From Basic to Humane Genomics Literacy: How Different Types of Genetics Curricula Could Influence Anti-Essentialist Understandings of Race

Brian M. Donovan[1]
Monica Weindling[1]
Dennis Lee[1]
[1]BSCS Science Learning

Correspondence to: Brian M. Donovan (bdonovan@bscs.org; 5415 Mark Dabling Blvd. Colorado Springs, CO 80918)

Abstract: Genetic essentialism of race is the belief that racial groups have different underlying genetic essences which cause them to differ physically, cognitively, or behaviorally. Apparently no published studies have explored if belief in genetic essentialism of race among adolescents differs after many weeks of formal instruction about different domains of genetics knowledge. Nor have any studies explored if such differences reflect a coherent change in students' racial schemas. We use a quasi-experimental design (N = 254 students in 7^{th}-12^{th} grade) to explore these gaps. Over the course of three months, we compared students who learned from a curriculum on multifactorial inheritance and genetic ancestry to students who learned from their business as usual (BAU) genetics curriculum that discussed Mendelian and molecular genetics without any reference to race, multifactorial genetics, or genetic ancestry. Relative to the BAU condition, classrooms that learned from the multifactorial genetics and ancestry curriculum grew significantly more in their knowledge of multifactorial genetics and decreased significantly more in their genetic essentialist perceptions, attributions, and beliefs. From a conceptual change perspective, these findings suggest that classrooms using a curriculum emphasizing genetic complexity are more likely to shift toward a coherent anti-essentialist understanding of racial difference.

Keywords: Genetics literacy, genetic essentialism, race, conceptual change, genetics education

Introduction

Genetic knowledge is used in sociopolitical debates about racial inequality in order to maintain or mitigate against structures that perpetuate oppression (Jackson Jr. & Depew, 2017). These debates boil down to assumptions that are made about the nature of race (Morning, 2011) and a growing body of research in science education demonstrates that school biology can affect the content of these assumptions, for better (Donovan et al., in review; Donovan, Semmens, et al., 2019), or worse (Donovan, 2014, 2015b, 2016, 2017). Specifically, this research has shown that belief in genetic essentialism—a cognitive form of prejudice—can be perpetuated or prevented by what is taught about race in genetics education (Donovan et al., forthcoming).

Estimates suggest that 20% of non-black US citizens believe in genetic essentialism of race (Morning et al., 2019). Psychologists (e.g. Dar-Nimrod & Heine, 2011) define genetic essentialism as the belief that people of the same race share some set of genes that make them physically, cognitively, and behaviorally uniform, but different from other races. Consequently, genetic essentialists believe that complex traits are influenced little by the social environment (Dar-Nimrod & Heine, 2011). Both of these beliefs make genetic essentialists prone to the naturalistic fallacy—that racial disparities need not be eliminated because they are immutable and natural (Lynch et al., 2018). Biology students are rarely taught why genetic essentialism is genetically inaccurate (Donovan, 2015b) or how biologists and anthropologists discredited and discarded essentialist models of human difference in the mid-20th century (Jackson Jr. & Depew, 2017).

Unsurprisingly, then, apparently no research has explored whether the development of different kinds of genetics knowledge affect students' genetic essentialist beliefs in different ways. Some scholars argue (e.g., Jamieson & Radick, 2013) and some studies tentatively suggest (Donovan, 2014, 2016, 2017; Jamieson & Radick, 2017) that when students develop knowledge of Mendelian genetics it can unintentionally increase their belief in genetic essentialism. Other studies have found correlations between greater knowledge of multifactorial genetics and decreased belief in genetic essentialism (Donovan et al., in review). And, randomized trials have demonstrated that belief in genetic essentialism can be prevented by helping students learn how patterns of human genetic variation refute genetic essentialism (Donovan et al., in review; Donovan, Semmens, et al., 2019), especially if students already have some prior knowledge of multifactorial genetics before they begin this learning (Donovan et al., in review).

To our knowledge, no comparative studies have estimated if belief in genetic essentialism of race among adolescents differs after many weeks of formal instruction about different domains of genetics knowledge. Nor have any studies explored if such differences reflect a coherent change in students' racial schemas. Filling these gaps in our knowledge could illuminate if and how the content of genetics education influences conceptualizations of race. Using a quasi-experimental design (QED) (N = 254, 7th-12th graders), we explore these gaps. Specifically, we estimate if learning about complex genetics topics produces greater declines in genetic essentialist perceptions, attributions, and beliefs compared to learning more simple patterns of inheritance organized around Mendelian and molecular genetics.

What is Genomics Literacy?

Genomics literacy is a domain specific form of scientific literacy. Norris and Phillips (2003) argue that scientific literacy has two interrelated components, fundamental literacy and derived literacy. Fundamental literacy is the ability to read and write in a particular scientific

discipline (Norris & Phillips, 2003). Derived literacy refers to a learner's domain specific knowledge, which is derived, in part, from fundamental literacy (Norris & Phillips, 2003). We contend that the type of derived genomics literacy that genetics education helps students to develop can either contribute to the prevention or the propagation of belief in genetic essentialism. These types are basic genomics literacy, standard genomics literacy, and humane genomics literacy.

Basic genomics literacy (BGL) refers to Mendelian and molecular genetics because these knowledges constitute the basis of most genetics curricula and standards (Dougherty et al., 2011; Dougherty, 2009; Jamieson & Radick, 2013; Stern & Kampourakis, 2017). When done well, a curriculum emphasizing basic genomics literacy tells a mechanistic story about the relatively rare human traits that are influenced by a single gene with two alleles. Having basic genomics literacy means knowing how meiosis, sexual reproduction, and homologous recombination generate genetic diversity, and how the structure of DNA encodes information needed for protein synthesis. In essence, basic genomics literacy is a basic understanding of the probabilistic processes that create genetic diversity within an individual or family (i.e., genotypes) and how this genetic diversity influences trait variation (i.e., phenotypes) from generation to generation within a family.

Table 1. *The derived sense of standard genomics literacy*

	Ideas	Supportive Scientific Evidence
1	**Multifactorial genetics**	Most forms of human variation are not discrete nor are they explained by a single gene (Dougherty, 2009). In nature, even Mendel's peas do not exhibit discrete forms of variation (Radick, 2015). In fact, most of the traits that are described as monogenic in biology textbooks are not actually well explained by a bi-allelic/monogenic model at all (*Myths of Human Genetics: Introduction*, n.d.). Rather, human variation, especially complex traits, are best explained by multifactorial models of inheritance where variation in a trait is influenced by a combination of environmental effects, polygenetic effects, and gene-by-environment interactions (Bush & Moore, 2012; Duncan & Keller, 2011; Keller, 2014; MacMahon, 1968). This means that complex human traits, like IQ, are malleable because they are influenced by cultural and ecological environments (e.g., see Devlin et al., 1997; Flynn, 1999; Turkheimer et al., 2003).
2	**Population thinking**	Population thinking is the idea that populations are not genetic types, rather they are aggregates of genetically varying individuals (Mayr, 1982). Genetic variation is a measurement of the amount of loci in variable DNA that differs, on average, when comparing the genomes of individuals of the same population (Rosenberg, 2011). Between group variation refers to the extra amount of loci that differ, on average, when comparing the genomes of individuals in different populations (Rosenberg, 2011). Patterns of genetic variation, in turn, affect causal inferences about the relationship between genetic variation and trait variation because multifactorial genetics is a statistical science based in population thinking. To establish that alleles influence a trait, Genome Wide Association Studies (GWAS) are performed. These studies attempt to establish that genetic variation within populations correlates with trait variation after controlling for factors that vary between populations, such as linkage disequilibrium, allele frequencies, and environmental factors (Bush & Moore, 2012). This means that GWAS can only explain the trait variance associated with alleles within a single population, and only if they have controlled for genetic and environmental factors that vary between populations.

Whereas BGL is a story about inheritance patterns within individuals or families, *standard genomics literacy* (SGL) is a story about the complex relationship between genetic and phenotypic

variation within populations (Table 1). SGL begins with the idea that most human traits are not determined by a single gene with two alleles (dominant and recessive) (Kampourakis, 2017). Rather most traits exhibit continuous variation within a population and they are polygenic, which means they are influenced by tens or even thousands of alleles (Bush & Moore, 2012). When combined, these alleles have a significantly smaller effect on trait variation than Mendelian genetics suggests. Furthermore, social or environmental factors strongly influence complex traits and they can also moderate gene expression (i.e., GxE), which means that complex traits are malleable (Moore & Shenk, 2017). Since this more complicated story is often underemphasized or even missing in most genetics standards and curricula (Dougherty et al., 2011), we refer to it as standard genomics literacy out of the need to emphasize these ideas more often in the standards that frame science curricula. Standard genomics literacy is inclusive of BGL, but it goes beyond BGL by incorporating population thinking and multifactorial models of inheritance.

Table 2. *The derived sense of humane genomics literacy.*

	Ideas	Description
1	**Population Thinking** makes it wrong to claim that people within a racial group are genetically uniform, and that racial groups are genetically discrete.	Only 0.1% of the human genome differs between any two randomly picked humans. When geneticists analyze variable DNA, they have found, repeatedly, that continental populations of humans exhibit low levels of genetic differentiation because there is proportionally more genetic variation within human populations (95.7%) than between them (4.3%) (Graves, 2015; Rosenberg et al., 2002). This finding results from three important patterns in the distribution of alleles in human populations (Rosenberg, 2011). First, across loci in the human genome, populations of people tend to have similar alleles, but they differ in allele frequency (Rosenberg, 2011). Second, private alleles that are found in only one human population are exceedingly rare (7.53% of alleles in the genome) and, on average, are only possessed by 1.65% of people in any single population (Rosenberg, 2011). Third, the amount of genetic variation within human populations declines as one samples indigenous populations living further from Africa because of the combined influences of migration out of Africa, the founder effect, and genetic drift (Rosenberg, 2011). While these patterns mean that there is a population structure in humans, it also means that genetic essentialism is inaccurate. For example, it is incorrect to assume that stereotypes are "genetically" true if most variation is found among individuals of the same group. Likewise, the fact that most alleles are widely distributed and that private alleles are rare means that "races" are not discrete.
2	**Multifactorial Genetics** makes it difficult to claim that racial disparities are simply the result of genes.	Since complex traits are not monogenic it is incorrect to argue that racial disparities occur because some "races" have certain genotypes that "other races" do not. Rather, complex traits are best explained by multifactorial models of inheritance (Kampourakis, 2017) where the association between polygenic variation and trait variation is contingent on the social environment (Bratsberg & Rogeberg, 2018; Moore & Shenk, 2017; Tucker-Drob & Bates, 2016). And, since polygenic contributions to group level differences in complex traits are predicted to be small, possibly spurious, and dependent on the environment (Rosenberg et al., 2018), it is a distortion of scientific knowledge to claim that racial disparities are simply caused by polygenic differences between races. For example, many studies demonstrate that racial disparities in education are caused by social factors like segregation and income inequality (Reardon et al., 2019), discriminatory beliefs/attitudes (Canning et al., 2019; Leslie et al., 2015), and other forms of structural racism (Markus & Moya, 2011). The differences in social environments between races that have resulted from systemic racism and discrimination in the US make it methodologically and ethically impossible to conduct a fair scientific experiment that would conclusively prove that genes are the cause of racial inequality (Donovan, 2015a; Feldman & Lewontin, 1975; Goldsby, 1973; Graves, 2015; Rosenberg et al., 2018). One can claim that genetic variation is the sole cause of trait variation among individuals of the *same race* if these individuals experience *the exact same* social-environment. However, one should be skeptical of anyone who claims that science shows that genes are the best explanation for disparities that exist between races.

The story of how population thinking and multifactorial genetics refutes genetic essentialism is what we call *humane genomics literacy (HGL)*. HGL is related to, yet distinct from, standard genomics literacy. Standard genomics literacy is the knowledge that trait variation is more complicated than the Mendelian explanation for it. This knowledge is more complex because it integrates molecular concepts, multifactorial concepts, and population thinking. HGL is more complicated because it is the story of how these concepts refute genetic essentialist assumptions about human variability and its causes. What makes HGL different from SGL is that it has the explicit aim of refuting genetic essentialism and it is predicated on the value of creating a society that is more equitable. Table 2 describes the derived sense of HGL.

BGL is the easiest form of genomics literacy to use for essentialist arguments. SGL is harder to use for essentialist arguments. And, HGL is impossible to use for essentialist arguments. A learner who develops HGL either has to ignore their HGL knowledge or believe that it is flawed in order to maintain their belief in genetic essentialism. Conversely, a learner who develops HGL and believes in it, should believe in genetic essentialism less. This means that learners who develop HGL have a greater probability of disbelieving genetic essentialism compared to learners who have developed SGL and/or BGL, but not HGL. For example, SGL does not have the explicit aim of refuting essentialism even though an anti-essentialist story latently exists within it. Thus, a learner could "cherry-pick" select pieces of the SGL story to support their belief in genetic essentialism. Alternatively, if they can abstract the anti-essentialist story that is latent within SGL, then a learner might be able to see how SGL refutes essentialism. BGL, on the other hand, lends itself to belief in genetic essentialism because it is a deterministic story that implies that there are disparate types of people because there are "genes for" different kinds of traits. We now explain how the development of these different types of genomics literacies could psychologically influence belief in genetic essentialism among learners.

Genomics Literacy and Genetic Essentialism Theory (GET)

We predict that the development of BGL, SGL, and HGL will create different effects on belief in genetic essentialism among students. We base this prediction in genetic essentialism theory (GET) (Dar-Nimrod & Heine, 2011). GET contends that exposure to genetic information that leads learners to believe that there is a *specific*, *proximate*, *stable* and *immutable* relationship between genes and traits tends to increase belief in genetic essentialism by influencing causal reasoning (Lynch et al., 2018). Also, GET contends that genetic information that leads learners to believe that individuals of the same group are genetically *uniform* and that different groups are genetically *discrete* tends to increase belief in genetic essentialism by influencing social categorization (Lynch et al., 2018). We now explain how these mechanisms of GET interface with different conceptions of genomics literacy to influence belief in genetic essentialism.

GET Hypothesis on Basic Genomics Literacy

A genetics education oriented toward BGL is designed to affect causal reasoning about genes by helping students construct a relatively simple understanding about genotype-phenotype relationships within an individual or family. When students are taught with a curriculum that is solely focused on BGL there could be unintended impacts on belief in genetic essentialism. For instance, Mendelian genetics education has been criticized for leading students to develop a model of inheritance in which most traits are monogenic (Jamieson & Radick, 2013; Lawson &

Thompson, 1988; Shaw et al., 2008; Venville et al., 2005) with no molecular mechanism to separate gene and trait (Duncan, Castro-Faix, & Choi, 2016; Duncan, Rogat, & Yarden, 2009) and no description of how this relationship is environmentally moderated (Jamieson & Radick, 2013).

From the standpoint of GET, the genetic models that students learn when developing their BGL imply that there is a *specific* causal relationship between a gene and trait, because the gene is conceptualized as the only cause of the trait (Lynch et al., 2018). Rarely, if ever, do students learn how the environment outside the body moderates gene expression inside the body when developing BGL (Jamieson & Radick, 2013; Stern & Kampourakis, 2017). Thus, the development of BGL places students at risk of believing that genes are a more *proximate* and *stable* cause of trait variation than any social/environmental factors operating outside of the body (Lynch et al., 2018), which could increase a student's belief in genetic essentialism. And, since people believe that traits influenced by social/environmental factors are more malleable than those influenced by genes (Lynch et al., 2018), the development of BGL could lead students to believe that human traits are *immutable*, thereby reinforcing belief in the naturalist fallacy (Lynch et al., 2018).

BGL education could also affect belief in genetic essentialism through impacts on beliefs that are tied to social categorization. For instance, it is estimated that 90% of biology textbooks discuss racial differences in monogenic disease prevalence (Morning, 2008; Willinsky, 2020). In this curriculum, students usually learn that sickle-cell-anemia is common among African-Americans and that cystic fibrosis is common among 'Caucasians' (Donovan, 2015b; Morning, 2008). Since the prevalence of both diseases in other ethnic or racial groups is rarely described by textbooks emphasizing BGL (Donovan, 2015b), these texts imply that racial groups are genetically *discrete* and that individuals of the same group are genetically *uniform*. This content could lead students to infer that trait variation between races is best explained by *discrete* genetic differences and/or the genetic *uniformity* within each group, thereby increasing genetic essentialism.

Mounting evidence from studies in different countries suggest that an emphasis on basic genomics literacy in the genetics curriculum–where environmental factors and population thinking are ignored–contributes to the development of genetic essentialism through causal reasoning. For example, experiments have shown that the blueprint metaphor for DNA used in textbooks can cause elevated levels of genetic essentialism in adults who read from these texts (Parrott & Smith, 2014). Lynch, Bevan, Achter, Harris, and Condit (2008) have also demonstrated that when adults (N = 104) read science texts that include "gene for" language they grow in belief in genetic essentialism. Then there are the results of randomized control trials (RCTs) by Donovan (2014, 2016, 2017) that are consistent with the social categorization mechanism of GET. Each of Donovan's studies demonstrated that when middle and/or high school biology students learn from curriculum describing the prevalence of monogenic disorders in different racial groups, it caused students to believe in genetic essentialism significantly more (Donovan, 2014, 2016, 2017). For example, one of these studies demonstrated that students (N = 135, 7^{th}-9^{th} graders) learning about genetic diseases with racial terminology (compared to those who did not) increasingly perceived more *discreteness* between racial categories and exhibited greater growth in belief genetic essentialism over a 3-month period. Altogether, this research suggests that when students learn the phenomena and concepts discussed in a curriculum oriented toward BGL there is a greater probability that their genetic essentialist beliefs will increase rather than decrease.

GET Hypothesis on Standard Genomics Literacy

In contrast, we hypothesize that standard genomics literacy influences causal reasoning and social categorization in a way that is more likely to reduce belief in essentialism than to increase it. For example, a curriculum oriented toward SGL would inform students that the relationship between genes and traits is not *specific* nor *proximate*. Developing this knowledge should then lead students to believe that the relationship between genes and traits is *unstable,* because the effect of genes varies across different environments. Since people believe that traits influenced by the environment are more malleable (Lynch et al., 2018), students who understand multifactorial models of inheritance should also believe that complex traits are *malleable* and not genetically *determined*. Less belief in the *proximity*, *stability*, *immutability*, and *determinative* power of genes should make genes a poor explanation for racial difference. Moreover, developing an understanding of multifactorial models of inheritance requires population thinking, because multifactorial genetics is a statistical science that is based on population variation (see Donovan et al., in review). Thus, the development of SGL could also reduce belief in genetic essentialism by influencing beliefs tied to social categorization, such as uniformity and discreteness beliefs. For example, developing the understanding that a tremendous amount of genetic or trait variation tends to exist within any human population could reduce beliefs about the uniformity of a group.

Evidence supporting the claim that SGL reduces belief in genetic essentialism is inconclusive, but promising. For example, Jamieson and Radick (2017) used a quasi-experimental design where undergraduates (N = 56) learned genetics from a BGL curriculum or from a SGL curriculum. Students completed surveys about their endorsement of genetic essentialism pre and post learning. Although there was selection bias of participants into treatment conditions in Jamieson and Radick (2017), students did not differ significantly in genetic essentialism before treatment. Yet, afterwards the students who learned from the SGL curriculum had significantly lower belief in essentialism than those who learned from the BGL curriculum. Further evidence for the negative relationship between SGL and belief in genetic essentialism comes from a study of US high school students (N = 721, 9^{th}-12^{th} grade) by Donovan et al. (in review). They found that SGL was negatively correlated with belief in genetic essentialism. In fact, they found that 11% of the between student variation in belief in genetic essentialism was associated with SGL. Conversely, in a sample of 427 Brazilian undergraduates, Gericke et al. (2017) found that SGL was not correlated with belief in genetic determination—a key component of genetic essentialism.

One possibility for these discrepancies is that the instrument that Gericke et al. (2017) used to measure SGL was more focused on BGL than SGL, and this created the null correlation in their study. For instance, their instrument primarily asked students about polygenic causation and Mendelian models of inheritance. Only two of the nine items in their assessment addressed how genes and environments independently affect trait variation. None of their items assessed knowledge of how genes and environments interact multiplicatively to affect trait variation. Nor did their instrument assess population thinking. Their inability to detect a statistically significant and negative correlation between SGL and belief in genetic determinism could therefore be an artifact of the content validity of their assessment. Thus, while research suggests that the relationship between SGL and belief in genetic essentialism is not always negative, the weight of evidence and genetic essentialism theory both suggest that this relationship should be negative.

GET Hypothesis on Humane Genomics Literacy

A handful of studies have explored whether a humane genetics education emphasizing humane population thinking decreases belief in genetic essentialism by affecting beliefs related to

social categorization, such as uniformity and discreteness beliefs. In three different RCTs, Donovan, Semmens, et al. (2019) demonstrated that teaching students about genetic variation within and between US census races in order to refute essentialism can significantly reduce belief in genetic essentialism. In the first RCT, Donovan, Semmens, et al. (2019) demonstrated that this kind of learning caused significant reductions in perceptions of genetic variation between racial groups and also in belief in genetic essentialism. They then replicated these findings in two more RCTs with adults (N = 176) and with biology students (N = 721, 9^{th}-12^{th} graders). Through a mediation analysis they also showed that learning about human genetic variation reduced belief in genetic essentialism through its impact on how students perceived genetic variation between races. Specifically, when students learned about genetic variation within and between human races, it reduced their perception of racial discreteness, which reduced their belief in genetic essentialism. These findings suggest that learning from a curriculum oriented toward the population thinking component of HGL can decrease student belief in genetic essentialism.

It has also been demonstrated that standard genomics literacy interacts with humane genomics literacy to reduce belief in genetic essentialism (Donovan et al., in review). The underlying hypothesis for this interaction is that students higher in SGL should be more capable of developing humane genomics literacy due to an expertise effect. In essence, high SGL students should have more of the relevant prior knowledge for making sense of information in a humane genomics curriculum, thereby making it easier for them to construct HGL. Additionally, possessing more prior SGL could increase feelings of self-efficacy while students learn from a humane genomics curriculum, thereby buffering against any emotional threat that students might experience when trying to make sense of how multifactorial genetics and population thinking refute genetic essentialism (Donovan et al., in review). Since emotional threat can thwart learning (Darner, 2019), greater SGL could help students to construct more HGL by mitigating against such threat. In a re-analysis of the third RCT in Donovan, Semmens et al. (2019), Donovan et al. (in review) produced evidence consistent with these hypotheses. Specifically, they found that 11% of the between student variation in belief in genetic essentialism was associated with SGL measured prior to learning, 4% was explained by whether students learned from a humane genomics curriculum emphasizing population thinking ($R^2 = .04$), and 3% was associated by the interaction between this curriculum and SGL ($R^2 = .03$). Thus, 18% of the total variance in belief in genetic essentialism was associated with SGL, HGL and their multiplicative interaction. This finding simultaneously demonstrates that genomics literacy matters for reducing belief in genetic essentialism and also that knowledge alone does not determine belief in genetic essentialism.

The Psychological Relationship of Genomics Literacy and Genetic Essentialism is Socio-Culturally Situated

To summarize, our hypotheses predict that students learning from a curriculum oriented only toward BGL will increase in belief in genetic essentialism. Conversely, students learning from a curriculum oriented toward SGL and/or HGL will decrease in belief in genetic essentialism. Thus, we predict that students learning from a curriculum oriented toward SGL and/or HGL will show greater declines in belief in genetic essentialism when compared to students who learn from a curriculum oriented toward BGL. Yet, there are reasons why these psychological changes may be inchoate, or even the opposite of what we have predicted. We now use the concepts of knowledge in pieces, socially-motivated reasoning, and the phenomenon of backfiring to describe the sociocultural boundary conditions on this predicted effect.

Knowledge in Pieces (KiP)

KiP is a theoretical account of how knowledge is organized and developed (diSessa et al., 2016; diSessa, 1993, 2018; diSessa, 1993), and thus, it has implications for conceptual change (diSessa, 1998; Gregoire, 2003; Strike & Posner, 1992). KiP is predicated on research indicating that it often takes students many years of learning to move from a naïve or lay understanding of a phenomenon to a scientifically coherent one that is organized schematically around theory (Committee on How People Learn II: The Science and Practice of Learning et al., 2018). Rather than assuming that students have coherent schemas, KiP posits that lay knowledge is composed of phenomenological primitives, or p-prims. P-prims are pieces of intuitive knowledge that people have derived from prior learning and/or life experiences, which they use to explain phenomena or evaluate their plausibility (diSessa, 2018). KiP predicts that conceptual change is time consuming and uneven in its rate of change because it requires the transformation, replacement, and/or reorganization of many p-prims (diSessa, 2018). It also assumes that conceptual change is context specific (e.g., see Ueno, 1993) because cognition is situated within cultural contexts composed of unique identities, conceptual tools, and norms (Brown et al., 1989). Therefore, if the activation of a p-prim depends on context, then changes to a p-prim will also be context specific (diSessa, 2018). Thus, KiP contends that long periods of structured sense-making in different contexts and with different phenomena are required to produce a coherent conceptual change in a learner.

To illustrate KiP in the domain of genetics, let us consider the p-prims related to SGL. A learner might begin their education on SGL with the p-prim that nature and nurture influence human traits (e.g., see Jayaratne et al., 2009; Meyer et al., 2020). They might also have the p-prim that DNA is a recipe for human difference (e.g., see Duncan et al., 2009; Lewis & Kattmann, 2004). A SGL curriculum would attempt to build upon the first p-prim about nature and nurture and integrate it into a coherent theoretical understanding of multifactorial genetics. It would also attempt to integrate the p-prim that DNA is a recipe for human difference into a coherent understanding of population thinking. KiP predicts that this integration will be incomplete, disorganized, and/or context specific. Thus, it is unlikely for a learner to move from these p-prims to a coherent understanding of SGL without a substantial amount of time devoted to sensemaking in many different social contexts. Instead, a learner is likely to develop the two components of SGL asymmetrically and inconsistently, which has consequences for their belief in essentialism. In different social contexts, these underdeveloped SGL components will be activated and used in different ways by different students. Thus, the relationship between SGL and essentialism will not be coherent within and between the social spaces in which learners are observed. Incoherent change could mean that a student who has developed SGL through their biology education will reason in an anti-essentialist manner in their biology classroom, but they might revert to genetic essentialism when making sense of novel racial phenomena outside of the classroom.

From the standpoint of KiP, moving a student from a naïve/lay understanding of genetics to BGL, then SGL, and finally HGL will take a lot of time and the end result of such learning will, in all likelihood, not resemble a schema that is internally consistent, organized by theory, and stable across contexts. Rather, KiP suggests that the best we can expect is that genetics education will produce inchoate understandings of BGL, SGL or HGL that are context specific. Thus, any relationship between BGL, SGL, or HGL and belief in genetic essentialism will likely be incoherent or inconsistent because the psychological mechanisms connecting this knowledge and belief are situated within a great deal of developmental and cultural-ecological complexity.

Socially Motivated Reasoning

One reason why the relationship between genomics literacy and belief in genetic essentialism could be incoherent is because of socially motivated reasoning. Studies have found that the reasoning connecting genetic concepts to belief in genetic essentialism is socially-motivated. If people feel a need to justify their social status, and/or if they are intolerant of social ambiguity, and/or if they perceive existential threat in a social situation, then they will be more likely to endorse essentialist thinking (Keller, 2005). For example, when a dominant group needs to justify the oppression of a minoritized group they tend to justify their discriminatory actions with genetic essentialism (Morton et al., 2009). But, when a member of their own group faces discrimination, they tend to negate essentialism (Morton et al., 2009).

If the reasoning connecting BGL, SGL, and/or HGL to belief in genetic essentialism is socially-motivated and contingent on identity, then we should not expect a genetics education oriented toward any form of genomics literacy to produce reliable changes in belief in genetic essentialism for all students. For instance, a BGL curriculum could increase belief in genetic essentialism among students who feel a need to justify their social status, but it might have no effect on essentialism among students who do not have this motivation or those who deeply value social justice. Likewise, a curriculum oriented toward SGL could decrease belief in genetic essentialism among students who have a motivation to confront and dismantle structural racism. But, this same curriculum could have the opposite effect, or no effect, on belief in genetic essentialism among students who feel a need to justify their racial privilege. Such effects could occur in this group because, unlike HGL, the knowledge constituting SGL is not organized for the purpose of refuting essentialism. For example, students with an inchoate understanding of the SGL concept of polygenic inheritance might use this idea to retrofit their essentialist beliefs, by claiming that each race has a polygenic essence instead of a Mendelian essence. However, since HGL has the explicit aim of refuting essentialism, it is impossible to use HGL to retrofit essentialist beliefs. An HGL curriculum should therefore create reductions in belief in genetic essentialism as long as the learner does not ignore it or believe that it is flawed. Either of these responses could occur because of socially motivated reasoning, and if they did, then a learner should exhibit no change in their belief in genetic essentialism when learning from an HGL curriculum. That said, an HGL curriculum could also produce more, not less, belief in genetic essentialism if it backfires.

Backfiring

Backfiring can occur when science interventions unintentionally threaten the worldviews or the core identities of a learner (Darner, 2019). When these threats are combined with instruction that thwarts a learner's psychological needs and elicits their negative emotions, it can inhibit conceptual change (Darner, 2019). Consequently, attempting to reduce belief in genetic essentialism by increasing HGL could backfire and produce more, not less, belief in genetic essentialism if it threatens the values or identities of students, especially those who are motivated by racial privilege. Additionally, backfiring is a risk whenever one uses science education to refute a scientific myth. For example, when an intervention uses scientific facts to discredit a myth, it creates a novel link between the myth and the fact, which can create a recall error where the learner believes the myth is the scientific fact (Lewandowsky et al., 2012). Thus, backfiring could occur among students even when their identities go unthreatened if there is a recall error and/or if instruction makes a learner feel incompetent, unsuccessful, or confused while learning.

Research Questions

Altogether, our review of the literature suggests that it is psychologically plausible that different types of genomics literacy produce differing effects on belief in genetic essentialism. At the same time, the plausibility of this psychological prediction is socio-culturally contingent. For these reasons we explore the following two questions:

1. Do students exhibit greater declines in belief in genetic essentialism while learning from a HGL and/or SGL curriculum compared to students learning from a BGL curriculum?
2. If so, then how coherent is the conceptual change in genetic essentialism associated with such learning?

Methods

To answer these questions, we conducted a quasi-experiment to explore if learning from a curriculum about multifactorial genetics and human genetic variation (i.e., a SGL/HGL curriculum) produces increases in students' genomics literacy and decreases in their genetic essentialist beliefs relative to business-as-usual (BAU) genetics instruction including only Mendelian and molecular genetics (i.e., a BAU/BGL curriculum) (see Figure 1). The SGL/HGL curriculum was a four-week unit that instructed students about multifactorial models of inheritance (Table 1: idea 1) before instructing them about population thinking and race (Table 2: idea 1). The curriculum was aligned with middle and high school level Next Generation Science Standards (NGSS, 2016). The unit taught about the core ideas of inheritance and variation of traits and genetic ancestry by engaging students in the scientific practices of data interpretation, argumentation, and modeling. This unit was framed with an anchoring phenomenon (*Phenomena | Next Generation Science Standards*, n.d.) and scaffolded with contrasting cases (Schwartz & Bransford, 1998), inductive learning (Shemwell et al., 2015), and academically productive talk participation structures oriented toward scientific argumentation (Osborne et al., 2016). The BAU/BGL curriculum was also a four-week unit aligned with NGSS genetics performance expectations.

Sample. Participants (N = 254, 7^{th}, 8^{th}, 9^{th}, and 12^{th} graders) were recruited from four schools. Two were public high schools in a high socio-economic status (SES) Colorado school district (74.9% of sample), and two were private, high SES middle schools in the San Francisco Bay Area (SFBA) (25.1% of sample). In the high schools, 71-79% of students identified as white, and 8-12% qualified for free and reduced-price lunch (FRL). In the middle schools, 40-60% of students self-identified as white and no students received FRL. Fifty-six percent of students in the sample identified with Liberal political values and 44% identified with conservative political values on the cultural theory of risk instrument (Kahan, 2012, 2016; Kahan et al., 2007)

Given our hypotheses, this sample provides a unique sociocultural context for generating "proof of concept" results that answer our research questions. Research indicates that the legacy of structural racism in America has led to a greater concentration of resources and highly qualified teachers in predominantly white, high SES schools (Darling-Hammond, 2010). These privileged schools are conducive to learning because the social-environment is enriched for learning. Genetics is one of the more difficult subjects to learn because it requires students to reason about phenomena not directly observable to the naked eye (i.e., DNA), it requires them to reason about probability and variation, and it requires them to reason across population, organismal, and cellular

phenomena (Duncan et al., 2009). From the KiP perspective, the privileged sociocultural contexts in which these students reside provides the optimal resources and teachers needed for developing a coherent understanding of genetics. Indeed, all of the teachers in our study possessed undergraduate degrees in biology. At the same time, these students are predominantly white and privileged. Thus, many students in this sample may have a social motivation to justify their privilege, which may make it more difficult to change their belief in genetic essentialism through genetics education. Altogether, this sample allows us to explore if different relationships exist between the development of BGL, SGL and/or HGL and belief in genetic essentialism.

Quasi-Experimental Design (QED). Our sample yielded us ten classrooms. All ten classrooms learned from their BAU/BGL genetics curriculum for four weeks to begin. Then, after week four, four of these classrooms (two 7th grade and two 12th grade) learned from the four-week genetics unit oriented toward SGL/HGL (treatment). The other six classrooms (two 8th grade and four 9th grade) continued to learn from their BAU/BGL genetics curriculum for four more weeks (control). To estimate if curricular differences were associated with different changes in genomics literacy and genetic essentialism, we surveyed students using four different instruments (Table 3) repeatedly, at two-week intervals, over a ten-week period. Figure 1 summarizes the design.

Table 3. Description of Dependent Variables

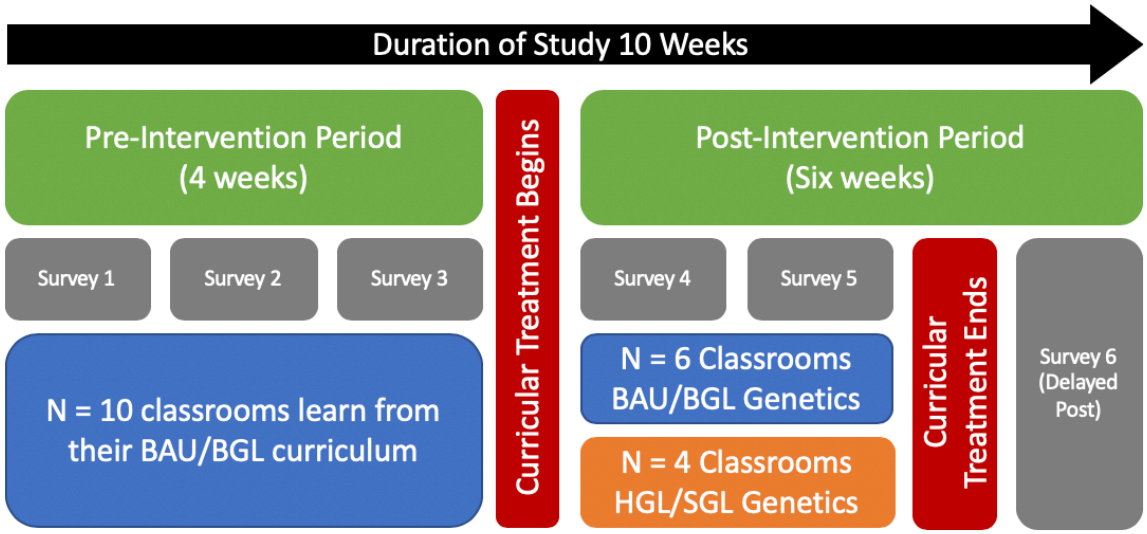
Instrument	Description	α
SGL	SGL was assessed through a subset of 16 multiple choice items from the Genetics Literacy Assessment Instrument (Bowling et al., 2008), five of which assessed BGL and eleven of which assessed SGL. Previous studies that have Rasch modeled these 16 items (e.g., Donovan et al., in review) have found that the BGL items are easier than the SGL items.	0.81
PHGV	This instrument had two item types, those about within group variation and those about between group variation. Six of the ten items in this instrument assessed how much genetic variation students perceived between races (BR). The remaining four assessed how much genetic variation students perceived within races (WR). The average of the between group items is divided by the sum of the averages of all items (i.e, $M_{BR}/(M_{BR}+M_{WR})$) yielding a proportion for each student that ranges between 0-1 (see Donovan, 2017). Higher scores on this instrument indicate that a student perceives a greater proportion of genetic variation between races relative to the total variation they perceive within and between races.	0.86
GARD	The four items in this instrument required students to use slider bars on a scale of 0-100% to show how much they thought races differed in (1) body structure and function; (2) brain structure and function; (3) intelligence; and (4) science ability; because of the social environments outside of bodies (E); the genes inside bodies (G); or personal choices (PC). For each trait (1-4), the value each student gives for the gene item is divided by the sum of the value for all the items (i.e., G/(G+E+PC)) (see Donovan, Stuhlsatz, et al., 2019). Then these proportions are averaged for all four traits. Higher scores on this 4-item instrument measure how much students attribute trait differences between races to genes versus environmental factors and personal choices on a scale of 0-100%.	0.81
BGE	Items were anchored on a scale of 1 (strongly disagree) to 7 (strongly agree). Thus, higher scores on this twelve-item, Likert-scaled, instrument indicate greater agreement with genetic essentialism. Examples of statements that students evaluate are: "Two Black people will always look more similar to each other than a Black person and a White person ever would"; "Racial differences in academic ability are caused by genetics".	0.84

The variables we assessed on each student were (1) standard genomics literacy (SGL); (2) perception of genetic variation between races (PHGV); (3) genetic attributions for racial differences (GARD); and (4) explicit belief in genetic essentialism (BGE). SGL was assessed using a subset of items from the Genetics Literacy Assessment Instrument (Bowling et al., 2008).

Genetic essentialism was triangulated using the PHGV (Donovan, Semmens, et al., 2019), GARD (adapted from Donovan, Stuhlsatz, et al., 2019), and a composite measure of BGE that included items from the genetically based racism instrument (Parrott et al., 2005) and the race conception scale (Williams & Eberhardt, 2008). Table 3 summarizes the characteristics of these instruments.

Measurements of these variables occurred at two-week intervals in all treatment and control classrooms through the Qualtrics platform. Three measurements occurred prior to when students began learning genetics and three more occurred after they began learning from the two different curricular treatments. We refer to these periods as the pre-intervention and post-intervention periods. Because we wanted to minimize the testing burden on students, a Solomon effect (Solomon & Lessac, 1968), and the impact of our study on instructional time, we did not ask students to respond to every item in each instrument. Rather, within classrooms we randomly assigned each student one item from each instrument, thereby reducing the probability of a Solomon effect. This means that students answered five different questions on each survey (the PHGV had two item types). So, instead of responding to all 42 items per survey (i.e., 252 responses per student), each student only responded to five items per survey (i.e., 30 responses per student), which cut the response burden greatly. The tradeoff of this method is that the content validity of our measurements is a group level phenomenon, because within any classroom, all of the items for an instrument were administered to students at random. Under a model of distributed cognition (Hutchins, 2000), our design allows for inferences about the conceptual change of classrooms.

Figure 1. Quasi-experimental design



Design Limitations

Four design limitations that need to be stated up front have to do with: (1) the heterogeneity *within* both the treatment and control conditions, (2) differences in the instructional frameworks *across* conditions, (3) differences in the student demographics *across* conditions, (4) difference in the teachers *across* conditions. These limitations will be further addressed in the discussion.

Limitation 1. One of the two teachers in our treatment condition had two classes of 12th grade Advanced Placement biology (n = 66) and they instructed students about the standard form of multifactorial genetics (Table 1: idea 1) before instructing students about the humane form of population thinking (Table 2: idea 1). However, the other teacher in our treatment condition did not believe that they had the time to instruct their 7th grade students about SGL before HGL. Thus, the middle school classrooms that received the treatment were only exposed to the story of humane population thinking outlined in Table 2 (idea 1), but they did not learn anything about multifactorial genetics laid out in Tables 1 or 2. The two control teachers, one from the SFBA with two classrooms of 8th grade biology (n = 32) and another from Colorado with four classrooms of 9th grade biology (n = 101) taught students using their business-as-usual (BAU) curriculum emphasizing basic genomics literacy. BAU comparison conditions are always heterogenous. Thus, the contrast in our QED consists of BGL versus SGL/HGL. It is a contrast that varies exposure to a curriculum emphasizing genetic simplicity versus one emphasizing genetic complexity. Since heterogeneity within experimental arms can increase the noise in statistical models of treatment effects, and since smaller samples yield less statistical power, our study is prone to a type 2 error.

Limitation 2. As stated earlier, the control curriculum was a BAU condition was not based in the exact same instructional frameworks as the SGL/HGL treatment curriculum. Yet, all students in both conditions still learned from NGSS oriented curricula that required students to engage in scientific practices to construct understandings of scientific concepts useful for explaining an anchoring phenomenon. The differences across conditions in instructional frameworks means that any treatment effect on the rate of change in the measured variables could be confounded by differences in instruction.

Limitation 3. As stated earlier, middle and high school students compose our sample. To the best of our ability, we attempted to create a group of control students that were developmentally and socio-demographically matched with the group of students in our treatment condition. Within the treatment arm of the QED, for example, there are two 8th grade classrooms and two 12th grade classrooms. Within the control arm, there are two 8th grade classrooms and four 9th grade classrooms. So, each condition has a mix of middle and high school aged students and across conditions students were sampled from the same school districts. For instance, all of the middle school classrooms in the treatment and control conditions were chosen from schools serving similar sociodemographic students in the same school district. Likewise, all of the high school classrooms in the treatment and control conditions were chosen from schools serving similar sociodemographic students in the same school district. Yet, there is a slight age difference across conditions. Thus, any treatment effect on the rate of change in the measured variables could be partially confounded by the grade levels of students or slight differences in the sociodemographic characteristics of students within each set of classrooms. For example, even though average levels of belief in genetic essentialism did not differ between treatment and control conditions prior to learning (β = -0.119, SE = 0.20, p = 0.572) the proportion of students who agreed with Liberal values on the cultural theory of risk scale was greater in treatment classrooms than it was in control classrooms (OR = 1.71, SE = 0.42, p = 0.028).

Limitation 4. Two different teachers implemented the treatment curricula, and two different teachers implemented the control curricula. Thus, any treatment effects on the rate of change in our measured variables could be partially confounded by a teacher effect.

Statistical Analysis Framework

In a quasi-experimental study with non-equivalence, such as ours, the common trends assumption needs to be met in order to make an argument about the influence of an intervention on a variable (Murnane & Willett, 2011; Somers et al., 2013). This is the assumption that treatment and comparison groups do not change over time in different ways before an intervention begins (Murnane & Willett, 2011; Somers et al., 2013). Violations of this assumption would indicate that any post intervention changes are simply the continuation of pre-treatment differences in change over time between groups (i.e., selection bias) (Murnane & Willett, 2011; Somers et al., 2013). If the common trends assumption is met, then it still needs to be established that, after receiving the treatment, the treatment condition changes over time in a significantly different manner than the counterfactual condition (Murnane & Willett, 2011; Somers et al., 2013; St.Clair et al., 2014).

When these two findings occur together, then one can make a stronger argument about the efficacy of a treatment using quasi-experimental data. For this reason, we measured student responses to the instruments targeting all four dependent variables (DVs) at two-week intervals before and after intervention. These repeated measurements allow us to test the common trends assumption with our data. They also allow us to estimate whether there is a difference-in-difference effect (see Somers et al., 2013) in the post intervention period that is suggestive of a treatment effect. Specifically, if we show that there is no treatment-by-time interaction in the pre-intervention period, then this suggests that the common trends assumption is met. Establishing evidence of the common trends assumption in the pre-intervention data would mean that students in both arms of the quasi-experiment are changing over time at the same rate. If so, then this would undercut the claim that the sociodemographic, developmental, and teacher effects described in limitations 3-4 are confounders. For example, if both groups (treatment and control) grow over time in the same way before treatment, then there is no evidence that the quality of teaching or developmental/social factors are confounding a treatment effect on the rate of change in our variables.

Then, if there is a significant treatment-by-time interaction in the post-intervention period, this difference-in-difference effect would suggest that the treatment and control conditions differed in their change over time on a variable after receiving different curriculum and instruction. We estimate both effects in the pre- and post-intervention period with EQ 1:

$$\text{EQ1: } DV_{ijk} = \beta_{0jk} + \beta_{1jk}(\text{TIME}) + \beta_2(\text{TRT}) + \beta_{3jk}(\text{TRTxTIME}) + \text{Error}$$

A significant effect on β_{3jk} in the pre-intervention period would violate the common trends assumption. An insignificant effect on β_{3jk} in the pre-intervention period would support the common trends assumption. If this insignificant effect on β_{3jk} in the pre-intervention period was observed with a significant effect on β_{3jk} in the post-intervention period, then these two results would tentatively suggest that the treatment and control conditions differed in their change over time on a variable after receiving different curriculum and instruction on genetics. This would be the case even if β_{1jk} and β_2 are statistically significant in the pre-intervention period (see Somers et al., 2013) because the difference-in-difference method we are using leverages critical assumptions in causal inference.

We also test if the post-intervention period β_{3jk} effect is robust to possible confounding created by differences between classrooms at baseline, such as those in limitation three. To do this, we generated a propensity score that regressed treatment condition onto student level controls that were measured during the first survey: (i) quantitative reasoning (refer to Donovan et al., in review

for items), (ii) stereotyping (refer to Donovan, Semmens, et al., 2019 for items), (iii) implicit person theories (IPT) of intelligence (Blackwell et al., 2007); (iv) IPT of science ability (Chen & Pajares, 2010); (v) IPT of group behavior (Halperin et al., 2011); and (vi) cultural theory of risk (CTR) (Kahan et al., 2007), which is a proxy for political orientation. Each of these covariates is either associated with belief in genetic essentialism (i.e., covariates ii-v; see Haslam et al., 2006) or genetics knowledge (i.e., covariate i; see Donovan et al., in review). Most of these variables are also implicated in socially-motivated reasoning (i.e., covariates ii-vi see Halperin et al., 2011; Kahan, 2016; Morin-Chassé et al., 2017). And, one of these variables is known to change with age (i.e., covariate ii; see see Bigler & Liben, 2007; Pauker et al., 2010). Thus, the propensity score we have created partially controls for selection bias in our QED created by classroom differences in the knowledge, racial beliefs, developmental levels, and social motivations of students.

Specifically, we model the main effect of propensity scores (PSCORE) and their interaction with time to check if the difference-in-difference estimate of the treatment effect in the post intervention period (β_{3jk}) is robust to potential selection bias. We do this through EQ 2:

$$\text{EQ2: } DV_{ijk} = \beta_{0jk} + \beta_{1jk}(\text{TIME}) + \beta_2(\text{TRT}) + \beta_{3jk}(\text{TRTxTIME}) + \beta_{4jk}(\text{PSCORE}) + \beta_{5jk}(\text{PSCORExTIME}) + \text{Error}$$

For genomics literacy treatment effects, we parameterize these equations using a generalized estimating equation (GEE) with a logit link to the binomial distribution and robust standard errors. For the other dependent variables, we use marginal models with an autoregressive lag and robust standard errors. Both types of models correct standard errors for the clustering of students within classrooms and the clustering of measurements within students. We use clustered-robust standard errors because the learning that occurs among students within classrooms is always correlated when students build new knowledge together through language, discourse, and academically productive talk (Michaels & O'Connor, 2015; Pearson et al., 2010; Vygotsky, 1978). Formal science learning is also culturally situated within the unique norms, beliefs, and identities of any single classroom (Brown et al., 1989). All of this creates correlated error at the classroom level that needs to be accounted for in statistical models of treatment effects on learning. Failing to account for this structure can result in downwardly biased standard errors and false conclusions about the statistical significance of results. Since we assigned treatment at the classroom level, and since our measurements are distributed across students within classrooms, our statistical models assume we have a sample size of 10 classrooms rather than a sample size of 254 students. We carry out our analysis using multiple imputation methods for missing data.

Results

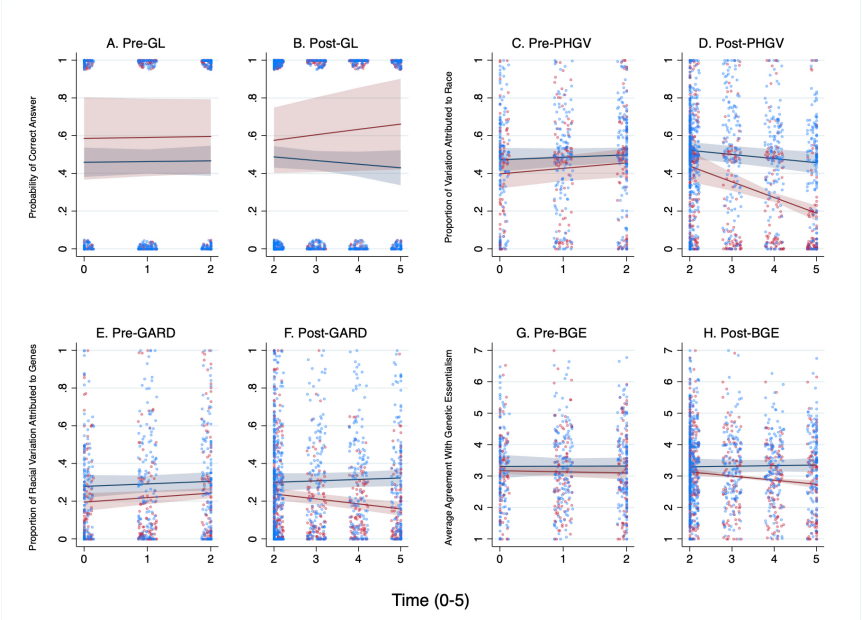
Research Question One

Standard Genomics Literacy (SGL). In the pre-test data (Figure 2A), classrooms that would be using the SGL/HGL curriculum in the future and those that would continue to use the BAU/BGL curriculum did not differ in genomics literacy at baseline (OR = 1.66, SE = .810, p = 0.295, 95% CI [.641, 4.32]) or in their growth rate in genomics literacy (OR = 1.006, SE = .121, p = 0.958, 95% CI [.794, 1.274]), thus supporting the common trends assumption. Notably, when treatment status was ignored, there was also no positive (or negative) growth rate in SGL during the pre-intervention period (OR = 1.02, SE = .046, p = 0.546) when all students were learning with

a BAU/BGL curriculum. In the post-test data (Figure 2B), classrooms that continued to use a BAU/BGL curriculum still exhibited no change in SGL (OR = .93, SE = .068, p = 0.293). However, classrooms using the SGL/HGL curriculum exhibited a greater increase in genomics literacy compared to classrooms that continued to use the BAU/BGL curriculum (OR =1.22, SE = .119, p = 0.041, 95% CI [1.007, 1.477]). The treatment effect on the post-test growth rate was statistically insignificant, but still positive, after controlling for the interaction of propensity scores with time (OR = 1.22, SE = .151, p = 0.115 , 95% CI [.953, 1.55]).

Perceptions of Human Genetic Variation (PHGV). In the pre-test data (Figure 1C), classrooms that would be using the SGL/HGL curriculum and those that would continue to use the BAU/BGL curriculum did not differ in their average PHGV at baseline (β = -0.074, SE = .050, p = 0.140, 95% CI [-.173, .024]) or in their growth rate in PHGV (β = .016, SE = .021, p = 0.448, 95% CI [-.025, .057]), thus supporting the common trends assumption. Notably, when treatment status was ignored there was also no positive (or negative) growth rate in PHGV during the pre-intervention period (β = .017, SE = .011, p = 0.114) when all students were learning with a BAU/BGL curriculum. In the post-test data (Figure 1D), classrooms that continued to use a BAU/BGL curriculum exhibited a slight reduction in PHGV (β = -0.21, SE = .006, p = 0.001). However, classrooms using the SGL/HGL curriculum exhibited a more negative decline in their PHGV compared to classrooms that continued to use the BAU/BGL curriculum (β = -.061, SE = .019, p = 0.002, 95% CI [-.100, -.021]). After controlling for the interaction of propensity scores with time, the decline in PHGV observed in BAU/BGL classrooms was insignificant (p = 0.862), however, the greater decline in PHGV observed in SGL/HGL classrooms remained significant and more negative (β = -.055, SE = .019, p = 0.004).

Figure 2. Difference-in-difference estimates in the pre-intervention and post-intervention period for SGL, PHGV, GARD, and BGE variables.



Note: Red = SGL/HGL, Blue = BAU/BGL; T0-T2 (Pre-Intervention Period), T3-T5 (Post-Intervention Period); GL = Genomics Literacy, PHGV = Perceptions of Humane Genetic Variation, GARD = Genetic Attributions for Racial Difference, BGE = Belief in Genetic Essentialism; Shaded areas are 95% CIs; Colored dots = student scores.

Genetic Attributions for Racial Difference (GARD). In the pre-test data (Figure 1E), classrooms that would be using the SGL/HGL curriculum and those that would continue to use the BAU/BGL curriculum differed in their average GARD at baseline (β = -0.083, SE = .039, p = 0.033, 95% CI [-.159, -.006]), but they did not differ in their growth rate in GARD (β = .010, SE = .018, p = 0.566, 95% CI [-.025, .047]). The average difference between quasi-experimental conditions at baseline was insignificant after controlling for propensity scores (p = 0.22). Thus, the common trends assumption was supported. Notably, when treatment status was ignored, there was also no positive (or negative) growth rate in GARD during the pre-intervention period (β = .012, SE = .009, p = 0.186) when all students learned with a BAU/BGL curriculum. In the post-test data (Figure 1F), classrooms that continued to use a BAU/BGL curriculum still exhibited no change in GARD (β = .007, SE = .007, p = 0.293). Yet, classrooms using the SGL/HGL curriculum exhibited a more negative decline in GARD scores compared to classrooms that continued to use the BAU/BGL curriculum (β = -.033, SE = .009, p < 0.001, 95% CI [-.052, -.015]). The treatment effect on the post-test growth rate remained statistically significant when controlling for the interaction of propensity scores with time (β = -.036, SE = .012, p = 0.004, 95% CI [-.061, -.011]).

Belief in Genetic Essentialism (BGE). In the pre-test data (Figure 1G), classrooms that would be using the SGL/HGL curriculum and classrooms that would continue to use the BAU/BGL curriculum did not differ in their average BGE at baseline (β = -0.132, SE = .197, p = 0.504, 95% CI [-.519, .255]), or in their growth rate in BGE (β = -.043, SE = .111, p = 0.694, 95% CI [-.262, .174]), thus supporting the common trends assumption. Notably, when treatment status was ignored, there was also no positive (or negative) growth rate in BGE during the pre-intervention period (β = -.012, SE = .070, p = 0.863) when all students were learning with a BAU/BGL curriculum. In the post-test data (Figure 1H), classrooms that continued to use a BAU/BGL curriculum exhibited no change in BGE (β = .018, SE = .055, p = 0.735). However, classrooms using the SGL/HGL curriculum exhibited a significantly more negative decline in their BGE compared to classrooms that continued to use the BAU/BGL curriculum (β = -.147, SE = .072, p = 0.041, 95% CI [-.289, -.006]). The treatment effect on the post-test growth rate remained statistically significant when controlling for the interaction of propensity scores with time (β = -.179, SE = .067, p = 0.008, 95% CI [-.313, -.046]).

Summary of Results. In the pre-intervention period when all students were learning from their BAU/BGL curriculum, the classrooms that would be treated with the SGL/HGL curriculum in the future, and those that would continue to learn from their BAU/BGL curriculum in the future, did not differ in their growth rate in SGL, PHGV, GARD or BGE. Even in models that disregarded treatment status by grouping all classrooms together, there was no statistically significant time change in any of these variables during the pre-intervention period despite the fact that all classrooms were learning with a BAU/BGL curriculum. Since the common trends assumption was supported by these findings, it suggests that pre-existing differences between treatment and control groups did not confound the treatment effects observed in the post-intervention period

In the post-intervention period, classrooms that learned with the SGL/HGL curriculum grew more in standard genomics literacy (SGL) and they decreased more in their essentialist perceptions (PHGV), attributions (GARD), and beliefs (BGE) compared to classrooms that continued to learn genetics with a BAU/BGL curriculum. In reference to our first research question, these findings tentatively suggest that students exhibited greater declines in belief in

genetic essentialism while learning from a HGL and/or SGL curriculum (relative to BAU/BGL).

Research Question Two

Now we address the question of whether or not the four classrooms that learned from the SGL/HGL curriculum exhibited a coherent conceptual change in their racial perceptions, attributions, and beliefs relative to the six BAU/BGL classrooms. To explore this question, we had to create a variable that categorized each student's set of perceptions, attributions and beliefs into five different levels that corresponded to essentialism (level 5), essentialism in pieces (level 4), incoherent beliefs (level 3), anti-essentialism in pieces (level 2), and anti-essentialism (level 1). However, our analysis of this data is aggregated up to the classroom level.

Level 5: Essentialism. Students who had BGE scores greater than or equal to four and PHGV and GARD scores greater than or equal to 0.5 were categorized as having essentialist schemas. The justification for such a categorization is that these students, on average, explicitly agreed with essentialism. Also, they perceived most genetic variation between races and they attributed trait differences between races mostly to genes. Consequently, these students had a coherent set of essentialist beliefs, perceptions, and attributions.

Level 4: Essentialism in pieces. Students who had BGE scores greater than or equal to four and either PHGV or GARD scores less than 0.5 were categorized as having essentialism in pieces. The justification for such a categorization is that these students explicitly agreed with essentialism, which makes them appear to be essentialist. However, they either perceived most genetic variation within races or they attributed trait differences between races mostly to the social environment and free-will. Consequently, these students had only two of the three "pieces" of essentialist thinking.

Level 3: Incoherent schemas. Students who had BGE scores greater than or equal to four and PHGV and GARD scores less than 0.5 were categorized as incoherent in their racial schema. Likewise, those who had BGE scores less than four and PHGV and GARD scores greater than 0.5 were categorized as incoherent. The justification for such a categorization is that the perceptions and attributions of these students contradicted their beliefs. For instance, they are the students who explicitly agreed with essentialism but still perceived most genetic variation within races and who attributed trait differences between races mostly to the social environment and free-will. Or, these were the students who disagreed with essentialism but they nevertheless perceived most genetic variation between races and they attributed racial differences in complex traits mostly to genes. Consequently, these students had contradictory beliefs, perceptions, and attributions.

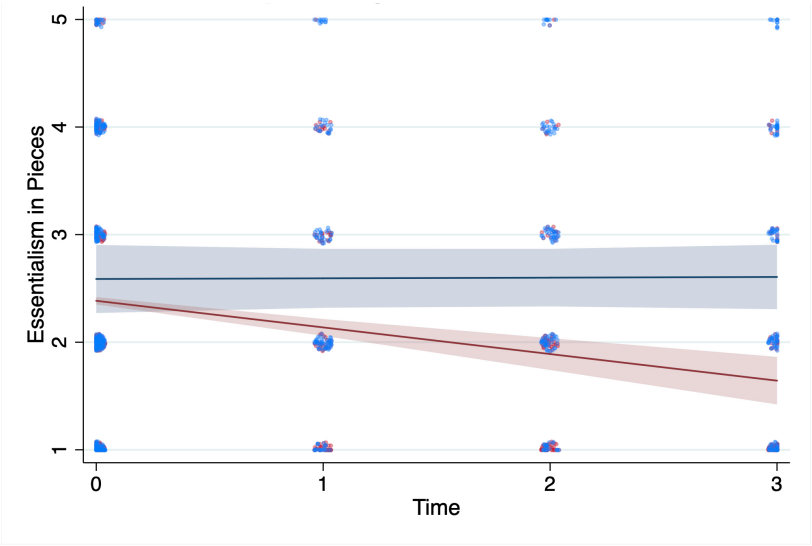
Level 2: Anti-essentialism in pieces. Students who had BGE scores less than four and either PHGV or GARD scores greater than 0.5 were categorized as having anti-essentialism in pieces. The justification for such a categorization is that these students, on average, explicitly stated that they disagreed with essentialism, making them appear anti-essentialist. However, they still perceived most genetic variation between races or they attributed trait differences between races mostly to genes. Consequently, these students had only two of the three "pieces" of anti-essentialist thinking.

Level 1: Anti-essentialism. Students who had BGE scores less than four and PHGV and GARD scores less than 0.5 were categorized as having anti-essentialist schemas. The justification for such a categorization is that these students, on average, explicitly stated that they disagreed with essentialism. Also, they perceived more genetic variation within races and they attributed trait differences between races mostly to the social environment and free-will. Consequently, these students had a coherent set of anti-essentialist beliefs, perceptions, and attributions.

For each student and time point in the data set, scores on the BGE, PHGV, and GARD were categorized in these ways. Modeling this variable allows us to estimate if classrooms using the SGL/HGL curriculum exhibited a more coherent decrease in essentialism over time relative to those classrooms using the BAU/BGL curriculum. Figure 3 summarizes the results of this analysis.

Students learning from the BAU/BGL curriculum did not exhibit any coherent change over time in their essentialist perceptions, attributions, and beliefs (β = .006, SE = .053, p = 0.906). However, classrooms using the SGL/HGL curriculum, over time, decreased more coherently in their essentialist perceptions, attributions, and beliefs relative to classrooms using the BAU/BGL curriculum (β = -.254, SE = .065, $p <$ 0.001, 95% CI [-.381, -.126]). To further understand the nature of the decline, we then examined how treatment status affected the proportion of students at each level of the essentialism in pieces variable per unit of time. There was no treatment effect on the odds of essentialism (OR = .403, SE = .309, p = 0.236), incoherence (OR = .914, SE = .110, p = 0.453), or anti-essentialism in pieces (OR = .867, SE = .144, p = 0.390). However, relative to the BAU/BGL classrooms, SGL/HGL classrooms exhibited a more negative reduction in the odds of essentialism in pieces (OR = .705, SE = .101, p = 0.014) and a more positive increase in the odds of anti-essentialism (OR = 1.474, SE = .1424, p < 0.001) per unit of time.

Figure 3. Treatment effect on essentialism in pieces variable



Note: Red = SGL/HGL, Blue = BAU/BGL; T0 (last measurement in pre-intervention period); T1-T3 (Post-Intervention Period); Shaded areas are 95% CIs; Colored dots refer to each student's scores.

Summary of results. In reference to our second research question, these findings suggest that there was a shift toward a coherent belief in anti-essentialism at the classroom level, because

as classrooms used the SGL/HGL curriculum, an increasing proportion of students within them began to exhibit all of the components of anti-essentialism (i.e., perceptions, attributions, beliefs).

Discussion

These quasi-experimental results are consistent with the hypothesis that developing standard and/or humane forms of genomics literacy can create a coherent reduction in belief in genetic essentialism at the classroom level. We base this claim in several results. First, when classrooms were learning with the BAU/BGL curriculum in the pre-intervention period there was no change in their SGL, nor was there any change in their essentialist perceptions, attributions, and beliefs. In the post-intervention period, too, the BAU/BGL classrooms did not change in their standard genomics literacy, or their essentialist perceptions, attributions, and beliefs. These results suggest that classrooms using the BAU/BGL curriculum did not go through any conceptual change. However, SGL/HGL classrooms increased more in standard genomics literacy than BAU/BGL classrooms in the post-intervention period. They also exhibited a more negative decrease in their essentialist perceptions, attributions, and beliefs. This change was also coherent because, over time, more students within SGL/HGL classrooms (relative to BAU/BGL) began to exhibit more coherence in their anti-essentialist perceptions, attributions and beliefs.

Were these effects causal? Unfortunately, the quasi-experimental nature of our study precludes strong causal claims. However, these results tentatively suggest a causal story. For example, the fact that the common trends assumption was supported means that the post-intervention treatment effects were not due to pre-intervention differences in how classrooms were already changing over time. While one might argue that the low statistical power of our study means that the null effects in the pre-intervention period were a type 2 error (i.e., limitation 1), a careful reading of Figure 2 shows that there was basically no time effect on any variable in either group prior to intervention. So, even if our sample was larger, and thus powered to create smaller standard errors (i.e., less noise) and smaller p-values, there would still be no signal to detect in the pre-intervention period. Consequently, it is difficult to argue that the treatment effects in the post-intervention period were only the result of confounding factors that already differed between conditions (i.e., limitation 3 described in the design section above).

Furthermore, in the post-intervention period, treatment effects on each variable were detected and robust to propensity score adjustment. The only effect that did not remain significant at a conventional level after adjusting for the main effect of propensity scores and their interaction with time was the treatment effect on SGL. This test was still on the border of being marginally significant (e.g., $p = 0.115$) and its statistical insignificance was likely the result of a degrees of freedom penalty that occurred from modeling too many variables (see EQ2) with a small sample. This penalty would have further impaired our statistical power and created inflated p-values. Nevertheless, the relationship between SGL and time was positive in the SGL/HGL classrooms. Given that there was no treatment effect on SGL in the pre-intervention period, a significant treatment effect without controls in the post-intervention period, and a marginally significant positive effect with controls, we contend that growth in SGL did differ between conditions. Since the learning of SGL apparently differed between conditions, our knowledge-based hypotheses are a plausible explanation for the treatment effects on belief in genetic essentialism that we observed.

Yet, one confounding factor that we cannot rule out entirely is differences in instructional quality between conditions (i.e., limitation 2). We cannot rule this factor out because there were different teachers in the treatment and control groups (i.e., limitation 4). But, if this alternative

explanation is correct, then why were there no differences between treatment and control classrooms in the pre-intervention period? If instructional differences drove these effects (i.e., limitation 2), and if teachers in disparate conditions differed systematically in their instructional abilities (i.e., limitation 4), then we would expect to see evidence of this effect in the pre-intervention period. Yet, we did not. One reason why we did not is that, in both conditions, teachers used 3-dimensional NGSS instruction. This factor may have attenuated any bias that was introduced by variation in teacher quality or instruction. Consequently, the best explanation for these results may still be the hypothesis described in our conceptual framework.

We hypothesized that when students develop understandings of the complex relationships between genetic variation and trait variation it can reduce their belief in genetic essentialism by triggering the mechanisms specified by genetic essentialism theory (Dar-Nimrod & Heine, 2011). GET contends that exposure to genetic information that leads learners to believe that there is a *specific*, *proximate*, *stable* and *immutable* relationship between genes and traits tends to increase belief in genetic essentialism through causal reasoning (Lynch et al., 2018). And, genetic information that leads learners to believe that same race people are *uniform* and that different racial groups are *discrete* tends to increase belief in genetic essentialism through social categorization (Lynch et al., 2018). We hypothesized that when the multifactorial component of SGL (Table 1; idea 1) or HGL (Table 2; idea 2) is developed, it could reduce belief in genetic essentialism by running the causal reasoning mechanism in reverse. We also hypothesized that developing the population thinking component of SGL (Table 1; idea 2) or HGL (Table 2; idea 1) could reduce belief in genetic essentialism by running the social categorization mechanism in reverse.

The middle school students in the treatment condition could only have developed the population thinking component of HGL (Table 2; idea 1) because that was the only learning objective in their curriculum. The high school students in the treatment condition could have developed both the multifactorial component of SGL (Table 1; idea 1) and the population thinking component of HGL (Table 2; idea 1) because those were the two learning objectives in their curriculum. Because of the heterogeneity within the treatment arm (i.e., limitation 1), we cannot be sure which form of literacy was causally active, or whether they interacted to reduce belief in genetic essentialism through both of the mechanisms specified by GET. Previous research from RCTs with middle and high school aged students already support the claim that the development of the population thinking component of HGL (Table 2; idea 1) reduces belief in genetic essentialism (Donovan, Semmens, et al., 2019). However, this research has also suggested that the population thinking component of HGL (Table 2; idea 1) is more easily developed among students who already have some knowledge of the multifactorial component of SGL (Table 1; idea 1) (Donovan et al., in review). Donovan et al. (in review) even suggests that the development of HGL will not lead to a reduction in genetic essentialism unless students already have some SGL.

Given this research, we think our results are the product of a two-step process based in an expertise effect (see Bransford, 2000). In high school classrooms, students developed SGL in the first half of their treatment curriculum, and this enabled them to develop more HGL from the second half of their treatment curriculum. As this occurred, students declined in their belief in genetic essentialism because both GET mechanisms were triggered sequentially. The evidence for this mechanism is the positive treatment effect on SGL and the negative treatment effect on BGE. In the middle school classrooms, students did not have the opportunity to develop the multifactorial component of SGL (Table 1; idea 1) because this content was removed due to time constraints faced by the teacher. Consequently, the treatment effect in these classrooms must have been driven by the population thinking component of HGL (Table 2; idea 1). If so, then previous research

suggests that students who had pre-existing knowledge of the multifactorial component of SGL (Table 1; idea 1) drove the treatment effect in the middle school classrooms (e.g., see Donovan et al., in review). Their prior knowledge allowed them to develop more HGL (Table 2; idea 1), which, in turn, reduced belief in essentialism by triggering the social categorization mechanism. Or, alternatively, the middle school students did not contribute to the treatment effect on BGE because they did not receive the SGL portion of the intervention curriculum that the high school students received. We realize that this explanation is making assumptions about the conceptual change of individual students even though we can only make inferences about conceptual change at the classroom level. However, the change of a classroom is based, in part, in the aggregate change of individuals within it. Future research will need to test the two-step hypothesis we have proposed.

Since our QED is underpowered, and since we do not have content valid measurements for individual students, we do not have the sample size or the instrumentation to discern which of these explanations better fits our data. Because of these limitations, and others, we can only assert that belief in genetic essentialism can be reduced when: (1) classrooms are taught that inheritance and variation in human populations is more complicated than the story told by BGL; and (2) when they also are taught why this complicated story refutes genetic essentialism. This explanation for the results is consistent with previous randomized trials (e.g., Donovan et al., in review; Donovan, Semmens, et al., 2019), and it is also warranted by conceptual change theory.

Under Gregoire's (2003) cognitive affective model of conceptual change, belief accommodation depends on whether learners have the prior knowledge and the motivation to understand information in scientific messages designed to change their beliefs. The fact that classrooms learning from the SGL/HGL curriculum increased in SGL and moved toward coherent anti-essentialist perceptions, attributions, and beliefs is consistent with this model of belief accommodation. It is also somewhat consistent with a KiP perspective on conceptual change.

From the standpoint of KiP, it takes a lot of time to move a student from a naïve lay understanding of genomics to a coherent understanding of BGL, then SGL, and finally HGL. KiP also predicts that any relationship between BGL, SGL, or HGL and belief in genetic essentialism will be incoherent or inconsistent and context specific. The students in our study's treatment classrooms, arguably, developed some BGL before learning about SGL. For example, they had four weeks of BGL genetics instruction before receiving four more weeks of SGL (2 weeks) and HGL instruction (2 weeks). Moreover, only the classrooms using the SGL/HGL curriculum exhibited an increase in SGL and a move toward a coherent set of anti-essentialist perceptions, attributions, and beliefs. Due to the time scale of learning and the coherence of the change observed within classrooms, a claim of conceptual change is partially warranted from a KiP perspective.

At the same time, we have no way of knowing whether these changes are context specific. If we measured these students' beliefs in different social spaces would they be the same? If we engaged students in a sociopolitical debate about whether racial inequality is genetic, then would they use their genetics knowledge to challenge essentialist claims made by other students? What would happen if, in the future, a student in this sample felt a need to justify their racial privilege in another social context? Would they revert to genetic essentialism? Did any of these students replace genetic essentialism with some other race conception, like social constructionism, or racial colorblindness? Of course, we have no way of answering these questions with our present data and a KiP perspective suggests that answers to them will not be clean cut, predictable, or generous to our hypothesis that genomics literacy matters for reducing belief in genetic essentialism. In all likelihood, the change that we have observed is limited to the social space of the biology classroom. Future studies need to explore if, how, when, and for whom such conceptual changes transfer out

of the biology classroom to influence how students make sense of new racial phenomena.

One last issue raised by this study has to do with the lack of any change in genetic essentialism in the BAU/BGL curriculum. Previous studies have found that the phenomena discussed in BGL curricula can lead to increased belief in genetic essentialism (Donovan, 2014, 2016, 2017; Parrott & Smith, 2014). Yet, we did not detect this effect in the pre- or post-intervention periods. One reason for this finding may be the phenomenon of race salience. Developmental psychologists have found that making race salient during social decision making can activate essentialist thinking and further engrain stereotype endorsement (Bigler & Liben, 2007; Pauker et al., 2010). Donovan (2014, 2016, 2017) has also found that when students learn about monogenic diseases with racial terminology their belief in genetic essentialism increases more rapidly than when they learn about these diseases without making race salient. Consequently, in our study, we asked the BAU/BGL teachers not to discuss racial differences in genetic disease prevalence in their curriculum. We did this to make the BAU/BGL curriculum inert so that we could attribute any treatment effect to the SGL/HGL curriculum. This decision may have been consequential and it is deserving of further research. We may have failed to detect a positive time effect on belief in genetic essentialism within the BAU/BGL learning condition because of a lack of race saliency. For this reason, readers should not conclude from our study that BGL is unrelated to belief in genetic essentialism. Alternatively, it could be the case that the BAU/BGL classrooms differed in curriculum and instruction in a way that introduced noise into the data, thereby preventing the detection of a BAU/BGL effect on essentialism (i.e., limitation 1, a type 2 error).

Research is needed to understand if and how BGL instruction affects belief in genetic essentialism. For example, is race saliency required for this relationship to exist? Does motivated reasoning affect this relationship? Which ideas in a BGL curriculum actually affect belief in genetic essentialism? Such questions need to be answered before drawing firm conclusions about the relationship between BGL learning and belief in genetic essentialism of race.

In closing, if our hypotheses are correct, then the type of genomics literacy that genetics education helps students to develop matters because of its impacts on belief in genetic essentialism. While the external validity of our study is limited to secondary schools serving predominantly white and high SES populations, it nevertheless shows that genetics education is a promising venue for helping students understand that racial inequality cannot be reduced to genes. By helping students to understand the complexity of inheritance (SGL) and how this complexity refutes essentialism (HGL), genetics educators can help students understand why it is scientifically problematic and socially prejudiced to use genes to rationalize racial oppression.

References

Bigler, R. S., & Liben, L. S. (2007). Developmental Intergroup Theory: Explaining and Reducing Children's Social Stereotyping and Prejudice. *Current Directions in Psychological Science*, *16*(3), 162–166. https://doi.org/10.1111/j.1467-8721.2007.00496.x

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*(1), 246–263.

Bowling, B. V., Acra, E. E., Wang, L., Myers, M. F., Dean, G. E., Markle, G. C., Moskalik, C. L., & Huether, C. A. (2008). Development and Evaluation of a Genetics Literacy

Assessment Instrument for Undergraduates. *Genetics*, *178*(1), 15–22. https://doi.org/10.1534/genetics.107.079533
Bransford, J. (2000). *How people learn: Brain, mind, experience, and school*. National Academies Press.
Bratsberg, B., & Rogeberg, O. (2018). Flynn effect and its reversal are both environmentally caused. *Proceedings of the National Academy of Sciences*, *115*(26), 6674–6678. https://doi.org/10.1073/pnas.1718793115
Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42.
Bush, W. S., & Moore, J. H. (2012). Genome-wide association studies. *PLoS Computational Biology*, *8*(12), e1002822.
Canning, E. A., Muenks, K., Green, D. J., & Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Science Advances*, *5*(2), eaau4734. https://doi.org/10.1126/sciadv.aau4734
Chen, J. A., & Pajares, F. (2010). Implicit theories of ability of Grade 6 science students: Relation to epistemological beliefs and academic motivation and achievement in science. *Contemporary Educational Psychology*, *35*(1), 75–87.
Committee on How People Learn II: The Science and Practice of Learning, Board on Behavioral, Cognitive, and Sensory Sciences, Board on Science Education, Division of Behavioral and Social Sciences and Education, & National Academies of Sciences, Engineering, and Medicine. (2018). *How People Learn II: Learners, Contexts, and Cultures*. National Academies Press. https://doi.org/10.17226/24783
Darling-Hammond, L. (2010). Structured for failure: Race, resources, and student achievement. In *Doing race: 21 Essays for the 21st Century* (pp. 295–318). W. W. Norton & Company.
Darner, R. (2019). How Can Educators Confront Science Denial? *Educational Researcher*, *48*(4), 229–238. https://doi.org/10.3102/0013189X19849415
Dar-Nimrod, I., & Heine, S. J. (2011). Genetic essentialism: On the deceptive determinism of DNA. *Psychological Bulletin*, *137*(5), 800–818. https://doi.org/10.1037/a0021860
Devlin, B., Daniels, M., & Roeder, K. (1997). The heritability of IQ. *Nature*, *388*, 908–912.
diSessa, A. (1998). What changes in conceptual change? *International Journal of Science Education*, *20*, 1115–1191.
diSessa, A. A. (1993). Ontologies in pieces: Response to Chi and Slotta. *Cognition and Instruction*, *10*(2/3), 272–280.
DiSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, *10*(2–3), 105–225.
diSessa, A. A. (2018). A Friendly Introduction to "Knowledge in Pieces": Modeling Types of Knowledge and Their Roles in Learning. In G. Kaiser, H. Forgasz, M. Graven, A. Kuzniak, E. Simmt, & B. Xu (Eds.), *Invited Lectures from the 13th International Congress on Mathematical Education* (pp. 65–84). Springer International Publishing. https://doi.org/10.1007/978-3-319-72170-5_5
diSessa, A., Sherin, B., & Levin, M. (2016). Knowledge analysis: An introduction. In A. diSessa, M. Levin, & N. Brown (Eds.), *Knowledge and interaction: A synthetic agenda for the learning sciences* (pp. 30–71). Routledge.
Donovan, B. M. (2014). Playing with fire? The impact of the hidden curriculum in school genetics on essentialist conceptions of race. *Journal of Research in Science Teaching*, *51*(4), 462–496. https://doi.org/10.1002/tea.21138

Donovan, B. M. (2015a). Putting humanity back into the teaching of human biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *52*, 65–75. https://doi.org/10.1016/j.shpsc.2015.01.011

Donovan, B. M. (2015b). Reclaiming Race as a Topic of the U.S. Biology Textbook Curriculum. *Science Education*, *99*(6), 1092–1117. https://doi.org/10.1002/sce.21173

Donovan, B. M. (2016). Framing the Genetics Curriculum for Social Justice: An Experimental Exploration of How the Biology Curriculum Influences Beliefs About Racial Difference. *Science Education*, *100*(3), 586–616. https://doi.org/10.1002/sce.21221

Donovan, B. M. (2017). Learned inequality: Racial labels in the biology curriculum can affect the development of racial prejudice. *Journal of Research in Science Teaching*, *54*(3), 379–411. https://doi.org/10.1002/tea.21370

Donovan, B. M., Salazar, B., & Weindling, M. (forthcoming). How can we make genetics education more humane? In *Genetics Education for the 21st Century*.

Donovan, B. M., Semmens, R., Keck, P., Brimhall, E., Busch, K. C., Weindling, M., Duncan, A., Stuhlsatz, M., Buck Bracey, Z., Bloom, M., & Salazar, B. (2019). Towards a More Humane Genetics Education: Learning about the social and quantitative complexities of human genetic variation research could reduce racial bias in adolescent and adult populations. *Science Education*, 1–32. https://doi.org/DOI: 10.1002/sce.21506

Donovan, B. M., Stuhlsatz, M., Edelson, D. C., & Buck Bracey, Z. (2019). Gendered Genetics: How reading about the genetic basis of sex differences in biology textbooks could affect beliefs associated with science gender disparities. *Science Education*, 1–31. https://doi.org/10.1002/sce.21502

Donovan, B. M., Weindling, M., Salazar, B., Duncan, A., Stuhlsatz, M., & Keck, P. (in review). Genomics Literacy Matters: Supporting the Development of Genomics Literacy Through Genetics Education Could Reduce Cognitive Forms of Racial Prejudice. *Journal of Research in Science Teaching*.

Dougherty, M. J., Pleasants, C., Solow, L., Wong, A., & Zhang, H. (2011). A comprehensive analysis of high school genetics standards: Are states keeping pace with modern genetics? *CBE—Life Sciences Education*, *10*(3), 318–327.

Dougherty, Michael J. (2009). Closing the Gap: Inverting the Genetics Curriculum to Ensure an Informed Public. *The American Journal of Human Genetics*, *85*(1), 6–12. https://doi.org/10.1016/j.ajhg.2009.05.010

Duncan, L. E., & Keller, M. C. (2011). A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry. *American Journal of Psychiatry*, *168*(10), 1041–1049. https://doi.org/10.1176/appi.ajp.2011.11020191

Duncan, R. G., Castro-Faix, M., & Choi, J. (2016). Informing a Learning Progression in Genetics: Which Should BE Taught First, Mendelian Inheritance or the Central Dogma of Molecular Biology? *International Journal of Science and Mathematics Education*, *14*(3), 445–472.

Duncan, R. G., Rogat, A. D., & Yarden, A. (2009). A learning progression for deepening students' understandings of modern genetics across the 5th-10th grades. *Journal of Research in Science Teaching*, *46*(6), 655–674. https://doi.org/10.1002/tea.20312

Feldman, M., & Lewontin, R. (1975). The heritability hang-up. *Science*, *190*(4220), 1163–1168. https://doi.org/10.1126/science.1198102

Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist*, *54*(1), 5–20.

Gericke, N., Carver, R., Castéra, J., Evangelista, N. A. M., Marre, C. C., & El-Hani, C. N. (2017). Exploring Relationships Among Belief in Genetic Determinism, Genetics Knowledge, and Social Factors. *Science & Education*, *26*(10), 1223–1259. https://doi.org/10.1007/s11191-017-9950-y

Goldsby, R. A. (1973). HUMAN RACES: THEIR REALITY AND SIGNIFICANCE. *The Science Teacher*, *40*(1), 14–18. JSTOR.

Graves, J. L. (2015). Great Is Their Sin: Biological Determinism in the Age of Genomics. *The Annals of the American Academy of Political and Social Science*, *661*(1), 24–50. https://doi.org/10.1177/0002716215586558

Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. *Educational Psychology Review*, *15*(2), 147–179.

Halperin, E., Russell, A. G., Trzesniewski, K. H., Gross, J. J., & Dweck, C. S. (2011). Promoting the Middle East Peace Process by Changing Beliefs About Group Malleability. *Science*, *333*, 1767–1769. https://doi.org/10.1126/science.1202925

Haslam, N., Bastian, B., Bain, P., & Kashima, Y. (2006). Psychological Essentialism, Implicit Theories, and Intergroup Relations. *Group Processes & Intergroup Relations*, *9*(1), 63–76. https://doi.org/10.1177/1368430206059861

Hutchins, E. (2000). Distributed cognition. *International Encyclopedia of the Social and Behavioral Sciences. Elsevier Science*, *138*.

Jackson Jr., J. P., & Depew, D. J. (2017). *Darwinism, Democracy, and Race: American Anthropology and Evolutionary Biology in the Twentieth Century*. Routledge.

Jamieson, A., & Radick, G. (2013). Putting Mendel in His Place: How Curriculum Reform in Genetics and Counterfactual History of Science Can Work Together. In K. Kampourakis (Ed.), *The Philosophy of Biology* (Vol. 1, pp. 577–595). Springer. http://link.springer.com/10.1007/978-94-007-6537-5_25

Jamieson, A., & Radick, G. (2017). Genetic Determinism in the Genetics Curriculum: An Exploratory Study of the Effects of Mendelian and Weldonian Emphases. *Science & Education*. https://doi.org/10.1007/s11191-017-9900-8

Jayaratne, T. E., Gelman, S. A., Feldbaum, M., Sheldon, J. P., Petty, E. M., & Kardia, S. L. R. (2009). The perennial debate: Nature, nurture, or choice? Black and White Americans' explanations for individual differences. *Review of General Psychology*, *13*(1), 24–33. https://doi.org/10.1037/a0014227

Kahan, D. M. (2012). Cultural cognition as a conception of the cultural theory of risk. In *Handbook of risk theory* (pp. 725–759). Springer. http://link.springer.com/10.1007/978-94-007-1433-5_28

Kahan, D. M. (2016). The politically motivated reasoning paradigm, part 1: What politically motivated reasoning is and how to measure it. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource.*, 1–24.

Kahan, D. M., Braman, D., Gastil, J., Slovic, P., & Mertz, C. K. (2007). Culture and identity-protective cognition: Explaining the white-male effect in risk perception. *Journal of Empirical Legal Studies*, *4*(3), 465–505.

Kampourakis, K. (2017). *Making Sense of Genes* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781316422939

Keller, J. (2005). In Genes We Trust: The Biological Component of Psychological Essentialism and Its Relationship to Mechanisms of Motivated Social Cognition. *Journal of Personality and Social Psychology*, *88*(4), 686–702. https://doi.org/10.1037/0022-3514.88.4.686

Keller, M. C. (2014). Gene × Environment Interaction Studies Have Not Properly Controlled for Potential Confounders: The Problem and the (Simple) Solution. *Biological Psychiatry*, *75*(1), 18–24. https://doi.org/10.1016/j.biopsych.2013.09.006

Lawson, A. E., & Thompson, L. D. (1988). Formal reasoning ability and misconceptions concerning genetics and natural selection. *Journal of Research in Science Teaching*, *25*(9), 733–746.

Leslie, S.-J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 262–265.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lewis, J., & Kattmann, U. (2004). Traits, genes, particles and information: Re-visiting students' understandings of genetics. *International Journal of Science Education*, *26*(2), 195–206. https://doi.org/10.1080/0950069032000072782

Lynch, J., Bevan, J., Achter, P., Harris, T., & Condit, C. M. (2008). A preliminary study of how multiple exposures to messages about genetics impact on lay attitudes towards racial and genetic discrimination. *New Genetics and Society*, *27*(1), 43–56. https://doi.org/10.1080/14636770701843634

Lynch, K. E., Morandini, J. S., Dar-Nimrod, I., & Griffiths, P. E. (2018). Causal Reasoning About Human Behavior Genetics: Synthesis and Future Directions. *Behavior Genetics*. https://doi.org/10.1007/s10519-018-9909-z

MacMahon, B. (1968). Gene-environment interaction in human disease. *Journal of Psychiatric Research*, *6*, 393–402.

Markus, H., & Moya, P. M. L. (2011). *Doing race: 21 essays for the 21st century*. WW Norton & Company. https://repository.library.georgetown.edu/handle/10822/548601

Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Belknap Press.

Meyer, M., Roberts, S. O., Jayaratne, T. E., & Gelman, S. A. (2020). Children's beliefs about causes of human characteristics: Genes, environment, or choice? *Journal of Experimental Psychology: General*.

Michaels, S., & O'Connor, C. (2015). Conceptualizing Talk Moves as Tools: Professional Development Approaches for Academically Productive Discussions. In L. B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socializing Intelligence Through Academic Talk and Dialogue* (pp. 347–361). American Educational Research Association. https://doi.org/10.3102/978-0-935302-43-1_27

Moore, D. S., & Shenk, D. (2017). The heritability fallacy: The heritability fallacy. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(1–2), e1400. https://doi.org/10.1002/wcs.1400

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference*. Cambridge University Press.

Morin-Chassé, A., Suhay, E., & Jayaratne, T. E. (2017). Discord Over DNA: Ideological Responses to Scientific Communication about Genes and Race. *The Journal of Race, Ethnicity, and Politics*, 1–40. https://doi.org/10.1017/rep.2017.17

Morning, A. (2008). Reconstructing Race in Science and Society: Biology Textbooks, 1952-2002. *American Journal of Sociology*, *114*, 106–137.

Morning, A. J. (2011). *The Nature of Race: How Scientists Think and Teach about Human Difference*. University of California Press.

Morning, Ann, Brückner, H., & Nelson, A. (2019). Socially Desirable Reporting and the Expression of Biological Concepts of Race. *Du Bois Review: Social Science Research on Race*, 1–17. https://doi.org/10.1017/S1742058X19000195

Morton, T. A., Hornsey, M. J., & Postmes, T. (2009). Shifting ground: The variable use of essentialism in contexts of inclusion and exclusion. *British Journal of Social Psychology*, *48*(1), 35–59. https://doi.org/10.1348/014466607X270287

Murnane, R. J., & Willett, J. B. (2011). *Methods Matter: Improving Causal Inference in Educational and Social Science Research*. Oxford University Press, USA. https://books.google.com/books/about/Methods_Matter_Improving_Causal_Inferenc.html?id=lA0qSsQk_AgC

Myths of Human Genetics: Introduction. (n.d.). Retrieved November 15, 2019, from http://udel.edu/~mcdonald/mythintro.html

NGSS. (2016). *Standards by Topic | Next Generation Science Standards*. http://www.nextgenscience.org/overview-topics

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, *87*(2), 224–240. https://doi.org/10.1002/sce.10066

Osborne, J., Donovan, B. M., Henderson, J. B., MacPherson, A. C., & Wild, A. (2016). *Arguing From Evidence in Middle School Science: 24 Activities for Productive Talk and Deeper Learning*. Corwin Press. https://us.corwin.com/en-us/nam/arguing-from-evidence-in-middle-school-science/book250184 description

Parrott, R. L., Silk, K. J., Dillow, M. R., Krieger, J. L., Harris, T. M., & Condit, C. M. (2005). Development and validation of tools to assess genetic discrimination and genetically based racism. *Journal of the National Medical Association*, *97*(7), 980–990.

Parrott, R., & Smith, R. A. (2014). Defining Genes Using "Blueprint" Versus "Instruction" Metaphors: Effects for Genetic Determinism, Response Efficacy, and Perceived Control. *Health Communication*, *29*(2), 137–146. https://doi.org/10.1080/10410236.2012.729181

Pauker, K., Ambady, N., & Apfelbaum, E. P. (2010). Race Salience and Essentialist Thinking in Racial Stereotype Development. *Child Development*, *81*(6), 1799–1813.

Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Literacy and Science: Each in the Service of the Other. *Science*, *328*(5977), 459–463. https://doi.org/10.1126/science.1182595

Phenomena | Next Generation Science Standards. (n.d.). Retrieved August 6, 2020, from https://www.nextgenscience.org/resources/phenomena

Radick, G. (2015). Beyond the "Mendel-Fisher controversy." *Science*, *350*(6257), 159–160. https://doi.org/10.1126/science.aad4140

Reardon, S. F., Kalogrides, D., & Shores, K. (2019). The Geography of Racial/Ethnic Test Score Gaps. *American Journal of Sociology*, *124*(4), 1164–1221. https://doi.org/10.1086/700678

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., & Feldman, M. W. (2002). Genetic structure of human populations. *Science*, *298*(5602), 2381–2385.

Rosenberg, Noah A. (2011). A Population-Genetic Perspective on the Similarities and Differences among Worldwide Human Populations. *Human Biology*, *83*(6), 659–684. https://doi.org/10.3378/027.083.0601

Rosenberg, Noah A., Edge, M. D., Pritchard, J. K., & Feldman, M. W. (2018). Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. *Evolution, Medicine, and Public Health*, *2019*(1), 26–34. https://doi.org/10.1093/emph/eoy036

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*(4), 475–522.

Shaw, K. R. M., Horne, K. V., Zhang, H., & Boughman, J. (2008). Essay Contest Reveals Misconceptions of High School Students in Genetics Content. *Genetics*, *178*(3), 1157–1168. https://doi.org/10.1534/genetics.107.084194

Shemwell, J., Chase, C. C., & Schwartz, D. L. (2015). *Seeking the General Explanation: ATest of Inductive Activities for Learning and Transfer*. *52*(1), 58–83.

Solomon, R. L., & Lessac, M. S. (1968). A control group design for experimental studies of developmental processes. *Psychological Bulletin*, *70*(3), 145–150.

Somers, M.-A., Zhu, P., Jacob, R., & Bloom, H. (2013). The Validity and Precision of the Comparative Interrupted Time Series Design and the Difference-in-Difference Design in Educational Evaluation. *MDRC*. http://eric.ed.gov/?id=ED545459

St.Clair, T., Cook, T. D., & Hallberg, K. (2014). Examining the Internal Validity and Statistical Precision of the Comparative Interrupted Time Series Design by Comparison With a Randomized Experiment. *American Journal of Evaluation*, *35*(3), 311–327. https://doi.org/10.1177/1098214014527337

Stern, F., & Kampourakis, K. (2017). Teaching for genetics literacy in the post-genomic era. *Studies in Science Education*, *53*(2), 193–225. https://doi.org/10.1080/03057267.2017.1392731

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. A. Duschl & R. J. Hamilton (Eds.), *Philosophy of Science, Cognitive Psychology, and Educational Theory and Practice* (pp. 147–176). State University of New York Press. https://books.google.com/books?hl=en&lr=&id=bJpW2WvfVkcC&oi=fnd&pg=PA147&dq=conceptual+change+strike+and+posner+1992&ots=1derTn4Azf&sig=EOck7eOdFcjrD9yshE6b_hoVbHU

Tucker-Drob, E. M., & Bates, T. C. (2016). Large Cross-National Differences in Gene × Socioeconomic Status Interaction on Intelligence. *Psychological Science*, *27*(2), 138–149. https://doi.org/10.1177/0956797615612727

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science*, *14*(6), 623–628.

Ueno, N. (1993). Reconsidering p-prims theory from the viewpoint of situated cognition. *Cognition and Instruction*, *10*(2–3), 239–248.

Venville, G., Gribble, S. J., & Donovan, J. (2005). An exploration of young children's understandings of genetics concepts from ontological and epistemological perspectives. *Science Education*, *89*(4), 614–633. https://doi.org/10.1002/sce.20061

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes*. Harvard University Press.

Williams, M. J., & Eberhardt, J. L. (2008). Biological conceptions of race and the motivation to cross racial boundaries. *Journal of Personality and Social Psychology*, *94*(6), 1033–1047.

Willinsky, J. (2020). The Confounding of Race in High School Biology Textbooks, 2014-2019. *Science & Education*.