the test was satisfying, the student was invited to the oral exam. He/she had to answer two questions out of a given list (known in advance). The teacher evaluated the answers by 1-10 points, and decided about the final grade. The points, the numbers of questions, and the final grade were recorded in the data matrix. The final grades could be 1 (the best), 2, 3, 4 (not passed).

2.5 Statistical Analysis

The statistical analysis was made by the simple and multiple dimensional statistical methods.

The simple analysis was based on describing characteristics (frequency, average, variation coefficient). The normal distribution of the sample file was tested by the Shapiro-Wilk test.

The non-parametric tests were applied to testing the hypotheses. These tests are connected with the hypotheses on the normality of the basic file without knowing its parameters. There are no (or nearly no) requirements on the nature of the investigated factors, and it is possible to use them even if the normality of distribution is disrupted (Jindrová et al., 2008).

The relation and course of the partial and final evaluation of the student was investigated using multiple dimensional statistical methods – regression analysis.

The regression analysis methods are designed for expression of dependency of quantitative continuous variable Y on one or more quantitative continuous variables X, so called regressors. It must be decided in advance which variable is independent and which is dependent.

The goal of the regression model was to describe the relations of the factors by appropriate mathematical model. Simple regression model describes the dependency of Y variable on one regressor; the multiple regressions reflect the dependency of Y on more regressors.

The multiple linear regression analysis models the dependent variable Y as a linear function of K independent variables X_1, X_2, \dots, X_K , as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K + \varepsilon,$$

where β_0 is the intercept term and $\beta_1, \beta_2, \dots, \beta_K$ are the partial regression coefficients; ε indicates random errors. Estimates of the unknown population parameters $\beta_0, \beta_1, \beta_K$ are obtained by the method of the least squares. The method of the least squares minimizes the sum of squares of the residuals. If the assumptions of linear regression are valid, the least squares estimates are unbiased estimates of the population parameters and have minimum variance (Hebák et al., 2006, Johnson & Wichern 2007).

The point estimates b_0, b_1, \dots, b_k are usually obtained from the input data by the least square method which

requires the minimal sum of squares of the differences $y_1, y_2, y_3, \dots, y_n$ of the recorded values of Y from the estimated regression function. The regression function then describes the course of dependency of Y values on the X values. The parameters b_1, \dots, b_K , called regression coefficients, reflects the average change of Y for a unit change of x_i (supposing all other parameters do not change). The regression coefficients acquires positive or negative values according the dependency is direct or indirect.

The statistical significance was tested with F-test which enables to test the model from the point of view of its predictive ability. In the case when the statistical significance of the model is proved, the model can be used for estimation of values Y based on given values of x_i .

The regression analysis is focused on an unilateral dependency; the double-sided dependency was investigated by the correlation analysis. The correlation discovers how strong the dependency between variables is. If there is a linear dependency between X and Y , it is possible to characterize the dependency by a sample correlation coefficient r , or by the Spearman's non-parametric coefficient. These coefficients could be from -1 to +1. The closer the value of the coefficient is to the 1 (in absolute value) the better quality can be reached in the estimations made by the regression model. The squared coefficient is called the coefficient of determination and it represents the important measure of tightness. Multiplied by 100, it indicates the percentage of changes that can be explained by the chosen linear regression model.

The statistical software SPSS 22 was used for the statistical analyses. The significance level for all tests was $\alpha = 0.05$.

3. Results

3.1 Descriptive Characteristics

The primary analysis of the quantitative signs was focused on basic descriptive characteristics—see Table 1.

Table 1. Basic descriptive characteristics

	<i>Descriptive characteristics</i>			
	Average	Standard deviation	Variation coefficient in %	N
Points from semester	984.33	147.46	14.98	257
Points from the theoretical part of the examination test	19.51	10.19	52.24	251
Points from the practical part of the examination test	33.97	14.86	43.75	269
Bonus points	11.13	8.77	78.77	267
Points from the oral part of the exam	7.22	6.25	86.50	269
Final grade	2.97	1.04	34.85	269

3.2 Statistics and Data Analysis

The values of the variation coefficient show that the highest variability was recorded for the results of the oral exam (86.50%). It means, the teacher saw the students' achievement as excellent or just satisfactory (not many in between). Majority of the bad results (exam failures) were eliminated by the necessity to pass the previous partial steps of the evaluation.

High variation was registered also for the bonus points (78.77 %). Only 186 out of 269 evaluated students exceeded the limit of 900 points from semester and got some bonus points. This fact is caused by the gap between the minimum points for the credit, 660, and the minimum points for the bonus points 900. Many of average students got to the zero zone, i.e. they worked enough to get the credit but not enough to get the bonus points. The bonus points again discriminated between excellent and satisfactory activity during the semester.

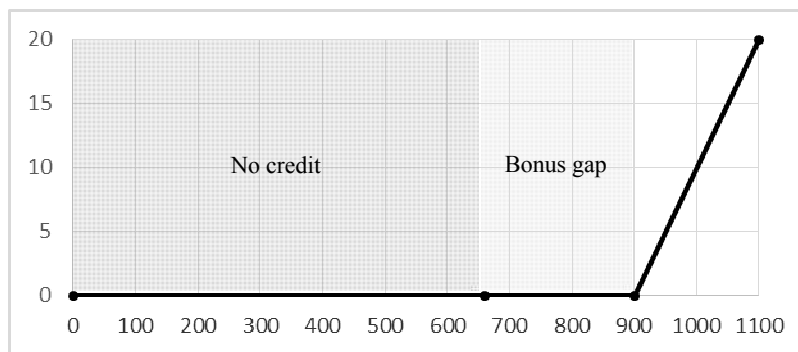


Figure 1. The points from the semester and the bonus points

The lowest variability was for the points from semester (15%). The lower difference was caused by no gap in the evaluation score. All the investigated students got to the exam, it means they all got the credit. They fulfilled their duties during the semester to get the credit; usually they exceed the minimum number of points (the average is slightly over 900 see Table 1) to get the credit for sure, to have some reserve.

For the same factors, the Shapiro-Wilk test of normal distribution followed. None of the factors manifested the normal distribution (see Table 2). None of the p-values exceeded the chosen level of significance 0.05.

The basic prerequisite for the parametric test (i.e. the normality of data) was not met, so that the Mann-Whitney test, which is a non-parametric variant of the t-test for two independent selections, was applied for testing the hypotheses.

Table 2. Results of the normality test

	<i>Shapiro-Wilk</i>		
	Statistic	df	p-value.
Points from semester	0.755	257	0.000
Points from the theoretical part of the examination test	0.981	251	0.002
Points from the practical part of the examination test	0.978	269	0.000
Bonus points	0.759	267	0.000
Points from the oral part of the exam	0.902	269	0.000
Final grade	0.822	269	0.000

Table 3. Results of the Mann-Whitney test

	<i>Gender</i>		<i>Branch of study</i>	
	Statistic	p-value	Statistic	p-value
Points from semester	6661.5	0.136	6911	0.275
Points from the theoretical part of the examination test	7661.5	0.127	8291	0.577
Points from the practical part of the examination test	6727	0.058	7097	0.164
Bonus points	6948.5	0.009	7496.5	0.083
Points from the oral part of the exam	6661.5	0.136	8100.5	0.380
Final grade	7273	0.290	6911	0.275

The results in the Table 3 show that the difference between genders was only in the number of bonus points. While the average number for the female students was 12.19; the average for male students was 9.51. The female students had worked harder during the semester and came to the exam with more bonus points. No other differences between the genders were proved.

There were no evidences of differences between the branches of study. All the p-value was over the level of significance ($p > 0.05$; see Table 3).

The hypotheses H_1 and H_2 were confirmed. The branch of study and the gender do not have important influence on the partial evaluation of the students neither on the final grade.

The correlation and regression analyses were focused on investigation of the strength and process of the dependencies. The reduction of the number of factors has been applied with respect to the goal of the study: to determine which partial evaluation has the strongest influence on the final grade. The factor "Points from semester" was removed because it does not have a direct relation to the final grade. It follows from the method of the students' evaluation that this factor is contained in another factor "Bonus points".

As mentioned above, the exploratory analysis of variables (Shapiro-Wilk test) showed that majority of investigated data do not have normal distribution. That was the reason for application of the Spearman correlation coefficient for explaining the linear dependency between the response variable (final grade – y) and explanatory variables (Points from the theoretical part of the examination test – x_1 ; Points from the practical part of the examination test – x_2 ; Bonus points – x_3 ; Points from the oral part of the exam – x_4).

The pair coefficient matrix (see Table 4) show the strongest negative correlation the variable x_4 – Points from the oral part of the exam ($r_s = -0.739$) and x_2 – Points from the practical part of the examination test. The negative direction of the dependency follows from the indirect relation between the factors; the better (the lower) the final grade is the higher is the number of points in the partial evaluation.

Table 4. Values of the Spearman's correlation coefficient

	<i>Final grade (y)</i>	x_1	x_2	x_3	x_4
y	1.000				
x_1	-0.523	1.000			
x_2	-0.714	0.283	1.000		
x_3	-0.345	0.141	0.253	1.000	
x_4	-0.739	0.342	0.100	-0.006	1.000

The model was in the beginning evaluated by the F-test with the result that the model as a whole had a statistical significance (tested even on the level $p < 0.0001$).

The regression model has been constructed using the estimated parameters of a linear function:

$$y = 4.764 - 0.013 x_1 - 0.018 x_2 - 0.012 x_3 - 0.108 x_4$$

The parameters b_1, \dots, b_k , (or the regression coefficients) express the average change of the response variable y in relation to a unit change of one explanatory variable x_i supposing all other parameters will not change. The parameters are positive or negative according to the fact that the dependency is direct or indirect. According to the model, the strongest influence on the final grade has the result of the oral exam ($b = -0.108$), followed by the practical part of the examination test ($b = -0.018$). This is in line with the results of the correlation coefficients which were the highest for the same factors (and negative).

The value of the coefficient of determination (R-square) states that the constructed regression model is able to explain the changes in the response variable in 82%.

The study confirmed that there is no influence of gender and branch of study (with the exclusion of gender differences for the number of bonus points) to the final grade. The investigation dealt also with the relation between the final grade and the partial evaluation during the semester and by the examination test. The detailed analysis of the explanatory variable has been made leading to the construction of a regression model. The research proved that the final grade is mostly dependent on the oral exam, i.e. the face to face evaluation of the student is deciding for passing or failing the exam, and the partial evaluation including on-line testing does not

replace the deciding role of the teacher.

4. Discussion

The subject is generally very difficult. The average grade 2.97 (with relatively high deviation of 1.04) indicates that many students failed or got the worst “passing” grade 3 (grades 1, 2, 3 mean passed, 4 means failed). In addition it is taught in the second year of study and can be a reason of school dropouts. There might be many other factors which influenced the results, but bad results in the mathematical subject in the beginning of studies can be taken as an important predictor of being not successful in the following years.

The group of investigated students was relatively homogenous because there were only students who got to the final exam. The poor students did not get credits and were not allowed to sit the final exam. These poor students have to repeat the subject next year, or to repeat the whole year (if they failed in more than 2 subjects). The school dropout is more probable for them in the current or in the following year. The high dropout rates in Western countries sharply contrast with the social and economic objectives and that is the reason why the potential predictors of non-graduation have generally been looked for (de Witte et al., 2013). The connection between dropout and unemployment is ambiguous (Dorn, 1996 in de Witte et al., 2013). On the other hand important differences were recorded by Rumberger and Lamb (2003). Almost a third of all dropouts who had finally completed high school were not working two years after high school, compared to 45% for dropouts who had never completed and only 8% for those who had never dropped out. Anyway there are definitely remarkable losses of the expenditures which society invests into unfinished education or repeated years. That is why the distinguishing of the predictor of school leaving is important. The poor academic ability manifested in the grade retention, sometimes bracketed together with an accumulation of credit deficits may be the strongest predictor (de Witte et al., 2013).

The students were homogenous bearing of skipped in the branch of study. No differences between the results were found and the hypothesis H2: “The branch of study has no influence to the final grade”, was confirmed.

The hypothesis H1: The gender has no influence to the final grade, was confirmed as well. There was a difference only in the work during the semester when the females were better than males. This finding is in line with the results of other authors (Falch & Naper, 2013; Machin & McNally, 2005). The subject is based on mathematics but the stereotype saying the girls are worse in maths was not confirmed. It is well established that negative stereotypes can undermine women's performance on mathematics tests but according to the data from the most difficult mathematics courses it is strongly suggested that women can even surpass men (Good et al., 2008). Bonnot and Croizet (2007) showed that women, who endorsed the stereotype of their group's inferiority in math relative to men, had a lower self-evaluation in math and lower grades. The lower self-evaluation was then more harmful than the gender stereotype. The math self-evaluation can be a goal of a following investigation.

The males in our investigation got the same results of the final exams as the females more often without (or with less) bonus points from the semester. It means they were more focused to the final exam or the examining method was more suitable for them: practical part of the test (it means calculations) and the oral exam that had a crucial influence on the final grade.

The oral exam finally appeared as the most important part of the whole evaluation. It is beneficiary to the students who prefer this method of evaluation. They are stable persons by Furnham, Batey, and Martin (2011), open personalities by Chamorro-Premuzic et al. (2005). It was found Chamorro-Premuzic et al. (2005) that males tended to have more positive attitudes towards written examinations than did females, whilst self-assessed intelligence was positively related to preference for written exams (notably in male students).

There are no doubts that the oral exam is the part of the evaluation which is to the highest degree influenced by the personality, experience, and all kinds of short time feelings or imperfections of the teacher. Teachers' grading styles are not necessarily uniform, but may be affected by personal-professional characteristics (e.g., gender, seniority, cultural background) (Resh, 2009; DeBoer et al., 2007) and other socio-cultural factors (Doran, Lawrenz, & Hegelson, 1994). Bonesrønning (2004) identifies teachers' grading as a potential teachers' tool by which student effort can be manipulated; students who are exposed to hard grading perform significantly better than students who are exposed to easy grading. Teachers do not always assign grades based on the achievement only. They either explicitly (i.e. effort on a rubric) or implicitly (homework completed, class participation) assign grades including the effort criteria and often also the proper behaviour (Randall & Engelhard, 2010).

The question is whether the partial evaluation is useful while the oral face to face exam is the most important for passing or failing the exam.

The examination scheme for the investigated subject is rather complicated; having oral exams only would be much easier. Important weak points are connected with evaluation based on oral exams only:

There is a high number of students (about 1000 with possible 3 attempts), the detailed and through going oral exam is highly demanding in time. The teacher usually cannot spend more than 10-15 minutes with one student and the examining teacher is not the seminary teacher (does not know the student personally from semester). This may harm students who have worked conscientiously during the semester but not able to utilize all their knowledge in a limited time under given conditions.

The subject is based on mathematical applications; the ability to calculate and solve practical examples cannot be fully proved at oral exam.

The evaluation by one teacher in one moment is not objective enough.

The motivation and willingness to work (even the presence to lectures and seminars) during the semester will not be high without partial evaluation.

Due to above mentioned reasons we do not recommend to skip any part of the evaluation scheme.

For improvement of the school effectiveness and economy it would be useful to find out clear indicators of the academic insufficiency of the students. As the subject is in the first second, the results can help prediction of the dropouts in the following years. The comparison with the results of other subjects and with other outputs and evaluations of the students could be a goal of the further research focused on possible future study success or failure.

Acknowledgements

The author gratefully acknowledges the support from the Faculty of Economic and Management, Czech University of Life Sciences, via IGA grant, no. 11170/1312/3164.

References

- Beleche, T., David, F., & Marks, M. (2012). Do course evaluations truly reflect student learning? Evidence from an objectively graded post-test. *Economics of Education Review*, 31(5), 709-719. <http://dx.doi.org/10.1016/j.econedurev.2012.05.001>
- Betts, J., & Grogger, J. (2003). The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review*, 22(4), 343-352. [http://dx.doi.org/10.1016/S0272-7757\(02\)00059-6](http://dx.doi.org/10.1016/S0272-7757(02)00059-6)
- Bonesrønning, H. (2004). Can effective teacher behavior be identified? *Economics of Education Review*, 23(3), 237-247. <http://dx.doi.org/10.1016/j.econedurev.2003.07.002>
- Bonnot, V., & Croizet, J.-C. (2007). Stereotype internalization and women's math performance: The role of interference in working memory. *Journal of Experimental Social Psychology*, 43(6), 857-866. <http://dx.doi.org/10.1016/j.jesp.2006.10.006>
- Chamorro-Premuzic, T., Furnham, A., Dissou, G., & Heaven, P. (2005). Personality and preference for academic assessment: A study with Australian University students. *Learning and Individual Differences*, 15(4), 247-256. <http://dx.doi.org/10.1016/j.lindif.2005.02.002>
- De Witte, K., Cabus, S., Thyssen, G., Groot, W., & van den Brink, H. (2013). A critical review of the literature on school dropout. *Educational Research Review*, 10, 13-28. <http://dx.doi.org/10.1016/j.edurev.2013.05.002>
- DeBoer, B. V., Anderson, D. M., & Elfessi, A. M. (2007). Grading styles and instructor attitudes. *College Teaching*, 55(2), 57-65. <http://dx.doi.org/10.3200/CTCH.55.2.57-64>
- Doran, R. L., Lawrenz, F., & Hegelson, S. (1994). Research on assessment in science. In D. L. Gabel (Ed.), *Handbook of research of science teaching and learning*. New York, NY: MacMillan Reference Books.
- Ewing, A. M. (2012). Estimating the impact of relative expected grade on student evaluations of teachers. *Economics of Education Review*, 31(1), 141-154. <http://dx.doi.org/10.1016/j.econedurev.2011.10.002>
- Falch, T., & Naper, L. R. (2013). Educational evaluation schemes and gender gaps in student achievements. *Economics of Education Review*, 36, 12-25. <http://dx.doi.org/10.1016/j.econedurev.2013.05.002>
- Furnham, A., Batey, M., & Martin, N. (2011). How would you like to be evaluated? The correlates of students' preferences for assessment methods. *Personality and Individual Differences*, 50(2), 259-263. <http://dx.doi.org/10.1016/j.paid.2010.09.040>

- Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology, 29*(1), 17-28. <http://dx.doi.org/10.1016/j.appdev.2007.10.004>
- Grant, D. (2007). Grades as information. *Economics of Education Review, 26*(2), 201-214. <http://dx.doi.org/10.1016/j.econedurev.2005.10.004>
- Hebák, P., Hustopecký, J., & Malá, I. (2006). *Vícerozměrné statistické metody [2]*. Prague, Czech Republic: Informatorium.
- Hinnerich, B. T., Höglin, E., & Johannesson, M. (2011). Are boys discriminated in Swedish high Schools? *Economics of education Review, 30*(4), 682-690. <http://dx.doi.org/10.1016/j.econedurev.2011.02.007>
- Jindrová, A., Prášilová, M., & Zeipelt, R. (2008). *Statistika I*. Prague, Czech Republic: Reprografické studio PEF ČZU v Praze.
- Jindrová, A., Vostrá Vydrová, H., & Dömeová, L. (2013). The students' evaluation of the e-support in the full time form of study. *Journal on Efficiency and Responsibility in Education and Science, 6*(2), 119-133. <http://dx.doi.org/10.7160/eriesj.2013.060205>
- Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis*. Upper Saddle River, N.J.: Pearson Prentice Hall.
- Krautmann, A. C., & Sander, W. (1999). Grades and student evaluations of teachers. *Economics of Education Review, 18*(1), 59-63. [http://dx.doi.org/10.1016/S0272-7757\(98\)00004-1](http://dx.doi.org/10.1016/S0272-7757(98)00004-1)
- Lai, F. (2010) Grades and students evaluations of teachers. *Economics of Education Review, 18*(1), 59-63. <http://dx.doi.org/10.1016/j.econedurev.2009.07.009>
- Lakhal, S., Sévigny, S., & Frenette, É. (2013). Personality and preference for evaluation methods: A study among business administration students. *Studies in Educational Evaluation, 39*(2), 103-115. <http://dx.doi.org/10.1016/j.stueduc.2013.02.002>
- Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics, 92*(10-11), 2083-2105. <http://dx.doi.org/10.3386/w10678>
- Niederle, M., & Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives, 24*(2), 129-144. <http://dx.doi.org/10.1257/jep.24.2.129>
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education, 26*(7), 1372-1380. <http://dx.doi.org/10.1016/j.tate.2010.03.008>
- Resh, N. (2009). Justice in grades allocation: Teachers' perspective. *Social Psychology of Education: An International Journal, 12*(3), 315-325. <http://dx.doi.org/10.1007/s11218-008-9073-z>
- Rumberger, R. W., & Lamb, S. P. (2003). The early employment and further education experiences of high school dropouts: A comparative study of the United States and Australia. *Economics of Education Review, 22*(4), 353-366. [http://dx.doi.org/10.1016/S0272-7757\(02\)00038-9](http://dx.doi.org/10.1016/S0272-7757(02)00038-9)

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).