

## **Can Racial Bias in Teachers' Decisions Be Reduced? Evidence from a Survey Experiment in French Middle Schools**

This study fills a key gap in France: although observational and qualitative work documents ethnic inequalities in education, few experiments test whether they partly stem from teachers' discriminatory judgments. Evidence is also limited on whether structured rubrics curb such bias and whether teacher characteristics—especially experience—moderate racial bias in grading, expectations of success, or tracking. Using a stereotype activation/attenuation framework, we test whether disparities arise under non-rubric grading and whether rubrics reduce them. Extending U.S. work (Quinn, 2020) to a high-stakes French middle-school setting, and linking teacher bias with rubric-based fairness (Quinn, 2020; Malouff & Thorsteinsson, 2016) and experience effects (Glock, Kovacs, & Pit-Ten Cate, 2019), we provide new evidence on mechanisms driving racial discrimination in grading and tracking.

### **Literature Review**

Teachers, like professionals in other domains, may hold biases that shape their expectations and judgments. Since grades play a central role in students' educational trajectories, understanding how bias influences evaluation is essential (Gil Hernández et al., 2024). Experimental studies confirm grading bias against minority students, e.g., Turkish-origin students in Germany (Sprietsma, 2013; Glock et al., 2013; Pit-Ten Cate et al., 2016; Tobisch & Dresel, 2017; Wenz & Hoenig, 2020), and African-American students in the U.S. (Elhoweris et al., 2005; Kunesh & Noltemeyer, 2019; Legette et al., 2021; Owens, 2022). A meta-analysis by Tenenbaum and Ruck (2007) found that teacher evaluations of performance, behavior, and placement tend to favor White students over Hispanic and African-American peers, with a small-to-moderate average effect size ( $d = 0.31$ ). To address such disparities, scholars have advocated for more structured and transparent assessment tools, notably rubrics (Merle, 2012). In an experimental study, Quinn (2020) found that ethnic bias in grading disappeared when teachers used a structured rubric. Similarly, Malouff and Thorsteinsson's (2016) meta-analysis showed that rubrics reduced bias effects ( $d = 0.24$ ) compared to unstructured grading ( $d = 0.39$ ). By standardizing evaluation criteria, rubrics help reduce subjectivity and support fairer grading (Jonsson & Svingby, 2007).

Like other professionals, teachers may hold biases that shape expectations and judgments. Because grades steer educational trajectories, identifying how bias enters evaluation is crucial (Gil Hernández et al., 2024). Experiments document grading bias against minority students—e.g., Turkish-origin students in Germany (Sprietsma, 2013; Glock et al., 2013; Pit-Ten Cate et al., 2016; Tobisch & Dresel, 2017; Wenz & Hoenig, 2020) and African-American students in the U.S. (Elhoweris et al., 2005; Kunesh & Noltemeyer, 2019; Legette et al., 2021; Owens, 2022). A meta-analysis shows teacher evaluations of performance, behavior, and placement favor White over Hispanic and African-American students ( $d = 0.31$ ; Tenenbaum & Ruck, 2007). To mitigate such disparities, scholars advocate structured, transparent assessments, notably rubrics (Merle, 2012). In an experiment, Quinn (2020) found ethnic grading bias disappeared with a structured rubric; Malouff & Thorsteinsson (2016) show rubrics reduce bias ( $d = 0.24$ ) relative to unstructured

grading ( $d = 0.39$ ). By standardizing criteria, rubrics reduce subjectivity and support fairer grading (Jonsson & Svingby, 2007).

## **Design and Methodology**

We use a between-subjects survey experiment with random assignment to conditions (Auspurg & Hinz, 2015; Bansak et al., 2021), presenting realistic ninth-grade profiles—a key stage for grading and tracking decisions.

### *Population Rationale*

We focus on French literature teachers, a central subject for tracking decisions; prior European work shows ethnic disparities in language grading (Sprietsma, 2013; Glock, 2016). Profiles are male only, given evidence that minority boys are especially vulnerable to stereotyping.

### *Experimental Manipulations & Materials*

Three randomized manipulations (**Table 1**):

(1) student name signaling French vs. North African origin; (2) student behavior (positive vs. disruptive); (3) grading condition (with rubric vs. without rubric).

Each teacher reviewed (a) a short vignette with name, behavior, academic record, and parental occupation (holding SES constant) and (b) a real student essay from national exam archives, originally graded 13/20 in blind evaluation—an average performance (Merle, 2018). Each respondent saw one profile combining name, behavior, and grading condition (e.g., North-African name, positive behavior, no rubric).

### *Outcome Variables*

Three primary dependent variables were measured:

1. Grade assigned to the essay (1–20 points), influenced by the rubric manipulation (see **Table 1**).
2. Perceived likelihood of passing the national lower-secondary exam (*Diplôme National du Brevet*), using the item:  
“Considering the description of this student and your evaluation of his work, do you think he is capable of passing the Brevet?”  
Response scale: 1 (Not at all capable) to 10 (Fully capable).
3. Academic tracking recommendation for tenth grade, using the item:  
“If you were the homeroom teacher of this student, which track would you recommend after ninth grade?”  
Response options: vocational, technological, or general track.

### *Measurement Validity*

Respondents rated the realism of the essay and vignette at 7.9 and 8.1 out of 10, respectively. The full instrument was reviewed by French literature teachers and academic inspectors to ensure validity.

### *Sample & Generalization Logic*

The target sample is 2,024 literature teachers determined via G\*Power (two-tailed t-test,  $d = 0.25$ ,  $\alpha = .05$ , power = .80). So far, 631 responses have been collected (data collection runs from February to December 2025). Institutional support from the French Ministry of Education enabled nationwide distribution to all public middle-school literature teachers. Findings concern French middle-school literature teachers. Findings concern French middle-school literature teachers; given the non-probabilistic, voluntary sample, external generalization beyond this population should be cautious and will be strengthened by completing national data collection.

### **Preliminary Results**

#### *Grade Assigned to the Essay*

**Figure 1** displays results from a linear regression predicting the grade assigned to the student essay. The left panel presents coefficients for both experimental and teacher-level variables. Most effects cluster near zero, suggesting no strong influence of teacher characteristics—such as gender or experience—on assigned grades in this subsample. The only significant effect concerns the use of a structured rubric: it is associated with grades approximately one point lower, indicating that rubrics may prompt more stringent evaluations. Complementary GLS analyses (not shown) suggest a slight reduction in grade dispersion when a rubric is used ( $\approx 3\%$  lower SD; SD ratio = 0.97, 95% CI [0.88, 1.08]), although this difference is not statistically significant (LRT  $p = 0.59$ ), likely due to limited statistical power. There is no significant effect of student origin: teachers assigned similar grades regardless of whether the student had a French- or North African-sounding name. Again, this may reflect the limited sample size and the difficulty of detecting small disparities.

The right panel of **Figure 1** shows the interaction between student origin and rubric use. Without a rubric, predicted scores for North African-named students are slightly lower, though not significantly so. With a rubric, scores decrease for all students, but the predicted grade for North African-named students becomes slightly higher than that for French-named peers. Although not significant, this interaction aligns with findings suggesting rubrics may reduce bias (Quinn, 2020; Malouff & Thorsteinsson, 2016). The full dataset will allow us to determine whether rubrics simply lower scores overall or actually mitigate origin-related disparities.

#### *Perceived Likelihood of Passing the Brevet*

As shown in **Figure 2**, ratings of the student's likelihood of passing the *Brevet* were heavily skewed toward the top of the scale (scores of 8, 9, or 10), suggesting a ceiling effect. To address this, the variable was dichotomized (1 = score  $\geq 8$ ; 0 = otherwise). Logistic regression reveals significant bias: students with North African-sounding names were 72% less likely to receive a high rating than otherwise identical French-named students (OR = 0.28; **Figure 3**). Interaction analyses (**Figure 4**) highlight that this gap is most pronounced among novice teachers. At 0 years of experience, the difference is statistically significant ( $p = 0.011$ ); it progressively narrows with experience and is no longer significant at 40 years ( $p = 0.257$ ).

### *Recommendation for the General Track*

**Figure 5** presents results from a logistic regression predicting whether the student would be recommended for the general track in Grade 10 (*seconde générale*), coded as a binary variable (1 = general track; 0 = all other tracks). The model indicates significant racial bias: North African–named students are 41% less likely to be recommended for the general track (OR = 0.59), controlling for all other variables. Rubric use is marginally significant (OR = 0.64,  $p < .10$ ), with a confidence interval that includes 1. The origin  $\times$  rubric interaction is positive but non-significant (OR = 1.55,  $p \approx .20$ ), consistent with a potential mitigating effect. The origin  $\times$  experience interaction (not shown) is slightly negative and non-significant ( $p \approx .40$ ).

### **Conclusion**

We report preliminary results from an ongoing survey experiment with French middle-school literature teachers: 631 responses to date (target = 2,024), with a national follow-up in Fall 2025 to boost power and reliability. Initial results show no significant effect of perceived origin on essay grades. Neither the rubric $\times$ origin nor the behavior $\times$ origin interaction is significant; the latter was therefore omitted from figures. These findings echo Gil Hernández et al. (2024) yet contrast with studies reporting minority penalties (e.g., Sprietsma, 2013; Glock et al., 2013; Tobisch & Dresel, 2017; Alesina et al., 2018).

By contrast, a significant bias emerges in judgments of the student’s likelihood of passing the lower-secondary exit exam: students with North African–sounding names are rated less capable. This bias is weaker among more experienced teachers, suggesting experience attenuates stereotype-driven judgments. Similarly, students with North African–sounding names are significantly less likely—*all else equal*—to be recommended for the general track, indicating systemic discrimination in guidance.

### **Discussion and Implications**

These early results highlight the role of teaching experience. Prior evidence is mixed (Pit-Ten Cate & Glock, 2019; Glock, Kovacs, & Pit-Ten Cate, 2019), but our data suggest less experienced teachers are more susceptible to stereotype activation. This concern mirrors the U.S., where students of color are disproportionately taught by novice teachers with limited equity training (Civil Rights Project, UCLA, 2024). Policy should prioritize targeted training, strategic assignment, and early mentoring to build bias awareness and equitable evaluation—especially before novices work in schools with large shares of immigrant-origin students. The role of rubrics in curbing discriminatory tracking remains unconfirmed; the full sample will test whether structured criteria buffer ethnic bias (Quinn, 2020; Malouff & Thorsteinsson, 2016).

## Table and Figures

### Table 1. Summary of Experimental Manipulations

Factors (dimensions/conditions)	Levels	Signals
Migration background	1. French origin 2. North African origin	Name of the student in the description and essay:  1. <b>French origin name:</b> Sébastien Morel 2. <b>North African origin name:</b> Karim Saadi
Behavior	1. Positive behavior 2. Disruptive behavior	<b>In the student description:</b>  1. <b>Positive behavior:</b> Karim/Sébastien does not participate enough orally. He may occasionally chat but stops when reminded, and his attitude is usually pleasant and disciplined. Karim/Sébastien is appreciated by peers and teachers. When encouraged, he shows enthusiasm and open-mindedness. Most teachers believe he can still make progress and needs regular encouragement to maintain motivation.  2. <b>Disruptive behavior:</b> Karim/Sébastien does not participate enough orally. Although this is occasional, his behavior is sometimes unpleasant and undisciplined. He must often be reminded to stop chatting and fidgeting. His behavior can disturb peers and teachers. He often refuses to participate when invited. He is not the most turbulent student, but teachers believe he needs frequent encouragement to remain motivated throughout the year.
Grading	1. Rubric (criteria-based) 2. Grade without rubric	Questions in the survey:  1. <b>With criteria (3 questions):</b> - How many points do you assign to the essay structure (introduction, conclusion, paragraphing, connectors)? (0-7) - How many points for argumentation quality (variety, development, relevance, examples)? (0-7) - How many points for language quality (clarity, syntax, spelling, grammar, vocabulary)? (0-6)  2. <b>Without criteria (1 question):</b> What grade out of 20 would you assign to this essay? (Numeric response 0-20, e.g., 12.5)

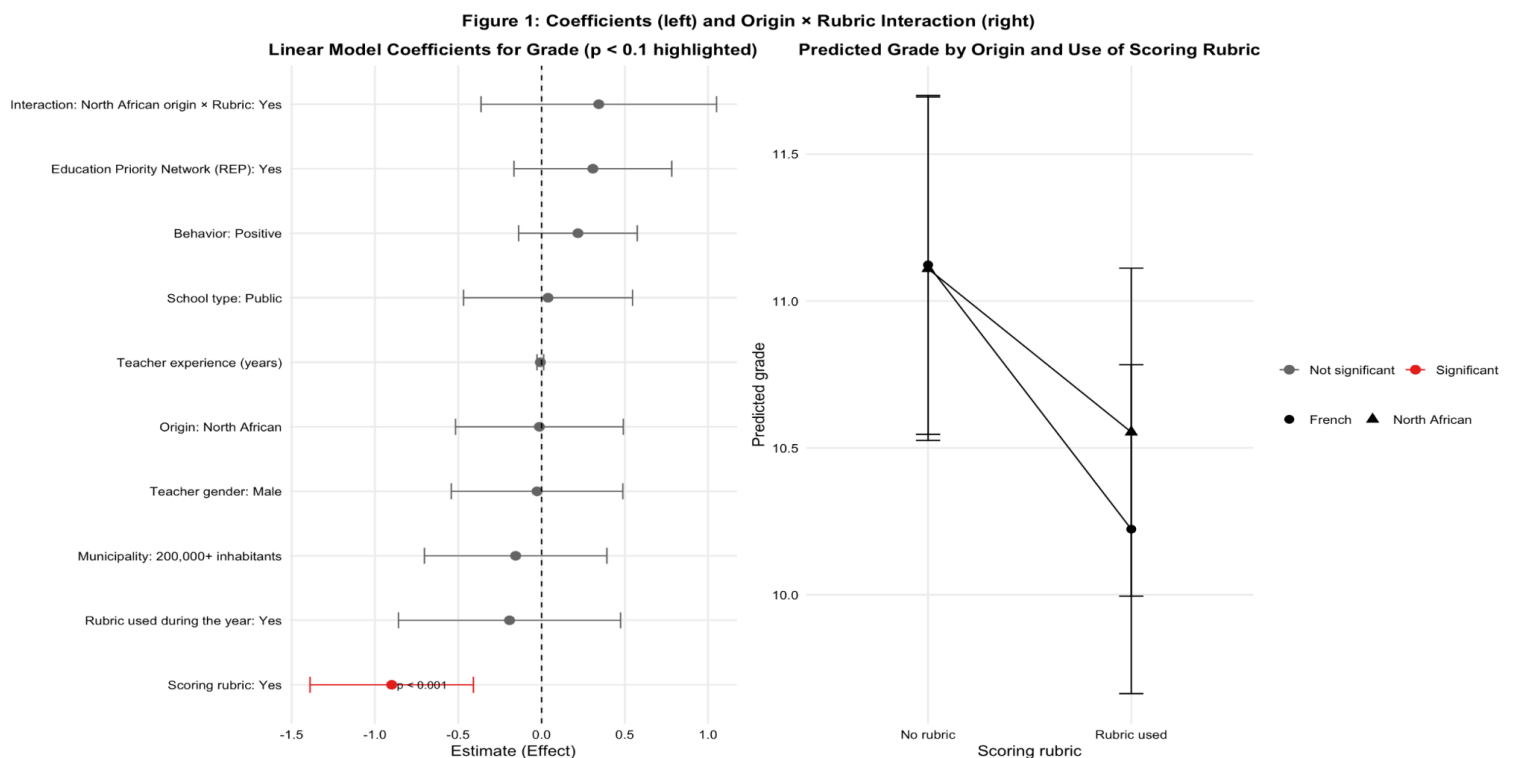


Figure 2 : Distribution of Responses to the 'Brevet' Question

Question: Based on the student's description and your assessment of their work,

do you think they are capable of passing the national diploma (Brevet)?

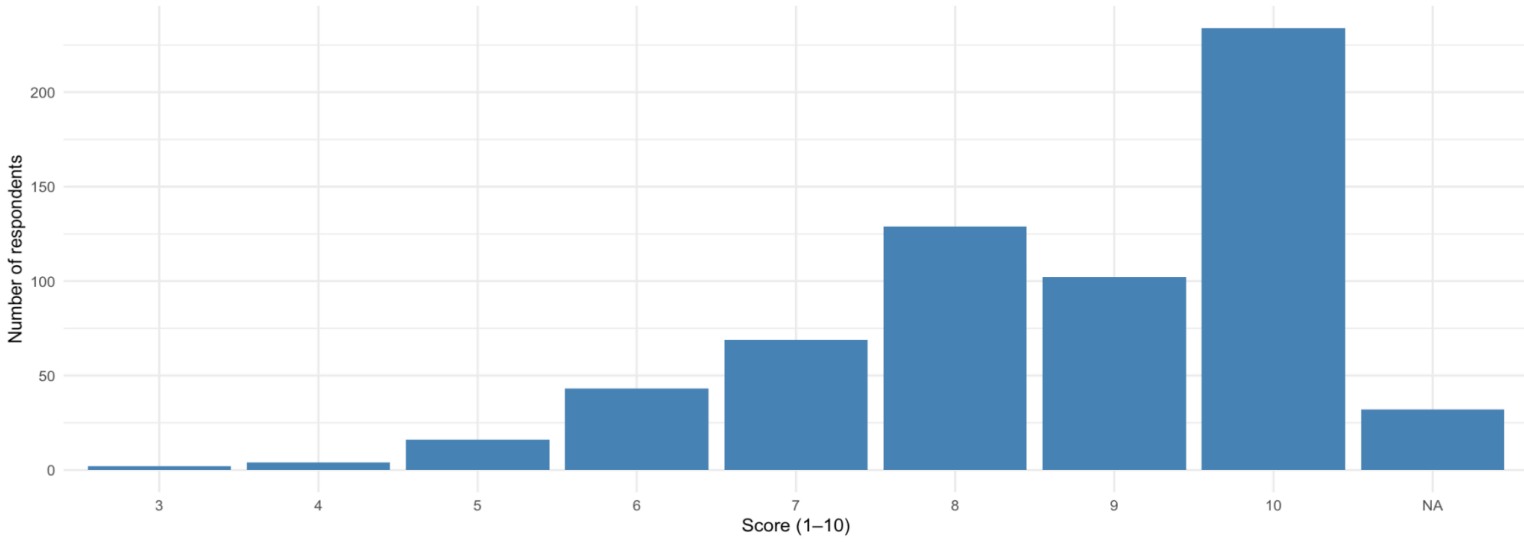


Figure 3: Perceived capability to pass the Brevet (score  $\geq 8/10$ )

Logistic regression odds ratios with 95% CIs

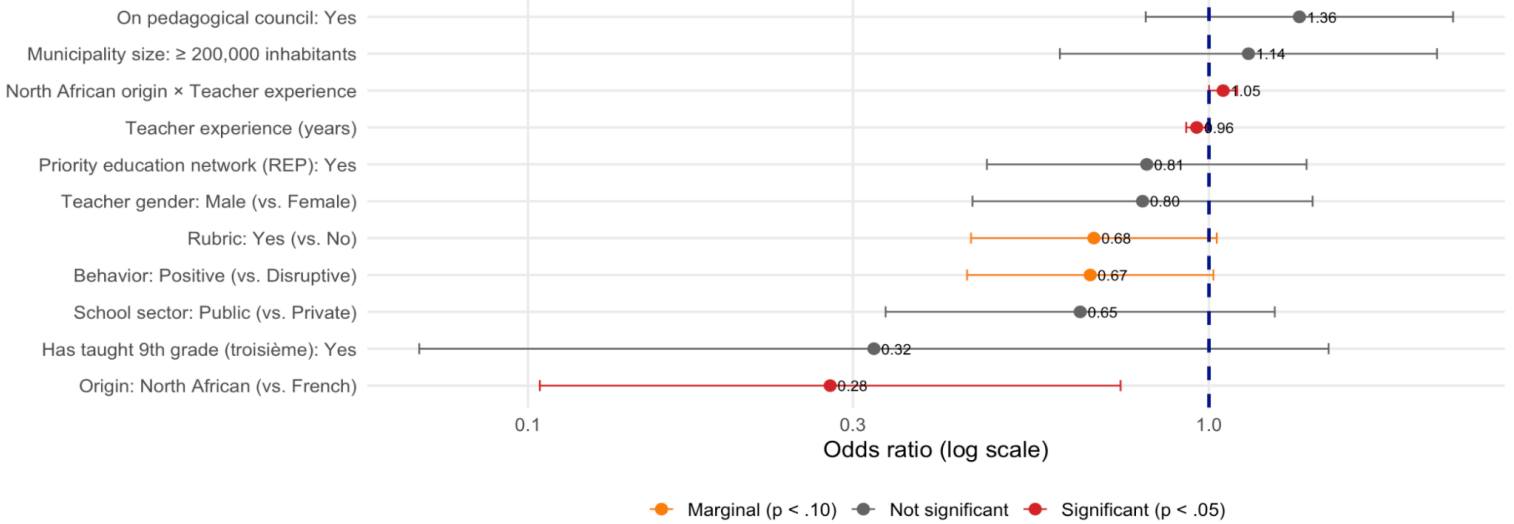


Figure 4: Predicted probability of scoring  $\geq 8/10$  by teacher experience and student origin

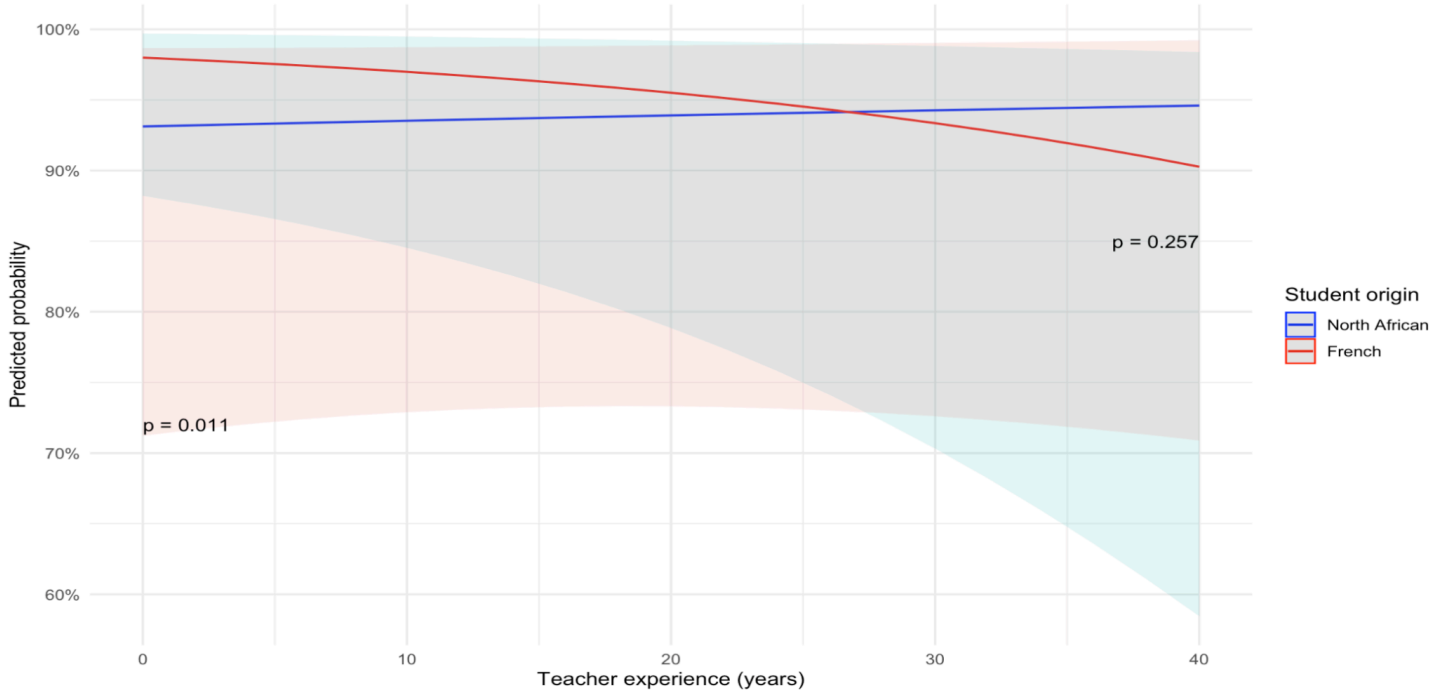
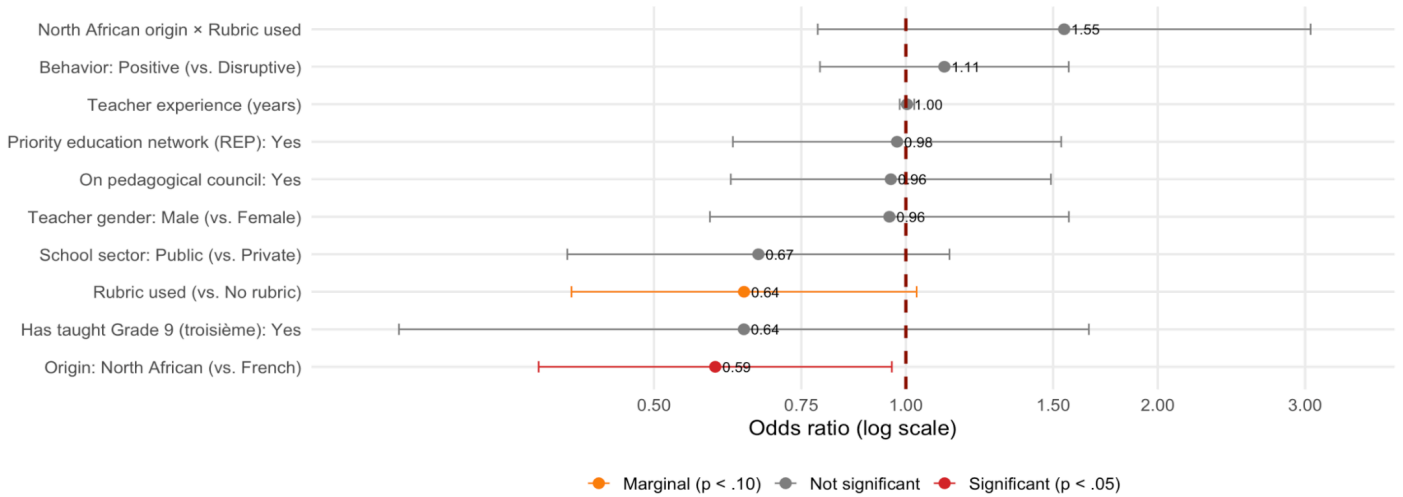


Figure 5: Placement into General Track (Seconde générale): Logistic model (odds ratios)

Question: If you were this student's homeroom teacher, which track would you recommend after Grade 9?



## References

- Alesina, Alberto, Michela Carlana, Eliana La Ferrara, & Paolo Pinotti. 2018. "Revealing Stereotypes: Evidence from Immigrants in Schools." *Working Paper*, National Bureau of Economic Research.
- Auspurg, Katrin, & Thomas Hinz. 2015. *Factorial Survey Experiments*. Los Angeles : SAGE Publications, Inc.
- Bansak, Kirk, Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. 2021. "Conjoint Survey Experiments." Pp. 19–41 in *Advances in Experimental Political Science*, edited by D. P. Green and J. N. Druckman. Cambridge : Cambridge University Press.
- Beauchemin, Cris, Christelle Hamel, & Patrick Simon. 2016. *Trajectoires et origines : Enquête sur la diversité des populations en France*. Paris: Ined Éditions.
- Belkacem, Lila, Gallot Fanny & Mosconi Nicole. 2019. "Penser l'intersectionnalité dans le système scolaire ?", *Travail, genre et sociétés*, 41(1), 147-152. <https://doi.org/10.3917/tgs.041.0147>.
- Ben Ayed, Choukri. 2023. *L'école discrimine-t-elle ? Le cas des descendants de l'immigration nord-africaine*. Paris : Éditions du Croquant.
- Brinbaum, Yaël, & Jean-Luc Primon. 2013. "Parcours scolaires des descendants d'immigrés et sentiments d'injustice et de discrimination.", *Economie et statistique*, n°464-466. Inégalités et discriminations : questions de mesure. pp. 215-243. doi: 10.3406/estat.2013.10239.
- Brinbaum, Yaël. 2019. "Education & Formations : La réussite des élèves : contextes familiaux, sociaux et territoriaux : N° 100. Chap. 4, Trajectoires scolaires des enfants d'immigrés jusqu'au baccalauréat : rôle de l'origine et du genre. Résultats récents.", *Publications de la DEPP*.
- Civil Rights Project, UCLA. 2024. New UCLA research highlights need to recruit and retain more teachers of color and Indigenous teachers in California. <https://www.civilrightsproject.ucla.edu/news/press-releases/press-releases-2024/new-ucla-research-highlights-need-to-recruit-and-retain-more-teachers-of-color-indigenous-teachers-in-california>
- Dhume, Fabrice, Suzana Dudic, Chauvel Séverine & Philippe Perrot. 2011. *Orientation scolaire et discrimination. De l'(in)égalité de traitement selon « l'origine »*. Paris : La Documentation française.
- Dhume, Fabrice, & Suzana Dukic. 2012. "Orientation scolaire et inégalités de traitement selon 'l'origine' : une synthèse critique des connaissances.", *Diversité : ville école intégration, Les jeunes des quartiers*, 167, p.165-175.
- Dhume, Fabrice, & Marguerite Cagnet. 2020. "Racisme et discriminations raciales à l'école et à l'université : où en est la recherche ?" *Le français aujourd'hui* 209(2):17–27. doi: 10.3917/lfa.209.0017.
- Elhoweris, Hala, Kagendo Mutua, Nigmeldin Alsheikh, & Pauline Holloway. 2005. "Effect of Children's Ethnicity on Teachers' Referral and Recommendation Decisions in Gifted and Talented Programs." *Remedial and Special Education* 26(1):25–31. doi: 10.1177/07419325050260010401.
- Felouzis, Georges. 2003. "La ségrégation ethnique au collège et ses conséquences." *Revue française de sociologie* 44(3):413–47. doi: 10.3917/rfs.443.0413.
- Gil Hernandez, Carlos, Irene Pañeda-Fernández, Leire Salazar, & Jonatan Castaño Muñoz. 2024. "Teacher Bias in Assessments by Student Ascribed Status: A Factorial Experiment on Discrimination and Cultural

Reproduction.” *JRC Publications Repository*.

<https://publications.jrc.ec.europa.eu/repository/handle/JRC136851>.

Glock, Sabine. 2016. “Does Ethnicity Matter? The Impact of Stereotypical Expectations on in-Service Teachers’ Judgments of Students.” *Social Psychology of Education* 19(3):493–509. doi: 10.1007/s11218-016-9349-7.

Glock, Sabine, Sabine Krolak-Schwerdt, Florian Klapproth, & Matthias Böhmer. 2013. “Beyond Judgment Bias: How Students’ Ethnicity and Academic Profile Consistency Influence Teachers’ Tracking Judgments.” *Social Psychology of Education* 16(4):555–73. doi: 10.1007/s11218-013-9227-5.

Glock, Sabine, Kovacs Carrie, & Pit-Ten Cate Ineke (2019). Teachers' attitudes towards ethnic minority students: Effects of schools' cultural diversity. *The British journal of educational psychology*, 89(4), 616–634. <https://doi.org/10.1111/bjep.12248>.

Guénif Souilamas, Nacira. 2000. *Des “beurettes” aux descendantes d’immigrants nord-africains*. Paris : B. Grasset.

Heath, Anthony, & Yaël Brinbaum. 2007. “Guest Editorial: Explaining Ethnic Inequalities in Educational Attainment.” *Ethnicities* 7(3):291–304. doi: 10.1177/1468796807080230.

Ichou, Mathieu. 2013. Différences d'origine et origine des différences : les résultats scolaires des enfants d'émigrés/immigrés en France du début de l'école primaire à la fin du collège. *Revue française de sociologie*. 54(1), 5-52. <https://doi.org/10.3917/rfs.541.0005>.

Jonsson, Anders, & Svingby Gunilla. 2007. The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130–144

Kunesh, Claire E., & Amity Noltemeyer. 2019. “Understanding Disciplinary Disproportionality: Stereotypes Shape Pre-Service Teachers’ Beliefs About Black Boys’ Behavior.” *Urban Education* 54(4):471–98. doi: 10.1177/0042085915623337.

Legette, Kamilah B., Amy G. Halberstadt, & Amber T. Majors. 2021. “Teachers’ Understanding of Racial Inequity Predicts Their Perceptions of Students’ Behaviors.” *Contemporary Educational Psychology* 67:102014. doi: 10.1016/j.cedpsych.2021.102014.

Lorcerie, Françoise. 2004. *Écoles et appartenances ethniques. Que dit la recherche ?* Rapport pour le PIREF, ministère de l’Éducation nationale - Direction de la recherche.

Malouff, John & Thorsteinsson, Einar. 2016. Bias in grading: A meta-analysis of experimental research findings. *Australian Journal of Education* 60. doi: 10.1177/0004944116664618.

Merle, Pierre. 2012. “14. L’évaluation par les notes : quelle fiabilité et quelles réformes ?” *Regards croisés sur l’économie* 12(2):218– 30. doi: 10.3917/rce.012.0218.

Merle, Pierre. 2018. *Les pratiques d’évaluation scolaire*. Paris : PUF.

Merle, Pierre. 2018. “Que Sait-on Des Pratiques d’évaluation Scolaires ?” *Le Café Pédagogique*. <https://cafepedagogique.net/2018/05/25/pierre-merle-que-sait-on-des-pratiques-devaluation-scolaires/>.

Owens, Jayanti. 2022. “Double Jeopardy: Teacher Biases, Racialized Organizations, and the Production of Racial/Ethnic Disparities in School Discipline.” *American Sociological Review* 00031224221135810. doi: 10.1177/00031224221135810.

Palheta, Ugo. 2012. *La Domination Scolaire. Sociologie de l'enseignement Professionnel et de Son Public*. Paris : PUF.

Payet, Jean-Paul. 2003. "École et immigration. Un bilan des travaux (1996-2002) un programme de recherche." *VEI enjeux*, n°135, 2003. La discrimination ethnique. Réalités et paradoxes. pp. 103-122. doi: 10.3406/diver.2003.1464.

Perroton, Joëlle. 2000. "Les Dimensions Ethniques de l'expérience Scolaire." *L'Année Sociologique*. 50(2),pp. 437-468.

Pit-Ten Cate, Ineke M., Sabine Krolak-Schwerdt, & Sabine Glock. 2016. "Accuracy of Teachers' Tracking Decisions: Short- and Long-Term Effects of Accountability." *European Journal of Psychology of Education* 31(2):225-43. doi:10.1007/s10212-015-0259-4.

Pit-Ten Cate, Ineke M., & Sabine Glock. 2019. "Teachers' Implicit Attitudes Toward Students From Different Social Groups: A Meta-Analysis." *Frontiers in psychology* vol. 10 2832. doi:10.3389/fpsyg.2019.02832

Quinn, David M. 2020. "Experimental Evidence on Teachers' Racial Bias in Student Evaluation: The Role of Grading Scales." *Educational Evaluation and Policy Analysis* 42(3):375-92. doi: 10.3102/0162373720932188.

Sprietsma, Maresa. 2013. "Discrimination in Grading: Experimental Evidence from Primary School Teachers." *Empirical Economics* 45(1):523-38. doi: 10.1007/s00181-012-0609-x.

Tenenbaum, Harriet R., & Martin D. Ruck. 2007. "Are Teachers' Expectations Different for Racial Minority than for European American Students? A Meta-Analysis." *Journal of Educational Psychology* 99:253-73. doi: 10.1037/0022-0663.99.2.253.

Tobisch, Anita, & Markus Dresel. 2017. "Negatively or Positively Biased? Dependencies of Teachers' Judgments and Expectations Based on Students' Ethnic and Social Backgrounds." *Social Psychology of Education* 20(4):731-52. doi: 10.1007/s11218-017-9392-z.

Vallet, Louis-André, & Jean-Paul Caille. 1996. "Niveau en français et en mathématiques des élèves étrangers ou issus de l'immigration." *Economie et Statistique* 293(1):137-53. doi: 10.3406/estat.1996.6053.

Van Zanten, Agnès. 2012. *L'école de la périphérie. Scolarité et ségrégation en banlieue*. Paris : PUF

Wenz, Sebastian E., & Kerstin Hoenig. 2020. "Ethnic and Social Class Discrimination in Education: Experimental Evidence from Germany." *Research in Social Stratification and Mobility* 65:100461. doi: 10.1016/j.rssm.2019.100461.