Reasoning about social relationships reduces class differences in tests of cognitive aptitude

Nicholas J. Fendinger[1]*, Andrea G. Dittmann[2], Eric D. Knowles[1]

[1]Department of Psychology, New York University

[2]Management and Organization, Marshall School of Business, University of Southern California

Data, code, and supplemental materials are available at the project's Open Science Framework page. We have no known conflicts of interest to declare. This work was supported by a National Science Foundation Graduate Research Fellowship Program awarded to Nicholas J. Fendinger (DGE-1839302). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

This work was supported in part through the NYU IT High Performance Computing resources, services, and staff expertise. Study 2 Data were collected by the Time-sharing Experiments for the Social Sciences program, NSF Grant 2017464, Maureen Craig, James Druckman, and Jeremy Freese, Principal Investigators.

*Correspondence concerning this article should be addressed to Nicholas J. Fendinger, New York University, 6 Washington Place, New York, NY, 10003, United States. Email: nfendinger@nyu.edu

Abstract

Class differences in aptitude testing are often attributed to working-class deficiencies in a host of basic cognitive skills. The present research challenges this deficit narrative in the context of logical reasoning. We argue that reasoning tests often privilege abstract styles of reasoning as a "neutral" benchmark and provide a poor match to social-cognitive competencies promoted in working-class contexts. We evaluate this hypothesis in archival data from an international study of mathematical aptitude (N = 158,216), a pre-registered nationally-representative experiment on logical deduction (N = 781), and an integrative data analysis (IDA) combining data from Study 2 with three supplemental studies (N = 2,269). In Study 1, reasoning items higher in social-relational content (i.e., mentions of people) yielded smaller class gaps among eighth-grade students from 46 countries. Study 2 and the IDA experimentally replicate and extend these findings, demonstrating that reframing an abstract reasoning task to recruit social-cognitive skills reduces class-based performance differences among adults. Study 2 and the IDA also highlight that these effects cannot be explained by the mere inclusion of real-world content (e.g., money, food), but are most robustly produced by content invoking social relations. Our results support a cultural-psychological account of reasoning performance, highlight working-class strengths in social reasoning, and point to inclusive approaches to testing as a way to attenuate social-class gaps in aptitude testing performance.

Keywords: social class; achievement; culture; theory of mind; logical reasoning

Significance Statement

Persistent class-based differences in academic outcomes are often attributed to working-class deficits in cognitive competencies. Rather than reflecting inherent deficits, we contend that popular tests of cognitive ability often rely on abstract content—defined as novel or decontextualized inputs—as a "neutral" measure of ability when in fact this content obscures the *social*-cognitive strengths fostered within working-class contexts. Supporting these claims, we find that reasoning problems including more (vs. fewer) references to people yield smaller class-based performance gaps in a global assessment of mathematical aptitude amongst eighth-grade students. Extending these findings in a synthesis of multiple experiments, we demonstrate that class performance gaps found on a typical, abstract deductive-reasoning task can be sizeably reduced when the same task is reframed to invoke social relationships. These findings challenge the assumption that reasoning ability is best captured by abstract tests, which may exclude the cognitive strengths of people from marginalized communities.

Reasoning about social relationships reduces class differences in tests of cognitive aptitude

Social-class inequality is partly rooted in achievement gaps that emerge prior to school entry and echo across the lifespan (Halle et al., 2009; Phillips et al., 2020; Stephens et al., 2017, 2019). Some researchers propose deficit models to explain such gaps, attributing them to impairments of reasoning and problem-solving abilities among members of poor and working-class communities (Bradley & Corwyn, 2002; DeGarmo et al., 1999; Gottfredson, 2004; Hall et al., 2014). To support these deficit perspectives, researchers often cite the early emergence of class disparities on formal tests of cognitive aptitude. Importantly, common aptitude tests tend to rely on a constrained set of *abstract* content to assess performance: inputs that are novel, symbolic, or decontextualized (Carpenter et al., 1990). For instance, test-takers might be asked to verify a logical rule based on relations between arbitrary numbers and letters (Ellis et al., 2022; Young et al., 2022).

The present research challenges a deficit-based understanding of class gaps in reasoning performance. We argue that reasoning questions rooted in abstract content tend to provide a poor cultural match to the cognitive competencies fostered within working-class contexts.[1] Owing to a relative lack of access to material resources, working-class contexts tend to promote interdependent patterns of behavior and cognition, whereas more affluent, middle-class contexts tend to instill independent ways of being and thinking (Fiske & Markus, 2012; Stephens et al., 2024). In particular, working-class cultures foster expertise in *social*-cognitive skills, as evidenced by superior performance on tests of face memory (Dietze et al., 2024), empathic accuracy (Dietze & Knowles, 2021; Kraus et al., 2010; Monroy et al., 2022), and visual perspective-taking (Dietze & Knowles, 2021). Importantly, these social-cognitive

[1]We use the term *working-class contexts* to refer to social milieus where people tend to have less than a four-year college degree, work in blue-collar occupations, and have relatively low incomes. We use the term *middle-class contexts* to refer to social locations in which people tend to have at least a four-year college degree, work in white-collar occupations, and have relatively high incomes (Stephens et al., 2025).

processes recruit many of the same underlying competencies allegedly hindered among those lower in social class (e.g., working-memory capacity and executive functioning; for a review, see Fendinger et al., 2023). Thus, it may be that class differences in cognitive achievement hinge less on *how* people reason than on *what* they reason about.

In the present work, we theorize that different types of test content vary in terms of their congruence to the cognitive competencies promoted within working-class cultures. As such, we hypothesize that social-class gaps in reasoning performance will be evident when tests rely on abstract content. In contrast, class-performance gaps will be attenuated when test content recruits working-class social-cognitive strengths. Before describing the studies, we make the case that abstract logic has been privileged as "pure" reasoning, thus discounting the types of thinking and problem-solving that individuals routinely perform in the real world. We then integrate two pertinent theoretical perspectives—the "hidden talents" approach and sociocultural models of social class—to identify which kinds of real-world content should most consistently attenuate class-based performance gaps in reasoning performance.

What Counts as Reasoning?

Intellectual ideals favoring logical abstraction are rooted in Greek philosophical traditions (Nisbett et al., 2001). Most prominently, Plato (c. 427–348 BC) argued that, at its most fundamental, reality consists in universal and unchanging ideas or "forms" rather than tangible particulars or "substances" (Sedley, 2016). Plato thus regarded abstract logic as the route to truth and saw perceptual experience as only serving to distort reasoning (Reeve, 2004). The Platonists were so committed to abstract forms and operations that they were willing to reject sensory evidence when it clashed with established logical rules (Lloyd, 1990). The Platonic ideal of reasoning is famously depicted in Rodin's *The Thinker*, whose hunched stature and downcast eyes emphasize that "good" thinking is an internal activity that must be protected from disruption by the outside world (Plaut & Markus, 2005).

The Platonic ideal of abstraction as “pure” reasoning is reflected in aptitude tests that privilege abstract content in assessing cognitive ability (Carpenter et al., 1990; Ellis et al., 2022; Evans, 2002; Young et al., 2022). Proponents of such approaches argue that abstraction minimizes the impact of prior knowledge and therefore most validly diagnoses underlying cognitive ability (Carpenter et al., 1990; Evans, 2002). Conversely, tests invoking real-world content are seen as inferior measures that might “introduce biases that could disguise the [test-taker's] underlying competence” (Evans, 2002, p. 981).

Despite their preeminence, perspectives that equate abstract thinking with the “pure process” of reasoning are heavily contested. Indeed, the opposing view has an equally ancient pedigree. Plato’s student Aristotle (384–322 BC) rejected his teacher’s theory of forms and instead proposed the doctrine of *hylomorphism*—the perspective that every object is a coequal compound of matter and form (Barnes, 1996). Consequently, Aristotle viewed both lay and scientific knowledge as created largely through direct interactions with the real world. Modern logical pragmatists echo Aristotle’s account of reasoning in their contention that human beings seldom reason abstractly in their day-to-day lives (Nickles, 2018; Simon, 1965). To the pragmatists, concrete problem-solving—such as attempting to infer who is friends with whom at a party—plays a critical role in human reasoning and is just as “pure” a form of reasoning as abstract logic (Johnson, 2015; Lakoff, 2012).

Anthropologists and cultural psychologists further emphasize that reasoning performance is fundamentally shaped by the concerns and values that predominate within a given cultural context (Nisbett et al., 2001). Rather than reflecting universal processes, even basic cognitive processes reflect long-standing cycles of socialization to particular cultures and habits of thought (Nisbett et al., 2001), with measures of aptitude tending to reflect the concerns prioritized within a given culture (Rogoff et al., 2017). For instance, traditional tests of intellectual ability correspond only partially to non-Westerners’ lay conceptions of intelligence (Grigorenko et al., 2001). Kenyan childrens’ scores on Western aptitude tests correlate poorly, and

sometimes negatively, with performance on tests of practical aptitude with herbal medicines (Sternberg et al., 2001). In the U.S., standard markers of language development (e.g., number of words spoken) may underestimate the linguistic competencies of children from working-class contexts whose speech patterns reflect interdependent norms of social interaction (e.g., listening before speaking; Rogoff et al., 2017). Given this evidence that aptitude tests do not seem to measure individuals' underlying abilities in a neutral manner, an alternative method to de-biasing testing instruments would be to better sample the universe of real-world content such that multiple communities and cultures can showcase their underlying strengths.

Seminal work using the Wason selection task—one of the most commonly used tests of deductive inference (Kellen & Klauer, 2020; Wason, 1968)—highlights how individuals' reasoning performance can vary drastically as a function of a test's content. In the typical, abstract version of the Wason task, participants are asked to test rules such as "If a card shows *5* on one face, then the word *Excellent* is on the opposite face" (see Figure 1). Although as few as 10% of test-takers overall reach the correct solution, scores on the abstract task increase with time spent in college as well as with mathematical training (Kellen & Klauer, 2020; Lehman & Nisbett, 1990). This suggests that performance may reflect experience in middle-class cultural spaces (e.g., university; Stephens, Fryberg, et al., 2012) as much as it does intrinsic reasoning ability.

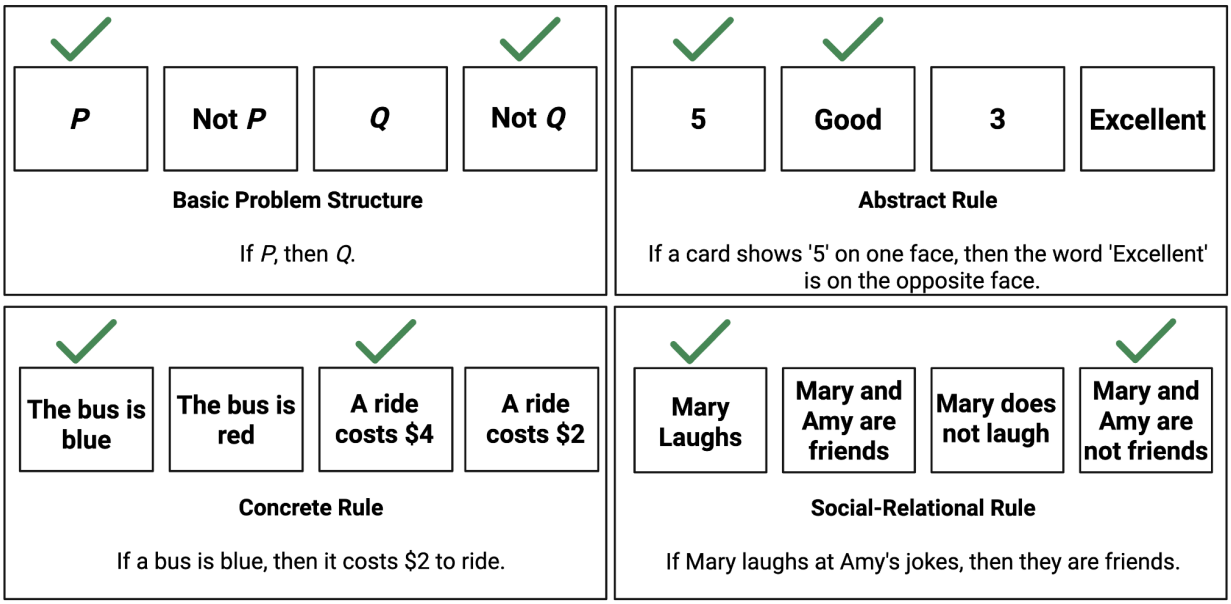

Figure 1. Examples of logical reasoning problems used in Study 2. Green check marks indicate the cards to turn to reach the correct solution. This figure was created using BioRender.com.

Since the task's inception, studies have found that performance can be improved by giving it substantive content. In one famous example, participants tested the rule "If a person drinks beer, then they must be over 18 years old" using cards showing a person's age (16 or 25) on one side and a beverage (beer or juice) on the other. In some studies, 70–80% of participants reach the correct solution in such social versions of the task (Braine, 1978; Cheng & Holyoak, 1985; Cosmides, 1989; Cox & Griggs, 1982). While various mechanisms have been proposed to explain the effects of domain-specific content on reasoning performance—including reasoning schemas (Cheng & Holyoak, 1985), natural logics (Braine, 1978), memory-cueing (Cox & Griggs, 1982), and evolved cognitive modules (Cosmides, 1989)—it is clear that people reason best about content that is familiar and important to them.

Aligning Testing Content with Working-Class Social Cognitive Strengths

Work from the "hidden talents" perspective—which posits that children raised in adverse contexts develop adaptive cognitive skills for solving problems in precarious environments (Ellis et al., 2022)—highlights how real-world content can attenuate class differences in performance. For instance, people raised in relatively unpredictable contexts during childhood exhibit better attention-shifting efficiency when test conditions evoke uncertainty (vs. control; Mittal et al., 2015). When tests of working memory incorporate ecologically-valid content (e.g., money, buses, and people), individuals exposed to violence and poverty perform as well as their advantaged peers (Frankenhuis et al., 2020; Young et al., 2022). These findings emphasize that *underperformance* is not the same as *inability*, and that typical aptitude tests may obscure the strengths of marginalized communities.

Importantly, past research adopting the hidden talents perspective emphasizes that the *type* of real-world content itself may matter. Whereas memory tests involving age relations did not impact performance, memory tests involving power hierarchies attenuated performance differences between individuals with low (vs. high) exposure to neighborhood violence (Frankenhuis et al., 2020). However, past efforts to examine the effects of content on class-based performance gaps have often compared abstract content (e.g., numbers, letters, shapes) against a broad category of real-world content (e.g., money, buses, and people; Muskens et al., 2024; Young et al., 2022). Although promising, we believe that integrating the hidden talents perspective with sociocultural theories of social class can provide granularity with respect to the specific kinds of real-world content that most consistently attenuate class-based performance gaps.

As cultural psychologists and sociologists have long noted, individuals from working-class contexts express a form of hard interdependence meant to cope with low-resourced, constrained, and risky environments (Carey & Markus, 2017; Kusserow, 1999). Owing to limited material resources and low geographic mobility, working-class contexts

promote smaller social networks of enduring ties that are seen as essential to navigating life's challenges (Carey & Markus, 2017). Those from working-class contexts are thus more likely to prioritize social ties and emphasize relationships as a source of strength when confronted with crises and life-altering decisions (Piff et al., 2012; Stephens et al., 2009). Members of working-class communities are also more self-protective and vigilant in their relationships, because it is crucial these relationships "pay off" in order to navigate their more chronically precarious social and material environments (Emery & Finkel, 2022).

Perhaps stemming from these relational realities, individuals from working-class contexts appraise other individuals as highly motivationally relevant and devote greater attention to other people relative to those from middle-class contexts (Dietze & Knowles, 2016). Greater attunement to others may contribute to working-class strengths in a host of social-cognitive skills such as emotion perception (Dietze & Knowles, 2021; Kraus et al., 2010; Monroy et al., 2022) and visual perspective-taking (Dietze & Knowles, 2021). Thus, building on these cultural approaches to studying social class, we suggest and test the idea that content prioritizing social relationships might be particularly effective in reducing class-based performance gaps.

As a secondary aim, we also sought to examine whether the effects of content on class-testing gaps generalizes to other testing domains outside of reasoning, given that prior work in the cognitive sciences suggests competing hypotheses regarding the effects of real-world content. On the one hand, real-world content should enhance performance by facilitating effective problem-solving strategies (Kotovsky et al., 1985; Schliemann & Carraher, 2002) and mental simulation (Glenberg et al., 2004). Real-world content may also increase test-takers' motivation and interest by phrasing problems in relatable terms rather than abstract games (Greer, 1997; Verschaffel et al., 2000). On the other hand, real-world content may harm performance by serving as a distractor, because such content could draw test-takers' attention away from the focal rules and principles embedded in the problem (Duquennois, 2022).

Importantly, whether real-world content facilitates problem solving depends on the type of problem. In one study, grounding mathematical content in salient everyday content led to a better conceptual understanding of core concepts but more arithmetic errors when those concepts were applied (McNeil et al., 2009). Hence, real-world content might facilitate complex reasoning by making unfamiliar concepts and logical rules more accessible while hindering performance on questions about low-level concepts and their applications.

The Present Research

The present research tested a cultural-mismatch account of class differences in tests of deductive and mathematical reasoning. We hypothesized that logical reasoning problems grounded in social relationships would attenuate class gaps in performance.

First, in an analysis of real-world global testing data (Study 1), we leverage linguistic analyses to test whether questions higher (vs. lower) in social-relational content yield smaller class-based performance gaps in mathematical reasoning. Given that the dataset also includes questions tapping other cognitive domains, we also probed whether our effects were unique to reasoning.

In Study 2, respondents from a nationally-representative sample completed modified versions of the standard abstract Wason task which rendered it more culturally relevant to test-takers from working-class contexts. This modified version required participants to infer other people's relationships and mental states.

Finally, we conducted an integrative data analysis (IDA) on Study 2 and Supplemental Studies 1-3 to provide a more precise estimate of how much content reframing strategies can attenuate class-based testing gaps.

Materials and Methods

Transparency and Openness. All data, materials, and analytic code for the studies can be found at https://osf.io/qnsxv/?view_only=f93fdd27e0f24030b838ee7d619fcd0d. Study 1 analyses were not pre-registered. Study 2 analyses were pre-registered at AsPredicted

(#229671). Data were analyzed using R, version 4.4.2 (R Core Team, 2024) and the package ggplot2, version 3.5.1 (Wickham, 2016).

Study 1

Sample and Materials. We obtained our sample from the Trends in International Mathematics and Science Studies (TIMSS; Martin & Mullis, 2013) 2011 eighth grade cohort (*N* = 158,216; 50% girls; 50% boys). TIMSS is an international assessment of student achievement in the math and sciences. In the present sample, data encompass 46 countries[2] from across the world. We restricted our analyses to the 2011 eighth grade cohort's math section, given that this subsample had full-text item information and a well-validated proxy for social class (i.e., parental education, dichotomized such that 1 = has a four-year degree or higher and 0 = does not have a four-year degree) analogous to that used in our experimental studies. Most of the eighth grade cohort's parents lacked a four-year degree (54% did not have a four-year degree; 26% had a four-year degree or higher; 20% did not have data for parental education). For detailed information about the countries sampled, we refer the reader to the TIMSS 2011 student information document linked in our supplemental materials.

Measure of Social Class. Of the items available to us in the TIMSS dataset, we used parental educational attainment (1 = has a four-year degree, 0 = does not have a four-year degree) to proxy adolescents' social class contexts. We see this measure as being the best representation of individuals' social class contexts in childhood, given past work demonstrating that students from working-class backgrounds (i.e., those whose parents lack a 4-year degree or higher) often experience a mismatch with middle-class norms prevalent in classroom settings from preschool to graduate school (Goudeau et al., 2024). This measure of social class is also consistent with our use of adults' own educational attainment in Study 2.

[2] The 2011 TIMSS dataset includes data from eighth-graders in 43 countries and from ninth-graders in 3 countries. For brevity, we refer to this sample simply as the “eighth-grade” cohort.

Measure of Social-Relational Content in Test Questions. To assess the relative amount of social-relational content within each testing item, we analyzed each full-item text using LIWC-22 software (Pennebaker et al., 2022). LIWC compares each word in the target texts to a dictionary and calculates the percentage of total words in the text that match those in various categories. Each category is meant to capture the relative frequency of specific psychological processes present in a given text sample. Thus, to proxy the relative amount of social-relational content in each item, we extracted the "social referents" category which computes the relative frequency of references to social others in a text (e.g, pronouns, names, and words indicating social relationships such as 'friends' or 'girlfriend'). As such, this category is a good index of the relative amount of social information students would need to parse and understand in order to reach the correct answer to a given question. For sample items that are higher (vs. lower) in social-relational content, see Table 1.

Testing Performance. The number of correctly answered questions served as our dependent variable. We calculated separate totals for each cognitive domain (knowing, applying, and reasoning).

***Analyses*.** We constructed a series of mixed-effects Poisson regressions to ascertain whether socially-relevant content reduces class gaps in testing outcomes. To understand whether this effect is constrained to logical reasoning or extends to other problem types, we ran separate models for each TIMSS cognitive domain (i.e., knowing, applying, and reasoning questions). In each model, we regressed answer correctness (1 = correct, 0 = incorrect) on an interaction between the social referents score and parental educational attainment (1 = 4-year degree or higher, 0 = no 4-year degree). In full models, we also adjusted for LIWC-22 measures of the amount of quantitative and spatial language, number of words, and number of big words (i.e., words with more than 6 characters) in each item, as well as student sex (1 = female, -1 = male). We allowed each of our control variables to replace parental education in the interaction with the social referents score (Yzerbyt et al., 2004). We also included random-intercepts to

account for the nested nature of the data at both the student and country levels. Given the size and complexity of our models, we ran these analyses on our institution's high performance computing cluster and recommend others do the same to ensure the models converge in a reasonable amount of time.

Table 1

Example items from the full set of mathematical items released by TIMSS. To measure social-relational content, we analyzed items using the 'social referent' dictionary in LIWC-22 software (Pennebaker et al., 2022).

Cognitive Domain	Low Social-Relational Content	High Social-Relational Content
Knowing (*M* =1.52%, *SD* = 3.24%)	**[Social References = 0%, Word Count = 61]** A piece of paper in the shape of a rectangle is folded in half as shown in the figure above. It is then cut along the dotted line, and the small piece that is cut is opened. What is the shape of the cutout figure? (A. an isosceles triangle, B. two isosceles triangles, C. a right triangle, D. an equilateral triangle)	**[Social References = 9.33%, Word Count = 75]** Pat and Chris were candidates for school president. Here are the election results: Pat 80% Chris 20% How likely would it be for a student asked at random to have voted for Pat? (A. It is certain that the student voted for Pat., B. It is likely that the student voted for Pat., C. It is unlikely that the student voted for Pat., D. It is certain that the student did not vote for Pat.)
Applying (*M* = 1.81%, *SD* = 3.42%)	**[Social References = 0%, Word Count = 44]** The figure above shows a shape made up of cubes that are all the same size. There is a hole all the way through the shape. How many cubes would be needed to fill the hole? (A. 6, B. 12, C. 15, D. 18)	**[Social References = 10.34%, Word Count = 29]** 480 students were asked to name their favorite sport. The results are shown in this table. Use the information in the table to complete and label this pie chart.
Reasoning (*M* = 1.71%, *SD* = 4.39%)	**[Social References = 0%, Word Count = 31]** Which of these is the reason that triangle PQR is a right angle triangle? (A. 32 + 42 = 52, B. 5 < 3 + 4, C. 3 + 4 = 12 - 5, D. 3 > 5 - 4)	**[Social References = 20.83%, Word Count = 96]** The results of a long jump competition were reported as follows: Average Length Team A 3.6 m Team B 4.8 m There were the same number of students in each team. Which statement about the competition MUST be true? A. Each student in team B jumped farther than any student in team A. B. After every student in team A jumped, there was a student in team B who jumped farther. C. As a group, team B jumped farther than team A. D. Some students in team A jumped farther than some students in team B.

Results

In the restricted model for the knowing domain, social-relational content did not have a significant effect on class-based performance gaps, as indicated by a nonsignificant Parental Education × Social Referents interaction. In the applying domain, social-relational content worsened class-based performance differences, as reflected by a positive Parental Education × Social Referents interaction—suggesting that social content might distract test-takers from working-class backgrounds from the focal task. Importantly, social-relational content *reduced* class-based performance differences in the reasoning domain, as indicated by the negative Parental Education × Social Referents interaction. This finding suggests that framing reasoning questions in terms of social-relational content provides a better cultural match to individuals from working-class contexts.

We visualize these interactions in two ways. First, in the top panel of Figure 2, we visualize the interaction on Poisson's logarithmic scale—that is, in terms of the log of correct answers. Second, in the bottom panel of Figure 2, we express the interaction in terms of incident rate ratios (IRRs) associated with attainment of a four-year college degree in each framing condition. IRRs allow us to compare the proportional increase in correct answers as a function of a college degree and thus the difference in the class-based performance gap. This approach is particularly useful for expressing the magnitude of performance differences at the low end of the scale. To illustrate, IRRs properly convey that an improvement in correct solutions from 5% to 10% is, proportionally, tenfold larger than an increase from 50% to 55%. These results remain substantively unchanged when covariates (e.g., quantitative language, spatial language, number of big words, word count, and student sex) are included in the model (see Table 2). For exploratory analyses probing country-level moderators, we refer the reader to supplemental tables S4-S5 and figures S2-S3.

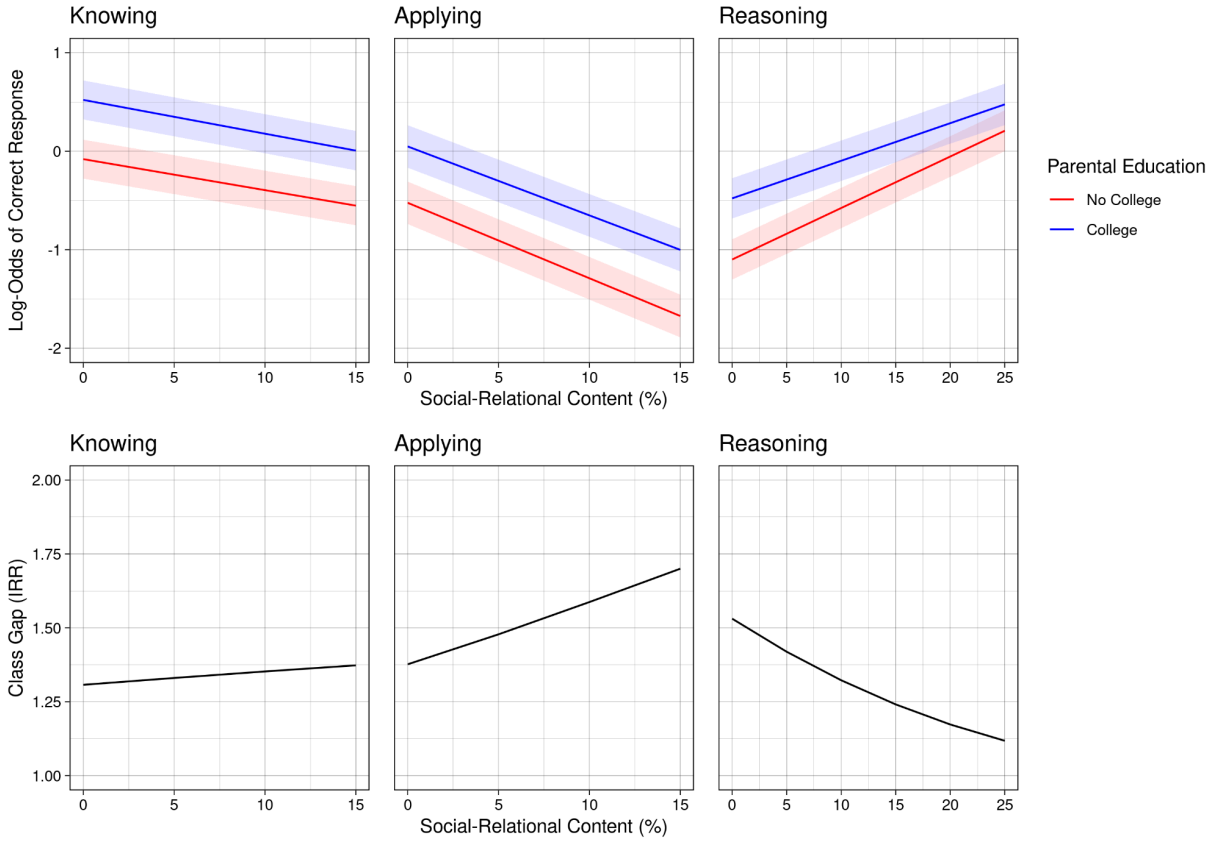

Figure 2. Questions high in social-relational content did not significantly impact the class gap in the knowing domain, worsened the class gap in the applying domain, and reduced the class gap in the reasoning domain (Study 3). Top panels: Scores are expressed in terms of Poisson's logarithmic scale. Bottom panels: The performance difference attributable to educational attainment is expressed in terms of incidence rate ratios (IRRs). For reference, an IRR of 1 represents the absence of an education-based performance gap.

Table 2

Study 1 results for each cognitive domain using parental education as a proxy for children's class contexts.

	Knowing		Applying		Reasoning	
	B (SE B) (restricted)	B (SE B) (full)	B (SE B) (restricted)	B (SE B) (full)	B (SE B) (restricted)	B (SE B) (full)
Social Referents (SR)	-0.03 (0.001)***	0.02 (0.001)***	-0.08 (0.001)***	-0.07 (0.001)***	0.05 (.001)***	0.04 (0.001)***
Parental Education (PE)	0.60 (0.01)***	0.63 (0.01)***	0.58 (0.01)***	0.59 (0.01)***	0.62 (0.01)***	0.65 (0.01)***
SR × PE	-0.003 (0.002)	0.01 (0.002)***	0.01 (0.002)***	0.01 (0.002)**	-0.01 (0.001)***	-0.01 (0.002)**
Sex		0.16 (0.01)***		-0.004 (0.01)		0.02 (0.01)
Quantitative Language		0.07 (0.003)***		0.13 (0.003)***		-0.12 (0.004)***
Spatial Language		0.02 (0.003)***		0.03 (0.003)***		0.10 (0.004)***
Big Words		-0.43 (0.004)***		-0.02 (0.004)***		-0.51 (0.005)***
Word Count		0.17 (0.003)***		-0.05 (0.004)***		0.41 (0.005)***
Sex × PE		-0.03 (0.01)*		-0.01 (0.02)		-0.02 (0.02)
Quantitative × PE		-0.01 (0.005)		-0.07 (0.01)***		0.06 (0.01)***
Spatial × PE		0.01 (0.01)		0.02 (0.01)***		-0.04 (0.01)***
Big Words × PE		-0.07 (0.01)***		0.02 (0.01)***		0.01 (0.01)*
Word Count × PE		-0.02 (0.01)**		-0.02 (0.01)**		-0.06 (0.01)***

Note. N_{total} = 158,216. Mixed-effects Poisson regressions were run for each cognitive domain, regressing answer correctness (1 = correct, 0 = incorrect) on an interaction with items' social content and parental education. Models included a random effect for students and country. Sex was coded as 1 = girl and -1 = boy. Results are presented for restricted models omitting controls and full models including controls. Significance codes: p<.001 '***', p<.01 '**', p<.05 '*'

Study 1 Discussion

Using linguistic analyses to quantify the amount of social-relational language in real-world testing data from 46 countries, we find that mathematical reasoning questions higher in social-relational content yielded smaller class-based performance differences. This effect did not generalize to questions probing students' basic mathematical knowledge. Moreover, we find that social-relational language worsens class gaps on items testing students' ability to apply mathematical concepts.

Although promising, using real-world testing data lacks the experimental control necessary to parse which kinds of content most consistently reduce class-gaps in logical aptitude. For example, we find that mentions of money and food similarly reduce class-based performance discrepancies (see tables S1-S2). However, of the 10 items which mention money, eight of those also mention other people. Of the 5 items that mention food, two of these include mentions of other people. Therefore, many of the items that invoke social relationships also contain references to other, salient concrete inputs (e.g., money, food), limiting our ability to differentiate between the sociocultural vs. hidden talents perspectives on social class. To address this limitation, we conducted a high-powered, pre-registered experiment using a nationally-representative sample in which participants evaluated the same underlying logical rule while manipulating its content. We also conducted an integrative data analysis (IDA)—combining data from every iteration of this paradigm we have ever run—to provide a more precise estimate of the effect of social-relational content on the class-based testing gap.

Study 2

Materials and Method

Participants. A probability sample of 781 Americans (M_{age} = 51.32; SD_{age} = 16.05) were recruited through NORC's AmeriSpeak Panel, as part of the NSF-funded Time-Sharing Experiments in the Social Sciences (TESS). The sample was nationally representative in terms of gender, age, education, and geography (Census division; for more detail regarding NORC's

sampling procedures and methods, see supplemental materials file 'TESS Final Sample Description' and/or https://amerispeak.norc.org/us/en/amerispeak/about-amerispeak/panel-design.html). Most participants were White (66% White, non-Hispanic; 10% Black, non-Hispanic; <1% other, non-Hispanic; 16% Hispanic; 3% more than one race; 4% Asian/Pacific Islander), about evenly split between individuals with (48%) and without (52%) a four-year college degree, and were evenly split between men (50%) and women (50%). The study reported here was approved by the Institutional Review Board at (institution removed for blind review).

Procedure. The data for Study 2 were collected May 20–June 10, 2025. All participants completed logic problems based on the Wason selection task (Kellen & Klauer, 2020; Wason, 1968). In the Wason task, participants are given a logical conditional of the form $if\ P,\ then\ Q$ ($P \rightarrow Q$) and shown four cards corresponding to P, not-P (i.e., $\neg P$), Q, and not-Q (i.e., $\neg Q$). Participants are then asked to select the cards that must be turned to test the conditional's accuracy. A response is considered correct if and only if a participant selects the P card to verify that the consequent is Q (per the *modus ponens* rule) and the $\neg Q$ card to verify that the antecedent is $\neg P$ (per the *modus tollens* rule).

To assess whether social-relational content attenuates class differences found in the typical, abstract version of the task (Lehman & Nisbett, 1990) and determine whether social-relational-content attenuates class-gaps in reasoning performance over and above general, real-world content, participants completed 12 Wason-style logic problems split across three categories: 4 *abstract* rules (e.g., "If a card shows *5* on one face, then the word *Excellent* is on the opposite face"), 4 *concrete* rules (e.g., "If a bus is blue, then it costs $2 to ride"), and 4 *social-relational* rules (e.g., "If Mary laughs at Amy's jokes, then they are friends").

The full set of Wason problems used in Study 2 may be found in the supplemental materials. Participants completed the card problems in randomized order. Upon completing the

problems, participants filled out a brief demographic questionnaire. They were then debriefed and compensated.

Measures of Social Class

Social class was proxied using educational attainment (1 = has a four-year degree or higher, 0 = does not have a four-year degree; Snibbe & Markus, 2005).

Task Performance

The number of logic questions answered correctly served as our dependent variable. Correct answers were summed separately for each condition.

Analyses

To probe the relationship between Wason task performance and social class, we conducted mixed-effects Poisson regressions. The number of correct solutions was regressed on participants' education, a contrast comparing the concrete condition to the abstract condition, a contrast comparing the social-relational condition to the abstract condition, and the interactions between each of these contrasts and participants' education. To adjust for differential selection probabilities and align the sample with U.S. population benchmarks, sampling weights were applied in all analyses.

Entered into analyses as controls were participant age (*z*-scored), participant race (1 = White, 0 = nonwhite), participant gender (1 = female, 0 = male), and sampling weight. Given that the effects of interest were interactions (e.g., Education × Social-Relational Cards), we also allowed each control variable to interact with the card contrast variables (Yzerbyt et al., 2004). A random intercept for participant was specified to account for the nested structure of our data (with 3 observations per participant).

Results

We hypothesized that social-class gaps in performance on the Wason selection task (Kellen & Klauer, 2020; Wason, 1968) would be attenuated when the task is framed in terms of social relations. As hypothesized, educational attainment interacted significantly with the

social-relational condition (but not the concrete condition) in predicting scores on the Wason task in restricted models; in a model that included control variables (i.e., age, race, and gender), this interaction remains in the predicted direction but fails to reach conventional levels of significance (p = .069). These models reveal that the social-relational or theory of mind version of the task yielded a significantly smaller class difference than the nonsocial version (see Table 2).

We again depict the significant interaction in two ways. First, in the left panel of Figure 2, we visualize the interaction on Poisson's logarithmic scale—that is, in terms of the log of correct solutions on the Wason task. Second, in the right panel of Figure 2, we express the interaction in terms of incident rate ratios (IRRs) associated with attainment of a four-year college degree in each framing condition. Participants with a four-year degree or higher correctly answered an average of 1.89 times more abstract questions and 1.57 times as many concrete questions than did participants without a four-year degree. Although the concrete condition reduced the class-performance gap by 35%, this difference was nonsignificant (p = .568).

Unlike the concrete condition, the social-relational condition reduced the class-performance gap to a significant degree. Participants with a four-year degree or higher correctly answered an average of 0.88 times as many social-relational questions as did participants without a four-year degree—a 113% reduction in the class-performance gap seen in the abstract condition (p = .009). The IRR observed in the mentalizing condition (0.88) was not significantly different from 1.0 (p = .436*),* indicating that working-class participants performed at parity with their higher-class counterparts on logic questions designed to recruit social-relational abilities.

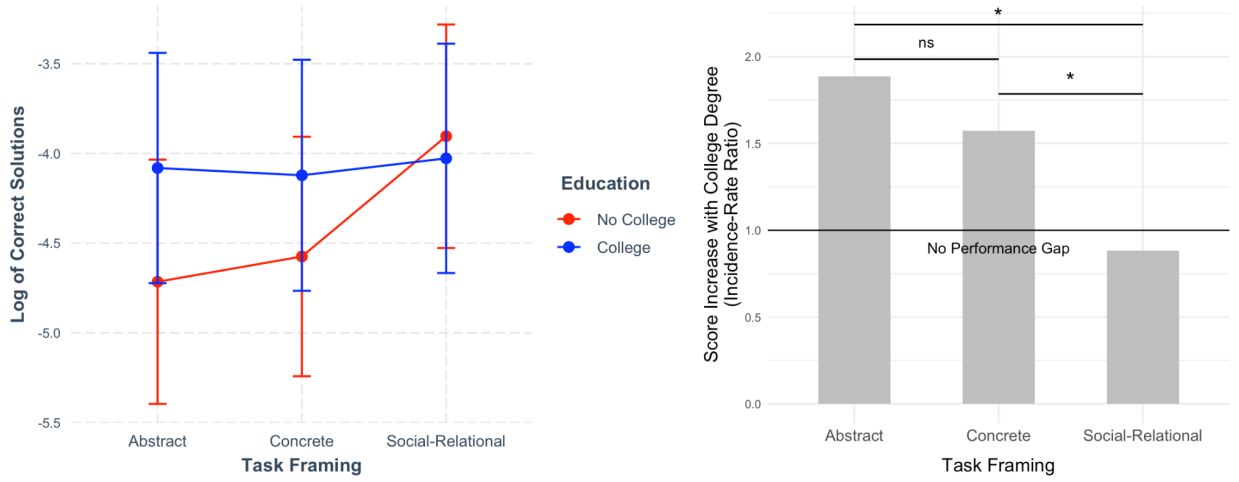

Figure 3. Educational attainment had a smaller effect on Wason performance when questions were about social relationships than when questions concerned abstract or concrete concepts (Study 4). Left panel: Scores in each condition are expressed in terms of Poisson's logarithmic scale. Lines show model-predicted log counts of correct solutions, and error bars represent 95% confidence intervals around the predicted values. Right panel: The performance difference attributable to educational attainment is expressed in terms of incidence rate ratios (IRRs); for reference, the black line represents the absence of an education-based performance gap (IRR = 1).

Table 3

Educational attainment and card type predicting performance on the Wason task in Study 2 (N = 781)

	Restricted Model	Full Model
	B (SE B)	B (SE B)
Concrete Cards (C)	0.14 (0.26)	-0.45 (0.33)
Social-Relational Cards (SR)	0.81 (0.23)***	0.41 (0.29)
Education	0.63 (0.33)	0.43 (0.34)
C × Education	-0.18 (0.32)	0.001 (0.34)
SR × Education	-0.76 (0.29) **	-0.55 (0.31)
Age		-0.26 (0.17)
Race		-0.67 (0.38)
Gender		-1.23 (0.36) ***
Age × C		-0.17 (0.17)
Age × SR		0.21 (0.15)
Race × C		0.54 (0.37)
Race × SR		0.62 (0.34)
Gender × C		0.88 (0.35) *
Gender × SR		0.65 (0.34)

Note. Table reports results of mixed-effects Poisson regressions. The number of correct solutions was regressed on educational attainment, a contrast comparing the real-world condition to the abstract condition, a contrast comparing the mentalizing (ToM) condition to the abstract condition, and the interactions between each of these contrasts and education. Full models adjust for participant age (z-scored), participant race (1 = White, 0 = nonwhite), and participant gender (1 = female, 0 = male). A random intercept for participant was specified to account for the nested structure of our data (with 3 observations per participant). Sampling weights were applied to both models to align the sample with U.S. population benchmarks. Significance codes: $p<.001$ '***', $p<.01$ '**', $p<.05$ '*'

Integrative Data Analysis

Prior to conducting our pre-registered TESS study, we conducted three studies to inform our TESS application and final study design. Some of these studies differed in design elements from our TESS study (e.g., some used different card stimuli categories) and provided mixed support for our core hypothesis that social-relational content reduces class-performance gaps on tests of logical aptitude. Specifically, only Study 2 (reported here) and Supplemental Study 1 provided statistically significant results that confirmed our primary hypothesis. We report these studies in their entirety in the supplemental materials (see Supplemental Studies 1-3).

These mixed patterns of results raise two related concerns. First, emphasizing only the preregistered, confirmatory study could create the impression of a file-drawer problem, insofar as conclusions would appear to rest on selectively reported evidence. Second, interpreting a set of individual studies in isolation increases the risk of inflated Type I error rates, given that multiple statistical tests were conducted across separate samples.

To address these issues, we employed integrative data analysis (IDA; Curran & Hussong, 2009). IDA is akin to a meta-analysis in that it integrates results across studies and samples, but is the preferred method when all original data are available (Curran & Hussong, 2009; Goh et al., 2016). In IDA, hypotheses are tested using pooled datasets, thereby avoiding the accumulation of independent significance tests which mitigates inflated Type I error rates, increasing statistical power, and providing an effect-size estimate based on the full available sample. Between-study heterogeneity can be modeled as fixed effects in instances where few individual datasets are available—as is our case with four total studies—to provide evidence that an effect in question generalizes across the combined set of samples (Curran & Hussong, 2009).

Our pooled dataset included all observations of abstract and social-relational card problems deployed across each of the four studies (4,358 complete observations across 2,269 participants). Sampling weights were again applied to align the sample with U.S. population

benchmarks—all convenience samples recruited via Prolific Academic received a weight of 1, given that those samples were not intended to be representative of a population. Study effects were adjusted with weighted-effects coding to fairly combine studies of different sizes, and our TESS study (Study 2) served as the referent category. We then reran our analyses from Study 2 in a restricted model, and allowed the study terms to interact with our hypothesized Education × Social-Relational Cards interaction in full models.

As can be seen in Table 3, social-relational content reduces the class-performance gap found in the typical, abstract version of the Wason task, although this effect was smaller in Supplemental Study 2. These results were unchanged with the addition of age, race, and gender as covariates.

As in Study 2, we visualize the pooled interaction effect both in Poisson's logarithmic scale and in terms of IRRs (see Figure 4). Across all datasets, participants with a four-year degree or higher answered an average of 2.24 times as many abstract questions correctly as did participants without a four-year degree. This class-performance gap is significantly reduced to a factor of 1.45 for social-relational questions—representing a 64% reduction of the class gap found in the typical, abstract version of the Wason task (p = .021).

Because it pools all relevant data collected by the present investigators, any conclusions made on the basis of the IDA are undistorted by potential file-drawer effects or inflated Type I error rates. Therefore, the IDA's significant SR × Education interaction implies that social-relational content reliably reduces class-based performance gaps in deductive reasoning.

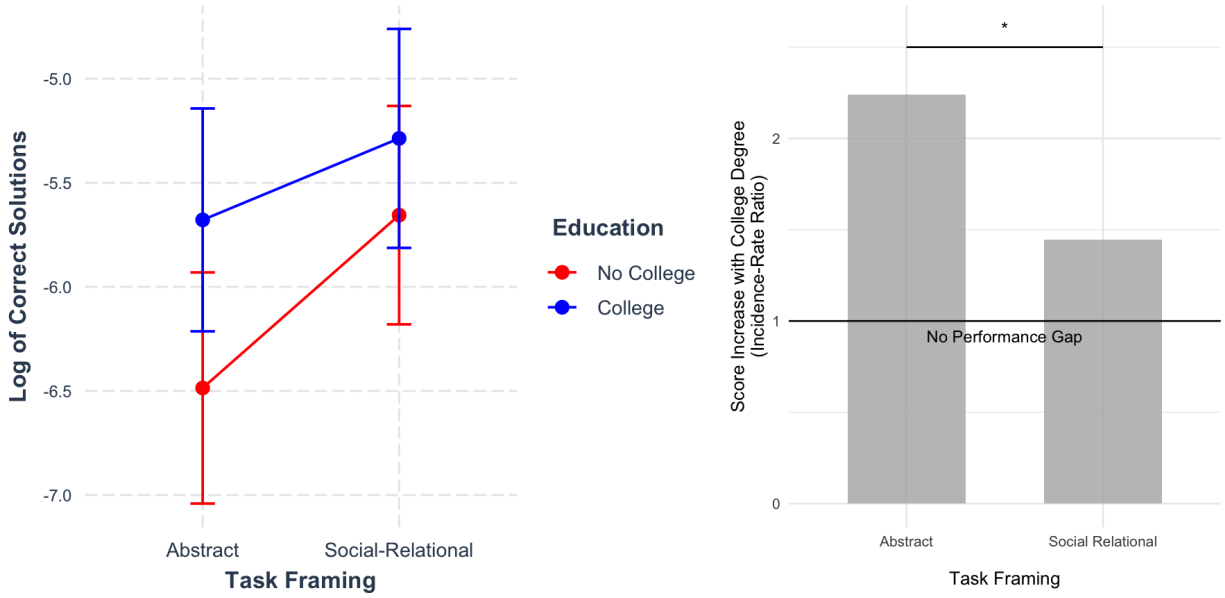

Figure 4. Educational attainment had a smaller effect on Wason performance when questions were about social relationships than when questions concerned abstract or concrete concepts (IDA). Left panel: Scores in each condition are expressed in terms of Poisson's logarithmic scale. Lines show model-predicted log counts of correct solutions, and error bars represent 95% confidence intervals around the predicted values. Right panel: The performance difference attributable to educational attainment is expressed in terms of incidence rate ratios (IRRs); for reference, the black line represents the absence of an education-based performance gap (IRR = 1).

Table 4

Educational attainment and card type predicting performance on the Wason task in the IDA (N = 2,269)

	Restricted Model	Full Model
	B (SE B)	B (SE B)
Social-Relational Cards (SR)	0.31 (0.05) ***	0.31 (0.05) ***
Education	0.29 (0.09) **	0.27 (0.09) **
SR × Education	-0.11 (0.05) *	-0.11 (0.05) *
Supplemental Study 1		0.62 (0.24) *
Supplemental Study 2		0.52 (0.24) *
Supplemental Study 3		-0.18 (0.24)
SS1 × Education		0.29 (0.24)
SS2 × Education		0.13 (0.24)
SS3 × Education		0.29 (0.24)
SS1 × SR		0.18 (0.13)
SS2 × SR		0.12 (0.13)
SS3 × SR		0.14 (0.13)
SS1 × SR × Education		0.06 (0.13)
SS2 × SR × Education		0.25 (0.13) *
SS3 × SR × Education		0.08 (0.13)

Note. Table reports results of mixed-effects Poisson regressions. The number of correct solutions was regressed on educational attainment, a contrast comparing the social-relational (SR) condition to the abstract condition, and the interaction between this contrast and education. Full models adjust for sample heterogeneity by including weighted-effects terms for study sample (Study 2 served as the referent category). A random intercept for participant was specified to account for the nested structure of our data. Sampling weights were applied to both models to align the sample with U.S. population benchmarks. Significance codes: $p<.001$ '***', $p<.01$ '**', $p<.05$ '*'

General Discussion

Prior work often attributes persistent class disparities in achievement outcomes to impaired reasoning and problem-solving skills among those from poor and working-class contexts (Bradley & Corwyn, 2002; DeGarmo et al., 1999; Gottfredson, 2004; Hall et al., 2014). The present research challenges these deficit narratives regarding class-based performance gaps by demonstrating that the content of a reasoning task can powerfully shape performance and skew conclusions about the reasoning abilities of individuals from working-class contexts.

Consistent with sociocultural models of social class (Fiske & Markus, 2012; Stephens et al., 2024), along with scholarship suggesting that typical aptitude tests may obscure the competencies fostered within working-class contexts (Dittmann et al., 2020; Ellis et al., 2022; Fryberg et al., 2013; Fryberg & Markus, 2007), we find that mathematical reasoning questions with greater mentions of other people and their relationships attenuates class-based testing gaps in data from 46 countries (Study 1). A nationally-representative experiment and an integrative data analysis (IDA) corroborate these findings. While adults from middle-class contexts outperformed their working-class counterparts on a deductive reasoning task devoid of social content (Study 2 and IDA), this class gap in reasoning performance was significantly reduced when the same task was experimentally reframed to center social relationships.

While class gaps were significantly reduced for reasoning questions, we did not observe this effect when questions tested knowledge of mathematical facts—and observed the opposite effect in questions requiring students to apply mathematical concepts (Study 1). We interpret these findings as evidence that social-relational content may render *reasoning* more accessible to test-takers from working-class contexts while serving as a distractor for questions testing core concepts and their application. Past work corroborates this possibility. In one study, salient real-world content (e.g., money) led to a better conceptual understanding of mathematical concepts but more arithmetic errors when those concepts were applied (McNeil et al., 2009). These findings suggest an important boundary condition on the present results and underscore

that structural interventions are not “silver bullet” solutions and must be implemented with attention to the specific cognitive processes that test questions invoke.

We derived our predictions from sociocultural models of social class, which emphasize that mismatches between the skills and competencies fostered within working-class contexts and the types of knowledge valued in academic spaces can contribute to class-based performance differences (Goudeau et al., 2024). When complex tasks are restructured to better match working-class cultural norms (e.g., by allowing students to work collaboratively rather than independently), class-based performance differences can be attenuated (Dittmann et al., 2020; Dietze & Dittmann, 2025). Extending this work, we show that—even when working independently—testing content framed to recruit social-cognitive competencies substantially reduces class gaps in performance. These findings echo prior research demonstrating that working-class contexts emphasize social interdependence (Carey & Markus, 2017) and promote social-cognitive strengths (e.g., emotion perception; Dietze & Knowles, 2021; Monroy et al., 2022). When tests are designed to recruit these competencies, individuals from working-class contexts display reasoning abilities on par with that of their more advantaged peers.

Although the introduction of social-relational content sizeably reduced class discrepancies in reasoning performance by up to 64% in the IDA, we did not observe a working-class *advantage* on problems framed in social terms. This supports the idea that cultural mismatch is less likely to influence those from middle-class contexts than their working-class counterparts (Stephens, Fryberg, et al., 2012; Stephens, Townsend, et al., 2012). This may reflect the fact that individuals from middle-class contexts (i.e., those with at least a four-year college degree) are more practiced with formal assessment instruments relative to their working-class counterparts, often translating into a large main effect of class (i.e., education) on performance across content framings. Indeed, even on the Wason selection task, researchers find that performance on the task increases with the time spent in university settings as well as mathematical training (Kellen & Klauer, 2020; Lehman & Nisbett, 1990). That

we were able to entirely eliminate this main effect in a nationally-representative experiment and reduce it by over half in an IDA speaks to the power of cultural match to counter large class-based practice effects in standardized testing.

Overall, our results closely align with research on cultural mismatch suggesting that underperformance on aptitude tests among members of disadvantaged groups often reflects the fact that testing instruments were designed by and for people from hegemonic (e.g., middle-class) cultural contexts. Central to this argument is that aptitude tests are rarely culture-neutral, and that typical test content may exclude the strengths and competencies of those from marginalized communities. Other approaches echo the view that content matters. Within evolutionary psychology, for instance, the “hidden talents” perspective argues that replacing abstract test content (e.g., shapes) with ecologically-relevant content (e.g., faces, buses, and money) can boost performance among children exposed to severe socioeconomic deprivation (Ellis et al., 2022; Young et al., 2022).

In the present work, we provide additional granularity as to which specific categories of real-world content yield the biggest performance improvements amongst economically disadvantaged individuals. For instance, we find that mentions of ecologically-relevant (but nonsocial) content such as money, buses, and food do not significantly reduce class-based performance gap relative to abstract content (Study 2). On the other hand, content invoking social relationships (i.e., inferring who is friends with whom) consistently attenuate class-based performance gaps. We see these results as complementing work in the hidden talents perspective by spotlighting the types of content which members of working-class communities tend to benefit from most.

These theoretical distinctions, as well as several methodological differences, can also explain why the results of our TIMSS analyses differ from a recent analysis inspired by the hidden talents perspective (Muskens et al., 2024). In that investigation, the authors found that lower-SES students (as indexed by number of books within children’s homes) performed worse

on items containing real-world content (i.e., money, food, and social relationships) than their higher-SES peers. We see several important differences between our investigation and the hidden talents work. First, although educational attainment and the number of books within one's home are modestly correlated (Heppt et al., 2022), these measures are not interchangeable.[3] We contend that parental educational attainment is the preferred proxy for children's social-class positioning, because it better captures their broader experiences of the norms and behaviors promoted within a given class-cultural context (Goudeau et al., 2024; Stephens et al., 2024).[4] Second, the hidden talents investigation did not subset analyses by TIMSS cognitive domain (i.e., knowing, applying, and reasoning) and thus does not speak to whether social-relational content closes the class gap in reasoning performance specifically. Indeed, the present work suggests that the effects of social-relational content are heterogeneous across cognitive domains.

In conclusion, we highlight that our results support strength-based, rather than deficit-based, analyses of class differences in educational performance. We show that when test content is carefully chosen to tap the sociocognitive strengths promoted within working-class contexts, it is possible to dramatically reduce class gaps in reasoning performance. This raises a broader possibility: that class differences across a range of reasoning assessments may stem more from a cultural mismatch with test content than from working-class deficits in reasoning ability. Of course, not all disparities are traceable to culturally-bound test content. Indeed, severe deprivation has well-documented physiological consequences that shape cognitive development and educational outcomes. Nevertheless, we caution that equating *underperformance* with *inability* subtly reinforces the idea that aptitude tests are neutral measures of achievement.

---

[3] Heppt and colleagues (2022) found that books at home and parental education were correlated at r = 0.30. In our analysis sample, we found that students from working-class backgrounds had fewer books at home (M_{books} = 2.39, SD_{books} = 1.14) than those from middle-class backgrounds (M_{books} = 3.28, SD_{books} = 1.24), $t(142{,}822) = -166.97$, $p < .001$, equivalent to r = 0.40.

[4] When we use books in the home to proxy social class, we see results consistent (though weaker) with those for parental education (see supplemental materials, Table S3).

Instead, making targeted structural changes to test content can enable the underlying strengths of those from marginalized communities to shine through (Ellis et al., 2022).

References

Barnes, J. (1996). *The Cambridge companion to Aristotle*. Cambridge university press.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*(1), 371–399. https://doi.org/10.1146/annurev.psych.53.100901.135233

Braine, M. D. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, *85*(1), 1–21. https://doi.org/10.1037/0033-295X.85.1.1

Carey, R. M., & Markus, H. R. (2017). Social class shapes the form and function of relationships and selves. *Current Opinion in Psychology*, *18*, 123–130. https://doi.org/10.1016/j.copsyc.2017.08.031

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*(3), 404–431. https://doi.org/10.1037/0033-295X.97.3.404

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*(4), 391–416. https://doi.org/10.1016/0010-0285(85)90014-3

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*(3), 187–276. https://doi.org/10.1016/0010-0277(89)90023-1

Cox, J. R., & Griggs, R. A. (1982). The effects of experience on performance in Wason's selection task. *Memory & Cognition*, *10*(5), 496–502. https://doi.org/10.3758/BF03197653

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*(2), 81–100. https://doi.org/10.1037/a0015914

DeGarmo, D. S., Forgatch, M. S., & Martinez, C. R., Jr. (1999). Parenting of divorced mothers as a link between social Ssatus and boys' academic outcomes: Unpacking the effects of

socioeconomic status. *Child Development*, *70*(5), 1231–1245. https://doi.org/10.1111/1467-8624.00089

Dietze, P., & Dittmann, A. G. (2025). When and why working together benefits first-generation college students: A registered report. *Journal of Experimental Psychology: General, 154*(10), 2860–2873. https://doi.org/10.1037/xge0001814

Dietze, P., & Knowles, E. D. (2016). Social class and the motivational relevance of other human beings. *Psychological Science*, *27*(11), 1517–1527. https://doi.org/10.1177/0956797616667721

Dietze, P., & Knowles, E. D. (2021). Social class predicts emotion perception and perspective-taking performance in adults. *Personality and Social Psychology Bulletin*, *47*(1), 42–56. https://doi.org/10.1177/0146167220914116

Dietze, P., Olderbak, S., Hildebrandt, A., Kaltwasser, L., & Knowles, E. D. (2024). A lower-class advantage in face memory. *Personality and Social Psychology Bulletin*, *50*(2), 285–298. https://doi.org/10.1177/01461672221125599

Dittmann, A. G., Stephens, N. M., & Townsend, S. S. M. (2020). Achievement is not class-neutral: Working together benefits people from working-class contexts. *Journal of Personality and Social Psychology*, *119*(3), 517–539. https://doi.org/10.1037/pspa0000194

Duquennois, C. (2022). Fictional Money, Real Costs: Impacts of Financial Salience on Disadvantaged Students. *American Economic Review*, *112*(3), 798–826. https://doi.org/10.1257/aer.20201661

Ellis, B. J., Abrams, L. S., Masten, A. S., Sternberg, R. J., Tottenham, N., & Frankenhuis, W. E. (2022). Hidden talents in harsh environments. *Development and Psychopathology*, *34*(1), 95–113. https://doi.org/10.1017/S0954579420000887

Emery, L. F., & Finkel, E. J. (2022). Connect or protect? Social class and self-protection in romantic relationships. *Journal of Personality and Social Psychology*, *122*(4), 683–699.

https://doi.org/10.1037/pspi0000368

Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin*, *128*(6), 978–996. https://doi.org/10.1037/0033-2909.128.6.978

Fendinger, N. J., Dietze, P., & Knowles, E. D. (2023). Beyond cognitive deficits: How social class shapes social cognition. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2023.03.004

Fiske, S. T., & Markus, H. R. (2012). *Facing social class: How societal rank influences interaction* (S. T. Fiske & H. R. Markus, Eds.). Russell Sage Foundation.

Frankenhuis, W. E., De Vries, S. A., Bianchi, J., & Ellis, B. J. (2020). Hidden talents in harsh conditions? A preregistered study of memory and reasoning about social dominance. *Developmental Science*, *23*(4), e12835. https://doi.org/10.1111/desc.12835

Fryberg, S. A., Covarrubias, R., & Burack, J. A. (2013). Cultural models of education and academic performance for Native American and European American students. *School Psychology International*, *34*(4), 439–452. https://doi.org/10.1177/0143034312446892

Fryberg, S. A., & Markus, H. R. (2007). Cultural models of education in American Indian, Asian American and European American contexts. *Social Psychology of Education*, *10*(2), 213–246. https://doi.org/10.1007/s11218-007-9017-z

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and Imagined Activity Can Enhance Young Children's Reading Comprehension. *Journal of Educational Psychology*, *96*(3), 424–436. https://doi.org/10.1037/0022-0663.96.3.424

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini Meta‐Analysis of Your Own Studies: Some Arguments on Why and a Primer on How. *Social and Personality Psychology Compass*, *10*(10), 535–549. https://doi.org/10.1111/spc3.12267

Gottfredson, L. S. (2004). Intelligence: Is It the epidemiologists' elusive "fundamental cause" of social class inequalities in health? *Journal of Personality and Social Psychology*, *86*(1),

174–199. https://doi.org/10.1037/0022-3514.86.1.174

Goudeau, S., Stephens, N. M., Markus, H. R., Darnon, C., Croizet, J.-C., & Cimpian, A. (2024). What causes social class disparities in education? The role of the mismatches between academic contexts and working-class socialization contexts and how the effects of these mismatches are explained. *Psychological Review*. https://doi.org/10.1037/rev0000473

Greer, B. (1997). Modelling reality in mathematics classrooms: The case of word problems. *Learning and Instruction*, *7*(4), 293–307. https://doi.org/10.1016/S0959-4752(97)00006-6

Grigorenko, E. L., Geissler, P. W., Prince, R., Okatcha, F., Nokes, C., Kenny, D. A., Bundy, D. A., & Sternberg, R. J. (2001). The organisation of Luo conceptions of intelligence: A study of implicit theories in a Kenyan village. *International Journal of Behavioral Development*, *25*(4), 367–378. https://doi.org/10.1080/01650250042000348

Hall, C. C., Zhao, J., & Shafir, E. (2014). Self-affirmation among the poor: Cognitive and behavioral implications. *Psychological Science*, *25*(2), 619–625. https://doi.org/10.1177/0956797613510949

Halle, T., Forry, N., Hair, E., Perper, K., Wandner, L., Wessel, J., & Vick Whittaker, J. (2009). Disparities in early learning and development: Lessons from the Early Childhood Longitudinal Study – Birth Cohort (ECLS-B). *Child Trends*.

Heppt, B., Olczyk, M., & Volodina, A. (2022). Number of books at home as an indicator of socioeconomic status: Examining its extensions and their incremental validity for academic achievement. *Social Psychology of Education*, *25*(4), 903–928. https://doi.org/10.1007/s11218-022-09704-8

Johnson, M. (2015). Embodied understanding. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00875

Kellen, D., & Klauer, K. C. (2020). Theories of the Wason Selection Task: A critical assessment of boundaries and benchmarks. *Computational Brain & Behavior*, *3*(3), 341–353. https://doi.org/10.1007/s42113-019-00034-1

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, *17*(2), 248–294. https://doi.org/10.1016/0010-0285(85)90009-X

Kraus, M. W., Côté, S., & Keltner, D. (2010). Social class, contextualism, and empathic accuracy. *Psychological Science*, *21*(11), 1716–1723. https://doi.org/10.1177/0956797610387613

Kusserow, A. S. (1999). De-Homogenizing American Individualism: Socializing Hard and Soft Individualism in Manhattan and Queens. *Ethos*, *27*(2), 210–234. https://doi.org/10.1525/eth.1999.27.2.210

Lakoff, G. (2012). Explaining Embodied Cognition Results. *Topics in Cognitive Science*, *4*(4), 773–785. https://doi.org/10.1111/j.1756-8765.2012.01222.x

Lehman, D. R., & Nisbett, R. E. (1990). A longitudinal study of the effects of undergraduate training on reasoning. *Developmental Psychology*, *26*(6), 952–960. https://doi.org/10.1037/0012-1649.26.6.952

Lloyd, G. E. R. (1990). *Demystifying mentalities*. Cambridge University press.

Martin, M. O., & Mullis, I. V. S. (Eds.). (2013). *Methods and procedures in TIMSS and PIRLS 2011*. TIMSS & PIRLS International Study Center.

McNeil, N. M., Uttal, D. H., Jarvin, L., & Sternberg, R. J. (2009). Should you show me the money? Concrete objects both hurt and help performance on mathematics problems. *Learning and Instruction*, *19*(2), 171–184. https://doi.org/10.1016/j.learninstruc.2008.03.005

Mittal, C., Griskevicius, V., Simpson, J. A., Sung, S., & Young, E. S. (2015). Cognitive adaptations to stressful environments: When childhood adversity enhances adult executive function. *Journal of Personality and Social Psychology*, *109*(4), 604–621. https://doi.org/10.1037/pspi0000028

Monroy, M., Cowen, A. S., & Keltner, D. (2022). Intersectionality in emotion signaling and

recognition: The influence of gender, ethnicity, and social class. *Emotion*, *22*(8), 1980–1988. https://doi.org/10.1037/emo0001082

Muskens, M., Frankenhuis, W. E., & Borghans, L. (2024). Math items about real-world content lower test-scores of students from families with low socioeconomic status. *Npj Science of Learning*, *9*(1), 19. https://doi.org/10.1038/s41539-024-00228-8

Nickles, T. (2018). Bounded rationality, scissors, crowbars, and pragmatism: Reflections on Herbert Simon. *Mind & Society*, *17*(1–2), 85–96. https://doi.org/10.1007/s11299-019-00206-3

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108*(2), 291–310. https://doi.org/10.1037/0033-295X.108.2.291

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Ireland, M. E. (2022). *Linguistic Inquiry and Word Count* (Version 2022) [Computer software]. Pennebaker Conglomerates.

Phillips, L. T., Stephens, N. M., Townsend, S. S. M., & Goudeau, S. (2020). Access is not enough: Cultural mismatch persists to limit first-generation students' opportunities for achievement throughout college. *Journal of Personality and Social Psychology*, *119*(5), 1112–1131. https://doi.org/10.1037/pspi0000234

Piff, P. K., Stancato, D. M., Martinez, A. G., Kraus, M. W., & Keltner, D. (2012). Class, chaos, and the construction of community. *Journal of Personality and Social Psychology*, *103*(6), 949–962. https://doi.org/10.1037/a0029673

Plaut, V. C., & Markus, H. R. (2005). The "inside" story: A cultural historical analysis of how to be smart and motivated, American style. In C. S. Dweck & A. Elliot (Eds.), *Handbook of competence and motivation* (pp. 457–487). Guilford Press.

R Core Team. (2024). *R: A Language and Environment for Statistical Computing.* (Version 4.4.0) [Computer software]. https://www.R-project.org

Reeve, C. D. C. (2004). *Plato: Republic* (3rd ed). Hackett Publishing Company, Inc.

Rogoff, B., Coppens, A. D., Alcalá, L., Aceves-Azuara, I., Ruvalcaba, O., López, A., & Dayton, A. (2017). Noticing learners' strengths through cultural research. *Perspectives on Psychological Science*, *12*(5), 876–888. https://doi.org/10.1177/1745691617718355

Schliemann, A. D., & Carraher, D. W. (2002). The Evolution of Mathematical Reasoning: Everyday versus Idealized Understandings. *Developmental Review*, *22*(2), 242–266. https://doi.org/10.1006/drev.2002.0547

Sedley, D. (2016). An Introduction to Plato's Theory of Forms. *Royal Institute of Philosophy Supplement*, *78*, 3–22. https://doi.org/10.1017/S1358246116000333

Simon, H. A. (1965). THE LOGIC OF RATIONAL DECISION. *The British Journal for the Philosophy of Science*, *16*(63), 169–186. https://doi.org/10.1093/bjps/XVI.63.169

Stephens, N. M., Dittmann, A. G., & Townsend, S. S. M. (2017). Social class and models of competence: How gateway institutions disadvantage working-class Americans and how to intervene. In *Handbook of competence and motivation: Theory and application, 2nd ed* (pp. 512–528). The Guilford Press.

Stephens, N. M., Emery, L. F., & Townsend, S. S. M. (2024). Social class. In D. T. Gilbert, S. T. Fiske, E. J. Finkel, & B. M. Wendy (Eds.), *The handbook of social psychology* (6th ed.). Situational Press.

Stephens, N. M., Fryberg, S. A., Markus, H. R., Johnson, C. S., & Covarrubias, R. (2012). Unseen disadvantage: How American universities' focus on independence undermines the academic performance of first-generation college students. *Journal of Personality and Social Psychology*, *102*(6), 1178–1197. https://doi.org/10.1037/a0027143

Stephens, N. M., Hamedani, M. G., Markus, H. R., Bergsieker, H. B., & Eloul, L. (2009). Why Did They "Choose" to Stay?: Perspectives of Hurricane Katrina Observers and Survivors. *Psychological Science*, *20*(7), 878–886. https://doi.org/10.1111/j.1467-9280.2009.02386.x

Stephens, N. M., Townsend, S. S. M., & Dittmann, A. G. (2019). Social-class disparities in

higher education and professional workplaces: The role of cultural mismatch. *Current Directions in Psychological Science*, *28*(1), 67–73. https://doi.org/10.1177/0963721418806506

Stephens, N. M., Townsend, S. S. M., Markus, H. R., & Phillips, L. T. (2012). A cultural mismatch: Independent cultural norms produce greater increases in cortisol and more negative emotions among first-generation college students. *Journal of Experimental Social Psychology*, *48*(6), 1389–1393. https://doi.org/10.1016/j.jesp.2012.07.008

Sternberg, R. J., Nokes, C., Geissler, P. W., Prince, R., Okatcha, F., Bundy, D. A., & Grigorenko, E. L. (2001). The relationship between academic and practical intelligence: A case study in Kenya. *Intelligence*, *29*(5), 401–418. https://doi.org/10.1016/S0160-2896(01)00065-4

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Swets & Zeitlinger Publishers.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*(3), 273–281. https://doi.org/10.1080/14640746808400161

Wickham, H. (with Sievert, C.). (2016). *ggplot2: Elegant graphics for data analysis* (Second edition). Springer.

Young, E. S., Frankenhuis, W. E., DelPriore, D. J., & Ellis, B. J. (2022). Hidden talents in context: Cognitive performance with abstract versus ecological stimuli among adversity-exposed youth. *Child Development*, *93*(5), 1493–1510. https://doi.org/10.1111/cdev.13766

Yzerbyt, V. Y., Muller, D., & Judd, C. M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of Experimental Social Psychology*, *40*(3), 424–431. https://doi.org/10.1016/j.jesp.2003.10.001