

RESEARCH ARTICLE

Racial and linguistic ideologies and formative assessment practices of U.S. science teachers: A preregistered conceptual replication and extension

Quentin C. Sedlacek^{1*}, Nickolaus Ortiz², Catherine Lemmi¹, Maricela León¹, Kimberly C. Feldman⁴

1 Department of Teaching & Learning, Annette Caldwell Simmons School of Education & Human Development, Southern Methodist University, Dallas, Texas, United States of America, **2** Department of Middle and Secondary Education, Georgia State University, Atlanta, Georgia, United States of America, **3** School of Education, California State University, Chico, Chico, California, United States of America, **4** Southwest Academy, Baltimore County Public Schools, Woodlawn, Maryland, United States of America

* qsedlacek@smu.edu



OPEN ACCESS

Citation: Sedlacek QC, Ortiz N, Lemmi C, León M, Feldman KC (2025) Racial and linguistic ideologies and formative assessment practices of U.S. science teachers: A preregistered conceptual replication and extension. PLoS One 20(7): e0327859. <https://doi.org/10.1371/journal.pone.0327859>

Editor: Mary Diane Clark, Lamar University, UNITED STATES OF AMERICA

Received: January 20, 2025

Accepted: June 23, 2025

Published: July 17, 2025

Copyright: © 2025 Sedlacek et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Pending publication, the materials used and fully de-identified dataset generated during this study will be available from the OSF repository at <https://osf.io/fjwbv/>.

Abstract

Language, race, and racism are deeply connected in the United States and throughout the world. Over the past 10 years, a growing body of research has argued that many language ideologies and racial ideologies are closely associated with each other. However, few studies have sought to quantify these proposed associations by directly measuring the language ideologies, racial ideologies, and language-related behaviors of participants. In a recently published study of K-12 preservice teachers, we identified substantial correlations between racial ideologies and seemingly race-neutral language ideologies, as well as correlations between these language ideologies and the written feedback participants generated during formative assessment of student science writing. These findings have important implications for the social sciences; however, their significance is contingent in part on their reproducibility and generalizability. For this reason, we conducted a preregistered conceptual replication and extension using a larger sample of current U.S. science teachers ($n=387$ teachers from 42 U.S. states). We demonstrate that all but one of the findings from our previous study are successfully replicated, and we identify several new findings. We discuss implications for research in science education, linguistic anthropology, social psychology, and sociology.

Introduction

Language, race, and racism are deeply connected in the United States and throughout the world. Over the 10 years since Flores and Rosa [1] coined the term *raciolinguistic ideology*, a growing body of research has argued that many language

Funding: The research reported in the article was made possible by grant #202100261 to author QS from the Spencer Foundation (www.spencer.org). Additional funding was provided by Southern Methodist University through a performance-based funding grant to author QS. The views expressed are those of the authors and do not necessarily reflect the views of the Spencer Foundation nor those of Southern Methodist University. The sponsors and funders played no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

ideologies and racial ideologies are closely associated with each other. Language ideologies are systems of beliefs about language(s) and their speaker(s) that can influence social interaction—for example, beliefs about the superiority of standardized forms of language over non-standardized forms [1–4]. Similarly, racial ideologies are systems of beliefs about the nature of race and racial identity, such as the belief that race is a “natural” category grounded in biology or the belief that race is an arbitrary category constructed by human societies [5,6]. Research exploring associations between racial ideologies and language ideologies has largely been qualitative in nature [7], although its key tenets are supported by past quantitative research (e.g., [8]). A small but growing body of quantitative research has sought to empirically explore raciolinguistic phenomena such as racially biased perceptions of language, usually spoken language (e.g., [8–10]). However, these studies typically do not directly measure the language ideologies and racial ideologies believed to be associated with such biases, instead typically measuring biased behavioral responses only.

Quantifying relationships between language ideology, racial ideology, and behavior could have important implications for multiple fields of social science. Language ideologies have been the subject of extensive qualitative (and some quantitative) research in fields such as sociolinguistics and linguistic anthropology [11], while racial ideologies and racially biased behaviors have been the subject of extensive qualitative and quantitative research in social psychology [12–14] and sociology [15–17]. Combining these fields of research using raciolinguistic ideology theory could prove highly generative. While qualitative research has convincingly demonstrated the existence and character of raciolinguistic ideologies, quantitative work could further our understanding of the scope and scale of these ideologies and the inequities they perpetuate. In doing so, quantitative work on raciolinguistic ideologies could help inform strategies for researchers, policymakers, and practitioners to address these ideologies, and could help develop tools for assessing the effects of specific pedagogical or policy interventions.

In a recently published study [18], our team began to address this gap in the specific context of U.S. K-12 science education, using quantitative surveys and performance tasks to investigate ideology and behavior among a population of K-12 preservice teachers. Abundant research suggests that seemingly race-neutral beliefs often mask covert forms of racism, as shown in a meta-analysis by Yi et al. [12], and our work similarly found substantial correlations between problematic racial ideologies and seemingly race-neutral prescriptive beliefs about science writing [18]. We also found associations between these ideologies and the written feedback participants generated during formative assessment of student science writing. Our findings go beyond previous quantitative raciolinguistics research that has measured accentedness judgments (e.g., [8,9]) or evaluative responses to language (e.g., [10]), and provides one of the first quantitative tests of the premise that behavioral responses to language are related to both language ideology and racial ideology. Our work thus has significant implications for sociolinguistics, linguistic anthropology, social psychology, and sociology, as well as science education and teacher education.

The significance of these findings is contingent in part on the idea that they are reproducible and generalizable to broader populations, such as in-service science teachers or American adults more generally. Many scholars believe science suffers from a reproducibility crisis [19], and the generalizability of our previous study was constrained by its small sample (65 pre-service teachers of science and other subjects from 2 college campuses). To better establish the reproducibility and generalizability of our findings, we sought to conceptually replicate and extend these findings in a larger preregistered study of current U.S. science teachers ($n=387$ teachers from 42 U.S. states).

Ideology and writing in science education

Language practices associated with racially marginalized communities, such as African American Language (AAL), are often heavily stigmatized in classrooms and other social contexts [1,20,21]. There is a broad consensus among linguists that such stigma is not grounded in any empirical research on language and instead is simply an ideology driven by historical racism [20,21]. As early as the 1970s, national professional organizations devoted to language arts education issued statements affirming students' right to use and maintain their language practices [22,23], and many linguists and educators assert that AAL is a valuable asset that can support students' language learning, content learning, and/or identity development in schools [18,20,22–27]. Despite this longstanding scholarly consensus, U.S. schools often implicitly or explicitly prohibit using AAL in classrooms, curtailing students' opportunities for sensemaking [20,23,24]. These prohibitions are harmful and counterproductive; research has shown that more expansive approaches to language, in which students are allowed or encouraged to draw on multiple aspects of their linguistic repertoire, may serve valuable academic and social purposes in the classroom, including for science and mathematics learning [28–32].

While teachers' language ideologies and responses to AAL have frequently been researched in language arts education (e.g., [27,33]), they have rarely been studied in other K-12 content areas such as STEM education [4,18,34]. However, language ideologies represent critical and necessary areas of focus even within these other content areas. For example, many science educators believe students should use formal scientific language in class, sometimes at the exclusion of other varieties of language [18,35], despite research showing that students benefit from opportunities to paraphrase or translate scientific ideas into their own words [30,36,37]. We refer to this notion—that students should exclusively or almost exclusively use relatively formal scientific language when writing or speaking in science courses—as a “language-exclusive ideology” [4], encompassing both prescriptive beliefs about language and restrictive pedagogical practices or policies aligned with those beliefs. Studying language-exclusive ideologies is important because these ideologies might impede the use of AAL as an asset for science learning by driving teachers to criticize or prohibit speaking or writing using AAL, shutting down valuable opportunities for student sensemaking and identity development [18].

Language-exclusive ideology might do more than simply prescribe the way teachers believe students ought to write or speak; it may also influence how teachers themselves read and listen to students' language use. In doing so, it may interfere with formative assessment in science and other content areas. In a recent mixed-methods study of 65 preservice teachers of science and other subjects [18], we found that participants who expressed a stronger language-exclusive ideology in a survey appeared more prone to making questionable inferences about students' scientific understanding based on purported samples of student writing during a formative assessment performance task. Participants who expressed stronger language-exclusive ideology also tended to write shorter feedback on all student science writing samples, despite often professing to believe that feedback on writing is essential to support student learning. Although this association does not necessarily imply a causal relationship, it suggests the possibility that language-exclusive ideologies might function in ways that limit the amount of useful feedback students receive on their science writing. Finally, these same language-exclusive participants appeared prone to giving feedback in racially biased ways—often “correcting” students' use of a valid and systematic grammatical feature associated with African American Language as though it were an “error,” yet often overlooking “errors” that were racially unmarked, i.e., deviations from standard spelling, grammar, or punctuation that are not specifically associated with African American Language in U.S. society. These patterns appeared to reflect

differences in ideology rather than differences in pedagogical content knowledge. We found that our survey scale measuring language-exclusive ideology was a stronger predictor of these patterns compared with a previously validated scale measuring science teachers' knowledge of language as an epistemic tool for teaching science [18,35].

We also found that seemingly race-neutral beliefs about science writing pedagogy measured through our language-exclusive ideology scale (e.g., the belief that students should avoid using personal pronouns in science writing) were, in fact, substantially correlated with teachers' racial ideologies and their asset-based attitudes (or lack thereof) towards African American Language. Teachers with stronger language-exclusive ideology tended to express stronger agreement with racial essentialism, the ideology that race is a biologically real and meaningful category rather than a social construct. Teachers with stronger language-exclusive ideology were also less likely to view African American Language as a valid form of language and a potential asset in STEM subjects such as science or mathematics. Such correlations between language ideologies and racial ideologies have been widely theorized and qualitatively studied in the decade since Flores and Rosa [1] coined the term "raciolinguistic ideology." However, the relationships between language ideologies and racial ideologies have rarely been studied quantitatively.

Motivation and research questions

These findings may have important implications for teacher education, linguistic anthropology, social psychology, and sociology. However, their significance is in part contingent on the idea that these findings are reproducible and generalizable to broader populations of science teachers, teachers of other subjects, or American adults more generally. Many scholars believe science suffers from a reproducibility crisis [19], and the generalizability of our previous study was limited by its relatively small and heterogeneous sample of 65 pre-service teachers of science and other subjects. To better establish the reproducibility and generalizability of our findings, we sought to conceptually replicate and extend them in a larger preregistered study of current U.S. science teachers ($n=387$). We pre-registered our study by posting our research design, sampling plan, instrumentation, and a detailed analysis plan through the Open Science Framework; pre-registering an analysis plan strengthens confidence in research findings by minimizing the risk of "p-hacking" and publication bias [38].

The current study is a *conceptual* replication, rather than an *exact* replication, because we used the same instrumentation and variables of interest as our original study [18] but replaced two covariates from our original study with covariates more appropriate for a larger sample of in-service (rather than pre-service) teachers. The original study contained a binary variable denoting whether pre-service teachers expected to primarily teach science or a different, non-science subject; our conceptual replication replaces this with a binary variable denoting whether in-service teachers *actually teach* biology, life science, or general science courses (i.e., the science courses most relevant to the study materials and performance task) or only teach other science courses such as physics or chemistry. The original study also contained a binary variable denoting which university a pre-service teacher attends (only two campuses were included in the original study); our conceptual replication replaces this with an integer variable that denotes the grade level most commonly taught by a participating in-service teacher.

We examined eighteen distinct preregistered hypotheses, which can be summarized by the following research questions:

RQ1. Are science teachers' language-exclusive ideologies, racial essentialism, and asset-based attitudes toward African American Language significantly correlated with each other? (Hypotheses 2–4)

RQ2. Are science teachers' language-exclusive ideologies associated with their self-reported grading practices (Hypothesis 1) and their behavior during a formative assessment performance task (e.g., grading science writing samples, writing feedback to students, writing inferences about student understanding) after controlling for the science discipline(s) they teach, the grade level of students they teach, and their knowledge of language as an epistemic tool for teaching science? (Hypotheses 5–18)

Positionality

In terms of our positionalities, various members of our team identify as White, Black, or Hispanic; as women or men; as cisgender and straight or LGBTQ; and as mother tongue speakers of African American Language or Black Language, English, and/or Spanish. Importantly, we view positionality statements not merely as a tool for informing readers about our identities but also as an avenue for critiquing and moving beyond the traditionally extractive and oppressive character of much social science research [39,40]. In this study, we were motivated by the Charity Hudley Rule for Liberatory Linguistics [41], which argues researchers studying marginalized communities and their language practices should explicitly describe efforts to “increase the participation of [insider] community members at your university, in your department, and in your research area” (p. 136). Some of our efforts include organizing and funding guest lectures by scholars who research African American Language from insider perspectives and citing and centering such scholars in our scholarship. We have also sought to support writing and publishing by African American community members inside and outside academia by offering research participants in some of our studies a choice among several forms of compensation, including a gift card to an online bookseller that centers African American writers and perspectives.

Materials and methods

We measured teachers’ formative assessment practices using a short performance task and measured ideologies, attitudes, and knowledge using a short survey. This study was approved by the Southern Methodist University Institutional Review Board under protocol 24–043. All participants provided informed written consent through the Qualtrics survey. The combined performance task and survey instrument are available online as part of our preregistration protocol (osf.io/fjwbv). To strengthen the design of our preregistration, we reviewed Bakker et al. [42] for a list of 29 elements of research design which, when left unspecified in preregistration plans, can act as “degrees of freedom” that result in low-quality or underspecified preregistrations. Each of these 29 elements was addressed within our preregistration protocol, and any deviations from the protocol are explicitly noted.

Sampling

By reviewing publicly available school-, district-, and state-level websites and databases, we curated a list of 9,031 email addresses for middle school or high school science teachers spread across all 50 U.S. states. Approximately 13% of these email addresses came from a single state that maintains a large publicly accessible database of teacher contact information; no other individual state represented more than 10% of the email corpus. Another ~29% of the emails came from four very populous states, all of which were among the six most populous states in the U.S. Another ~29% came from ten additional states of varying sizes. The remaining ~29% came from the remaining 35 states. The corpus included email addresses from at least three different schools or districts in every state.

We recruited participants via email using a Qualtrics mailmerge, meaning each prospective participant received an individualized survey link. This enabled us to verify participants’ identities and prevent duplicate submissions or submissions from spambots. We sent email invitations in April and May of 2024 and continued recruiting participants until after we had surpassed 377 responses, the minimum required based on a power calculation described in our preregistration protocol. We ultimately had 387 completed responses, a response rate of 4.3%. Participants taught in 42 of the 48 contiguous lower U.S. states (no responses were received from Alaska or Hawai‘i). All participants received an Amazon.com gift card after completing the study.

As shown in [Table 1](#), racial and gender demographics of our sample were broadly similar to those of science teachers in the United States, of whom approximately 88% identify as White, 5% as Black, and 7% as Hispanic [43] and of whom approximately 60.5% identify as female and 39.5% as male [44].

Table 1. Participant Demographics.

	N	%
Science course(s) taught (non-exclusive; may total more than 100%)		
Biology and/or other Life Science	196	50.6
Chemistry, Physics, and/or other Physical Science	192	49.6
General Science	78	20.2
Another type of science	58	15.0
Grade level most commonly taught		
6th	17	4.4
7th	22	5.7
8th	29	7.5
9th	87	22.5
10th	98	25.3
11th	100	25.8
12th	34	8.8
Racial identity (non-exclusive; may total more than 100%)		
White	316	81.7
Black or African American	14	3.6
Hispanic, Latina, Latino, or Latinx	13	3.4
Southeast Asian	12	3.1
East Asian	10	2.6
American Indian or Alaska Native	9	2.3
Multiracial	9	2.3
South Asian	4	1.0
Middle Eastern	3	0.8
Native Hawaiian or Pacific Islander	2	0.5
Other	16	4.1
Prefer not to say	16	4.1
Gender identity		
Female	240	62.0
Male	128	33.1
Non-binary, third gender, more than one gender, or prefer not to say	19	4.9
Multilingual		
Yes	77	19.9
No	310	80.1
Currently interacts (or previously interacted) regularly with AAL users		
Yes	141	36.4
Maybe/ not sure	42	10.9
No	204	52.7
Dialect region^a		
New England	105	27.1
Midland/North	139	35.9
South	78	20.2
West	65	16.8

^aBased upon approximate boundaries of dialect regions of the United States as defined by Labov [45]. For purposes of this study, New England includes the 6 states north and/or east of Connecticut; South includes the 13 states south and/or east of Virginia, West Virginia, Kentucky, Arkansas, and Texas; West includes the 11 contiguous states west of Montana, Wyoming, Colorado, and New Mexico; Midland/North includes the remaining 18 contiguous states.

<https://doi.org/10.1371/journal.pone.0327859.t001>

The language backgrounds of our sample also appeared similar to those of the general U.S. population, approximately 22% of whom use a language other than English at home [46]. However, we did not gather data on the exact languages used by participants or the contexts in which they use these languages.

Survey measures

We measured language-exclusive ideologies using a five-item Likert-type scale (Cronbach's alpha=0.752) developed by the authors, which exhibited concurrent validity with both self-reported formative assessment practices and performance task outcomes of preservice teachers in a prior study [18]. This scale measured agreement or disagreement with items such as "Students need to be fairly skilled writers before writing activities will be useful in a science class" and "If a student does not know the scientific vocabulary terms to describe a concept, they do not yet understand the concept." We argue that agreement with such items conveys a broader prescriptive belief that there is only one valid way (or a narrow range of valid ways) to convey scientific ideas in writing. None of the items on this scale included any reference to race, ethnicity, or racism.

We measured racial essentialism using a twenty-two item Likert-type scale (Cronbach's alpha=0.869) developed and validated by Williams and Eberhardt [47]; this instrument has previously been used to study racial ideologies in science education contexts (e.g., [48]). This scale measured agreement or disagreement with items such as "Racial groups are primarily determined by biology" and "It's natural to notice the racial group to which people belong." These items were originally designed to measure whether respondents believe race is "biological...natural...easily discernible...stable across time...stable across contexts...and stable within individuals" [47]. We argue that agreement with these items reflects a set of widely debunked and harmful notions about race [5,48].

We measured teachers' asset-based attitudes toward African American Language using eight modified items from a Likert-type scale (Cronbach's alpha=0.793) developed by Blake and Cutler [27] and also used in our prior study [18]. This scale measured agreement or disagreement with items such as "African American English (sometimes called AAE, AAVE, or Ebonics) is a form of English" and "It is educationally sound to use a student's first language or dialect as a way of teaching that student the standard language of a community," and also measured agreement or disagreement with reverse-coded items such as "Standard English is dominant in schools and business because it is the best form of English" and "It is inappropriate for teachers or students to use AAE in school subjects such as science or math." We argue these items measure the degree to which respondents value linguistic diversity in general and African American Language in particular as an asset, both in STEM classrooms and in other environments.

We measured teachers' knowledge of language as an epistemic tool for teaching science using a fifteen-item Likert-type scale (Cronbach's alpha=0.675) developed and validated by Fulmer et al. [35]. This scale measured agreement or disagreement with items such as "Writing to different audiences helps students to deepen conceptual understanding" and with reverse-coded items such as "As students get older, science class should not encourage students to use everyday language for science concepts." These items were designed to measure teachers' knowledge of language as it relates to teaching and to strategies for facilitating students' development of conceptual understanding [35].

We measured teachers' self-reported formative assessment practices with a five-item Likert-type scale (Cronbach's alpha=0.856), asking how frequently they believe they should grade students' spelling, grammar, punctuation, use of scientific vocabulary, and use of passive voice in the context of evaluating student science writing.

Tavakol and Dennick [49] report that researchers typically accept values of alpha ranging between 0.70 and 0.95; lower values may reflect items that are not sufficiently interrelated to measure a single construct, while higher values may reflect redundancy. In our study, all scales had alphas between 0.752 and 0.869, except for the scale measuring teachers' knowledge of language as an epistemic tool for teaching science [35], which exhibited an alpha of 0.675. This scale's reliability was problematically low, though it was higher than sometimes exhibited by this scale in prior research (e.g., [50]). Readers should keep this issue in mind when interpreting our analyses.

Formative assessment performance task

We also measured teachers' formative assessment practices using a performance task which asked participants to respond to three samples of purported student writing that describe mitosis. (All three samples were actually composed by teachers and then edited by the researchers; see Sedlacek et al. [18] for details and see preregistration materials at osf.io/fjwbv for the writing samples themselves). Writing Sample #1 was short (22 words) and used many technical scientific vocabulary terms, similar to a description students might copy out of a textbook. Writing Sample #2 was longer (54 words) and used fewer technical vocabulary terms; it also included a well-researched grammatical feature known as habitual *be*, which denotes an action or process habitually repeated over a long period of time (in this case, "Cells always be dividing to heal cuts, grow, and reproduce..."). Habitual *be* is often perceived as a grammatical error, yet it is in fact a well-documented and grammatically valid feature associated with African American Language [51]. Writing Sample #3 was longer still (110 words) and included few technical vocabulary terms and no instances of habitual *be*. Instead, it included several mechanical errors that are not associated with African American Language: two misspellings, one of which was repeated twice; a missing comma prior to an appositive, which can be interpreted as either a punctuation error or grammatical error; and multiple uses of sentence-initial "so," which sometimes elicits criticism from prescriptivists as grammatically undesirable despite its increasing usage in English [52,53]. All three writing samples conveyed accurate information about mitosis. All are the same writing samples used in Sedlacek et al. [18] and are also available in the preregistration materials.

Participants were asked to assign a grade to each writing sample (on an eleven-point scale from 0 to 10) and to quantify the level of student understanding they could infer from each writing sample (on an eleven-point scale from 0 to 10). In two open-ended items, participants were also asked first to compose feedback to the author of each writing sample, and then to write any inferences they could make about student understanding based on each writing sample.

Following our preregistered protocol, the outcomes we examined in this analysis included: the grade teachers assigned to each writing sample; the level of student understanding teachers based on each sample; the length of feedback and length of inferences (measured in characters) teachers wrote in responses to each sample; and a binary outcome measure denoting whether or not each teacher included mechanics-focused comments in their feedback (e.g., whether a teacher wrote that a student author needed to change their grammar, spelling, punctuation, etc.).

Analysis

Following our preregistered protocol, we tested eighteen distinct hypotheses using correlation coefficients, multiple linear regression, and logistic regression.

Tests of model assumptions

Following our preregistered protocol, we initially tested our data for possible violations of regression assumptions. Collinearity diagnostics showed that all four predictor variables included in our models exhibited variance inflation factors (VIFs) below 1.2, well within acceptable ranges. We initially tested for normality using Shapiro-Wilk tests. These indicated that *none* of the variables of interest were normally distributed except for the racial essentialism scale; in retrospect, however, we recognized that the Shapiro-Wilk test is not necessarily appropriate for Likert-type data, particularly with large sample sizes [54]. Thus, in a minor deviation from our preregistration protocol, we also visually inspected Q-Q plots; these suggested that *all* variables of interest were approximately normally distributed, except for the six variables quantifying the length of feedback and the length of inferences written by participants, all of which appeared to exhibit negative skew. Thus, we retained all variables in their original forms for analysis except for the six variables describing the length of feedback and inferences, which were transformed using a square root function that successfully addressed the normality problem. All analyses involving these variables were conducted twice, once using the transformed variables and once using the untransformed variables. All coefficients or odds ratios that were statistically significant

following a Benjamini-Hochberg correction on analyses of the untransformed variables were also significant following a Benjamini-Hochberg correction [55] on analyses of the transformed variables. (One additional effect displayed statistical significance in the analysis of transformed variables but not untransformed variables: an effect of teacher grade level on the length of feedback written on Writing Sample #2. This effect is denoted with a + in the relevant table.) We report the analyses of untransformed variables here for readers' convenience.

Also, following our preregistration protocol, we used a Breusch-Pagan test that identified heteroskedasticity in two of the linear regression models: the model predicting the grade assigned to Writing Sample #2, and the model predicting the frequency with which participants believed they should grade mechanics and vocabulary in student writing. However, neither of these dependent variables exhibited outliers, which our preregistration defined as data points with $|Z| > 3.5$. For these reasons, following our preregistration protocol, we used weighted least squares regression for these two models.

Combining qualitative and quantitative data

We used quantitizing mixed methods [56] to code the qualitative feedback teachers had written in response to each of the three writing samples and integrate these codes into the quantitative dataset. We used a binary variable to denote whether feedback included comments focused on the mechanics of student writing (e.g., spelling, grammar, punctuation, sentence structure), excluding comments about scientific vocabulary. First, the lead authors and two coders briefly discussed the “mechanics” code included in the preregistration protocol and discussed how this code might be applied to approximately 15 samples of feedback teachers had written in our prior study [18]. Second, the lead author separated the formative assessment performance task data in the *current* study from the rest of the dataset (so that coders would not know which responses were written by teachers with higher or lower levels of language-exclusive ideology). Next, the lead author reviewed all responses and determined that none of the 1,161 individual pieces of feedback needed to be excluded based on preregistration criteria (specifically, none of the feedback consisted of text strings which were both irrelevant to the study and longer than 5 characters in length). Then, following our preregistration protocol, the two coders independently applied the “mechanics” code to all 1,161 pieces of feedback. The lead author checked intercoder agreement, which was 96.0% across all responses; Cohen’s kappa was 0.835. Although our intercoder agreement was 96.0% across all responses, it was only 75.1% among the small subset of responses that were coded Mechanics = 1 by one or both raters.

Our preregistration protocol stated that if we failed to achieve at least 80% simple agreement across the subset of responses coded Mechanics = 1 by one or both raters, we would have coders meet to discuss the coding scheme, revise the code, and then assign two new coders to re-code the entire dataset using this new code. However, we recognized at this stage that a measure of reliability across the full dataset would be more appropriate, and that Cohen’s kappa is a more appropriate measure of reliability than simple agreement given the relatively low proportion of responses that received the mechanics feedback code (between 13% and 15%, depending on the coder). Cohen’s kappa was 0.835 for the full dataset, a value considered to be high [57]. Thus, in a deviation from our preregistered protocol, we did not revise the code definition and re-code the data. Instead, we continued with the next preregistered stage of our analysis.

The two coders met to discuss and resolve disagreements, and the coded values were reintegrated into the dataset as a binary variable (1 = feedback included one or more comments about writing mechanics such as spelling, grammar, punctuation, or sentence structure; 0 = feedback did not include comments about writing mechanics). Finally, we conducted a logistic regression of the form-focused feedback variable on language-exclusive ideology, controlling for the science discipline(s) participants teach, the grade level of students they teach, and their knowledge of language as an epistemic tool for teaching science.

Correction for multiple comparisons

We corrected for multiple tests of statistical significance using a Benjamini-Hochberg correction [55] calibrated for 63 tests of statistical significance. This enabled us to test our eighteen hypotheses while also looking for additional potentially meaningful findings among all coefficients in the linear and logistic regression models. Because the Benjamini-Hochberg

correction controls the False Discovery Rate rather than the Familywise Error Rate, it is inappropriate to report adjusted *p*-values for individual statistics; instead, we simply indicate whether each individual statistic was significant following the Benjamini-Hochberg procedure with the False Discovery rate set to .05.

Results

Relationships among language ideology, racial ideology, and attitudes toward African American Language

In our preregistration, we hypothesized that measures of language-exclusive ideology, racial essentialism, and asset-based attitudes towards African American Language would all have significant correlations. Ultimately, the Pearson correlation between language-exclusive ideology and racial essentialism was $r=0.366$, the Pearson correlation between racial essentialism and asset-based attitudes towards African American Language was $r=-0.514$, and the Pearson correlation between asset-based attitudes towards African American Language and language-exclusive ideology was $r=-0.409$. All three correlations were significant following a Benjamini-Hochberg correction.

Formative assessment practices

[Table 2](#) shows that most of our preregistered hypotheses related to quantitative formative assessment practices were borne out. The language-exclusive ideology measure was positively related to the frequency at which teachers self-reported grading the mechanics and vocabulary of student writing ($\beta=0.546$), positively related to the level of student understanding they inferred from Writing Sample #1 ($\beta=0.236$), and negatively related to both the grade teachers gave ($\beta=-0.161$) and the level of student understanding they inferred ($\beta=-0.147$) based on Writing Sample #2, and was negatively related to both the grade teachers gave ($\beta=-0.151$) and level of student understanding they inferred ($\beta=-0.255$) based on Writing Sample #3. All of these relationships were statistically significant after controlling for the science discipline(s) and grade level(s) participants teach and their knowledge of language as an epistemic tool for teaching science. Several of these relationships (e.g., lower grades and lower levels of inferred understanding based on Writing Sample #2) appeared in our previous smaller study but did not exhibit statistical significance, perhaps due to insufficient statistical power.

As shown in [Table 3](#), teachers who expressed higher levels of language-exclusive ideology wrote shorter feedback and generated shorter inferences in response to all three writing samples. We also found that teachers of life science or general science courses (who may be more likely to teach mitosis content) tended to write longer feedback ($\beta=0.160$) and longer inferences ($\beta=0.187$) on Writing Sample #1 and longer feedback ($\beta=0.132$) and longer inferences ($\beta=0.146$) on Writing

Table 2. Standardized Coefficients from Multiple Linear Regression^a of Grades and Inferences of Student Understanding on Predictor Variables.

Variable	Frequency of grading mechanics and vocabulary ^a	Grade assigned to the writing sample			Understanding inferred from the writing sample		
		#1	#2 ^a	#3	#1	#2	#3
Language-Exclusive Ideology	0.546* (0.043)	0.102 (0.051)	-0.161* (0.051)	-0.151* (0.051)	0.236* (0.050)	-0.147* (0.051)	-0.255* (0.050)
Teaches Life Science or General Science Course(s)	-0.052 (0.044)	-0.019 (0.053)	-0.040 (0.053)	-0.085 (0.053)	0.088 (0.052)	0.018 (0.053)	-0.114 (0.051)
Grade Level of Students	0.008 (0.045)	-0.093 (0.053)	0.004 (0.053)	0.020 (0.053)	-0.051 (0.052)	0.031 (0.053)	0.075 (0.051)
Knowledge of Language as an Epistemic Tool for Teaching Science	0.022 (0.043)	0.017 (0.051)	0.003 (0.051)	0.023 (0.051)	-0.075 (0.050)	0.019 (0.051)	0.017 (0.049)

Note. * denotes coefficients that were statistically significant following a Benjamini-Hochberg correction with the False Discovery Rate set at .05.

^aDue to heteroskedastic errors in the models predicting the frequency participants believe they should grade mechanics and the actual grades they gave on Writing Sample #2, weighted least squares regression was used for these models.

<https://doi.org/10.1371/journal.pone.0327859.t002>

Table 3. Standardized Coefficients from Multiple Linear Regression of Length of Feedback and Inferences about Student Understanding on Predictor Variables.

Variable	Length of Feedback on Writing Sample			Length of Inferences made from Writing Sample		
	#1	#2	#3	#1	#2	#3
Language-Exclusive Ideology	-0.255* (0.050)	-0.187* (0.051)	-0.230* (0.050)	-0.183* (0.050)	-0.229* (0.050)	-0.253* (0.050)
Teaches Life Science or General Science Course(s)	0.160* (0.051)	0.132* (0.052)	0.079 (0.052)	0.187* (0.052)	0.146* (0.052)	0.042 (0.052)
Grade Level of Students	0.141* (0.051)	0.119* (0.052)	0.128* (0.052)	0.147* (0.052)	0.140* (0.052)	0.041 (0.052)
Knowledge of Language as an Epistemic Tool for Teaching Science	0.042 (0.050)	0.037 (0.050)	0.053 (0.050)	0.020 (0.050)	-0.020 (0.049)	0.035 (0.050)

Note. * denotes coefficients that were statistically significant following a Benjamini-Hochberg correction with the False Discovery Rate set at 0.05.

+ denotes a finding that exhibited statistical significance in the analysis of transformed variables but not untransformed variables.

<https://doi.org/10.1371/journal.pone.0327859.t003>

Sample #2, but not on Writing Sample #3, and that teachers of older students (who may expect their students to write more complex texts) tended to write longer feedback on Writing Sample #1 ($\beta=0.141$) and Writing Sample #3 ($\beta=0.128$), and tended to generate longer inferences in response to Writing Sample #1 ($\beta=0.147$) and Writing Sample #2 ($\beta=0.140$).

In our preregistration protocol, we also hypothesized that science teachers' language-exclusive ideologies would be associated with the likelihood that they included mechanics-focused feedback in their responses to Writing Samples #2 and #3 after controlling for other variables. (In our previous study, language-exclusive ideology was positively related to the use of mechanics-focused feedback in response to Writing Sample #2 but not Writing Sample #3.)

[Table 4](#) shows that language-exclusive ideology was *not* significantly related to whether or not science teachers gave mechanics-focused feedback on either Writing Sample #2 (Odds Ratio 1.033, 95% CI {0.991, 1.077}) or Writing Sample #3 (Odds Ratio 1.042, 95% CI {0.996, 1.090}) after controlling for other variables. Thus, our previous finding that the language-exclusive ideology scale was associated with biased responses to AAL features was not replicated.

We noted that 24% of respondents gave mechanics-focused feedback on a writing sample with a single AAL feature, while only 20% gave mechanics-focused feedback on the writing sample with deviations from standard spelling, grammar, or punctuation that were not specifically associated with African American Language. In an additional exploratory, non-preregistered analysis, an asymptotic McNemar's test [58] determined that there was not a statistically significant difference in the proportion of respondents giving mechanics-focused feedback on Writing Samples #2 and #3, $p=.098$ (not included in our Benjamini-Hochberg correction).

Discussion

Collectively, these findings have important implications for multiple fields. In sociolinguistics and linguistic anthropology, our findings support arguments by Flores and Rosa [1], Martin et al. [59], and others that suggest attempts to contest

Table 4. Binary Logistic Regression of Mechanics-Focused Feedback Practices Observed in Performance Task on Predictor Variables.

Variable	Gave mechanics-focused feedback on Writing Sample #2				Gave mechanics-focused feedback on Writing Sample #3			
	B (SE)	Wald	Odds Ratio Exp(B)	95% CI for Odds Ratio	B (SE)	Wald	Odds Ratio Exp(B)	95% CI for Odds Ratio
Language-Exclusive Ideology	0.033 (0.021)	2.340	1.033	(0.991, 1.077)	0.041 (0.023)	3.258	1.042	(0.996, 1.090)
Teaches Life Science or General Science Course(s)	-0.028 (0.263)	0.012	0.972	(0.580, 1.628)	-0.120 (0.278)	0.186	0.887	(0.514, 1.530)
Grade Level of Students	-0.067 (0.082)	0.662	0.935	(0.796, 1.099)	0.033 (0.090)	0.136	1.034	(0.867, 1.233)
Knowledge of Language as an Epistemic Tool for Teaching Science	-0.017 (0.020)	0.657	0.984	(0.945, 1.024)	0.027 (0.022)	1.435	1.027	(0.983, 1.073)

<https://doi.org/10.1371/journal.pone.0327859.t004>

oppressive *language* ideologies must attend to issues of *racial* ideology as well, because these ideologies are often deeply interconnected. While theories and methods from linguistic anthropology are used extensively in research on first language and second language education, they are more rarely invoked in science education or in STEM education more generally (with the notable exception of systemic functional linguistics, e.g., [60–62]). STEM education would benefit from further application of linguistic anthropological insights on language *ideologies*, particularly raciolinguistic ideologies; for initial examples of this work, see Sedlacek et al. [18], Lemmi et al. [4], and Harper and Kayumova [63].

In social psychology, our findings advance researchers' ongoing agenda to understand how essentialism is related to other psychological constructs and behaviors [64,65]. The link we have shown between racial essentialism, pre-scriptivist beliefs about language, and behavioral responses to language may suggest a new line of inquiry for studying racial stereotyping and prejudice, both in STEM education and beyond. For example, do teacher education programs designed to disrupt racial essentialism among biology teachers and biology students (e.g., [66,67]) indirectly lead to changes in language ideologies, as well? Could language ideologies somehow moderate or mediate processes of ideological change related to racial essentialism? Collaborative investigations of such questions, combining methods and expertise from science education and social psychology, could help psychologists understand the deeper structure of essentialist beliefs while also extending science educators' understanding of strategies to disrupt essentialism in biology education.

In sociology, our findings suggest that beliefs about language prescriptivism may function as a previously under-researched form of color-evasive ideology, whereby oppressive beliefs about race are covertly embedded in seemingly race-neutral discourses and beliefs [68,69]. Color-evasive ideologies remain common among U.S. STEM educators at the college level [70] and evidence suggests these ideologies might impede educators' adoption of effective, inclusive teaching practices [71]. Understanding color-evasive ideologies in greater detail—especially those that directly relate to STEM education, such as the language-exclusive ideologies measured in our study—could generate insights that might inform future efforts to implement effective, inclusive teaching practices in STEM education.

In STEM education, our findings suggest that language-exclusive ideology may be an impediment to effective formative assessment in science and other disciplines. Teachers of science (and perhaps other content areas as well) appear to exhibit better formative assessment practices if they also espouse more language-*inclusive* ideology—that is, if they value multiple varieties of language in the classroom, including African American Language. If the observed relationships between language ideology and formative assessment turn out to be causal, then promoting language-inclusive ideology in science teacher education might lead teachers to generate more extensive feedback for students and might lead them to devote additional time or effort to making well-justified inferences about student understanding. It is also possible that promoting language-inclusive ideology may lead teachers of science, mathematics, and other subjects to provide students with more substantive, content-focused feedback on their writing. In both our original study and this conceptual replication, participants with lower levels of language-exclusive ideology reported grading the mechanics and vocabulary of student writing less frequently and tended to provide lengthier feedback and lengthier inferences about student understanding compared with their more language-exclusive colleagues. Our findings echo those of earlier work suggesting that ideologies about language in general, and AAL in particular, may shape the ways that teachers select and implement effective or high quality pedagogical practices such as providing students with feedback [72].

However, one finding did fail to replicate: in our original study, more language-exclusive teachers were also more likely to critique habitual *be* in student writing, yet in the replication, both more- and less-language-exclusive teachers appeared to critique habitual *be* at comparable rates. While teachers with more language-*inclusive* ideologies might be expected to critique this AAL feature at lower rates, our replication data suggests otherwise. Interestingly, this pattern was not clearly attributable to the behavior of either language-exclusive or language-inclusive teachers; language-inclusive teachers critiqued habitual *be* at substantially higher rates in the replication than in our original study, while language-exclusive teachers critiqued habitual *be* at substantially lower rates in the replication than in our original study. This may indicate

that our initial finding of a correlation between language-exclusive ideology and anti-AAL bias in feedback was spurious, and disappeared in the replication due to simple regression to the mean. (To be clear, we do find evidence of anti-AAL bias in both studies—in both studies, an AAL feature which is not an error is “corrected” by a substantial fraction of participants—but this bias was uncorrelated with language ideology in the replication.)

Alternatively, this failed replication may indicate that the relationship between language-exclusive ideology and biased patterns of feedback differs between a population of pre-service teachers of various subjects and a population of in-service teachers of science only, or between teachers with different amounts of teaching experience. The latter possibility is supported by the fact that language-exclusive ideology was correlated with racial essentialism in both studies; perhaps racial essentialism was associated with biased responses to AAL, but this bias is less severe among more experienced teachers. A study of gender bias in science assessment practices in E.U. countries found that such bias tended to be stronger or more common among early-career teachers rather than experienced teachers [73]. While we did not collect data on participants’ years of teaching experience in our replication, our sampling of email addresses from publicly available websites and databases was likely to preferentially include teachers with at least one and likely multiple years of experience (since these teachers have been at their current workplaces long enough to be added to school websites).

Further research should explore these phenomena using causal methods. These might include experiments that control for all features of student writing except for the presence or absence of AAL features, or tests of interventions that seek to change science teachers’ language ideologies and measure change(s) in teachers’ formative assessment practices. Additional qualitative and mixed-methods research should also explore *how* science teachers make sense of information about African American Language and apply such knowledge in their classrooms.

Limitations

Our methods had important limitations. For example, our sampling procedure relied on the public availability of teacher contact information, meaning it preferentially identified science teachers who (1) are employed by schools large enough to offer lists of teacher contact information on their websites and (2) have been in their jobs long enough to be listed on these websites. These may introduce important biases into the results; for example, as discussed above, the overrepresentation of teachers in larger schools and teachers with multiple years of experience (compared with teachers in smaller schools and first-year teachers) may have shaped our findings in important ways. Our instruments also had limitations—for example, the low reliability of the scale measuring teachers’ knowledge of language as an epistemic tool, which may have contributed to the lack of observed relationships between this construct and other constructs measured in our survey.

Given our goal of conceptually replicating a previous correlational study, this time with a larger and more externally valid sample, we did not conduct a controlled experiment (though our performance task could easily be adapted into an experiment in future studies—for example, by varying which of the three writing samples contained an instance of habitual *be*, or by varying the verb tenses in each of writing samples #1, #2, and #3). With this in mind, we cannot make causal claims about the relationship between language-exclusive ideology and formative assessment practices. We can merely point to correlational findings and argue that causal claims about these relationships are *plausible* based on convergent evidence from other fields. Further research is needed to explore relationships among language ideologies, racial ideologies, and classroom formative assessment practices in greater depth, both in science education and other fields.

Conclusions

In this conceptual replication, we reproduce our previous finding that seemingly race-neutral beliefs about writing pedagogy are in fact correlated with essentialist ideologies of race. We also reproduce and extend our previous findings about the formative assessment practices of teachers with more- or less-prescriptive beliefs about student writing, showing that teachers with relatively language-inclusive ideology appear to respond differently to student writing samples compared with their more language-exclusive colleagues.

Our findings provide quantitative evidence in support of raciolinguistics, a set of linguistic anthropological theories which postulate connections between racial ideology and seemingly race-neutral beliefs about language. These findings suggest that efforts to improve formative assessment in science education should address not only teachers' knowledge of language, but their language ideologies and racial ideologies as well. Our findings also have implications for the fields of social psychology and sociology, and point to future areas for interdisciplinary investigation and collaboration between these fields and STEM education. In the meantime, further research is needed to develop and improve measures of language-exclusive ideology and to identify and test strategies for disrupting such ideologies. Doing so could support broader efforts to advance racial and linguistic justice in STEM education and in education more broadly.

Acknowledgments

The authors wish to thank Anne Charity Hudley, Christine Mallinson, Anthony Petrosino, AJ Alvero, Candace Walkington, Annie Wilhelm, and Meredith Richards for their useful suggestions and feedback at various stages of the project.

Author contributions

Conceptualization: Quentin C. Sedlacek, Nickolaus Ortiz, Catherine Lemmi, Maricela León, Kimberly C. Feldman.

Data curation: Quentin C. Sedlacek.

Formal analysis: Quentin C. Sedlacek, Nickolaus Ortiz, Catherine Lemmi.

Funding acquisition: Quentin C. Sedlacek.

Investigation: Quentin C. Sedlacek, Maricela León.

Methodology: Quentin C. Sedlacek, Nickolaus Ortiz, Catherine Lemmi.

Project administration: Quentin C. Sedlacek.

Writing – original draft: Quentin C. Sedlacek.

Writing – review & editing: Quentin C. Sedlacek, Nickolaus Ortiz, Catherine Lemmi, Maricela León.

References

1. Flores N, Rosa J. Undoing Appropriateness: Raciolinguistic Ideologies and Language Diversity in Education. *Harvard Educational Review*. 2015;85(2):149–71. <https://doi.org/10.17763/0017-8055.85.2.149>
2. Silverstein M. Monoglot “standard” in America: Standardization and metaphors of linguistic hegemony. In: Brenneis D, Macaulay RKS, editors. *The matrix of language*. 1st ed. Boulder, CO: Westview. 1996. p. 284–306.
3. Horner K, Weber JJ. Introducing multilingualism: A social approach. New York: Routledge. 2012.
4. Lemmi C, Brown BA, Wild A, Zummo L, Sedlacek Q. Language ideologies in science education. *Science Education*. 2019;103(4):854–74. <https://doi.org/10.1002/sce.21508>
5. Morning A. Reconstructing race in science and society: biology textbooks, 1952–2002. *AJS*. 2008;114 Suppl:S106-37. <https://doi.org/10.1086/592206> PMID: 19569402
6. Tawa J. Does Social Constructionist Curricula Both Decrease Essentialist and Increase Nominalist Beliefs About Race?. *Sci & Educ*. 2020;29(6):1513–40. <https://doi.org/10.1007/s11191-020-00125-7>
7. Chávez-Moreno LC. A Literature Review of Raciolinguistics in Dual-Language Bilingual Education. *The Handbook of Dual Language Bilingual Education*. Routledge. 2023. p. 254–65. <https://doi.org/10.4324/9781003269076-20>
8. Rubin DL. Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Res High Educ*. 1992;33(4):511–31. <https://doi.org/10.1007/bf00973770>
9. Kutlu E. Now You See Me, Now You Mishear Me: Raciolinguistic accounts of speech perception in different English varieties. *Journal of Multilingual and Multicultural Development*. 2020;44(6):511–25. <https://doi.org/10.1080/01434632.2020.1835929>
10. Shepherd MA. Effects of Race/Ethnicity, Gender, and Intonation on Secondary Science Teachers' Evaluation of Spoken Responses. *Urban Education*. 2016;55(5):730–52. <https://doi.org/10.1177/0042085916660346>
11. Rosa J, Burdick C. *Language Ideologies*. García O, Flores N, Spotti M, editors. Oxford University Press. 2016. <https://doi.org/10.1093/oxfor-dhb/9780190212896.013.15>

12. Yi J, Neville HA, Todd NR, Mekawi Y. Ignoring race and denying racism: A meta-analysis of the associations between colorblind racial ideology, anti-Blackness, and other variables antithetical to racial justice. *J Couns Psychol*. 2023;70(3):258–75. <https://doi.org/10.1037/cou0000618> PMID: 35604716
13. Whitley BE Jr, Webster GD. The Relationships of Intergroup Ideologies to Ethnic Prejudice: A Meta-Analysis. *Pers Soc Psychol Rev*. 2019;23(3):207–37. <https://doi.org/10.1177/1088868318761423> PMID: 29616588
14. Whitley BE Jr, Luttrell A, Schultz T. The Measurement of Racial Colorblindness. *Pers Soc Psychol Bull*. 2023;49(11):1531–51. <https://doi.org/10.1177/01461672221103414> PMID: 35818309
15. Quillian L, Midtbøen AH. Comparative Perspectives on Racial Discrimination in Hiring: The Rise of Field Experiments. *Annu Rev Sociol*. 2021;47(1):391–415. <https://doi.org/10.1146/annurev-soc-090420-035144>
16. Doane A (“Woody”). Beyond Color-blindness: (Re) Theorizing Racial Ideology. *Sociological Perspectives*. 2017;60(5):975–91. <https://doi.org/10.1177/0731121417719697>
17. Hartmann D, Croll PR, Larson R, Gerteis J, Manning A. Colorblindness as Identity: Key Determinants, Relations to Ideology, and Implications for Attitudes about Race and Policy. *Sociological Perspectives*. 2017;60(5):866–88. <https://doi.org/10.1177/0731121417719694>
18. Sedlacek QC, Lemmi C, Feldman K, Ortiz N, León M. African American Language in science education: A translanguaging perspective. *J Res Sci Teach*. 2024;62(1):228–69. <https://doi.org/10.1002/tea.22011>
19. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature*. 2016;533(7604):452–4. <https://doi.org/10.1038/533452a> PMID: 27225100
20. Rickford JR, Rickford RJ. Spoken soul: the story of Black English. New York: John Wiley & Sons. 2000.
21. King S. From African American Vernacular English to African American Language: Rethinking the Study of Race and Language in African Americans’ Speech. *Annu Rev Linguist*. 2020;6(1):285–300. <https://doi.org/10.1146/annurev-linguistics-011619-030556>
22. Committee on CCCC Language Statement. Students’ right to their own language. *Coll Engl*. 1975;:709–26.
23. Smitherman G. “Students’ Right to Their Own Language”: A Retrospective. *English J*. 1995;84(1):21. <https://doi.org/10.2307/820470>
24. Baker-Bell A. Linguistic Justice. Routledge. 2020. <https://doi.org/10.4324/9781315147383>
25. Delpit L. The Silenced Dialogue: Power and Pedagogy in Educating Other People’s Children. *Harvard Educational Review*. 1988;58(3):280–99. <https://doi.org/10.17763/haer.58.3.c43481778r528qw4>
26. Kynard C. Writing while Black: the colour line, Black discourses and assessment in the institutionalization of writing instruction. *Engl Teach-Pract Cri*. 2008;7(2):4–34.
27. Blake R, Cutler C. AAE and Variation in Teachers’ Attitudes: A Question of School Philosophy?. *Linguistics and Education*. 2003;14(2):163–94. [https://doi.org/10.1016/s0898-5898\(03\)00034-2](https://doi.org/10.1016/s0898-5898(03)00034-2)
28. Lemmi C, Pérez G. Translanguaging in elementary science. *International J Science Education*. 2023;46(1):1–27. <https://doi.org/10.1080/09500693.2023.2185115>
29. Brown BA, Reveles JM, Kelly GJ. Scientific literacy and discursive identity: A theoretical framework for understanding science learning. *Sci Ed*. 2005;89(5):779–802. <https://doi.org/10.1002/sce.20069>
30. Brown BA, Ryoo K. Teaching science as a language: A “content-first” approach to science teaching. *J Res Sci Teach*. 2008;45(5):529–53. <https://doi.org/10.1002/tea.20255>
31. Muhammad GE, Ortiz NA, Neville ML. A Historically Responsive Literacy Model for Reading and Mathematics. *The Reading Teacher*. 2021;75(1):73–81. <https://doi.org/10.1002/trtr.2035>
32. Ortiz NA. Let our rejoicing rise without bound: a focus on black language and mathematics learning. *International Journal of Qualitative Studies in Education*. 2022;:1–13. <https://doi.org/10.1080/09518398.2022.2061623>
33. Godley AJ, Reaser J, Moore KG. Pre-service English Language Arts teachers’ development of Critical Language Awareness for teaching. *Linguistics and Education*. 2015;32:41–54. <https://doi.org/10.1016/j.linged.2015.03.015>
34. Mallinson C, Charity Hudley AH. Partnering Through Science: Developing Linguistic Insight to Address Educational Inequality for Culturally and Linguistically Diverse Students in U.S. STEM Education. *Language and Linguist Compass*. 2014;8(1):11–23. <https://doi.org/10.1111/lnc3.12060>
35. Fulmer GW, Hwang J, Ding C, Hand B, Suh JK, Hansen W. Development of a questionnaire on teachers’ knowledge of language as an epistemic tool. *J Res Sci Teach*. 2020;58(4):459–90. <https://doi.org/10.1002/tea.21666>
36. McDermott MA, Hand B. A secondary reanalysis of student perceptions of non-traditional writing tasks over a ten year period. *J Res Sci Teach*. 2009;47(5):518–39. <https://doi.org/10.1002/tea.20350>
37. Morrison JR, Bol L, Ross SM, Watson GS. Paraphrasing and prediction with self-explanation as generative strategies for learning science principles in a simulation. *Education Tech Res Dev*. 2015;63(6):861–82. <https://doi.org/10.1007/s11423-015-9397-2>
38. Brodeur A, Cook NM, Hartley JS, Heyes A. Do pre-registration and pre-analysis plans reduce p-hacking and publication bias?. 2022. <https://hdl.handle.net/10419/262738>
39. Boveda M, Annamma SA. Beyond Making a Statement: An Intersectional Framing of the Power and Possibilities of Positioning. *Educational Researcher*. 2023;52(5):306–14. <https://doi.org/10.3102/0013189x231167149>
40. Kouritzin S, Nakagawa S. Toward a non-extractive research ethics for transcultural, translingual research: perspectives from the coloniser and the colonised. *Journal of Multilingual and Multicultural Development*. 2018;39(8):675–87. <https://doi.org/10.1080/01434632.2018.1427755>

41. Hudley AH, Mallinson C, Bucholtz M. Talking college: Making space for Black language practices in higher education. New York: Teachers College Press. 2022.
42. Bakker M, Veldkamp CLS, van Assen MALM, Crompvoets EAV, Ong HH, Nosek BA, et al. Ensuring the quality and specificity of preregistrations. *PLoS Biol.* 2020;18(12):e3000937. <https://doi.org/10.1371/journal.pbio.3000937> PMID: 33296358
43. National Center for Education Statistics, U.S. Department of Education. Percentage distribution of teachers by ethnicity, race, school type, and selected main teaching assignment: 2020–21; 2021. Available from: https://nces.ed.gov/surveys/ntps/estable/table/ntps/ntps2021_7182a_t12n
44. National Center for Education Statistics, U.S. Department of Education. Percentage distribution of teachers' selected main teaching assignment, age, and sex and average age in years, by school type and selected main teaching assignment: 2020–21; 2021. Available from: https://nces.ed.gov/surveys/ntps/estable/table/ntps/ntps2021_7181_t12n
45. Labov W. The organization of dialect diversity. In: ICSLP4, November 1996. Philadelphia. Available from: http://www.ling.upenn.edu/phono_atlas/home.html
46. Dietrich S, Hernandez E. Language use in the United States: 2019. American community survey reports; 2022 Aug. Available from: <https://www.census.gov/library/publications/2022/acs/acs-50.html>
47. Williams MJ, Eberhardt JL. Biological conceptions of race and the motivation to cross racial boundaries. *J Pers Soc Psychol.* 2008;94(6):1033–47. <https://doi.org/10.1037/0022-3514.94.6.1033> PMID: 18505316
48. Donovan BM. Playing with fire? The impact of the hidden curriculum in school genetics on essentialist conceptions of race. *J Res Sci Teach.* 2013;51(4):462–96. <https://doi.org/10.1002/tea.21138>
49. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ.* 2011;2:53–5. <https://doi.org/10.5116/ijme.4dfb.8dfdf> PMID: 28029643
50. Ding C, Lammert C, Fulmer GW, Hand B, Suh JK. Refinement of an instrument measuring science teachers' knowledge of language through mixed method. *Discip Interdiscip Sci Educ Res.* 2023;5(1). <https://doi.org/10.1186/s43031-023-00080-7>
51. Rickford JR. Phonological and grammatical features of African American Vernacular English. *African American Vernacular English: Features, evolution, educational implications.* Malden, MA: Blackwell. 1999. p. 3–14.
52. Culpeper J. Foreword. *Studies in Corpus Linguistics.* John Benjamins Publishing Company. 2020. p. xi–xiv. <https://doi.org/10.1075/scl.97.for>
53. Graves S. New use of an old discourse marker: The rise of prefacing answers to questions with so. *Journal of Pragmatics.* 2024;225:69–86. <https://doi.org/10.1016/j.pragma.2024.01.002>
54. Leung S-O. A Comparison of Psychometric Properties and Normality in 4-, 5-, 6-, and 11-Point Likert Scales. *J Social Service Res.* 2011;37(4):412–21. <https://doi.org/10.1080/01488376.2011.580697>
55. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology.* 1995;57(1):289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
56. Creamer EG. An Introduction to Fully Integrated Mixed Methods Research. SAGE Publications, Inc. 2018. <https://doi.org/10.4135/9781071802823>
57. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb).* 2012;22(3):276–82. PMID: 23092060
58. Pembury Smith MQR, Ruxton GD. Effective use of the McNemar test. *Behav Ecol Sociobiol.* 2020;74(11). <https://doi.org/10.1007/s00265-020-02916-y>
59. Martin KM, Aponte GY, García O. Countering Raciolinguistic Ideologies: The role of translanguaging in educating bilingual children. *Cahiers internationaux de sociolinguistique.* 2019;N° 16(2):19–41. <https://doi.org/10.3917/cisli.1902.0019>
60. Fang Z. Scientific literacy: A systemic functional linguistics perspective. *Sci Ed.* 2005;89(2):335–47. <https://doi.org/10.1002/sce.20050>
61. Matthiessen CMIM. Systemic Functional Linguistics as applicable linguistics: social accountability and critical approaches. *DELTA.* 2012;28(spe):435–71. <https://doi.org/10.1590/s0102-44502012000300002>
62. O'Hallaron CL, Palincsar AS, Schleppegrell MJ. Reading science: Using systemic functional linguistics to support critical language awareness. *Linguistics and Education.* 2015;32:55–67. <https://doi.org/10.1016/j.liinged.2015.02.002>
63. Harper A, Kayumova S. Invisible multilingual Black and Brown girls: Raciolinguistic narratives of identity in science education. *J Res Sci Teach.* 2022;60(5):1092–124. <https://doi.org/10.1002/tea.21826>
64. Haslam N, Bastian B, Bain P, Kashima Y. Psychological Essentialism, Implicit Theories, and Intergroup Relations. *Group Processes & Intergroup Relations.* 2006;9(1):63–76. <https://doi.org/10.1177/1368430206059861>
65. Mandalaywala TM. Does essentialism lead to racial prejudice? It is not so Black and White. *Adv Child Dev Behav.* 2020;59:195–245. <https://doi.org/10.1016/bs.acdb.2020.05.007> PMID: 32564794
66. Donovan BM, Weindling M, Lee DM. From Basic to Humane Genomics Literacy. *Sci & Educ.* 2020;29(6):1479–511. <https://doi.org/10.1007/s11191-020-00171-1>
67. Donovan BM, Weindling M, Salazar B, Duncan A, Stuhlsatz M, Keck P. Genomics literacy matters: Supporting the development of genomics literacy through genetics education could reduce the prevalence of genetic essentialism. *J Res Sci Teach.* 2020;58(4):520–50. <https://doi.org/10.1002/tea.21670>
68. Bonilla-Silva E. Racism without racists: Color-blind racism and the persistence of racial inequality in America. 3rd ed. Lantham, MD: Rowman & Littlefield. 2021.

69. Annamma SA, Jackson DD, Morrison D. Conceptualizing color-evasiveness: using dis/ability critical race theory to expand a color-blind racial ideology in education and society. *Race Ethnicity and Education*. 2016;20(2):147–62. <https://doi.org/10.1080/13613324.2016.1248837>
70. Russo-Tait T. Color-blind or racially conscious? How college science faculty make sense of racial/ethnic underrepresentation in STEM. *J Res Sci Teach*. 2022;59(10):1822–52. <https://doi.org/10.1002/tea.21775>
71. Aragón OR, Dovidio JF, Graham MJ. Colorblind and multicultural ideologies are associated with faculty adoption of inclusive teaching practices. *Journal of Diversity in Higher Education*. 2017;10(3):201–15. <https://doi.org/10.1037/dhe0000026>
72. Ortiz NA, Ruwe D. Black English and Mathematics Education: A Critical Look at Culturally Sustaining Pedagogy. *Teachers College Record: The Voice of Scholarship in Education*. 2021;123(10):185–212. <https://doi.org/10.1177/01614681211058978>
73. Hofer SI. Studying Gender Bias in Physics Grading: The role of teaching experience and country. *Int J Sci Education*. 2015;37(17):2879–905. <https://doi.org/10.1080/09500693.2015.1114190>