

Does grading bias affect indigenous college students in Mexico? Evidence from a field experiment

Abstract

Overcoming ethnic grading bias is one of the many challenges education systems may face in addressing educational inequities. Based on the results from a field experiment, we show that indigenous students in a higher education institution in Mexico obtain lower grades than nonindigenous students with similar academic performance, with a negative difference equivalent to 0.26 s.d. Our findings suggest that stereotypes and social prejudices among faculty are the main explanatory factors after other individual characteristics are controlled. Options to reduce grading bias are examined in the final section of this article to suggest potential interventions to overcome this problem.

Introduction

Evidence on ethnic grading bias in Mexico is still scarce, even though available research indicated that indigenous students are discriminated against in higher education institutions.¹ Discrimination against students is usually associated with race, language, and cultural characteristics.² Among the specific discriminatory practices toward indigenous students observed in educational institutions, researchers have documented differential treatment from educational administrators (Pérez 2012), a lack of integration or involvement of students in academic activities (Carnoy et al. 2002), and even increased opacity regarding students' assessments (Baltazar 2012).

Identifying whether grading bias against indigenous students is observed in higher education institutions is relevant since this group is usually underrepresented, mainly in terms of their socioeconomic status. This article aims to identify whether college instructors show ethnic bias in the assessment of the academic performance of minority students in a higher education institution in Mexico. In addition, it seeks to explore how different factors suggested by previous conceptual studies might be associated with lower expectations of students from disadvantaged backgrounds, including indigenous backgrounds (e.g., Jussim and Harber 2005; Sprietsma 2009; Rema and Leigh 2012).

To address our main research question, we conducted a field experiment in which instructors were invited to grade six equivalent essays, two of which were attributed to fictitious indigenous students. By grading these comparable essays and collecting information on the instructors' personal characteristics, we were able to estimate the magnitude of grading bias and

¹ Chávez 2008; Conrado and Sánchez 2011; Baltazar 2012; Pérez 2012; Castro and Santos 2015

² Carnoy et al. 2002; Gallart and Henríquez 2006; Chávez 2008; Casillas 2012

identify variations in the potential impact of different discriminatory practices identified in previous research on the assessment of students' academic performance.

This article is divided into three sections. The first section presents the main findings from the literature review on grading bias and other discriminatory practices observed in educational institutions. In the second section, we describe our analytical strategy and main findings, pointing out some of the decisions we made in conducting our field experiment. Finally, in the last chapter, we present our main conclusions, as well as possible policy options and potential future research.

Prior studies

Grading bias affecting minorities has been reported across different countries and educational levels. For instance, Rema and Leigh (2012) found that Indian instructors graded members of lower castes differently than those from higher castes with similar academic performance, with a difference varying from -0.03 to -0.08 s.d. Another study reported a similar situation in Switzerland, where instructors favored national over foreign students, with a difference equivalent to - 0.20% s.d. (Tyrefors et al. 2011). In Germany, national students obtained, on average, higher grades than students of Turkish origin (0.10% s.d.) (Sprietsma 2009), confirming the existence of grading biases based on national origin. Additionally, Botelho et al. (2013) found that in Brazil, Afro-descended students received lower grades than students of other ethnic origins with similar academic performance, with a difference of -0.04 to -0.08 s.d.

More than 50 years ago, Rosenthal and Jacobsen (1968) described the "Pygmalion effect"—or the self-fulfilling prophecy—a concept used to describe a process by which instructors' expectations regarding students' abilities effectively determine students' academic

performance. This effect is explained by a differential treatment in terms of access to pedagogical content, opportunities for participation, the type of response, and the socioemotional climate, among others.³ This effect may affect not only students' academic performance but also their intellectual development, self-conception, and motivation (Trouilloud and Sarrazin 2002).

Multiple factors influence instructors' expectations of students (Jussim and Harber 2005), such as instructors' preconceptions regarding students' academic performance, behaviors, and motivations (e.g., Jussim et al. 1994; Trouilloud and Sarrazin 2003; Trouilloud et al. 2016), as well as information obtained from tests administered during application processes, grades, or even feedback from previous instructors.⁴ Other factors influencing instructors' perceptions of students include early course assessments, instructors' insights into students' behavior, and instructors' observations of students' commitment during teaching sessions, as well as instructors' perceptions of the support students receive from their own families.⁵ Another factor that has been reported is the stereotypes and prejudices that instructors develop (Trouilloud and Sarrazin 2002), which are apparently associated with instructors' own characteristics, such as their professional experience, gender, personality, perceived self-efficacy, dogmatism, and authoritarianism (Jussim et al. 1994; Rosenthal 1994; Trouilloud and Sarrazin 2003).

Factors related to the organization of the school, such as the class size, subjects, and instructional practices, also affect learning interactions. In addition, similarities between instructors and students are important. Factors such as socioeconomic background, ethnicity, and opportunities to participate in professional development activities all have a potential impact on

³ Williams 1976; Rosenthal 1994; Jussim et al. 1996; Trouilloud and Sarrazin 2002; Jussim and Harber 2005; Tyrefors et al. 2011

⁴ Cooper 1985; Jussim 1986; Good 1997; Trouilloud and Sarrazin 2002; Jussim and Harber 2005

⁵ Jussim 1986; Good 1997; Good and Brophy 1997; Trouilloud et al. 2002; Trouilloud and Sarrazin 2003; Jussim and Harber 2005; Ragland 2012

expectations (Jussim 1986; Jussim et al. 1994; Trouilloud and Sarrazin 2003). In addition, low social awareness among instructors may result in differential treatment of students, either by gender or by ethnic origin (Dickman 1993; Aksoy 2007; Murillo and Hidalgo 2015). In Table 1, we summarize some of the main mechanisms through which the factors described in this section may result in grading bias.

Analytical strategy

To assess whether grading bias affects indigenous students in Mexico, we conducted a field experiment. This methodology allowed us to capture the environment in which instructors regularly work and make decisions, which is a methodological advantage (Coates 1972; Jacobson and Effertz 1974; DeMeis and Turner 1978).

To conduct this experiment, we replicated the selective admission process of a higher education academic program, which is a process commonly implemented in Mexico. To detect and estimate the magnitude of the expected grading bias, we first obtained “contrast grades” by inviting 20 nonindigenous higher education students to write and submit an academic essay answering one open question that was drawn from items included in a national standardized test. These students were enrolled in a higher education institution located in Mexico City, and their participation was voluntary; the students were offered only a symbolic award for the student with the best submitted essay. None of the students were informed about the characteristics of our study to avoid any potential bias affecting their writing and answers.

Once we obtained the 20 written essays, we then selected and invited 12 external evaluators, all of whom enrolled graduate students, to grade the essays on a scale from 1 to 10 points, which is a commonly used scale in Mexican education institutions. In the

evaluation process, we did not provide the external evaluators with any information on the authors or the goals of our study; we provided them with only a grading rubric to define common evaluation standards. When we had received 240 different grades, we conducted an interrater reliability test (Ripoll 2014; Vituri and Évora 2014) to explore multiple combinations of evaluators and grades to identify the set of six academic essays, grades, and evaluators with the highest Landis and Koch index value; we ultimately selected a set with a value of 0.83.⁶

The higher education institution where our experiment was conducted is located in a region of the country where a large indigenous population resides. According to the surveyed instructors, approximately one-third of their current students were identified as indigenous. A large proportion of indigenous students was one of the criteria used to select this institution since it was expected that the instructors would be familiar with teaching indigenous students.

The participation of the instructors in this experiment was voluntary. With support from the director of this institution, we distributed an invitation to participate in an exercise to evaluate essays submitted during a selective admission process to an anonymous higher education institution. To guarantee that assessments by the instructors would be conducted properly, we informed the instructors that their collaboration would be essential for estimating the validity of the test for adequately selecting potential college students. We invited the instructors to attend any of the six sessions we organized to grade essays in situ, which we believed would guarantee that all academic essays would be graded by instructors working in a similar environment. These evaluation sessions were held in classrooms located in the same institution.

⁶ Identifying a set of comparable tests using the intraclass correlation coefficient was a key component of this experiment since it allowed us to assign the same cover page to the same test in all cases. A random distribution of covers (with the same test sometimes having a cover page with an indigenous name and other times having a cover page with a nonindigenous name) would have required a greater number of evaluators to ensure the validity of the results of the experiment.

In each of these sessions, we distributed the same set of six comparable essays we had previously selected based on the Landis and Koch index to all of the participating instructors. A cover page with prefilled data was attached to each of the essays to be graded that included information about the fictional students (full name, language, and gender). For the cover pages depicting fictitious indigenous students, the given and last names were chosen from a national catalog of indigenous names (Gálvez and Embriz 2006), and the appropriate selection of names was validated through a local instructor who spoke the language of the predominant indigenous group in the region.

The nonindigenous student names were chosen based on the most frequent names reported in the voter list statistics for the state where the higher education institution is located that were published by the National Electoral Institute (2015). Regarding language, for the cover pages with nonindigenous student names, Spanish was reported as the mother tongue, while for the cover pages with indigenous names, the language spoken by the largest indigenous group in the region was reported. Finally, regarding gender, we included only male fictitious students to avoid variations in expectations that might be attributable to gender differences. Two of the six cover pages had indigenous student names (treatment group), and four had nonindigenous names (control group). We also included a grading rubric to guarantee that every instructor used the same evaluation criteria.

Finally, to prevent any bias due to the order in which the instructors graded the six essays (e.g., being stricter at the end or beginning of the grading process), tests with cover pages with indigenous names were randomly ordered within the set of six essays to be reviewed.

In addition to grading the essays, the instructors responded to a context questionnaire designed to collect information on their perceptions of four different aspects: a) factors related to

information about students' academic performance (grades from previous school periods, admissions exams, comments from other instructors, the first exams of the period, good behavior, motivation and effort, students' families attendance of university and families' interest in students' performance); b) factors related to prejudices and stereotypes (personal presentation, intellectual potential, customs, accent, use of traditional dress, expectations regarding academic performance and behavior), including a subgroup of questions related to affective or emotional aspects (Ruiz 2012); c) factors indirectly affecting instructor expectations (personal perceptions, self-image of instructors and perceived self-efficacy); and d) other instructor characteristics (academic achievement, college major, participation in academic mobility programs, and social awareness).

Estimating grading bias

Once the data were collected, Student's t-test and three different linear regression models were conducted to identify and estimate the magnitude of grading bias potentially affecting the fictitious indigenous students.

In our first model, Y_{ie} is a continuous variable representing the difference between the grades given by the 52 instructors i to each of the six fictitious students and the contrast grades. The "Group" variable indicates whether an essay belonged to the control or treatment group (0 = cover page with a nonindigenous name, 1 = cover page with an indigenous name).

$$Y_{ie} = \alpha_0 + \beta_1(\text{Group}) + \epsilon_{ie} \quad (1)$$

Model two is a multiple linear regression designed to identify whether any effect remains significant after the essay order during the grading sessions is controlled. The "Test" variable is a

continuous variable indicating the order in which essays were distributed (one to six), while “Prof” is a unique identifier assigned to each instructor (1 to 52).

$$Y_{ie} = \alpha_0 + \beta_1(\text{Group}) + \beta_2(\text{Test}) + \beta_3(\text{Prof}) + iui \quad (2)$$

Model three was designed to verify whether any treatment effect remains statistically significant after the characteristics of the instructors are controlled. If the effect found with β_1 remains and the covariance (β_1, β_j) tends to zero, then the grading bias does not depend on the included observable variables.

$$Y_{ie} = \alpha_0 + \beta_1(\text{Group}) + \beta_2(\text{Gender}) + \beta_3(\text{Experience}) + \beta_4(\text{SES}) + \beta_5(\text{Achievement}) + \beta_6(\text{Major}) + iui \quad (3)$$

Causal mechanisms

Unlike the first three models, which we designed to focus on identifying a treatment effect, the following models were designed to explore whether different sets of direct, indirect, and modulatory factors are related to the observed treatment effect. These factors were identified in our literature review. To identify potential associations, we conducted a multiple linear regression analysis for clustered data since our database contained six observations for each instructor. In our model, we grouped data into 50 clusters since only 50 of the 52 instructors responded to the context questionnaire.

$$Y_{ie} = \alpha_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 \quad (4)$$

In this model, Y_{ie} is a continuous variable representing the contribution of each essay to the general grading bias.⁷ To estimate individual contributions to the general grading bias, the

⁷ General bias was disaggregated into 312 values of individual contributions representing each graded essay.

difference between the contrast and given grades for each essay was divided by the number of observations corresponding to either the control or treatment group.

Since differences in the treatment group increased the overall grading bias in a different direction than differences in the control group, individual differences were multiplied by -1.⁸ In this way, the sum of the observations for each variable was equivalent to the bias observed in the overall experiment, thus representing the individual contribution of each variable. To facilitate interpretation, data were represented on a scale from 1 to 10 points, which is a common grading scale in Mexico. Due to the large number of potential covariates based on our theoretical framework, additional independent variables were selected based on the results of 10 different regression models. These covariates were grouped into three categories: direct, indirect, and modulatory factors.

Direct factors

X_1 is a vector of eight factors corresponding to the instructors' perceptions regarding student behavior and academic performance: grades from previous school years, the results from admissions processes, feedback provided from other faculty, initial assessments during the course, discipline, motivation, effort, and parental involvement. Similar to other variables measuring agreement levels, these factors were represented on a 5-point Likert scale. As will be later explained, only two of the seven explored variables influenced grading bias, which is with previous research and theories.

Indirect factors

⁸ We adopted a difference-in-differences approach, considering the contribution of each observation to general grading bias.

A vector of this group of factors (variables capturing the instructors' personal prejudices and stereotypes) is represented by X_2 . This vector includes five variables representing characteristics of students' families, variables representing the instructors' perceptions of stereotypes and prejudices, and a variable representing affective attitudes (Ruíz 2012). These variables were constructed in a similar way to the variables of the direct factors (5-level, categorical), except for the last variable, which was represented as a three-level categorical item.

Modulatory factors

X_3 is a vector representing the instructors' personal characteristics potentially influencing their expectations of students. It includes variables such as gender and teaching experience.

Additionally, a socioeconomic status index and a measure of parents' academic achievement were included since these factors are commonly used in studies in Mexico. Indigenous language was not included in this estimation since only one instructor reported being able to speak an indigenous language. In addition, as a proxy for self-efficacy, variables measuring the instructors' perceptions of their teaching effectiveness and their pride in their profession were included.

Additional factors

Vector X_4 includes other potential explanatory factors, such as academic achievement and college major. We also explored whether the instructors participated in mobility programs. Finally, we included five variables to capture the instructors' beliefs about whether discrimination against indigenous people is a current problem, their history of witnessing discriminatory practices based on ethnicity, their perceptions of the quality of education that

indigenous students usually have access to, and their beliefs about whether assessment criteria should consider the cultural differences of ethnic minorities. Although these factors could also be considered modulatory, there are limited references in the available literature on their potential effects. Therefore, before these variables were included in the general model, their associations with the dependent variable were verified separately and then jointly with other modulatory factors.

Main findings

As mentioned, 52 instructors participated in this experiment, with each instructor grading six essays, for a total of 312 graded essays. A total of 104 of these essays belonged to the treatment group (cover page with an indigenous first language and given name), while 208 were considered part of the control group (cover page with a nonindigenous name and Spanish as the first language). To estimate the potential effect of the instructors' bias, first, a Student's t-test was conducted (Table 2). A comparison of the mean differences in grades between the control and treatment groups helped identify a statistically significant effect of - 0.50 points for the group of fictitious indigenous students ($p= 0.01$). In other words, essays with cover pages with indigenous names were given, on average, 0.50 fewer points than essays with cover pages with nonindigenous names, which was equivalent to 26% of the standard deviation. Although this effect could be considered small on a grading scale of 0 to 10 points, this difference would be enough to affect a student's academic trajectory by causing him or her to fail a course or lose a scholarship.

These initial results from our Student's t-test support findings from previous research. These results also indicate that the grades given by the instructors who participated in the

experiment to both indigenous and nonindigenous students were higher, on average, than the contrast scores. That is, in general, the instructors were more generous than the external evaluators who provided the contrast grades. However, the instructors who participated in the experiment were more generous when grading essays from the fictitious nonindigenous students, as the nonindigenous student experimental grades were an average of 0.91 points higher than the nonindigenous student contrast grades, while the indigenous student experimental grades were an average of only 0.41 points higher than the indigenous student contrast grades. These differences represent a bias affecting the fictitious indigenous students. The results included in Table 3 help to explain some of the associations between grading bias and instructor characteristics.

As explained, we aimed to identify other mechanisms explaining grading bias based on previous research. In model four, we grouped all these factors into different categories, with the results described in Table 4. Our main findings are summarized below.

a) Perceptions of students' performance and motivation

Previous research suggests that some students' behavior and characteristics may predict academic performance (Trouilloud and Sarrazin 2003). Seven factors representing perceptions of students' performance were included in model four to explore their potential associations with grading bias. Of these factors, only two were positively associated with grading bias: scores from previous tests (0 .28) and parental attendance of college (0 .27). One possible explanation for these associations is that instructors might interpret both factors as reliable predictors of better academic performance. These factors may also reflect some stereotypes regarding indigenous students' trajectories as low performers and their limited cultural capital in their families.

b) Stereotypes and prejudices

Factors included with the aim of determining how stereotypes and prejudices may influence instructors' perceptions and grades were strongly associated with bias in our model. This finding is consistent with previous research that pointed out how stereotypes contribute to the overstatement of differences between population groups (Jussim et al. 1996; Trouilloud and Sarrazin 2003; Jussim and Harber 2005). Despite the potential social desirability problems of our methodology, the instructors' perceptions of the personal presentation of indigenous students were associated with grading bias due to ethnicity (-.26). In other words, there was an association between the instructors' perceptions of indigenous students' personal presentation and the instructors' interpretation of their presentation as an indicator of a lack of academic commitment. The greatest association we observed in this category, however, was with the instructors' nonexplicit attitudes (Ruíz 2012), which were expressed as their willingness to live with indigenous people (-.73). This finding suggests that at least part of the grading bias was explained by the instructors' implicit traits or unconscious decisions (Spruietsma 2009). One explanation for the role of implicit traits is that a person may consciously disapprove of stereotypes or prejudices but unintentionally have a different attitude regarding specific characteristics of students (Ruíz 2012). Finally, the observable characteristics of students' families, such as the student's parents being legally married (-0.16), seemed to influence the instructors' perspectives. One possible explanation for this association is that instructors may positively value this type of "traditional" family setting, but more research is needed to understand this connection.

c) Factors regulating expectations

We found statistically significant associations for three different factors related to the instructors' characteristics. Gender was one of the variables that explained grading bias, with women instructors showing less bias (-0.59). As mentioned, it is unclear how gender may affect academic expectations and therefore result in grading bias, but some studies have suggested that women may be more optimistic about the academic performance of ethnic minorities (Gershenson et al. 2015).

Teaching experience reduced grading bias (0.012). A possible explanation of this relationship is that instructors increase their skills and competencies over time and therefore improve their performance of student assessments (Trouilloud and Sarrazin 2003). However, even if experience results in a better alignment of expectations and performance, it may not be enough to avoid the amplification of expected differences across population groups, particularly when expectations result from stigmatization, as appears to be the case for unconscious behavior (Chu and Huan 2015). Finally, we found a statistically significant association of grading bias with self-efficacy. These results are consistent with previous research, showing that greater perceived self-efficacy reduces grading bias (-.28), perhaps because of instructors having better expectations of minority students' performance (Trouilloud and Sarrazin 2003).

d) Other factors

Educational attainment was another factor that was found to be negatively associated with grading bias (-.37). Although there is no consensus on this association, this result is consistent with the literature indicating that schooling can reduce bias or discriminatory attitudes towards students (Clotfelter et al. 2007; Moreno et al. 2011; Chu and Huan 2015).

Our model also identified associations related to college majors. For instance, having a major in arts and humanities (-.85), agronomy (-1.2), or health (-1.6) was negatively associated with ethnic grading bias. This finding is consistent with previous research suggesting that studying humanities may provide instructors with a humanistic perspective, helping them avoid engaging in unfair discrimination (ILO-UNESCO 2008).

Regarding the association of college majors in agriculture and health with grading bias, one possible explanation could be that internships or professional experience, particularly interaction with the neediest populations, may help instructors develop a greater sensitivity to social inequalities. An additional factor was related to participation in academic mobility in college. Having experiences living or studying abroad increased grading bias (0.35). This finding, however, contradicts previous research,⁹ which suggests that exposure to other cultures and values does not always result in a broader appreciation of native cultures. However, this contradictory finding requires validation based on larger populations.

An important additional factor that was explored was social awareness. Our findings align with previous research, as the association of grading bias with variables related to the instructors' awareness of social inequalities was statistically significant. For instance, an awareness of discrimination based on skin color in one's institution was associated with a reduction in grading bias (-0.16). Similarly, potential agreement with the assumption that indigenous students' the previous educational trajectory was of adequate quality increased grading bias (.24), indicating that when the instructors perceived that students had been less exposed to educational inequalities in previous educational levels, the effect of grading bias of grading bias was more likely to increase.

⁹ Gacel 2015; Jacobone 2015; Maldonado 2015; Reisberg 2015

Conclusions

This study aimed to contribute to the literature by providing evidence on factors that reproduce social inequalities within higher education systems. Through a field experiment, we showed how the instructors' ethnic grading bias affected fictitious indigenous students, whose essays, on average, received lower grades than equivalent essays from nonindigenous students.

These results update previous findings on grading bias reported in other countries and for different educational levels.¹⁰ In addition, we found that higher education instructors' stereotypes and prejudices are key explanatory factors, a finding also consistent with previous research (Jussim et al. 1996; Trouilloud and Sarrazin 2003; Jussim and Harber 2005).

Our experiment also confirmed that other instructor characteristics are associated with ethnic grading bias. For instance, on average, a female instructor is likely to give higher grades to students considered ethnic minorities than male instructors (Gershenson et al. 2015), and teaching experience is also negatively associated with grading bias. In addition, self-efficacy is also related to a reduction in grading bias. These findings may confirm the importance of the influence of instructors' perceptions on the formation of expectations and the unintended effects of these expectations on instructors' behavior while grading students' work (Chu and Huan 2015). Furthermore, the study suggests that onboarding processes could be key activities for informing and sensitizing novice instructors regarding how their own perceptions could affect ethnic minorities.

This study also has several implications for public policy and future research. First, it suggests the urgency of analyzing and estimating the extent to which grading bias is present

¹⁰ Sprietsma 2009; Tyrefors et al. 2011; Rema and Leigh 2012; Botelho et al. 2013

across higher education institutions, including institutions currently training future instructors. Furthermore, the study indicates that it is necessary to identify and estimate the magnitude of grading bias based on not only ethnicity but also other factors, such as gender, skin color, or any other personal trait.

A second implication for policy is the importance of implementing teaching policies that explicitly define a reduction in educational inequities as a relevant goal. This would require the redesign of selection processes and the implementation of in-service professional development programs. In other words, since instructors seem to translate their expectations of students into differential treatment, it is necessary to sensitize instructors regarding the potential negative effects that their expectations may have in the evaluation of vulnerable populations. An additional suggestion is the promotion of permanent interventions aimed at identifying and exploring how grading bias may affect factors other than course grades, such as students' self-efficacy, motivation, participation, cooperation and attendance (Jussim 1986).

Grading bias is only one of several potential sources of educational inequities in higher education systems in developing countries. Although our field experiment has limited external validity, importantly, the organization where this experiment was conducted shares multiple characteristics with many other public higher education institutions in Mexico and the region, such as hiring processes, performance evaluations, student selection processes, professional development activities, management procedures, instructional practices, and learning goals.

Given the current trends whereby access to higher education is usually associated with higher socioeconomic status, promoting additional efforts to identify potential sources of educational inequalities in colleges seems to be a necessary step to support vulnerable populations in their efforts to access college education. Beyond promoting access, identifying,

analyzing, and implementing effective interventions to support the engagement of ethnic minorities is fundamental to promoting a better distribution of educational opportunities.

References

- Aksoy, Hasan H. 2007. "Teacher Education Student Perceptions and Views about Equity and Discrimination Practiced in Universities in Ankara." *World Applied Sciences Journal* 2: 1-11.
- Baltazar, E. 2012. "My School Experience as an Indigenous Student in Higher Education."
- Botelho, Fernando, Ricardo Madeira, and Marcos A. Rangel. 2013. *Discrimination Goes to School? Racial Differences in Performance Assessments by Teachers*. Working Paper, Website of Rangel M., Professor at Duke University. <http://marcos-a-rangel.com/files/77033474.pdf>.
- Carnoy, M., L. Santibañez, A. Maldonado, and I. Ordorika. 2002. "Barriers to Entry to Higher Education and Professional Opportunities for the Mexican Indigenous Population." *Latin American Journal of Educational Studies* 32: 9-43.
- Casillas, M. 2012. "Cultural Diversity in the University: Indigenous and Afro-Descendant Students in Higher Education, Need for Visibility and Recognition." In *Higher Education for Indigenous People and People of African Descent in Latin America*, ed. J. Badillo, M. Casillas and V. Ortiz. Mexico: Universidad Veracruzana.
- Castro, D., and Santos, A. (2012. "Los indígenas son mis padres, yo ya no. El día que Ángel decidió dejar de ser indio" (The Indigenous Are My Parents, I Am No Longer: The Day Angel Decided to Stop Being an Indian)." *Revista Latinoamericana de Educación Inclusiva*, 6 (1): 23-37.
- Chávez, M. 2008. "Being indigenous in Higher Education: Real or Assigned Disadvantages?" *Journal of Higher Education* 37: 31-55.

- Chu, James, and Wang Huan. 2015. *Discrimination Based on Gender and Prior Grades: Results from a Field Experiment*. Working Document.
<https://pdfs.semanticscholar.org/d46d/9b6dc6afed54729e4e6ed7c33dc50388e18b.pdf>.
- Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2007. "Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects." *Economics of Education Review* 26 (December): 673-682.
- Coates, Brian. 1972. "White Adult Behavior toward Black and White Children." *Child Development* 43 (March): 143-154.
- Conrado, A., and C. Sánchez. 2011. "The Words Themselves Matter. Perspective of Social Discrimination Institutions of Higher Education in San Luis Potosi." In *XI National Congress of Educational Research*. Mexico Federal District.
- Cooper, H. 1985. *Models of Teacher Expectation Communication*. New Jersey, NJ: Erlbaum.
- DeMeis, Debra K., and Ralph R. Turner. 1978. "Effects of Students' Race, Physical Attractiveness, and Dialect on Teachers' Evaluations." *Contemporary Educational Psychology* 3 (January): 77-86.
- Dickman, Carolyn B. 1993. "Gender Differences and Instructional Discrimination in the Classroom." *Journal of Invitational Theory and Practice* 2 (Winter): 35-42.
- Gacel, J. 2015. "Internationalization and Institutional capacities of Higher Education." In *Seventh Meeting of University Networks and Councils of Rectors of Latin America and the Caribbean*. Mexico: Congress held in Guanajuato.
- Gallart, M., and C. Henríquez. 2006. "Indigenous People in Higher Education: Some Reflections." *Universities* 32: 27-37.

- Gálvez, X., and A. Embriz. 2006. *And What is Your Name? The Voices of Indigenous Peoples to Name the People*. Mexico: CDI.
- Gershenson, Seth, Stephen B. Holt, and Nicholas W. Papageorge. 2015. *Who Believes in Me? The Effect of Student-Teacher Demographic Match on Teacher Expectations*. Upjohn Institute Working Paper 15-231. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- Good, T (1987). Two Decades of Research on Teacher Expectations: Findings and Future Directions. *Journal of Teacher Education*, 38(4), 32 – 47.
- Good, Thomas L., and Jere E. Brophy. 1997. *Looking in Classrooms*. New York: Longman.
- ILO-UNESCO. 2008. *Joint ILO-UNESCO Recommendation Concerning the Status of Teaching Personnel (1966) and UNESCO Recommendation Concerning the Status of Higher Education Teaching Personnel (1997)*. Paris: ILO-UNESCO.
- Jacobone, V. 2015. "Academic Mobility in Europe: Evaluative Studies in the South of Italy." In *Seventh Meeting of University Networks and Councils of Rectors of Latin America and the Caribbean*. Mexico: Congress held in Guanajuato.
- Jacobson, Marsha B., and Joan Effertz. 1974. "Sex Roles and Leadership." *Organizational Behavior and Human Performance* 12 (December): 383-396.
- Jussim, Lee. 1986. "Self-Fulfilling Prophecies: A Theoretical and Integrative Review." *Psychological Review* 93 (October): 429-445.
- Jussim, Lee, Jacquelynne Eccles, and Stephanie Madon. 1996. "Social Perception, Social Stereotypes, and Teacher Expectations: Accuracy and the Quest for the Powerful Self-Fulfilling Prophecy." *Advances in Experimental Social Psychology* 28 (January): 281-388.

- Jussim, Lee, and Kent D. Harber. 2005. "Teacher Expectations and Self-Fulfilling Prophecies: Knowns and Unknowns, Resolved and Unresolved Controversies." *Personality and Social Psychology Review* 9 (May): 131-155.
- Jussim, Lee, Stephanie Madon, and Celina Chatman. 1994. "Teacher Expectations and Student Achievement." In *Applications of Heuristics and Biases to Social Issues*, ed. L. Heath, R. S. Tindale, J. Edwards, E. J. Posavac, F. B. Bryant, E. Henderson-King, Y. Suarez-Balcazar and J. Myers. Boston, MA: Springer US.
- Maldonado, A. 2015. "Why Short-Term Mobility Matters in Low Income Individuals?" In *Seventh Meeting of University Networks and Councils of Rectors of Latin America and the Caribbean*. Mexico: Congress held in Guanajuato.
- Moreno, Juan, Celestina M. Galindo, and Eduardo Cervelló. 2011. "Predictive Relationship between Students' Perception of the Teacher's Disciplinary Strategies and of Equalitydiscrimination in Physical Education Classes." *Journal of Education* 355 (January): 381-403.
- Murillo, J., and N. Hidalgo. 2015. "Fundamental Approaches to Student Assessment for Social Justice." *Ibero-American Journal of Educational Evaluation* 8: 43-61.
- National Electoral Institute. 2015. *Statistics Nominal List and Electoral Register*. Mexico: National Electoral Institute.
- Pérez, S. 2012. "My School Experience as an Indigenous Student in Higher Education." In *Higher Education for Indigenous People and People of African Descent in Latin America*, ed. J. Badillo, M. Casillas and V. Ortiz. Mexico: Veracruzana University.

- Reisberg, L. 2015. "Mobility: Challenges for Latin America." In *Seventh Meeting of University Networks and Councils of Rectors of Latin America and the Caribbean*. Mexico: Congress held in Guanajuato.
- Rema, Hanna N., and Linden L. Leigh. 2012. "Discrimination in Grading." *American Economic Journal: Economic Policy* 4 (November): 146-168.
- Ripoll, J. 2014. *Evidence-Based Reading Comprehension: Practical Questions about Intraclass Correlation*. Author's Personal Website. <https://clbe.wordpress.com/tag/coeficiencia-de-correlacion-intraclase/>.
- Rosenthal, Robert. 1994. "Interpersonal Expectancy Effects: A 30-Year Perspective." *Current Directions in Psychological Science* 3 (December): 176-179.
- Rosenthal, Robert, and Lenore Jacobson. 1968. *Pygmalion in the Classroom*. New York: Holt, Rinehart, and Winston.
- Ruiz, E. 2012. "Attitudes, Stereotypes and Prejudices Its Influence on Down Syndrome." *Down Syndrome Magazine* 114: 110-121.
- Sprietsma, Maresa. 2009. "Discrimination in Grading: Experimental Evidence from Primary School Teachers." *Empirical Economics* 45 (June): 523-538.
- Trouilloud, David O., Philippe G. Sarrazin, Thomas J. Martinek, and Emma Guillet. 2002. "The Influence of Teacher Expectations on Student Achievement in Physical Education Classes: Pygmalion Revisited." *European Journal of Social Psychology* 32 (June): 591-607.
- Trouilloud, David, and Philippe Sarrazin. 2002. "L'Effet Pygmalion Existe-t-il en Education Physique et Sportive ? Influence des Attentes des Enseignants sur la Motivation et la Performance des Élèves." *Science et Motricité* 46 (January): 69-94.

- Trouilloud, David, and Philippe Sarrazin. 2003. "Les Connaissances Actuelles sur l'Effet Pygmalion: Processus, Poids et Modulateurs." *French Journal of Pedagogy* 145 (May): 89-119.
- Trouilloud, David, Damien Tessier, and Philippe Sarrazin. 2016. "Stéréotypes et Engagement des Élèves in EPS." In *Les Stereotypes*, ed. P. Fontayne and A. Chalabaev. Paris: Editions EP&S.
- Tyrefors, Björn, Erik Höglin, and Magnus Johannesson. 2011. *Ethnic Discrimination in High School Grading: Evidence from a Field Experiment*. Working Paper. Switzerland: Stockholm University.
- Vituri, Dagmar W., and Yolanda D. M. Évora. 2014. "Reliability of Indicators of Nursing Care Quality: Testing Interexaminer Agreement and Reliability." *Revista Latino-Americana de Enfermagem* 22 (March): 234-240.
- White S. y Locke E. (2000). Problems with the Pygmalion effect and some proposed solutions. *The Leadership Quarterly*, 11 (3), Pages 389-415.
- Williams, Trevor. 1976. "Teacher Prophecies and the Inheritance of Inequality." *Sociology of Education* 49 (July): 223-236.

Tables

Table 1. Factors influencing instructors' expectations.

Variable	Explanation
1. Subject	Negative instructor expectations increase grading bias, more often in reading than in mathematics courses.
2. Gender	Some authors have argued that there is no significant influence of gender (White and Locke 2000), but recent evidence suggests that women tend to be more optimistic about the outcomes of ethnic minorities (Gershenson et al. 2015), thus making them less likely to show grading bias toward minorities.
3. Instructors' beliefs about teaching skills	When an instructor's sense of self-efficacy is weak, he or she will spend less time working with students for whom he or she has low expectations (Trouilloud and Sarrazin 2003).
4. Personal beliefs	Instructors considered to be authoritarian, dogmatic, or prejudiced tend to develop rigid expectations over time (Jussim 1986).
5. Awareness of social inequalities	Instructors who are less aware of inequalities may reproduce unfavorable conditions for some minorities, such as ethnic minorities. Student gender might also be a factor associated with grading bias (e.g., Dickman 1993; Aksyo 2007; Murillo and Hidalgo 2015).
6. Cultural diversity/Travel capital	Exposure to different cultures and values increases intercultural understanding and respect for diversity (Gacel 2015). National or international mobility expands appreciation and respect for cultural diversity, thus reducing the likelihood of showing grading bias toward ethnic minorities (Gacel 2015; Jacobone 2015; Maldonado 2015; Reisberg 2015).

Table 2. Summary table: experimental grades and their differences with contrast grades.

Group	Average contrast grade	Average experimental grade	Observations	Average difference
Nonindigenous	5.99	6.90	208	0.91 (1.70)
Indigenous	7.03	7.44	104	0.41 (0.15)
Both groups	6.51	7.08	312	-
Diff in Diff				-0.50
	Difference p value (<i>t test</i>)			0.01

Table 3. Treatment effects (indigenous cover page).

	Model 1	Model 2	Model 3
Group	-0.490** (0.199)	-0.490*** 0.182)	-0.543*** (0.201)
Test	_____	-0.292*** (0.05)	_____
Prof	_____	0.030*** (0.006)	_____
Gender	_____		-0.104 (0.231)
Teaching experience	_____		-0.01 (0.022)
SES	_____		-0.996* (0.578)
Academic Achievement	_____		0.055 (0.181)
Arts and Humanities	_____		1,173 (0.452)
Natural Sciences	_____		0.572 (0.255)
Engineering	_____		0.255 (0.32)
Agronomy	_____		-0.978 (0.726)
Health	_____		0.15 (0.722)
High school	_____		0.557 (0.521)

	Model 1	Model 2	Model 3
Constant	0.904 ^{***}	1,107 ^{***}	1,141 [*]
	-0.115	-0.256	(0.613)

The results were statistically significant at the 1%, 5% and 10% levels, as indicated by ^{***}, ^{**} and ^{*}, respectively.

Table 4. Models exploring different factors associated with grading bias.

	Direct		Indirect				Other Factors		General Model		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Previous grades	-0.122 (0.084)	-0.143* (0.085)									0.004 (0.093)
Admissions tests	0.022 (0.088)	0.041 (0.081)									0.032 (0.077)
Colleagues	0.034 (0.078)	0.067 (0.072)									0.036 (0.051)
Diagnostic tests	0.033 (0.086)	0.088 (0.109)									0.284** (0.118)
Good behavior	-0.036 (0.095)	-0.023 (0.098)									-0.175* (0.093)
Motivation or commitment	-0.05 (0.124)	-0.141 (0.11)									-0.236** (0.102)
Parental involvement	0.033 (0.095)	0.137 (0.096)									0.272*** (0.079)
Family (1)			0.091 (0.138)	0.138 (0.107)							
Family (2)			-0.234* (0.124)	-0.346*** (0.119)							-0.159** (0.063)
			0.036	0.083							

	Direct		Indirect				Other Factors		General Model		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Family (3)			(0.085)	(0.071)							
Family (4)			0.054 (0.065)	0.031 (0.074)							
Family (5)			0.092 (0.088)	0.162 (0.104)							
Family (6)			0.026 (0.091)	-0.04 (0.073)							
Family (7)			-0.052 (0.082)	-0.12 (0.074)							
Outstanding performance					0.097 (0.076)	0.145 (0.093)					
Personal presentation					-0.119* (0.06)	-0.175** (0.074)					-0.265*** (0.063)
Perception as Problematic					0.042 (0.073)	0.033 (0.074)					
Isolation					-0.112 (0.078)	-0.054 (0.069)					
Potential capacities					-0.062 (0.09)	-0.069 (0.084)					

	Direct		Indirect				Other Factors		General Model		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Indigenous customs					0.155*** (0.054)	0.113** (0.051)					0.054 (0.058)
Indigenous accent					0.015 (0.055)	-0.024 (0.058)					
Traditional attire					-0.053 (0.076)	-0.114* (0.066)					0.024 (0.062)
Diversity in college							-0.053 (0.098)	-0.055 (0.111)			
Tolerance							-0.137 (0.328)	-0.537** (0.262)			-0.733*** (0.25)
Relationship							0.035 (0.089)	0.043 (0.081)			
Values and culture							-0.194 (0.132)	-0.159 (0.143)			
Institutionalized discrimination									0.093 (0.065)	0.144** (0.068)	0.124 (0.082)
									-0.150**	-0.213***	-0.157**

	Direct		Indirect					Other Factors		General Model	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Discrimination based on skin color									(0.067)	(0.07)	(0.067)
Discrimination Awareness									-0.014 (0.062)	-0.068 (0.05)	-0.047 (0.059)
Perceived trajectory									0.004 (0.057)	0.068 (0.053)	0.239*** (0.049)
Cultural differences									-0.102 (0.079)	-0.088 (0.066)	0.01 (0.051)
Academic mobility programs									0.064 (0.201)	0.323 (0.193)	0.350** (0.158)
Academic achievement		-0.168 (0.109)		-0.374*** (0.107)		-0.161 (0.112)		-0.187 (0.133)		-0.184* (0.106)	-0.373*** (0.097)
Humanities		-0.455 (0.301)		-0.568** (0.22)		-0.464 (0.298)		-0.523* (0.281)		-0.405 (0.271)	-0.852*** (0.144)
Science		-0.144 (0.188)		-0.099 (0.172)		-0.199 (0.18)		-0.204 (0.176)		-0.2 (0.186)	0.049 (0.148)
Engineering		-0.520** (0.232)		-0.473** (0.231)		-0.475*** (0.172)		-0.430** (0.182)		-0.531*** (0.189)	-0.299* (0.159)
Agronomy		-1.463*** (0.304)		-0.952*** (0.332)		-1.137*** (0.319)		-1.343*** (0.212)		-1.440*** (0.192)	-1,240*** (0.255)
Health		-0.775*** (0.266)		-0.869*** (0.311)		-1.288*** (0.412)		-1.138*** (0.355)		-0.046 (0.33)	-1.553*** (0.304)

	Direct		Indirect				Other Factors			General Model	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
High school		-0.477 (0.332)		-0.376 (0.26)		-0.291 (0.326)		-0.534 (0.393)		-0.527 (0.357)	-0.38 (0.232)
Effectiveness as an instructor		-0.088 (0.092)		-0.043 (0.075)		-0.076 (0.07)		-0.03 (0.067)		-0.115** (0.054)	-0.284** (0.11)
Proud professional		-0.119 (0.167)		0.013 (0.164)		-0.167 (0.182)		-0.126 (0.161)		0.029 (0.189)	0.282 (0.189)
Gender		-0.513*** (0.176)		-0.351** (0.131)		-0.428*** (0.156)		-0.440*** (0.155)		-0.513*** (0.153)	-0.590*** (0.135)
Teaching experience		0.015 (0.014)		0.007 (0.008)		0.013 (0.01)		0.013 (0.011)		0.028** (0.013)	0.058*** (0.012)
SES		-0.051 (0.44)		0.258 (0.286)		-0.052 (0.387)		-0.15 (0.382)		-0.212 (0.356)	-0.342 (0.247)
Constant	4,916*** (0.875)	6,956*** (1,122)		4,403*** (0.328)	4,745*** (0.517)	7,315*** (0.993)	5,131*** (0.945)	8,385*** (1,199)	4,888*** (0.441)	6,310*** (1,076)	8,068*** (1,205)

Note. The results are statistically significant at the 1%, 5% and 10% levels, as indicated by ***, ** and * respectively.