Creating birds of similar feathers:
Leveraging similarity to improve teacher-student relationships and academic achievement

Hunter Gehlbach\textsuperscript{a}
Maureen E. Brinkworth\textsuperscript{a}
Aaron King\textsuperscript{a}
Laura Hsu\textsuperscript{b}
Joe McIntyre\textsuperscript{a}
Todd Rogers\textsuperscript{c}

Author affiliations:
\textsuperscript{a} Harvard Graduate School of Education, 13 Appian Way, Cambridge, MA 02138
\textsuperscript{b} Merrimack College, School of Education and Social Policy, 315 Turnpike St., North Andover, MA 01845
\textsuperscript{c} Harvard Kennedy School of Government, 79 John F. Kennedy St., Cambridge, MA 02138

Corresponding author:
Hunter Gehlbach
Longfellow 316, 13 Appian Way, Cambridge, MA 02138
(617) 496-7318
Hunter_Gehlbach@harvard.edu

(Word count: 5926)

Acknowledgements
This research was supported by funding from the Spencer Foundation and the National Science Foundation—Grant #0966838. The conclusions reached are those of the investigators and do not necessarily represent the perspectives of the funder.
Creating birds of similar feathers:
Leveraging similarity to improve teacher-student relationships and academic achievement

Abstract:

By fabricating common characteristics between individuals, psychologists have shown that similarity is a powerful mechanism that can promote liking and improved relationship outcomes in laboratory studies. Correlational studies have shown that similarity has strong associations with these same relationship outcomes in real-world settings. Yet, no studies have leveraged veridical similarities to improve real-world relationships. Because teacher-student relationships are central to a host of important student outcomes, we tested whether an intervention that used real similarities might enhance these relationships. Thus, our manipulation represents the first field experiment to use actual similarities as a lever to improve relationships. We examine these teacher-student relationships in a sample of ninth grade students (n = 315) and their 25 teachers. We find that, five weeks after our intervention, teachers and students in the treatment conditions perceive greater degrees of similarity with the other party. Furthermore, when teachers received feedback about being similar to their students, they perceived better relationships with those students, and those students earned higher class grades. Sub-group analyses indicate that these effects are primarily localized to relationships where teachers learn that they are similar to their “underserved” students. This brief intervention appears to close the achievement gap at this school by over 60%.

Authors: Hunter Gehlbach, Maureen E. Brinkworth, Aaron King, Laura Hsu, Joe McIntyre, Todd Rogers

Keywords: Matching, Similarity, Social Processes/Development, Teacher-student relationships

Please cite as:

Creating birds of similar feathers:

Leveraging similarity to improve teacher-student relationships and academic achievement

Fostering social connections with others is increasingly appreciated as a fundamental, intrinsic human motivation – an activity people are hard-wired to do (Lieberman, 2013; Ryan & Deci, 2000). The extent to which individuals successfully relate to others affects a broad constellation of outcomes ranging from happiness (Gilbert, 2006) to health (Taylor et al., 2004). For children to thrive depends, in part, upon cultivating positive relationships with parents, peers, and teachers (Scales, Sesma, & Bolstrom, 2004). Even for adolescents, achieving positive teacher-student relationships (TSRs) is an important outcome in its own right and may catalyze important downstream benefits (Eccles et al., 1993).

Thus, how to improve TSRs is a question of tremendous interest for those who study positive youth development and schooling. One promising approach is to leverage individuals’ perceptions of similarity as a means to promote a sense of relatedness. Numerous basic social psychological texts underscore some version of the basic message that “likeness begets liking” (Myers, 2015, p. 330). Similarity along various dimensions (style of dress, background, interests, personality traits, hobbies, attitudes, etc.) has been connected to a wide array of relationship-related outcomes (such as attraction, liking, compliance, and prosocial behavior) in scores of studies (Cialdini, 2009; Montoya, Horton, & Kirchner, 2008). Theoretically, interacting with similar others is presumed to reinforce one’s sense of self and one’s values. Because people’s identities are positively supported by interacting with similar others, people are particularly motivated to affiliate with these individuals (Montoya et al., 2008; Myers, 2015).

Unfortunately, similarity as a potential lever for improving relationships is plagued by a fundamental problem: people either share something in common or they do not. Thus, scholars can develop experimental manipulations of similarities but these interventions typically rely upon fabricated similarities. While these studies enable causal inferences to be made, the fact that the similarities are fictitious minimizes their utility for real-world interventions. On the other hand, numerous correlational studies have identified real similarities between individuals in real relationships and have shown that these similarities correspond with improved relationship outcomes. However, the correlational nature of these studies precludes causal inferences from being made. Thus, how scholars might successfully leverage real similarities to improve real-world relationships, such as TSRs, remains a vexing challenge.

In this study, we test the effects of an intervention that potentially mitigates these trade-offs. Specifically, we experimentally manipulate perceptions of veridical similarities as a means to try and improve TSRs between ninth graders and their teachers. In addition to examining TSRs as a key outcome, we note that these relationships have shown robust associations with consequential student outcomes (McLaughlin & Clarke, 2010). Thus, we also test whether the intervention affects students’ classroom grades. To our knowledge, this is the first experimental study to use actual similarities as a means to improving real, ongoing relationships.
Similarity and Relationships

Of the research that connects similarity with interpersonal relationships, two main types of studies proliferate: those that have fabricated similarities for the sake of experimental manipulations and those that have investigated actual similarities. Both types of studies have enhanced scientific understanding of the importance and potency of similarity in relationships.

Across both the experimental and correlational approaches, two notable themes emerge. First, the types of similarities that are associated with improved relationships outcomes can pertain to almost anything. Experimentally manipulating the similarity of names has been shown to boost liking and compliance – e.g., getting people to return questionnaires more frequently if the name on the cover letter was similar to their own (Garner, 2005). In a series of primarily correlational studies, Mackinnon, Jordan, and Wilson (2011) found that students who are physically similar to one another (e.g., both wearing glasses) will tend to sit next to one another in class. Using both experimental and correlational approaches, Boer et al. (2011) found that shared music preferences helped foster closer social bonds between people.

Second, even the most trivial similarities can lead to positive sentiments toward another person. Laboratory experiments informing participants that they and another participant share: a preference for Klee versus Kandinsky paintings (Ames, 2004), the tendency to over- or under-estimate the number of dots on a computer screen (Galinsky & Moskowitz, 2000), or purported similarity in fingerprint patterns (Burger, Messian, Patel, del Prado, & Anderson, 2004), have all resulted in enhanced relationship-related outcomes. Correlational studies show comparably surprising findings. For example, people who have similar initials are disproportionately likely to get married (Jones, Pelham, Carvallo, & Mirenberg, 2004).

Despite their contributions, these two approaches to studying the connections between similarity and relationships leave two important gaps in our knowledge. First, this work leaves open the crucial scientific question of whether real similarities cause improved outcomes in real relationships. Certainly, the preponderance of this experimental and correlational evidence, generalized across so many types of similarities – including ones that seem especially unimportant – suggests that this causal relationship should exist, but without direct experimental evidence some doubt remains.

A second gap in our knowledge is particularly salient for educational practitioners. Without some way to leverage real similarities between individuals within a classroom, the associations between similarity and relationship outcomes have limited practical applications. Car salespeople may be well-served by suggesting that they too enjoy camping, golf, or tennis if they notice tents, clubs, or rackets in the trunk of your car (Cialdini, 2009). However, teachers who lie about what they might have in common with individual students are much more likely to be found out over the course of an ongoing relationship (to say nothing of the ethically dubious nature of this tactic). One could argue that teachers might be able to leverage similarity by learning what students have in common with each other and assigning them to collaborative groups with like-minded classmates. However, many would argue that an important socialization goal of schools is to help students work effectively with those from different backgrounds. In sum, as compelling and robust as the similarity-relationship research is, there are important scientific and applied gaps in our understanding of these associations that need to
be answered before similarities might be used as a lever to improve important relationships such as TSRs.

Teacher-Student Relationships and Student Outcomes

In addition to healthy relationships as an important outcome in their own right (Leary, 2010), TSRs are especially important to examine in large part because they are associated with a broad array of valued student outcomes including: academic achievement, affect, behavior, and motivation.

Many studies have shown that students with better TSRs tend to achieve more highly in school (Cornelius-White, 2007; Roorda, Koomen, Split, & Oort, 2011). Roorda et al. (2011) also report associations between negative aspects of TSR and worse academic achievement in their meta-analysis. For example, Wentzel (2002) found that middle-school students’ perceptions of their teachers on relational dimensions such as fairness and holding high expectations predicted their end-of-year grades. Estimated effect size of TSRs on achievement range from $r = .13$ to $.28^1$ for positive relationships at the secondary level (Roorda et al., 2011).

With respect to students’ affect towards school, students in classes with more supportive middle school teachers have more positive attitudes toward school (Roeser, Midgley, & Urddan, 1996; Ryan, Stiller, & Lynch, 1994) and their subject matter (Midgley, Feldlaufer, & Eccles, 1989). Conversely, middle school students who lack a bond with their teacher are more likely to disengage or feel alienated from school (Murdock, 1999). Cornelius-White’s meta-analysis (2007) showed that TSRs were correlated with students’ satisfaction with school ($r = .44^2$).

Associations between TSRs and students’ behavior include middle school students who are more willing to pay attention in class when they think their teacher cares more (Wentzel, 1997). Conversely, adolescents’ who perceived more disinterest and/or criticism from their teachers were more likely to cause discipline problems (Murdock, 1999). Cornelius-White’s (2007) findings show that more positive student perceptions of their TSRs corresponded with increased student participation ($r = .55$) and attendance ($r = .25$), and decreased disruptive behavior ($r = .25$).

Studies of TSRs and student motivation follow similar patterns. Adolescents’ perceptions of teacher support and caring predict student effort as reported by both teachers (Goodenow, 1993; Murdock & Miller, 2003) and students (Sakiz, Pape, & Hoy, 2012; Wentzel, 1997). Meta-analyses (Cornelius-White, 2007; Roorda et al., 2011) show that TSRs are associated with motivation ($r = .32$) and secondary school engagement ($r = .30$ to $.45$).

Of this array of important outcomes, we chose to focus on students’ classroom grades. Among the associations between TSRs and these outcomes, we felt grades were (arguably) the most important for students’ futures – potentially affecting advancement/retention decisions, tracking, graduation, college placement, etc.

---

1 This range represents the lower and upper bounds of the confidence intervals across both the fixed and random effects models the authors used.

2 Cornelius-White (2007) does not report elementary and secondary student results separately for his outcomes.
Scientific Context of the Study

We strove to structure this study so as to make an important theoretical contribution to the scientific study of similarity and its effect on relationships. Specifically, we hoped to learn whether the causal associations between similarity and relationships found in laboratory studies generalized to real, ongoing relationships. Furthermore, if successful, our intervention would have important applications for classrooms. Specifically, it would offer a tangible example of how similarities might be leveraged to actually improve relationships in the classroom. Simultaneously, we hoped to evaluate the effects of our intervention in as rigorous a manner as was feasible in a naturalistic setting and to err on the side of being conservative in the inferences we made from our data.

We evaluated our intervention using a 2 X 2 design. Through this design each party within every teacher-student dyad was randomly assigned to receive feedback (or not) from a “get-to-know-you” survey. Specifically, students were randomly assigned to either learn what they had in common with their teacher (i.e., students in the “Student Treatment” group), or not learn any information they had in common with their teacher (i.e., students in the “Student Control” group). Teachers found out what they had in common with about half of their students (i.e., students in the “Teacher Treatment” group) but not the other half (i.e., students in the “Teacher Control” group). Thus, all randomization occurred at the student level.

In the spirit of recent recommendations (Cumming, 2014; Simmons, Nelson, & Simonsohn, 2011), we identified six “pre-specified hypotheses” prior to analyzing our data. Specifically, we hypothesized that students in the Student Treatment group would (1) perceive themselves as more similar to their teachers and (2) report a more positive TSR as compared to those in the Student Control group. For students in the Teacher Treatment, we expected that, (3) their teacher would perceive the students as more similar, (4) their teacher would rate their TSR more positively, and the students’ (5) mid-quarter grade, and (6) final quarter grade would be higher than students whose teacher was in the Teacher Control group. As described in the Statement of Transparency in our supplemental online materials we also collected additional variables and conducted other analyses that we treat as exploratory.

Two explanatory notes about these hypotheses are in order. First, we hypothesized that students’ grades would be affected by the Teacher Treatment (but not the Student Treatment) based on previous correlational work. Gehlbach, Brinkworth, and Harris (manuscript under review) showed that when both teachers’ and students’ perceptions of their TSR are accounted for, the teachers’ perceptions (but not students’ perceptions) of the TSR are associated with students’ grades. Second, similar studies of brief interventions that have impacted students’ grades have found that the effect of the intervention was concentrated within a sub-population of students, such as African-American students (Cohen, Garcia, Apfel, & Master, 2006; Walton & Cohen, 2011), Latino students (Sherman et al., 2013), or low self-efficacy students (Hulleman & Harackiewicz, 2009). However, in the absence of information about which sub-groups might react most positively to the intervention, we made no predictions about potential sub-group effects of the intervention.
Methods

Participants

We conducted the study at a large, suburban high school in the southwestern part of the United States. We focused on ninth graders because they were just transitioning to high school and might particularly benefit from connecting with an adult in a school where they did not know any of the other authority figures. The students in our final sample \( N = 315 \) were 60% female, 51% White, 19% Latino, 11% Asian, 6% Black, and 10% reporting multiple categories or other. These proportions of different races/ethnicities are similar to the school as a whole (54% White, 20% Latino, 13% Asian, and 10% Black). These students were mostly native English speakers (81%) and came from families where college graduation represented the median educational level of the mothers and fathers (though the range included mothers and fathers who had not attended school to those who completed graduate school).

The teachers in our sample \( N = 25 \) were 52% male, 80% White, and 92% native English speakers. The mean age of the teachers was 47.5 years old \( (sd = 10.42) \), and the mean years of experience was 18.0 \( (sd = 9.5) \). Most teachers (72%) had completed a graduate degree and came from families where 1 year of college represented the median educational level for both their mothers and fathers (though the range extended from completing fourth grade to those who completed graduate school).

Measures

Our main measures were borrowed from Gehlbach, Brinkworth, and Harris (2012). Students’ perceptions of their degree of similarity to their teachers were assessed through a six-item scale \( (\alpha = .88) \), which included items such as “How similar do you think your personality is compared to your teacher’s?” Students’ perceptions of their TSR were measured with a nine-item scale \( (\alpha = .90) \) that asked students to evaluate their overall relationship with their teachers, e.g., “How much do you enjoy learning from \(<teacher's name>\)?” To minimize the burden on teachers, we asked them a single item to assess their perceptions of similarity to each student, “Overall, how similar do you think you and \(<student's name>\) are?” However, they did complete the parallel nine-item teacher-form of the TSR scale \( (\alpha = .86 \text{ for teachers}) \). Mid-quarter and final quarter grades were collected from student records.

Our exploratory analyses employed several additional measures. Teachers rated the amount that they interacted with their students by answering, “Compared to your average student, how much have you interacted with \(<student's name>\) this marking period?” Teachers and students rated the amount of homework the student submitted by responding to, “Approximately what percentage of the assigned homework \(<does student's name / do you> complete for your class?” We also collected attendance and tardy data and (eventually) end-of-semester grades from school records.

Procedure

The study unfolded over the course of the first marking period at the school. Just prior to the beginning of the school year, the principal helped our research team recruit as many ninth grade teachers as were willing to participate. In turn, during the first week of school these 27
consenting teachers helped us collect consent forms from their students. Throughout the following week of school, these students and teachers visited their computer lab and completed the initial get-to-know-you survey. By the middle of the third week of classes, the school had received the feedback forms, which were then completed by 315 students and 24 teachers over the course of the next two weeks. An additional teacher submitted her feedback sheet late (though her students completed their sheets on time); this teacher and her students were retained in the sample. Two teachers and their classes never completed the feedback forms, thereby reducing the final sample size to 315 students and 25 teachers. Mid-quarter progress grades were finalized at the end of the fifth week of classes. During the eighth and ninth weeks of classes, students and teachers took the follow-up survey. (Because teachers were allowed to take the survey on their own time, some teachers completed the follow up survey up to one month later). The quarter concluded at the end of the tenth week of classes.

Students and teachers took the 28 item get-to-know-you survey during their first period class. The survey asked teachers and students what they thought the most important quality in a friend was, which class format is best for student learning, what they would do if the principal announced that they had a day off, which foreign languages they spoke, and so on (See Figure 1). From these surveys we composed the feedback sheets that comprised the core of the intervention.

On these feedback sheets, students were assigned to either receive feedback about five things they had in common with their teacher (in the Student Treatment group3) or to receive five commonalities they shared with students at a school in another state (in the Student Control group). Each teacher received five items that they had in common with each student who was randomly selected from half of their participating first-period students (in the Teacher Treatment group). Teachers were informed that in the interest of providing prompt feedback, we could not provide reports on their remaining first-period students (the Teacher Control group). The five similarities were chosen based on an approximate rank ordering of the similarities that had seemed to be most important for generating perceptions of similarity from the pilot test in the previous year. Students and teachers responded to a series of brief questions on their feedback sheets such as, “Looking over the five things you have in common, please circle the one that is most surprising to you.” Our hope was that by completing these questions on their feedback sheets, students and teachers would more deeply consider and better remember their points of commonality with one another. Copies of the measures and materials are available from the first author upon request.

Results and Discussion

Pre-specified Hypotheses

As detailed in our “Statement of Transparency” (see the supplemental online materials), we pre-specified six hypotheses (Cumming, 2014). Specifically, we anticipated that (as compared to those in the Student Control group) students in the Student Treatment group (1) would perceive more similarities and (2) a more positive TSR with their teacher. As compared to those in the Teacher Control group, we hypothesized that teachers would perceive students in

---

3 We were able to generate five similarities for all but one teacher-student dyad, where only four similarities were present. This dyad was retained in our analyses.
the Teacher Treatment group as (3) being more similar, and (4) teachers would develop a more positive TSR with these students. Finally, we expected that the students in the Teacher Treatment group would earn (5) higher mid-quarter and (6) higher end-of-quarter grades than their counterparts in the Teacher Control group.

As described in the Statement of Transparency we expected to test these hypotheses through a combination of multi-level modeling (i.e., hypotheses 3, 5, and 6 when the outcome was a single item) and multi-level structural equation modeling (i.e., hypotheses 1, 2, and 4 when the outcome was a latent variable). However, our statistical consultant advised us that the number of teachers (i.e., level 2 clusters) was inadequate for Mplus to provide trustworthy estimates for the models using latent variables. Our models for latent variables had more parameters to be estimated than clusters, making multilevel SEM impossible. Due to this nested structure of our data, we relied on mean- and variance-adjusted weighted least squares for complex survey data (WLSMV-complex) estimation, using the CLUSTER option in Mplus. WLSMV-complex, which uses a variance correction procedure to account for clustered data, provides corrected standard errors, confidence intervals, and coverage (Asparouhov, 2005). We used full information maximum likelihood (FIML) to address missing data. The maximum proportion of missing data for any variable was .012. However, we used Mplus’ robust standard error approach when our outcomes were latent. To evaluate each hypothesis, we regressed the outcome on the condition as described above. Because random assignment produced equivalent groups between both treatment groups and their respective control groups on key demographic characteristics (specifically gender, race, English language status, and parents’ educational level), no covariates were used in these analyses. Consistent with Cumming’s (2014) recommendation, we evaluated our hypotheses using 95% confidence intervals rather than p-values and report standardized $\beta$ to provide an estimate of effect size (except for grade-related outcomes where the original 0 to 4.0 scale provides meaningful equivalents of an F through an A). Descriptive statistics are presented in Table 1.

Our results are congruent with the similarity hypotheses (i.e., hypotheses 1 and 3). Each treatment made students and teachers feel more similar to one another by the end of the marking period ($\beta = 0.33, SE = 0.12, CI: 0.10, 0.56$ for students; and $\beta = 0.33, SE = 0.11, CI: 0.11, 0.55$ for teachers). However, the students perceived their TSRs to be approximately the same regardless of the condition to which they were assigned ($\beta = 0.09, SE = 0.14, CI: -0.18, 0.36$). In other words, we found little support for hypothesis 2. Within the Teacher Treatment, teachers perceived a more positive relationship with these students ($\beta = 0.21, SE = 0.11, CI: 0.00, 0.42$). For students in the Teacher Condition, we found no support for an effect on mid-quarter grades ($\beta = 0.04, SE = 0.10, CI: -0.15, 0.23$). Yet, we saw evidence of a modest boost in end-of-quarter grades ($\beta = 0.21, SE = 0.11, CI: 0.00, 0.37$), although the confidence interval does include 0. Figures 1-4 in the supplementary online materials show how the unadjusted means are distributed when the Teacher and Student Conditions are separated into their four unique groupings of the 2 X 2 design.

The first pair of findings is that the intervention appears to have successfully enhanced teachers’ and students’ perceptions of similarity. On the one hand, the effects do not seem particularly potent – perhaps reflecting only a mildly-to-moderately strong intervention. On the other hand, students processed their feedback sheets for approximately fifteen minutes before handing them back in, and yet, still perceived themselves as being more similar to their teacher
over a month later. Teachers presumably spent even less time on each feedback sheet given that most teachers had several to complete. Thus, while one might argue that the effects of the intervention were weak, this interpretation should be calibrated against the brevity of the intervention and the amount of time that elapsed before the outcomes were collected (Cumming, 2014).

Although the intervention appeared to improve teachers’ perceptions of their relationships with students, we find no evidence that the intervention improved TSRs from students’ perspectives. To the extent that this result reflects a genuine difference in the effect of the intervention, one plausible explanation is that teachers view part of their role as needing to foster positive relationships with students. Thus, they are motivated to perceive students who they view as similar in a positive light. By contrast, students may not feel any particular obligation to form a positive relationship with their teachers. Learning that they share common ground with their teacher may do nothing to change their perception of their teacher because students have no particular motivation to cultivate this social relationship.

Our findings for students’ academic achievement seem paradoxical: we found positive effects of the intervention after finding no effects. However, we think this apparent paradox is relatively easily explained. In an unfortunate oversight, we finalized our pre-specified hypotheses prior to reviewing the timing of each key aspect of the study. Although direction of the estimate for students’ mid-quarter grades is the same as the end-of-quarter grade, we suspect that the intervention occurred too close to the deadline for teachers to submit their grades to have a meaningful effect. In other words, students did not have a sufficient opportunity to do enough graded work between the time that they (and their teachers) completed their feedback sheets and the date that mid-quarter grades were due. As a result, we do not discuss this outcome further. Students’ performance on their final quarter grades, by contrast, showed that the intervention did cause students’ grades to increase. Our point estimate of this increase corresponds to a little less than a fifth of a letter grade.

To better understand our initial pattern of results, we examined whether our intervention might have had differential effects on different sub-populations of students. By having Mplus fit a series of multi-level models (for observed outcomes) and models with robust standard errors (for latent outcomes), we conducted a series of exploratory analyses on different student subgroups.

Exploratory Analyses

A number of previous studies that employ relatively brief, social psychological interventions (Cohen et al., 2006; Hullemen & Harackiewicz, 2009; Sherman et al., 2013; Walton & Cohen, 2011) suggest that certain subgroups of students often benefit disproportionately from the interventions. Specifically, we thought that the school might serve some students better than others, or that there might be a dominant culture at the school that was more inclusive of some students than others. After speaking with the principal about this possibility, he suggested that the White and Asian students were typically well-served by the school, while Black and Latino students typically faced more challenging circumstances at home, school, and throughout their community. Thus, we re-examined our data by analyzing the White and Asian students as a separate group from the remaining “underserved” students. Because these are exploratory analyses, we do not retain the same level of confidence in these findings as
our pre-specified hypotheses. However, we argue that these results are likely to be instructive for generating future hypotheses (Cumming, 2014).

When fitting our models, we found little evidence for any effects of the intervention on the White and Asian students. We find no evidence that White and Asian students in the Student Treatment group perceived different levels of similarity with their teachers ($\beta = 0.17, SE = 0.15, CI: -0.13, 0.46$) or felt their relationships to be different ($\beta = -0.12, SE = 0.17, CI: -0.46, 0.21$) as compared to those in the Student Control group. We find a comparable lack of evidence that the intervention affected teachers’ perceptions of their similarity to their White and Asian students ($\beta = 0.11, SE = 0.16, CI: -0.20, 0.41$) and teachers’ perceptions of their relationships with these students ($\beta = 0.00, SE = 0.15, CI: -0.29, 0.29$). Finally, we find no evidence that the intervention affected White and Asian students’ end-of-quarter grades ($\beta = -0.01, SE = 0.15, CI: -0.29, 0.27$).

For the underserved students, the story differed. Underserved students who received feedback about commonalities with their teachers felt much more similar to their teachers ($\beta = 0.56, SE = 0.20, CI: 0.18, 0.96$) than their counterparts who did not receive this feedback. It was less clear whether these students felt more positive about their relationships with their teachers ($\beta = 0.39, SE = 0.24, CI: -0.08, 0.86$), though the estimated effect size was moderate and in the expected direction. When teachers received feedback about similarities with their underserved students, they perceived greater levels of similarity with those students as compared to their control counterparts ($\beta = 0.56, SE = 0.24, CI: 0.08, 1.04$). Although the confidence interval does include 0, our results suggest that they felt more positive about their TSRs with these students ($\beta = 0.43, SE = 0.27, CI: -0.11, 0.96$). Finally, underserved students’ end-of-quarter grades ($\beta = 0.36, SE = 0.20, CI: -0.04, 0.75$) were higher when their teacher received feedback about their commonalities as compared to students in the Teacher Control condition. As depicted in Figure 2, this difference translates into about .4 of a letter grade on a 4.0 scale and corresponds to the difference between a C+/B- versus a B.

These effects on grades are substantial. If we compare the White and Asian students versus the underserved students in Figure 2, we can estimate the achievement gap between well-served and underserved ninth graders at this school to be approximately .6 of a letter grade. When teachers learned about the similarities that they shared with their underserved students, the achievement gap was reduced by two-thirds to only .2 of a letter grade. This reduction is in line with other relatively brief interventions that have closed the achievement gap. For example, Cohen et al. (2006) report a 40% closure with an even briefer intervention; Walton and Cohen (2011) report a 52% to 79% reduction (depending upon the time period examined) from their more intensive intervention.

Given the potential importance of these differences, we carried out two final sets of analyses. First, in order to see the extent to which these results persisted over time, we obtained students’ grades in their focal class for the full semester. These analyses showed that the effects of the intervention on the underserved students trended in the same direction as the results for students’ end-of-quarter grades ($\beta = 0.33, SE = 0.22, CI: -0.11, 0.77$).

Second, in anticipation of trying to understand more about the effect of the intervention, we tested whether the intervention appeared to affect other variables we had collected. In particular, we examined attendance and tardy data from school records; how much teachers
reported interacting with each student as compared to the average student; and how much homework the student submitted. The results from these analyses suggest that the intervention did not affect students’ attendance in their focal class (see Figures 4a and 4b in the supplemental online materials). However, the previously noted subgroup differences emerged in how much teachers interacted with their students. Specifically, teachers reported no differences by condition in how much they interacted with their White and Asian students ($\beta = -0.13, SE = 0.16, CI: -0.43, 0.17$), but they interacted more with their underserved students if in the Teacher Condition ($\beta = 0.43, SE = 0.16, CI: 0.12, 0.74$).

**Conclusion**

Our study builds on the robust social psychological research showing that similarity fosters liking and more positive relationships. By experimentally manipulating teachers’ and students’ perceptions of actual similarities, our study allows for causal inferences to be made about the effects of similarity on real-world, ongoing relationships. Results from our pre-specified hypotheses suggest that the intervention alters students’ and teachers’ perceptions of how much they have in common, benefits TSRs (from the teacher’s perspective), and bolsters students’ classroom grades.

Like any study, ours includes several limitations that warrant readers’ attention. First, the implementation of the various steps of the intervention was imperfect (e.g., a teacher failing to complete the feedback sheets on time, two other teachers responding to the final survey late, etc.). We hope that future studies can remedy these problems and design systems to administer the intervention consistently. However, we also note that implementation of all manner of interventions (new curricula, disciplinary systems, web portals for parents, and so on) in schools tend to be imperfect. The fact that our intervention was effective despite the flaws in execution is an important footnote for practitioners.

Second, our analyses (particularly the exploratory analyses) lacked the statistical power we desired. This caused us to shift to a different statistical approach than the one we had originally planned on in our statement of transparency. Our statistical consultant also noted that the multi-level model and clustered standard error approaches we employed, may still result in too many Type-I errors when the number of clusters is small, i.e., fewer than 50 (see, for example, Bertrand, Duflo, & Mullainathan, 2004; Donald & Lang, 2007). To address this potential limitation, we employed a wild cluster bootstrap-t (Cameron, Gelbach, & Miller, 2008). As shown in the Table 1 in the appendix, our findings using that approach were generally consistent with those we obtained from our multi-level model and robust standard error models. Particularly given the emerging hypothesis that the effectiveness of the intervention may be localized to underserved students, future replications should try to obtain substantially larger samples with more clusters across a variety of schools to better evaluate this possibility.

Third, our exploratory findings suggested differences between well-served (White and Asian) and under-served (primarily Black and Latino) students. However, this division of students may mask a more accurate understanding of what moderates the effects of the intervention. For example, we lacked a reasonable measure of socio-economic status in our data set. Given the correspondence between race and socio-economic status in this country, it may be that we actually detected a moderating effect that would be more accurately described as
economic- rather than race-based. Thus, future studies that can collect a wider array of more precise demographic measures would also be particularly beneficial.

Other future directions emerge out of the data themselves. First, the Teacher Treatment seemed to yield a greater effect on our outcomes than the Student Treatment. When teachers learned what they had in common with their students, they felt they had more in common with those students, perceived better relationships with them, and those students got better grades. Although more speculative, it appears that the effects of the Teacher Treatment were concentrated among the underserved students. Thus, one set of future studies might investigate whether the effects of the intervention are really concentrated on teachers and underserved students, or whether this finding varies by context or population. Another set of studies could investigate whether the intervention might be adapted to improve students’ perceptions of the relationship or to make it effective for all students rather than just a subset of students. Additional research might investigate the role of teachers’ race and/or the congruence between students’ and teachers’ race on the effectiveness of the intervention.

Second, future studies should investigate why these effects are occurring. One potential explanation is rooted in interactions. Many teachers may see it as a part of their role to connect with students and form a positive working relationship. Knowing what they have in common with their students provides them with a lever through which they can begin developing this relationship. For a group of predominantly white teachers, learning what they have in common with their underserved students may be critically important. Indeed, we find that teachers report interacting with these students more frequently. By contrast, ninth graders (regardless of race) may have little interest in connecting with their teachers or having any more interactions than necessary. They might be much more focused on connecting with their peers during this developmental stage.

An alternative explanation is rooted in perceptual biases. Perhaps teachers typically perceive their students – particularly their underserved students – in stereotypical fashion. However, when they realize several domains in which they share some common ground with these students, the teachers begin to perceive their relationship more like members of their own in-group (Hewstone, Rubin, & Willis, 2002) and reward them with higher grades as a consequence.

Our exploratory analyses suggest that the possibility of perceptual biases will also be an important, challenging area of future investigation. On the one hand, we might expect students, who are welcomed into a classroom where the teacher more frequently interacts with them in positive ways, to attend class more regularly and arrive on time more often. While we did not find much evidence congruent with this conjecture, there are many factors that affect a student’s presence in class.

On the other hand, the perceptual bias story is not completely congruent with the finding that teachers report interacting more frequently with students in the Teacher Treatment condition than with their control group peers. Thus, research that can begin to shed light on the mechanisms – be they teacher-student interactions, teacher perceptions, a combination of both, or other factors – through which this intervention affects these important outcomes of teacher-student relationships and grades will be especially fruitful.
In closing, this study contributes an important new finding that (perceptions of) real similarities can be manipulated to affect real relationships in a consequential setting like a high school. Our findings offer suggestive evidence that the improvements in TSRs may, in turn, cause downstream benefits for students’ grades. Finally, these results generate strong hypotheses that similar interventions in the future may be effective in helping to close achievement gaps between subgroups of students.
References


Gehlbach, H., Brinkworth, M. E., & Harris, A. D. (manuscript under review). Understanding teacher-student relationships and student outcomes: The positives and negatives of assessing both perspectives.


Table 1: Descriptive statistics for key variables in the study (unadjusted *mean*, *sd*, and Pearson (*r*) correlations).

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Mean</th>
<th>sd</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>1) Students' similarity</td>
<td>2.68</td>
<td>0.73</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2) Teachers’ similarity</td>
<td>2.90</td>
<td>0.91</td>
<td>.13</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Students’ TSR</td>
<td>3.68</td>
<td>0.68</td>
<td>.69</td>
<td>.18</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Teachers' TSR</td>
<td>3.85</td>
<td>0.55</td>
<td>.29</td>
<td>.63</td>
<td>.32</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5) Mid-quarter Grade</td>
<td>3.26</td>
<td>0.99</td>
<td>.34</td>
<td>.23</td>
<td>.35</td>
<td>.41</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6) End-of-quarter Grade</td>
<td>3.16</td>
<td>1.10</td>
<td>.30</td>
<td>.18</td>
<td>.35</td>
<td>.43</td>
<td>.76</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7) Semester grade</td>
<td>2.79</td>
<td>1.11</td>
<td>.24</td>
<td>.31</td>
<td>.28</td>
<td>.47</td>
<td>.67</td>
<td>.79</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8) Tardies</td>
<td>0.26</td>
<td>0.66</td>
<td>-.13</td>
<td>-.01</td>
<td>-.08</td>
<td>-.05</td>
<td>-.20</td>
<td>-.22</td>
<td>-.13</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>9) Absences</td>
<td>1.29</td>
<td>1.61</td>
<td>-.15</td>
<td>-.08</td>
<td>-.06</td>
<td>-.16</td>
<td>-.20</td>
<td>-.15</td>
<td>-.10</td>
<td>.15</td>
<td>--</td>
</tr>
<tr>
<td>10) Teacher reported</td>
<td>4.74</td>
<td>1.10</td>
<td>.21</td>
<td>.37</td>
<td>.17</td>
<td>.46</td>
<td>.16</td>
<td>.22</td>
<td>.21</td>
<td>-.11</td>
<td>-.10</td>
</tr>
<tr>
<td>interactions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11) Student reported homework</td>
<td>94.14</td>
<td>12.63</td>
<td>.16</td>
<td>.14</td>
<td>.15</td>
<td>.19</td>
<td>.38</td>
<td>.36</td>
<td>.30</td>
<td>-.17</td>
<td>-.07</td>
</tr>
<tr>
<td>completion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12) Teacher reported</td>
<td>91.90</td>
<td>12.31</td>
<td>.29</td>
<td>.29</td>
<td>.29</td>
<td>.50</td>
<td>.64</td>
<td>.71</td>
<td>.59</td>
<td>-.06</td>
<td>-.18</td>
</tr>
<tr>
<td>homework completion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: 1) Ns ranged from 262-362.
2) Correlations are unadjusted for the nesting of students within classrooms.
3) Approximate significance levels are as follows: for |rs| ranging from 0 to .12, *p* = *ns*; for |rs| ranging from .13 to .16, *p* < .05; for |rs| ranging from .17 to .20, *p* < .01; for |rs| .21 and greater, *p* < .001.
Figure 1: Screen shot of the get-to-know-you survey.

The most important quality in a friend is:
- Being there when you need him/her
- Listens to you and understands you
- Always has your back

If you could have one thing in common with your teacher, which of the following would it be?
- Sense of humor
- Interest in the same subject matter
- Mutual respect
- Similar personality

If you could go to one sporting event, which of the following would you go to:
- World Cup Soccer
- Olympics
- NBA Championship
- Super Bowl
- World Series
Figure 2: Mean differences and 95% confidence intervals for underserved students by Teacher Condition in teachers’ perceptions of similarity, perception of their teacher-student relationships (TSR), and students’ end-of-quarter grades in their focal class. Means for White and Asian students are presented for comparison.

Notes: The 65% reduction in the achievement gap shown in the right-hand triad of bars corresponds to the difference between less than a B- to a B.
### Table 1: Results from re-analyses using a wild cluster bootstrap-t.

<table>
<thead>
<tr>
<th></th>
<th>Full sample</th>
<th>Well-served students</th>
<th>Under-served students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta$</td>
<td>[95% CI]</td>
<td>$\beta$</td>
</tr>
<tr>
<td>Students’ Perceived Similarity for Student Treatment</td>
<td>0.18</td>
<td>[0.05, 0.33]</td>
<td>0.11</td>
</tr>
<tr>
<td>Teachers’ Perceived Similarity for Students in Teacher Treatment</td>
<td>0.28</td>
<td>[0.07, 0.49]</td>
<td>0.05</td>
</tr>
<tr>
<td>Students’ Perceived Teacher-student Relationship for Student Treatment</td>
<td>0.05</td>
<td>[-0.13, 0.23]</td>
<td>-0.08</td>
</tr>
<tr>
<td>Teachers’ Perceived Teacher-student Relationship for Students in Teacher Treatment</td>
<td>0.15</td>
<td>[-0.01, 0.31]</td>
<td>0.00</td>
</tr>
<tr>
<td>End-of-quarter Grades for Students’ in Teacher Treatment</td>
<td>0.21</td>
<td>[-0.01, 0.41]</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

Notes: The standardized bootstrap relies on the bootstrap-implied distribution of a t-statistic rather than a beta estimate (Cameron et al., 2008), and so we do not report the standard errors of the t-statistic; the bootstrap makes no assumptions about the normality or even symmetry of the sampling distribution, and so standard errors cannot be used to calculate confidence intervals or conduct hypothesis tests.
Table 2: Raw (unadjusted for nesting) means of key variables by Student and Teacher Conditions: **Mean**, (Standard Errors), and [95% Confidence Intervals].

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Student Treatment Only</th>
<th>Teacher Treatment Only</th>
<th>Student &amp; Teacher Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 60)</td>
<td>(n = 63)</td>
<td>(n = 69)</td>
<td>(n = 62)</td>
</tr>
<tr>
<td>Students’ Perceived Similarity</td>
<td><strong>2.56</strong> (0.09)</td>
<td><strong>2.74</strong> (0.09)</td>
<td><strong>2.58</strong> (0.09)</td>
<td><strong>2.78</strong> (0.09)</td>
</tr>
<tr>
<td></td>
<td>[2.37, 2.74]</td>
<td>[2.56, 2.92]</td>
<td>[2.41, 2.75]</td>
<td>[2.61, 2.96]</td>
</tr>
<tr>
<td>Teachers’ Perceived Similarity</td>
<td><strong>2.61</strong> (0.11)</td>
<td><strong>2.88</strong> (0.11)</td>
<td><strong>2.99</strong> (0.10)</td>
<td><strong>3.01</strong> (0.11)</td>
</tr>
<tr>
<td></td>
<td>[2.39, 2.83]</td>
<td>[2.67, 3.10]</td>
<td>[2.78, 3.19]</td>
<td>[2.86, 3.30]</td>
</tr>
<tr>
<td>Students’ Perceived Teacher-student Relationship</td>
<td><strong>3.70</strong> (0.09)</td>
<td><strong>3.71</strong> (0.09)</td>
<td><strong>3.59</strong> (0.08)</td>
<td><strong>3.71</strong> (0.09)</td>
</tr>
<tr>
<td></td>
<td>[3.53, 3.88]</td>
<td>[3.54, 3.88]</td>
<td>[3.43, 3.75]</td>
<td>[3.53, 3.88]</td>
</tr>
<tr>
<td>Teachers’ Perceived Teacher-student Relationship</td>
<td><strong>3.64</strong> (0.06)</td>
<td><strong>3.86</strong> (0.06)</td>
<td><strong>3.83</strong> (0.06)</td>
<td><strong>3.90</strong> (0.06)</td>
</tr>
<tr>
<td></td>
<td>[3.52, 3.77]</td>
<td>[3.73, 3.98]</td>
<td>[3.71, 3.95]</td>
<td>[3.77, 4.02]</td>
</tr>
<tr>
<td>Students’ Mid-quarter Grades</td>
<td><strong>3.16</strong> (0.12)</td>
<td><strong>3.21</strong> (0.12)</td>
<td><strong>3.23</strong> (0.11)</td>
<td><strong>3.22</strong> (0.12)</td>
</tr>
<tr>
<td></td>
<td>[2.93, 3.39]</td>
<td>[2.98, 3.44]</td>
<td>[3.01, 3.44]</td>
<td>[2.99, 3.45]</td>
</tr>
<tr>
<td>Students’ End-of-quarter Grades</td>
<td><strong>2.95</strong> (0.13)</td>
<td><strong>2.99</strong> (0.13)</td>
<td><strong>3.16</strong> (0.12)</td>
<td><strong>3.13</strong> (0.13)</td>
</tr>
<tr>
<td></td>
<td>[2.69, 3.20]</td>
<td>[2.74, 3.24]</td>
<td>[2.92, 3.40]</td>
<td>[2.88, 3.39]</td>
</tr>
<tr>
<td>Number of Student Tardies</td>
<td><strong>0.29</strong> (0.08)</td>
<td><strong>0.28</strong> (0.08)</td>
<td><strong>0.30</strong> (0.08)</td>
<td><strong>0.21</strong> (0.09)</td>
</tr>
<tr>
<td></td>
<td>[0.12, 0.45]</td>
<td>[0.13, 0.44]</td>
<td>[0.14, 0.46]</td>
<td>[0.03, 0.38]</td>
</tr>
<tr>
<td>Number of Student Absences</td>
<td><strong>1.32</strong> (0.19)</td>
<td><strong>1.18</strong> (0.19)</td>
<td><strong>1.41</strong> (0.19)</td>
<td><strong>1.24</strong> (0.20)</td>
</tr>
<tr>
<td></td>
<td>[0.94, 1.70]</td>
<td>[0.81, 1.54]</td>
<td>[1.03, 1.79]</td>
<td>[0.84, 1.64]</td>
</tr>
</tbody>
</table>

Notes: To facilitate the review process, we are presenting these more comprehensive tables in place of the series of figures described in the Statement of Transparency. We are happy to include either for the final publication.
Figures 1a and 1b: Students’ and teachers’ perceptions of similarity to one another by condition (Mean and 95% CI).
Figures 2a and 2b: Students’ and teachers’ perceptions of their teacher-student relationship by condition (Mean and 95% CI).
Figures 3a and 3b: Students’ mid-quarter and end-of-quarter grades in their focal class by condition (Mean and 95% CI).
Figure 4a and 4b: Students’ tardiness and attendance by condition (Mean and 95% CI).