Intelligence and academic performance: Is it all in your head?

Katherine L. Bottenhorn[1,2], Jessica E. Bartley[1,3], Michael C. Riedel[1,3], Taylor Salo[1,2], Elsa I. Bravo[2], Rosalie Odean[4], Alina Nazareth[5], Robert W. Laird[1,3], Erica D. Musser[1,2], Shannon M. Pruden[2], Eric Brewe[6,7,8], Matthew T. Sutherland[1,2], Angela R. Laird[1,3]*

[1]Center for Imaging Science, Florida International University, Miami, FL, USA
[2]Department of Psychology, Florida International University, Miami, FL, USA
[3]Department of Physics, Florida International University, Miami, FL, USA
[4]Department of Human Development and Family Sciences, Ohio State University, Columbus, OH, USA
[5]Department of Psychology, Temple University, Philadelphia, PA, USA
[6]Department of Physics, Drexel University, Philadelphia, PA, USA
[7]Department of Education, Drexel University, Philadelphia, PA, USA
[8]Department of Teaching and Learning, Florida International University, Miami, FL, USA

Abstract

Academic performance relies, in part, on intelligence; however, intelligence quotient (IQ) is limited in predicting academic success. Furthermore, while the search for the biological seat of intelligence predates neuroscience itself, its findings remain conflicting. Here, we assess the interplay between IQ, academic performance, and brain connectivity with behavioral and functional MRI data collected from undergraduate students as they completed an active learning or lecture-based semester-long university physics course. IQ (i.e., full-scale WAIS scores) increased significantly pre- to post-instruction, were associated with physics knowledge and reasoning measures, but were unrelated to overall course grade. IQ was related to brain connectivity during physics-related cognition, but connectivity did not mediate IQ's association with task performance. These relations depended on students' sex and instructional environment, providing evidence that physics classroom environment and pedagogy may have a gendered influence on students' performance. Discussion focuses on opportunities to improve physics reasoning skills for all students.

Keywords: cognitive neuroscience; STEM education; intelligence; functional neuroimaging

Corresponding Author: Dr. Angela R. Laird, Professor of Physics, Florida International University, Miami, FL, USA | Email: alaird@fiu.edu

Introduction

Intelligence testing has been the subject of substantial scientific inquiry over the past century, especially with respect to academic performance and overall success.[1–4] These lines of inquiry have a pernicious history, as intelligence testing and IQ have been used to deny educational and employment opportunities and as a foundation of eugenics in the United States and abroad in the 20th century, [5,6] though updates to these tests have mitigated bias with respect to racial, ethnic, and gender differences.[7,8] Both in research and in popular discourse, intelligence is often treated as an inherent, stable, trait-like quality, equated with an intelligence quotient (IQ).[9–13] While intellectual abilities are moderately heritable,[14] IQ and other psychometric measures of intellectual ability are also influenced by a number of experiential factors, and the relative influences of genes and environment on IQ change across the lifespan.[15,16] In general, intelligence is psychometrically assessed via a range of verbal and nonverbal cognitive tests, capturing a general view of ability across domains. One such test is the Wechsler Adult Intelligence Scale (WAIS[17]), which demonstrates moderate stability across adulthood, with increasing stability for shorter intervals and with increasing age.[18–28] A history of research and popular discourse presupposes that IQ predicts one's predisposition to academic and life success,[29–33] though in reality, the picture is much more complicated due to a variety of sociocultural factors.[34–37] Decades of research suggest education and psychometric intelligence are entwined in a bidirectional relationship, as intellectual ability predicts access to and extent of education, through a variety of socioeconomic factors,[38,39] while years of education predict modest increases in intellectual abilities,[38,40] and some educational interventions likely to improve one's ability to acquire and apply knowledge and skills.[22,41–44]

University students who pursue science, technology, engineering, and mathematics (STEM) disciplines are exposed to a rigorous curriculum designed to transform their problem-solving skills,[45–48] which engages students' perceptual and verbal abilities.[49–51] The fourth edition of the WAIS (WAIS-IV) provides a full-scale measure of intellectual ability (FSIQ) and four component index scores: Processing Speed, Perceptual Reasoning, Working Memory, and Verbal Comprehension.[52] The WAIS is widely used as an extensively validated and researched clinical tool, but less often applied in educational research, though it may be particularly well-suited for exploring associations between education and skill development. Introductory physics presents a prime opportunity for such study, as a gateway course for STEM majors with a relatively standard curriculum across universities, including instruction on classical Newtonian mechanics and emphasizing the development of quantitative, visuospatial reasoning and problem-solving skills likely captured by WAIS-IV index scores. Unfortunately, female students often perform worse on specific conceptual evaluations in these courses, though not necessarily on overall course grades, due to a host of socioaffective and -cultural factors present in education and physics classrooms, specifically.[53–58] Female students also constitute a smaller proportion of the student body, compared to their male counterparts[53,56,57,59,60] and ultimately, such disparities can propagate across courses, leading to higher rates of STEM degrees among male students as compared to female students (64% male vs. 36% female in 2015-2016).[61] Recently, institutions have sought to improve STEM student success using active learning instructional approaches that yield improved student performance outcomes[62,63] and impact socioemotional aspects of university education, including self-efficacy, science-related anxiety, and identity.[54,59] Together, these effects may mitigate existing sex differences in performance,[55,58] though this is contradicted by some findings.[64] Altogether, there is a need to better understand sex differences in physics education and potential avenues for mitigating these differences, to ensure all students have the opportunity to succeed.

For as long as we have been trying to understand intelligence, we have been searching for its biological substrates. Recently, neuroscience research has studied the underlying neurobiology of intelligence using neuroimaging techniques such as functional magnetic resonance imaging (fMRI).[65–69] Task-based fMRI research has focused on understanding how differences in brain activation during cognition differs relates to intelligence, yielding two theories of the neurobiology of intelligence: the parieto-frontal integration theory (P-FIT[70]) and the neural efficiency hypothesis (NEH[71]). The PFIT suggests that interactions between frontal and parietal regions

underlie intelligence, while the NEH suggests that higher intelligence is reflected by more "efficient" brain activation during cognitively demanding tasks. An alternative view on "neural efficiency" comes from the application of network science to functional connectivity, to show that more intelligent individuals exhibit greater topological efficiency, which describes ease of information transfer across the brain,[66–68,72] rather than activation efficiency. Much of this work has focused on functional brain connectivity during the resting state, i.e., in the absence of a task or externally-directed cognition,[73–75] often referred to as "intrinsic" connectivity,[76,77] mirroring the notion of intelligence as an inherent trait. However, recent work shows that individual differences in intelligence are better predicted by task-evoked connectivity,[78] presenting an opportunity to merge these two lines of research to better understand individual differences in the neurobiology of intellectual abilities.
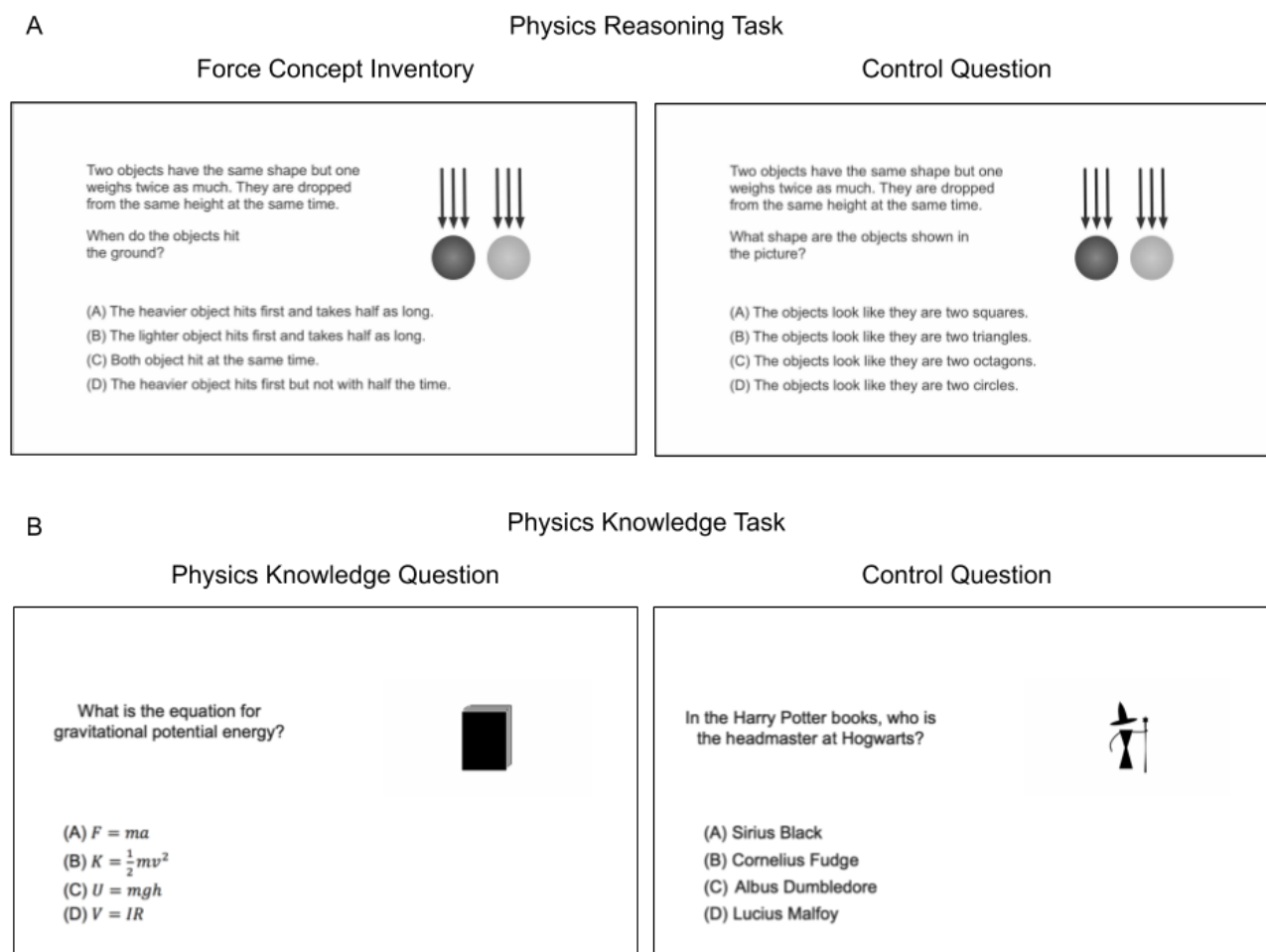

Figure 1. fMRI tasks performed by students at both pre- and post-instruction time points. During MRI scanning sessions, students completed tasks that probed students' physics-related cognition. The *physics reasoning task* (A) included questions from the Force Concept Inventory (FCI) and engaged students' conceptual understanding of Newtonian mechanics (left), with perceptually similar control questions (right). The *physics knowledge task* (B) included questions about definitions and equations that students learn in class (left), with perceptually matched general knowledge control questions (right).

Here, we build on the knowledge that the WAIS measures cognitive abilities that are subject to influence by educational interventions, and leverage the WAIS to study individual differences across student performance in an introductory physics course, and potential roles of sex and pedagogy therein. Then, we build on prior neuroimaging research suggesting task-evoked brain organization can explain individual differences in intellectual abilities, to search for a biological substrate for associations between ability and student performance. To do so, we collected data from undergraduate students enrolled in either a lecture-based or active learning

section of an introductory physics course. Both pre- and post-instruction, students completed the WAIS-IV alongside a robust fMRI protocol, including two physics-related tasks (Figure 1) with different demands on cognition. The first task engaged students' reasoning skills and conceptions about forces at work in the natural world (Figure 1A), based on the Force Concept Inventory (FCI[79]; see Bartley et al.[80] for detailed FCI task results). In this physics reasoning task, students viewed questions about forces on and the movement of objects, along with answer choices that included the correct (e.g., Newtonian) explanation, and choices that reflect common but incorrect (e.g., non-Newtonian) conceptions about forces and motion. The second task required students to recall concepts and equations taught in an introductory physics course (Figure 1B). In this physics knowledge task, students engage semantic memory to recognize equations or definitions of physics concepts learned in the course from a list of possible answer choices presented. Here, we used these data to, first, assess changes in WAIS-IV scores (both FSIQ and index scores) over the course of the semester, then applied a series of linear regressions to assess associations between post-instruction and pre- to post-instruction changes in WAIS-IV scores and post-instruction student performance (i.e., task accuracy and final course grade). Finally, we assessed associations between WAIS-IV scores, and changes therein, and post-instruction functional brain organization during the two tasks, and the degree to which these associations provide a common neural substrate supporting the role of cognitive abilities in student performance. We hypothesized that, while FSIQ itself is stable, different WAIS-IV index scores are differentially associated with performance on physics-related assessments with different cognitive demands. Further, we hypothesized these d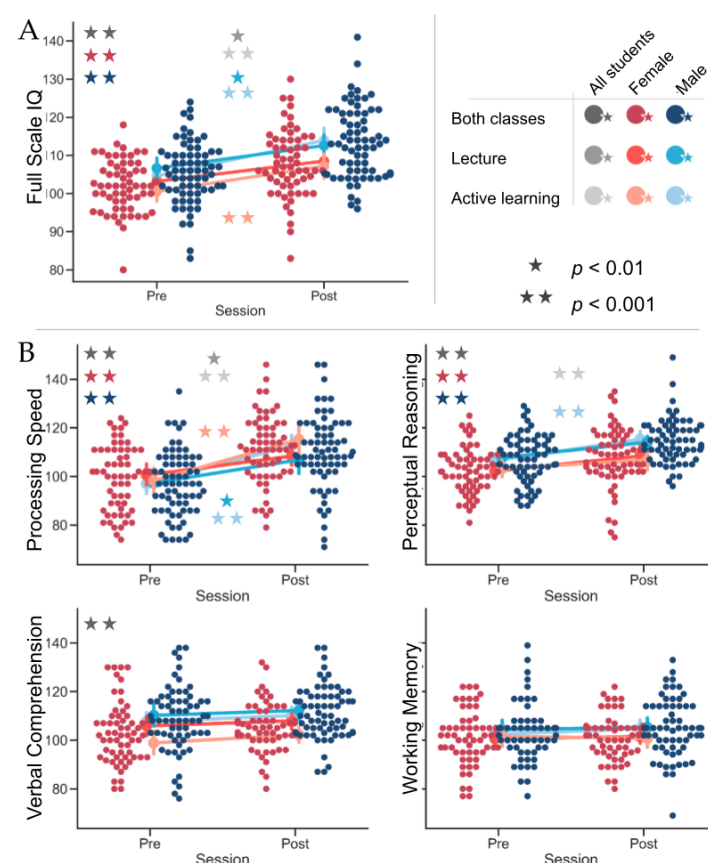ifferences would be reflected in brain organization during these domain-specific tasks (e.g., physics reasoning and physics knowledge), providing a neurobiological explanation for ability-performance relationships across physics-related cognition. Finally, including sex and pedagogy in these assessments may provide an insight into sex differences in physics education and the potential role of active learning in their mitigation, though we did not expect different brain-IQ relations across active learning and lecture classrooms.

Results

WAIS-IV scores increased over a semester of physics learning regardless of sex or classroom.

First, to understand characteristics of WAIS-IV scores (i.e., FSIQ and index scores) in this sample, we administered the WAIS-IV before and after a semester of physics instruction from 110 students who completed undergraduate introductory physics in lecture-based (25 female, 28 male) or active-learning (24 female, 33 male) classrooms. Pre- to post-instruction changes in WAIS-IV scores were assessed via Wilcoxon signed-rank tests, due to the presence of outliers (Supplementary Figure 1). Full-scale WAIS-IV scores increased an average of 7.03 points (W^+ = 822, $p < 0.001$; Table 1; Figure 2A). This change was driven by significant

Figure 2. Change in WAIS scores pre- to post-instruction. Full-scale WAIS scores (A) significant increased across students within the sample, for male students in both class types (i.e., active learning and lecture), and for female students in lecture classes. Of the WAIS subscores (B), processing speed (PSI; top left), perceptual reasoning (PRI; top right), and verbal comprehension (VCI; bottom left) showed significant increases across the sample. When separated by sex, only PSI and PRI significantly increased for both sexes, driven by increases in lecture classes. Working memory (bottom right) did not significantly increase for any group in this sample.

increases in three of the four index scores of the WAIS (Table 1; Figure 2B), greatest in Processing Speed (PSI) and Perceptual Reasoning (PRI). Importantly, however, there were no significant differences in the change in WAIS-IV scores with respect to sex and only PSI changes varied with respect to classroom (i.e., active learning, lecture), evidenced by a significant time by class interaction (Supplementary Table 1). The changes in WAIS-IV scores were commensurate with previously reported retest gains among college students across a similar time period.[81]

Table 1. Average change in WAIS scores pre- to post-instruction

WAIS Score	Pre-instruction	Post-instruction	Pre to Post Change	Estevis et al., 2012	Change - Estevis *t*-test (*p*)
Full score WAIS-IV Score	103.9 ± 7.6	110.8 ± 9.4	**7.0** ± 7.3	**6.7** ± 5.2	-0.12 (0.90)
Perceptual Reasoning Index	105.0 ± 9.6	111.5 ± 11.1	**6.0** ± 9.5	**3.6** ± 5.6	2.39 (0.02)
Processing Speed Index	98.3 ± 13.3	110.6 ± 14.5	**12.6** ± 17.5	**10.5** ± 9.5	0.71 (0.48)
Verbal Comprehension Index	106.0 ± 13.0	108.5 ± 10.6	**2.5** ± 9.0	**4.2** ± 7.6	-1.28 (0.20)
Working Memory Index	102.3 ± 10.8	103.2 ± 10.5	2.0 ± 9.1	3.2 ± 4.8	-2.36 (0.02)

Note. Bolded changes indicate significant Wilcoxon Signed Rank tests at $p_{FWE\text{-}corr} < 0.05$ ($\alpha_{adjusted}$ = 0.014), after controlling for familywise error using the Šidák correction.[81] The *Change - Estevis* t-*test (p)* column (far right) provides the results of a Student's *t*-test for independent samples, comparing the changes in WAIS scores seen in this sample with those reported by Estevis et al. (2012) to determine whether the changes in our sample are comparable with those reported elsewhere.

Physics task accuracy, but not course grade, is related to WAIS-IV scores differently for male and female students.

To assess associations between course performance and intellectual ability, we separately regressed each WAIS-IV score (both post-instruction and pre- to post-instruction changes) on post-instruction course grade and physics-related task accuracies (denoted "performance" below), controlling for students' sex, classroom environment, and other demographics, for a total of 30 separate regressions (controlling for familywise error rate with the Šidák correction).

Equation 1

$performance = \beta_1 IQ + \beta_2 IQ \times Sex + \beta_3 IQ \times Class + \beta_4 IQ \times Class \times Sex + \beta_5 Sex + \beta_6 Class + \beta_7 Sex \times Class + \beta_8 Age + \beta_9 Years\ in\ Univ.$
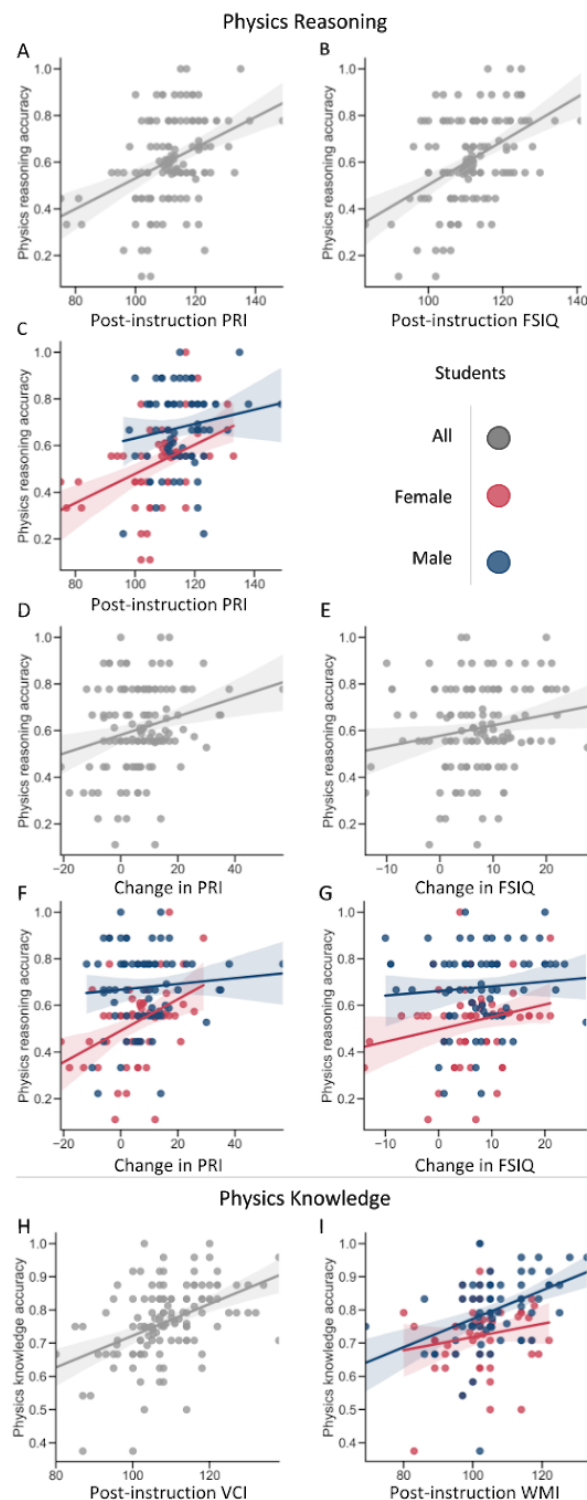
Figure 3. Physics task performance is related to WAIS scores. Post-instruction accuracy on the physics reasoning task was significantly related to post-instruction PRI scores (A) and differently for female and male students (C), and to post-instruction FSIQ scores (B), as well as changes in PRI scores overall (D) and differently for female and male students (F), and changes in FSIQ scores overall (E) and differently for female and male students (G). Accuracy on the physics knowledge task was associated with post-instruction VCI scores (H) and WMI scores differently for male and female students (I).

Here, task accuracy refers to the proportion of correct answers given while participants performed a physics task in the MRI scanner (Figure 1). In addition to the full models (Equation 1), we assessed nested models without the WAIS interaction terms (Supplementary Table 2).

Post-instruction accuracy on the physics reasoning task was significantly related to post-instruction PRI scores (post PRI; $F_{(9, 120)} = 5.122$, $p < 0.001$; Figure 3A, 3C) and FSIQ scores (post FSIQ; $F_{(9, 120)} = 5.770$, $p < 0.001$; Figure 3B), in addition to the pre- to post-instruction changes in both PRI (ΔPRI; $F_{(9, 120)} = 5.034$, $p < 0.001$; Figure 3D, 3F) and FSIQ scores (ΔFSIQ; $F_{(9, 120)} = 4.498$, $p < 0.001$; Figure 3E, 3G). After controlling for demographics and interactions, post PRI, post FSIQ, ΔPRI, and ΔFSIQ all significantly predicted physics reasoning task accuracy, implying that they were not wholly dependent on students' sex and classroom environment. However, relations between each post PRI, ΔPRI, and ΔFSIQ and performance were moderated by students' sex, such that female students exhibited a more positive relationship between WAIS-IV scores and performance than male students (Figure 3B, 3C, and 3D; Table 2, Physics Reasoning Accuracy). Conversely, male students demonstrated overall higher accuracy on the task, in line with previous research.[82] Of these, the regression of task accuracy on ΔPRI that included class- and sex-interaction terms (i.e., per Equation 1) explained significantly more variance in physics reasoning task accuracy than did the corresponding model without interaction terms. Thus, the relation between task accuracy and ΔPRI is better understood in the context of students' sex and classroom environments.

Post-instruction accuracy on the physics knowledge task was associated with post-instruction Verbal Comprehension Index (post VCI; $F_{(9, 120)} = 6.474$, $p < 0.001$; Figure 3H) and Working Memory Index (post WMI; $F_{(9, 120)} = 6.008$, $p < 0.001$; Figure 3I). Physics knowledge accuracy was related to post VCI after controlling for potential moderations of this relation by sex and class type (Figure 3H). The converse was true of relations between accuracy and post WMI. Post WMI displayed sex-dependent relations with task accuracy, such that greater accuracy in male students' was associated with greater increases in WAIS scores than those of female students (Figure 3I). Notably, this moderation was in the opposite direction of those between WAIS scores and physics reasoning accuracy noted above. Of these, regression of task accuracy on WAIS-IV scores with class- and sex-interaction terms (i.e., of the form displayed in Equation 1) did not explain significantly more of the

variance in physics knowledge task accuracy than did the models without interaction terms. This indicates that, unlike in the case of physics reasoning, these sex and class interactions do not significantly add to our understanding of these relations.

We found no relations between students' final course grade and any WAIS-IV score or change therein. Parameters and test statistics for all regressions, including non-significant regressions, can be found in Supplementary Tables 2 and 3.

Table 2. Significant relations between physics task performance and WAIS scores.

	Physics Reasoning Accuracy				*Physics Knowledge Accuracy*	
	Post PRI	Post FSIQ	ΔPRI*	ΔFSIQ	Post VCI	Post WMI
WAIS	**0.005**	**0.007**	*0.006*	**0.005**	**0.003**	0.001
WAIS X Sex (M)	**-0.009**	-0.003	*-0.010*	*-0.010*	0.004	**0.005**
WAIS X Class (A)	0.002	0.003	-0.001	-0.001	-0.002	0.002
WAIS X Sex X Class	0.006	-0.001	0.008	0.008	-0.002	-0.004
Sex X Class	-0.764	0.038	-0.109	-0.109	0.270	0.498
Sex (M)	**1.190**	0.458	*0.232*	*0.225*	-0.434	**-0.507**
Class (A)	-0.212	-0.319	0.004	-0.028	0.142	-0.235
Age	**-0.022**	-0.014	**-0.026**	**-0.029**	-0.001	-0.004
Year in University	0.011	0.001	0.020	0.016	-0.014	-0.013

Regression coefficients are shown for each variable on which physics reasoning and physics knowledge were regressed, for OLS regressions of the form shown in Equation 1 that were significant at $p_{FWE-corr} < 0.05$ and in which a WAIS term or interaction parameter significantly related to performance. Bold text indicates parameters significant at $p < 0.05$; bold and italicized text, parameters significant at $p < 0.01$. *Indicates that the full interaction model (i.e., of the form shown in Equation 1) explained significantly more variance than a smaller model without interactions between each WAIS scores, class (A for active learning), and sex (M for male).

These data indicate that WAIS-IV scores were clearly, but differentially associated with performance on physics-related assessments, suggesting a distinction between skills related to physics conceptual reasoning and content knowledge recall. Significant associations between physics reasoning accuracy and each ΔPRI and ΔFSIQ suggest that the development of perceptual reasoning ability and general intellectual ability underscore performance in physics reasoning. Similarly, relations between post-instruction VCI and WMI scores and physics knowledge accuracy suggest that working memory and verbal comprehension at post-instruction support successful physics knowledge retrieval, not necessarily the development of those skills (i.e., pre- to post-instruction). Nonetheless, post PRI and full-scale WAIS scores remain relevant for students' performance on physics conceptual reasoning and problem solving tasks, in addition to the development of such skills.

Functional brain network efficiency and connectivity differentially support component intelligence across contexts.

To further investigate associations between intellectual ability and task performance, we assessed brain organization during physics-related cognition using regressions of the same form as Equation 1 above. Specifically, we combined theories of the neurobiology of intelligence (i.e., P-FIT and NEH) with methods for studying individual differences in the brain organization (i.e., connectomics and network science) to search for a common neural substrate underlying the relations between WAIS-IV scores and accurate physics cognition. Measures of functional connectivity and network efficiency (denoted "topology" below) were regressed on WAIS-IV scores, while students' sex, classroom environment, demographics, and head movement (i.e., framewise displacement, "fd").

Equation 2

$$topology = \beta_1 IQ + \beta_2 IQ \times Sex + \beta_3 IQ \times Class + \beta_4 IQ \times Class \times Sex + \beta_5 Sex + \beta_6 Class + \beta_7 Sex \times Class + \beta_8 Age + \beta_9 Years\ in\ Univ. + \beta_{10} fd$$

Topological measures were calculated from functional connectivity graphs computed from fMRI data collected while participants performed the physics reasoning and physics knowledge tasks, using two brain parcellations to ensure that results are not parcellation-induced artifacts. In these graphs, individual brain regions comprise nodes and the pairwise correlation of their BOLD signals comprise edge weights, representing functional connectivity. We regressed, separately, (a) global efficiency calculated during each task, (b) local efficiency of each brain region during each task, and (c) connectivity between each pair of brain regions, during each task, on only the WAIS-IV scores significantly related to performance on said task. Significance thresholds of $\alpha < 0.05$ were adjusted to control the familywise error rate using the Šidák procedure, adjusted to account for dependence of correlated measures (Li & Ji, 2005; Sidak, 1967). Regression test statistics, fit statistics, and parameter estimates for all regressions calculated here are shown in Supplementary Tables 4 - 6.

Global and Local Efficiency

These analyses found that WAIS-IV scores were not significantly associated with global efficiency or with local efficiency across the brain during either task. Across both tasks, only head movement was associated with brain network efficiency.

Connectivity

Of the WAIS-IV scores associated with physics reasoning task accuracy, only post FSIQ was additionally associated with task
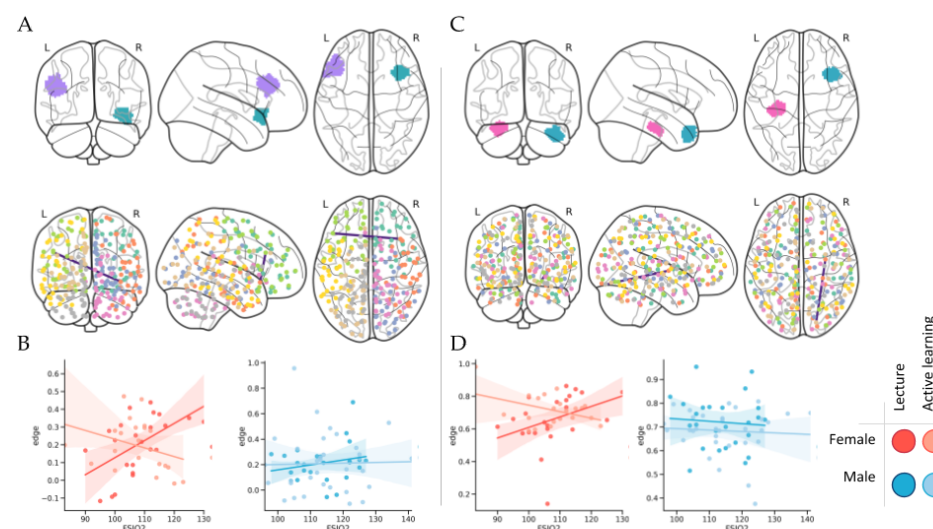

Figure 4. Functional connectivity during physics-related cognition is contextually related to WAIS scores following a semester of physics instruction. Regions (top row) demonstrating functional connectivity (second row) during the physics reasoning task that is significantly related to full-scale IQ scores, color-coded according to anatomical position in the brain to highlight similarities and differences between the parcellations. This connectivity is associated with FSIQ scores differently for male and female students (bottom row) in the Shen (left column) and Craddock (right column) parcellations. Each includes a portion of the anterior insula, though the region to which it is connected differs between parcellations.
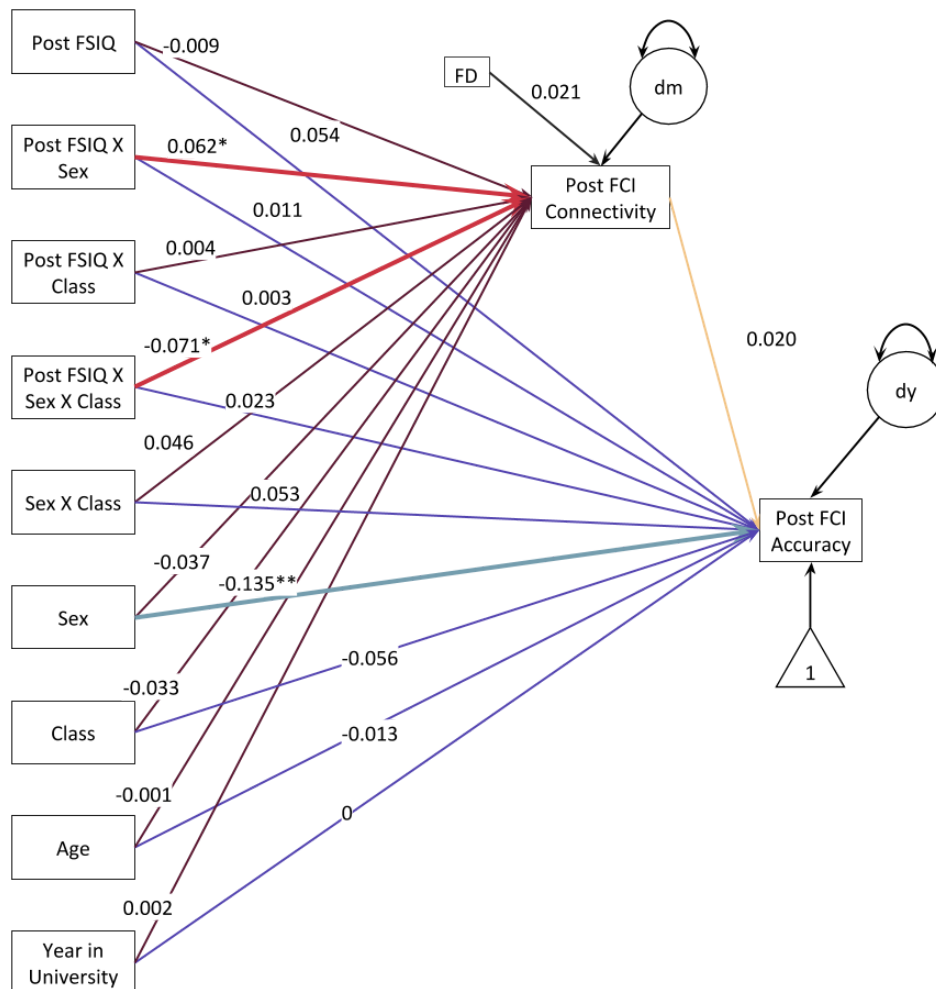
Figure 5. Functional connectivity is related to FSIQ, but does not mediate its relationship with physics reasoning accuracy. This model combines Equations 1 and 2 into a mediation model to assess whether task-based connectivity that is related to full-scale WAIS scores explains the relations between full-scale WAIS scores and students' accuracy on the physics reasoning task. Overall, this was not the case, although several exogenous variables that significantly explained variance in accuracy (see Equation 1, **Table 2**) were no longer significantly associated with accuracy in this model, where their relations with connectivity are simultaneously being considered. Lighter red paths indicate significant predictors of post-instruction physics reasoning connectivity, while lighter blue paths indicate significant predictors of post-instruction physics reasoning accuracy. *Indicates significance at $p < 0.05$; **, at $p < 0.01$, model is significant at $p_{FWE-corr} < 0.01$. Covariance between exogenous variables is not shown in this model, but was assessed (see **Supplementary Table 4**).

connectivity, of the right anterior insula (Figure 4A, 4C), depending on students' sex and classroom environment (Figure 4B, 4D). For female students greater FSIQ was associated with increased connectivity for those enrolled in lecture-based classes, but decreased connectivity for those enrolled in active learning classes. Meanwhile, for male students, the direction of associations between post-instruction FSIQ and connectivity did not differ due to classroom environment, though it did across parcellations (Figure 4B, 4D). Furthermore, the two parcellations both indicate significant associations of FSIQ and functional connectivity during the physics reasoning task; they did not converge on a particular network or region, providing no specific neuroanatomical locus. Across both parcellations, there was no consistent association between functional connectivity during the physics knowledge task and either post WMI or VCI scores, only with students' head movement during this task (see Supplementary Figures 2 and 3 for all significant edges).

Functional brain networks are not a common neural substrate supporting the role of intelligence in physics-related cognition.

We sought to explore possible common neural substrates for WAIS-IV and physics reasoning task accuracy. To this end, we assembled mediation models to assess whether brain connectivity significantly associated with WAIS-IV scores (Table 3, Figure 4) explains shared variance between WAIS-IV scores and task accuracy (Table 2, Figure 3), accounting for the interactions and covariates in Equations 1 and 2 (all model statistics reported in Supplementary Table 7). These models (Figure 5) indicated that the functional connectivity was unrelated to students' accuracy on the task (yellow path), and while FSIQ scores continued to explain a significant proportion of variability in connectivity (red paths), they no longer significantly accounted for variability in physics reasoning task accuracy (blue paths). Across parcellations, anterior insula connectivity (Figure 4) was significantly associated with FSIQ, but not task accuracy, and provided no significant mediation of the

IQ-accuracy relationship (Supplementary Table 3). While WAIS scores and task accuracy seem to capture related behavioral phenomena, FSIQ-accuracy relations were weakened by the inclusion of functional connectivity in the model, and connectivity did not significantly mediate the FSIQ-accuracy association (Table 3).

Table 3. Mediation of the relations between changes in full-scale IQ and physics reasoning accuracy by functional connectivity.

	Connectivity		Accuracy	
	Parameter estimate	*P*-value	**Parameter estimate**	*P*-value
FSIQ	-0.009	0.695	0.054	0.123
FSIQ X Sex	**0.062**	0.025	0.011	0.805
FSIQ X Class	0.004	0.894	0.003	0.938
FSIQ X Sex X Class	**-0.071**	0.043	0.023	0.691
Sex X Class	0.046	0.199	0.053	0.355
Sex	-0.037	0.26	***-0.135***	0.001
Class	-0.033	0.154	-0.056	0.191
Age	-0.001	0.931	-0.013	0.168
Year in University	0.002	0.467	0	0.967
Head motion	0.021	0.098	--	--
Connectivity	--	--	0.020	0.891

Bold indicates significant parameters at $\alpha < 0.05$, bold and italicized, at $\alpha < 0.01$. Values shown are for the connectivity significant in the Craddock parcellation. Values for mediation with functional connectivity significant in the Shen parcellation can be found in Supplementary Table 4.

Discussion

We present evidence of complex relations between intellectual ability and physics learning, behaviorally and neurally. Among these are significant increases in cognitive ability (i.e., WAIS-IV scores) over a semester of physics instruction, corresponding with previously reported increases in college students tested twice over a similar three-month period. These changes did not differ based on students' sex or based on course pedagogy and classroom environment. Gains in PRI were among the largest across index scores, and positively related to physics reasoning accuracy, suggesting the WAIS-IV components driving FSIQ increases represent physics-related skill development. Our data indicate that WAIS-IV measures of cognitive ability related to task performance were also related to brain connectivity, but not efficiency, during the task. Although we found no evidence of common neural underpinnings for the performance-ability relationship, we did uncover a moderation of brain-ability associations by students' sex and physics classroom environment. Therefore, the neurobiology supporting skill acquisition is context-, and perhaps pedagogy-dependent, instead of intrinsic, underscoring the importance of experience and environment in associations between ability and performance.

Are WAIS gains related to physics education or general college education?

The study of intelligence lacks concrete definitions,[83,84] but compensates with extensive psychometrics.[38,84–89] Here, we observed gains in intellectual ability, per the WAIS-IV, that outpaced retest gains reported in the measure's standardization sample.[17,90,91] These gains were seen across WAIS-IV index scores, greatest in PSI and PRI though minimal in WMI, but did not significantly differ from retest effects over a similar time period in an independent sample of college students.[92] WAIS-IV score increases were not related to overall class performance (i.e., course grade), but instead to physics reasoning ability at course completion (i.e., post-instruction FCI accuracy). This suggests they capture skills related to students' grasp of Newtonian mechanics. While these relations are moderated by students' class type, we cannot definitively link WAIS-IV score increases to physics instruction, and pedagogical differences therein, without a control group of participants who were not exposed to a semester of physics instruction. Perceptual reasoning seems directly related to skills developed by the specific demands of the course, though it and other components of intelligence are likely developed and engaged broadly across university coursework. To clarify how university instruction may impact different components of intelligence across disciplines, future studies should extend assessments across a range of curricula in the sciences, arts, and humanities and compare retest effects across a semester of education with those seen in the broader population. This is especially relevant given (a) the present evidence of education-related gains in WAIS-IV scores that appear to capture domain-specific skills and, importantly, (b) the widespread belief in the significance of intelligence and IQ for student and life success.

Perceptual reasoning improvements underlie physics-related cognition.

Following a semester of physics instruction, physics knowledge task accuracy was related to post-instruction verbal comprehension and working memory abilities, while physics reasoning task accuracy was related to pre- to post-instruction changes in perceptual reasoning and full-scale WAIS-IV scores (i.e., intellectual or cognitive ability), in addition to their values post-instruction. These differences between the tasks may reflect their unique cognitive demands and the manner in which the associated skills are acquired and exercised.

While physics *knowledge* reflects memorization of formulae and definitions of physics concepts, physics reasoning reflects the development of accurate conceptual understanding of Newtonian mechanics and the macro-scale forces at work in the physical world (e.g., gravity, friction). Students are unlikely to know definitions and formulae learned in class before enrolling in a physics class, but working memory and reading comprehension skills captured by WMI and VCI are not domain-specific but domain-general abilities exercised across curricula.[93–101] Therefore, it is the associated post-instruction scores that capture the skills utilized in accurately recalling physics knowledge, i.e., recalling definitions and formulae learned in class, that were not necessarily developed during the class.

Conversely, physics *reasoning* draws on the conceptions of physical phenomena that students bring into their first physics class, including pre-existing "common-sense beliefs" acquired over a lifetime of interacting with the physical world that are at odds with scientific explanations.[102] Decades of physics education research suggests that these pre-existing conceptions are difficult for students to overcome, even with formal instruction.[102–104] As this instruction relies heavily on visual representations of the movement of macroscale objects to teach Newtonian mechanics, encouraging students to rely on visuospatial skills and mental imagery, likely developing their perceptual reasoning skills throughout the course.[50,105–109] In this respect, our results indicate that students with more accurate conceptions of Newtonian mechanics following physics instruction were those demonstrating larger increases in perceptual reasoning skills and greater absolute perceptual reasoning ability, post-instruction. It may be that either (a) students who acquire more perceptual reasoning skills in the course are better able to align their conceptions of how the world works with Newtonian explanations or (b) students who are able to overcome previous conceptions about the mechanisms of the physical world gain better perceptual reasoning skills than their peers who have more difficulty doing so. In the former case, these findings might illustrate

another approach for physics instructors to help students develop more accurate conceptions of Newtonian mechanics: focusing on students' perceptual reasoning skills.

Insofar as changes in WAIS-IV index scores reflect skill acquisition, it follows that physics reasoning accuracy, reflecting conceptualizations of Newtonian mechanics honed throughout physics instruction, is associated with *changes* in perceptual reasoning and full-scale intelligence, though *absolute* levels of these skills remain relevant. Conversely, physics knowledge accuracy, reflecting correct recall of definitions and formulae learned throughout physics instruction, is associated with absolute levels of verbal comprehension and working memory skills, but not the changes therein. This may highlight an opportunity for correcting students' pre-existing conceptions of physical phenomena, by focusing on students' perceptual reasoning skills throughout their physics instruction. However, the sex- and classroom-differences suggest that this is not a one-size-fits-all opportunity, and that students may benefit differently from such learning interventions. For female students, who historically perform poorer on the Force Concept Inventory (our physics reasoning task) due to a host of sociocognitive factors, increases in PRI might provide a means of "leveling the playing field" compared to their male counterparts.

Brain network organization during physics cognition is related to intelligence, but does not explain its relationship with physics task accuracy.

Contrary to previous research,[73–75] our data did not indicate associations between students' WAIS-IV scores and either global or local brain network efficiency and did not support the neural efficiency hypothesis, which suggests more intelligent individuals benefit from "more efficient" brains. The concept of neural "efficiency" is, at best, unclear, with varying definitions across the years, and at worst, empty and misleading.[110] Here, we use the network science definition of efficiency, defined mathematically in terms of connections between nodes of a graph (i.e., connectivity between brain regions).[111] These findings add domain-specific insight to what has previously been a study of intrinsic abilities and neurobiological processes, failing to find support in a sample demonstrating intelligence gains related to an extrinsic manipulation (i.e., physics instruction). Our data do not support the parieto-frontal integration hypothesis, either, as we uncovered no parieto-frontal connectivity underlying intelligence during physics-related cognition.

Representing a small proportion of all possible connections in the brain, sparse connectivity during physics reasoning was related to post-instruction full-scale WAIS-IV scores, the full complement of measured verbal, perceptual, working memory, and processing speed skills. Different roles of pedagogy and classroom environment on brain network connectivity, cognitive abilities, and relations between the two with respect to sex point out a potentially significant sex difference in classroom experience. In a heavily male-dominated field like physics, it is a reasonable assumption that male and female students would have differential classroom experiences.[60,112–115] For example, it is a commonly-held stereotype that men are good at math and that women are not,[116,117] subjecting women and female students to stereotype threat in physics classrooms, where beliefs negatively affect classroom experience and performance.[118,119] The data presented here indicate no sex differences in overall course performance, but persistent sex differences across classroom environments in associations between cognitive abilities and not only performance on physics assessments, but brain connectivity during those assessments. Together, this literature and our findings indicate meaningful neurobiological consequences of classroom experience on the basis of sex. However, our data show no sex- or classroom-related difference in changes in intelligence or cognitive abilities, though the brain-WAIS relationships point to a multitude of neurobiological representations of intelligence. Neural phenomena related to intelligence differ based not only on cognitive context, but on sociological and pedagogical contexts, as well.

Although intelligence measures associated with physics task accuracy also explained certain functional connectivity during these tasks, we found no evidence of a common neural substrate for intelligence and accuracy. On the whole, task-based connectivity related to intelligence was unrelated to participants' accuracy on said task, and no connectivity associated with intelligence accounted for any significant portion of the relations

between intelligence and accuracy. This may indicate a lack of a common neural substrate for these two phenomena, suggesting that, while they are measuring a similar skill or capacity, they are not measuring the same skill or capacity. On the other hand, the neural substrate of a common skill measured by PRI or full-scale WAIS scores and physics reasoning accuracy may merely lie beyond a linear relation with brain network connectivity or topology. In any case, understanding the neural instantiation of a common reasoning skill to perceptual reasoning and physics reasoning requires further study.

Finally, rather than a global property of brain network organization, as indicated in prior research,[66,68,120] these data indicate that sparse, coordinated interactions of disparate brain regions underlie intelligence, in this domain-specific context. That connections across the brain during physics-related cognition are related to changes in students' overall intellectual skills, but differently with respect to their classroom environment, casts further doubt on the notion of IQ or intelligence as a fixed, innate measure and, instead, highlights the role of environment and experience. Although differences in these relationships between female and male students support the substantial body of literature supporting this notion of sex differences in the biological representations of intelligence,[71,121–128] here we suggest a potential sociological explanation. These findings indicate, too, neural support for intelligence exhibits domain-specific relations in the context of STEM education. Not only does cognitive context matter to the relations between intelligence and brain network organization, our data indicate that sex and learning environment matter, too.

Limitations and future directions

Here, we demonstrate increases in intellectual ability over a semester of physics instruction. While students' sex and classroom environment did not affect the extent of these increases, the data suggest meaningful consequences of classroom environment on the relations between intellectual ability and underlying brain network organization. The implication that the learning environment affects male and female students differently, both cognitively and neurobiologically, in a field as male-dominated as physics demands attention and further study. However, as there were no observed sex differences in final course grade or in change in intellectual ability, any differences in classroom experience are not differentially affecting female and male students' academic performance in the course. Further work should assess whether differences in experience and associated brain function are linked to long-term success for male and female students. This assessment should consider factors beyond overt measures of success and focus on variables related to self-efficacy and in-classroom experiences, both previously been shown to affect men and women differently in physics education.[58]

While we have identified differences in physics-related brain organization and its relation to WAIS-IV scores based on class type, this study is unable to distinguish whether these differences are due to differences in pedagogy or social classroom environment, or to practice effects across a short assessment period. Future research should include in-classroom assessments of social climate and the possibility of gender differences in social interactions during physics instruction. Furthermore, control groups in (1) another, less male-dominated, domain and (2) an age- and sex-matched groups outside of university would provide insight into both sex differences in and the degree to which changes in intellectual ability are associated with STEM education.

Conclusion

While our data indicate clear relations between domain-specific components of intellectual ability and performance on assessments of physics conceptual reasoning and content knowledge, we found no association between academic outcomes and intelligence. Likewise, intelligence was related to functional brain connections during these assessments, but none that explain its association with performance. Our multifaceted approach to studying the neurobiological underpinnings of intelligence did not uncover a single, robust aspect of brain network organization that was consistent across cognitive contexts and experiential influences. However, relations between intelligence and, separately, physics task performance and task-based functional connectivity

were moderated by students' sex and classroom environment. Ultimately, both the magnitude and development of perceptual reasoning skills is meaningful over the course of a semester of physics instruction. As these data show significant increases in perceptual reasoning over a semester, we present an optimistic and more plastic view of intelligence, as a set of skills to be developed, rather than an innate capacity that students either have or don't have. Together, these data highlight the complicated nature of relations between intelligence and classroom successes, which vary with students' sex and domain-specific experiences.

Methods

Participants and Study Design

One hundred and thirty healthy right-handed undergraduate students (mean age = 20.03 ± 2.25 years, range = 18-25 years; 61 females) who completed a semester of introductory calculus-based physics at Florida International University (FIU), a Hispanic Serving Institution, took part in this study. Participants were not currently using psychoactive medications and reported that they had not been diagnosed with any cognitive impairments or neurological or psychiatric conditions. The physics course emphasized problem solving skill development and covered topics in classical Newtonian mechanics, including motion along straight lines and in two and three dimensions, Newton's laws of motion, work and energy, momentum and collisions, and rotational dynamics. Students were either enrolled in a lecture class or an active learning, "Modeling Instruction", class, which bases course content in conceptual scientific models and instructs students to appropriate scientific models for their own use. Students completed behavioral assessments and MRI scans in separate appointments at two time points: at the beginning ("pre-instruction") and conclusion ("post-instruction") of the 15-week semester. Pre-instruction data collection sessions were acquired no later than the fourth week of classes and post-instruction sessions were completed no more than two weeks after the final exam. Written informed consent was obtained in accordance with FIU's Institutional Review Board approval.

Missing Data

A missing value analysis indicated that, of the variables of interest in this study, missingness ranged from 2% to 17%. Data were more often missing from MRI data than behavioral or demographic data and more often missing from post-instruction data than from pre-instruction. Assessment of the relations between missingness on each variable and values of each other variable of interest revealed that data were likely missing completely at random (MCAR). Behavioral and brain network efficiency data were imputed using iterated Bayesian ridge regression implemented in scikit-learn (v. 0.23.1; scikit-learn.org/). Due to its high dimensionality, missingness in edgewise functional connectivity data was addressed using distance-weighted *K*-Nearest Neighbors approach (*K* = 100, where p = 71,824) implemented in scikit-learn, which is robust to missingness up to 20% [129].

Behavioral Measures

During pre- and post-instruction behavior sessions, participants were administered the fourth edition of the Wechsler Adult Intelligence Scale (WAIS-IV[17]), a standardized intelligence test for adults, in addition to other assessments not used here. The WAIS-IV provides scores in four domains, in addition to an overall score of intellectual functioning. The Verbal Comprehension index measures application of verbal skills in problem solving. The Perceptual Reasoning index measures the ability to detect the underlying conceptual relationship among visual objects and use reasoning to identify and apply rules. The Working Memory index measures short-term memory with auditory and visual stimuli. The Processing Speed index measures speed of mental operations and visual-motor coordination. Lastly, the Full-Scale intelligence quotient (IQ) represents a global estimate of intellectual or cognitive ability. All instruments were administered by researchers for the purpose of this research and not professionally or clinically (i.e., for diagnostic or instructional purposes).

fMRI Tasks

In the scanner, participants performed two different physics tasks, each probing different aspects of physics learning and problem solving (Figure 1).

Participants completed three runs of a *physics reasoning* task, which uses questions from the Force Concept Inventory (FCI; [79]) to assess domain-specific problem solving. This task includes two conditions, FCI and control, presented in a block design with self-paced trials. FCI and control questions were presented in three screens (Figure 1A, SI Figure 3), between which participants advanced by the press of a button. The first screen presented a written description of a physical scenario and corresponding figure; the second, a question relating to the scenario; and the third, four answer choices from which the participants were instructed choose the correct answer while mentally justifying their choice.

Participants additionally completed two runs of a *physics knowledge* task, which probed physics-related memory retrieval and included physics, general, and control conditions. Participants were asked a series of multiple-choice questions and instructed to respond by indicating their choice with the press of a button (Figure 1B). In the physics condition, participants were asked to recall definitions and formulas taught in the physics course (e.g., "What does the 'SI' in SI units stand for?" or "What is the value of the acceleration due to gravity?"). In the general condition, participants were asked to recall general trivia (e.g., "Which of these is not an automobile brand?" or, "Who is the President of the United States?"). The low-level control condition asked participants to press the button corresponding to a letter or symbol. Conditions were organized into blocks and each run included three blocks per condition.

fMRI Acquisition and Pre-Processing

Neuroimaging data were acquired on a GE 3T Healthcare Discovery 750W MRI scanner at the University of Miami. Functional MRI (fMRI) data were acquired with an interleaved gradient-echo, echo planar imaging (EPI) sequence (TR/TE = 2000/30ms, flip angle = 75°, field of view [FOV] = 220x220mm, matrix size = 64x64, voxel dimensions = 3.4×3.4×3.4mm, 42 axial oblique slices). A T1-weighted series was also acquired using a 3D fast spoiled gradient recall brain volume (FSPGR BRAVO) sequence with 186 contiguous sagittal slices (TI = 650ms, bandwidth = 25.0kHz, flip angle = 12°, FOV = 256x256mm, and slice thickness = 1.0mm). A 2-mm isotropic MNI152 template image was nonlinearly oriented to each participant's structural T1-weighted image using FMRIB's Software Library's (FSL; https://fsl.fmrib.ox.ac.uk/fsl/fslwiki[130]) nonlinear registration tool (FNIRT[131]). Then, each participant's T1-weighted image was coregistered to the middle volume of each functional run, using FSL's linear registration tool (FLIRT[132]). These two transformations were concatenated and used to align regionwise parcellations to each subject's functional images. Tissue-type masks for white matter, gray matter, and cerebrospinal fluid (CSF) were created from each subject's T1-weighted images using FSL's automated segmentation tool (FAST[133]).

Task-based fMRI preprocessing began with FSL's MCFLIRT with spline interpolation, per run per functional task, to align all volumes of each subject's fMRI time series with that middle volume. To further correct for in-scanner motion effects, functional volumes unduly affected by motion were identified using fsl_motion_outliers, with a framewise displacement threshold of 0.9mm for functional scans.[134] Data were standardized, detrended, and high-pass filtered, according to the period of each task. The physics knowledge task was high-pass filtered at 0.018Hz and the physics reasoning task was thresholded according to each participant's individual timing.

Regionwise Parcellation and Brain Connectivity Analyses

Each participant's fMRI data were parcellated according to two functionally-derived, whole-brain parcellations with similar numbers of regions. Here, we used a 268-region parcellation computed via multigraph k-way

clustering, without spatial constraints, henceforth referred to as the Shen parcellation.[135] To ensure that our results were not artifacts of node definition, we additionally performed all analyses with an atlas generated from resting-state fMRI data by performing normalized-cut spectral clustering on voxelwise functional connectivity data to define homogeneous, spatially-constrained clusters (i.e., regions), henceforth referred to as the Craddock parcellation,[136] which includes a range of atlas sizes from which we chose a 268-region solution, to match the granularity of the Shen parcellation. Connectivity graphs were computed using two different brain parcellations (Supplementary Figure 4). Following preprocessing, data analysis continued in two parallel streams, one with the Craddock parcellation and one with the Shen parcellation, to ensure that any results were not artifacts of parcellation schema. Presented results (e.g., Figures 4, 5) include brain-IQ relations observed for both parcellations, but values displayed are derived from the Shen parcellation.

For each region, a single time series was computed as an average of the fMRI time series from all voxels within the region, after further regressing out six motion parameters (from MCFLIRT) and censoring high-motion volumes (framewise displacement >0.9mm), as well as the immediately preceding volume and two following volumes, following recommendations from Power et al. [137]. Functional tasks' regionwise time series were standardized (i.e., z-scored), divided by condition per task per run and spliced together across runs, creating separate time series per condition per task for each participant. Adjacency matrices were constructed with each parcellation per participant, per functional task, per session (pre- and post-instruction) using Nilearn (v. 0.3.1, http://nilearn.github.io/index.html), a Python (v 2.7.13) module, built on scikit-learn, for the statistical analysis of neuroimaging data,[138,139] by computing the pairwise Pearson's correlations between each pair of regions, resulting in a 268x268 region-wise correlation matrix for each subject per condition per task per session (pre- and post-instruction). Graph theoretic, topological measures were calculated across a range of density-based thresholds. The lowest thresholds at which each network became (a) scale-free and (b) became node-connected were calculated for each adjacency matrix. From these values, a lower- and upper-bound for network thresholding were estimated, following recommendations from Lynall[140] and Ginestet[141], such that networks would remain node-connected, meaning there are no brain regions completely separate from the rest of the brain, and spurious connections would be removed while maintaining the scale-free degree distribution expected of the brain per prior research.[142–144]

All topological measures were calculated using bctpy, a Python toolbox intended to replicate the functionality of the Brain Connectivity Toolbox, a MATLAB toolbox for graph theoretic analysis of functional and structural brain connectivity (brain-connectivity-toolbox.net[111]). From each correlation matrix described above, we calculated global efficiency, characteristic path length, and modularity for across the range of proportional thresholds as calculated above (τ = [0.21, 0.31] at steps of 0.01), then calculated the area under the curve (AUC) of each measure.[145,146] These AUCs were used in all following statistical tests assessing the relationship between brain network organization and IQ and will henceforth be referred to per the topological measure from which they were calculated. All topology-related results reported here are significant per topology values calculated from graphs generated from both parcellations.

Statistical Analyses

Statistical inference was performed using the lavaan R package, and the Python modules SciPy (v. 1.2.1; scipy.org[147–153]) statsmodels (v. 0.9.0; statsmodels.org), and nilearn (v. 0.6.2; nilearn.github.io).

Paired t-tests were used to assess changes in WAIS-IV scores pre- to post-instruction. Two-sample t-tests were used to assess differences in the changes in WAIS-IV scores (post- minus pre-instruction) between male and female students, as well as between students enrolled in the active learning and lecture classes.

Ordinary least squares (OLS) regressions implemented in R were used to regress measures of academic and task performance on WAIS-IV scores, sex, class, age, and years in university, per Equation 1. Significance of individual

models was assessed by comparing models' p-values to a significance threshold adjusted for multiple comparisons via Šidák correction.[81]

A similar procedure, using mass-univariate OLS regressions with permutation testing as implemented by the Python package Nilearn,[154–156] was used to regress topological measures and functional connectivity (thresholded at τ = 0.31) on WAIS-IV scores, sex, class, age, years in university, head size, average framewise displacement (calculated by fsl_motion_outliers, per run, per task), per Equation 2. To correct for multiple comparisons across these regressions, we used the Šidák correction as mentioned above. All reported results were significant in both parcellations, to minimize the effects of brain parcellation on our interpretations.

Mediation models to assess whether brain connectivity explained the relationship between WAIS-IV scores and task performance were run using the R package lavaan.[157]

Data Availability

A GitHub repository was created at github.com/62442katieb/physics-learning-iq to archive the code and source files for this study, including data preprocessing and analysis scripts and behavioral data. Significant neuroimaging results are available at neurovault.org/collections/9385/.

Acknowledgments

Competing Interests

The authors declare no competing interests.

Author Contributions

ARL, EB, SMP, MTS, RWL conceived and designed the project. JEB, EIB, RO, AN acquired behavioral and fMRI data. KLB, JEB, ARL analyzed data. MCR, TS contributed scripts and pipelines. KLB, JEB, ARL wrote the paper and all authors contributed to the revisions and approved the final version.

References

1. Alloway, T. P. & Alloway, R. G. Investigating the predictive roles of working memory and IQ in academic attainment. *J. Exp. Child Psychol.* **106**, 20–29 (2010).

2. Gottfredson, L. S. Where and why g matters: Not a mystery. *Hum. Perform.* **15**, 25–46 (2002).

3. Gottfredson, L. S. The General Intelligence Factor. *Sci. Am. Presents Explor. Intell.* **9**, 24–30 (1998).

4. Harris, D. Factors affecting college grades: a review of the literature, 1930-1937. *Psychol. Bull.* **37**, 125–166 (1940).

5. Frisby, C. L. & Henry, B. Science, politics, and best practice: 35 years after Larry P. *Contemp. Sch. Psychol.* **20**, 46–62 (2016).

6. Reddy, A. The eugenic origins of IQ testing: Implications for post-Atkins litigation. *DePaul Rev* **57**, 667 (2007).

7. Reynolds, C. R. & Suzuki, L. A. Bias in Psychological Assessment. in *Handbook of Psychology, Second Edition* (American Cancer Society, 2012). doi:10.1002/9781118133880.hop210004.

8. Weiss, L. G., Chen, H., Harris, J. G., Holdnack, J. A. & Saklofske, D. H. CHAPTER 4 - WAIS-IV Use in Societal Context. in *WAIS-IV Clinical Use and Interpretation* (eds. Weiss, L. G., Saklofske, D. H., Coalson, D. L. & Raiford, S. E.) 97–139 (Academic Press, 2010). doi:10.1016/B978-0-12-375035-8.10004-7.

9. Dzirasa, K. A brilliant approach to study the basis of intelligence? *Sci. Transl. Med.* **9**, eaao0978 (2017).

10. Bartels, M., Rietveld, M. J. H., Van Baal, G. C. M. & Boomsma, D. I. Genetic and environmental influences on the development of intelligence. *Behav. Genet.* **32**, 237–249 (2002).

11. Canivez, G. L. & Watkins, M. W. Long-term stability of the Wechsler Intelligence Scale for Children—Third Edition. *Psychol. Assess.* **10**, 285–291 (1998).

12. Hertzog, C. & Schaie, K. W. Stability and change in adult intelligence: 1. Analysis of longitudinal covariance structures. *Psychol. Aging* **1**, 159–171 (1986).

13. Herrnstein, R. J. & Murray, C. *The Bell Curve: Intelligence and Class Structure in American Life*. (Free Press, 1994).

14. Plomin, R. & von Stumm, S. The new genetics of intelligence. *Nat. Rev. Genet.* **19**, 148–159 (2018).

15. Reynolds, C. A. *et al.* Quantitative Genetic Analysis of Latent Growth Curve Models of Cognitive Abilities

in Adulthood. *Dev. Psychol.* **41**, 3–16 (2005).

16. *Behavior Genetics of Cognition Across the Lifespan*. (Springer-Verlag, 2014). doi:10.1007/978-1-4614-7447-0.

17. Wechsler, D., Coalson, D. L. & Raiford, S. E. *WAIS-IV: Wechsler adult intelligence scale*. (Pearson San Antonio, TX, 2008).

18. Begovac, I., Begovac, B., Majić, G. & Vidović, V. LONGITUDINAL STUDIES OF IQ STABILITY IN CHILDREN WITH CHILDHOOD AUTISM – LITERATURE SURVEY. *Psychiatr. Danub.* **21**, 310–319 (2009).

19. Johnstone, B. & Wilhelm, K. L. The longitudinal stability of the WRAT-R Reading subtest: Is it an appropriate estimate of premorbid intelligence? *J. Int. Neuropsychol. Soc.* **2**, 282–285 (1996).

20. Moffitt, T. E., Caspi, A., Harkness, A. R. & Silva, P. A. The Natural History of Change to Intellectual Performance: Who Changes? How Much? Is it Meaningful? *J. Child Psychol. Psychiatry* **34**, 455–506 (1993).

21. Pietschnig, J. & Voracek, M. One Century of Global IQ Gains: A Formal Meta-Analysis of the Flynn Effect (1909–2013). *Perspect. Psychol. Sci.* **10**, 282–306 (2015).

22. Ritchie, S. J. & Tucker-Drob, E. M. How Much Does Education Improve Intelligence? A Meta-Analysis. *Psychol. Sci.* **29**, 1358–1369 (2018).

23. Schneider, W., Niklas, F. & Schmiedeler, S. Intellectual development from early childhood to early adulthood: The impact of early IQ differences on stability and change over time. *Learn. Individ. Differ.* **32**, 156–162 (2014).

24. Schuerger, J. M. & Witt, A. C. The temporal stability of individually tested intelligence. *J. Clin. Psychol.* **45**, 294–302 (1989).

25. Tuma, J. M. & Appelbaum, A. S. Reliability and Practice Effects of Wisc-R Iq Estimates in a Normal Population. *Educ. Psychol. Meas.* **40**, 671–678 (1980).

26. Verhaeghen, P. Aging and vocabulary score: A meta-analysis. *Psychol. Aging* **18**, 332–339 (2003).

27. Yang, P., Jong, Y.-J., Hsu, H.-Y. & Lung, F.-W. Role of assessment tests in the stability of intelligence scoring of pre-school children with uneven/delayed cognitive profile. *J. Intellect. Disabil. Res.* **55**, 453–461 (2011).

28. Yu, H., McCoach, D. B., Gottfried, A. W. & Gottfried, A. E. Stability of intelligence from infancy through adolescence: An autoregressive latent variable model. *Intelligence* **69**, 8–15 (2018).

29. Hunter, J. E. & Hunter, R. F. Validity and utility of alternative predictors of job performance. *Psychol. Bull.*

96, 72–98 (1984).

30.   Deary, I. J., Strand, S., Smith, P. & Fernandes, C. Intelligence and educational achievement. *Intelligence* **35**, 13–21 (2007).

31.   Strenze, T. Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence* **35**, 401–426 (2007).

32.   Zax, J. S. & Rees, D. I. IQ, Academic Performance, Environment, and Earnings. *Rev. Econ. Stat.* **84**, 600–616 (2002).

33.   Murray, C. A. *Income Inequality and IQ*. (AEI Press, 1998).

34.   von Stumm, S. Socioeconomic status amplifies the achievement gap throughout compulsory education independent of intelligence. *Intelligence* **60**, 57–62 (2017).

35.   Day, L., Hanson, K., Maltby, J., Proctor, C. & Wood, A. Hope uniquely predicts objective academic achievement above intelligence, personality, and previous academic achievement. *J. Res. Personal.* **44**, 550–553 (2010).

36.   Zuffianò, A. *et al.* Academic achievement: The unique contribution of self-efficacy beliefs in self-regulated learning beyond intelligence, personality traits, and self-esteem. *Learn. Individ. Differ.* **23**, 158–162 (2013).

37.   Bergold, S. & Steinmayr, R. Personality and Intelligence Interact in the Prediction of Academic Achievement. *J. Intell.* **6**, 27 (2018).

38.   Neisser, U. *et al.* Intelligence: Knowns and unknowns. *Am. Psychol.* **51**, 77–101 (1996).

39.   Busato, V. V., Prins, F. J., Elshout, J. J. & Hamaker, C. Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personal. Individ. Differ.* **29**, 1057–1068 (2000).

40.   Walker, A. J., Batchelor, J. & Shores, A. Effects of education and cultural background on performance on WAIS-III, WMS-III, WAIS-R and WMS-R measures: Systematic review. *Aust. Psychol.* **44**, 216–223 (2009).

41.   Ding, L. Verification of causal influences of reasoning skills and epistemology on physics conceptual learning. *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **10**, 023101 (2014).

42.   Jackson, J., Dukerich, L. & Hestenes, D. Modeling Instruction: An Effective Model for Science Education. *Sci. Educ.* **17**, 10–17 (2008).

43.   Roberts, J. D., While, A. E. & Fitzpatrick, J. M. Problem solving in nursing practice: application, process, skill acquisition and measurement. *J. Adv. Nurs.* **18**, 886–891 (1993).

44.   Taasoobshirazi, G. & Sinatra, G. M. A structural equation model of conceptual change in physics. *J. Res. Sci. Teach.* **48**, 901–918 (2011).

45.   Bicer, A. *et al.* Moving from STEM to STEAM: The effects of informal STEM learning on students' creativity and problem solving skills with 3D printing. in *2017 IEEE Frontiers in Education Conference (FIE)* 1–6 (2017). doi:10.1109/FIE.2017.8190545.

46.   Etkina, E., Murthy, S. & Zou, X. Using introductory labs to engage students in experimental design. *Am. J. Phys.* **74**, 979–986 (2006).

47.   Gabel, D. L. & Sherwood, R. D. Facilitating problem solving in high school chemistry. *J. Res. Sci. Teach.* **20**, 163–177 (1983).

48.   Stieff, M. & Uttal, D. How Much Can Spatial Training Improve STEM Achievement? *Educ. Psychol. Rev.* **27**, 607–615 (2015).

49.   Hegarty, M., Stieff, M. & Dixon, B. L. Cognitive change in mental models with experience in the domain of organic chemistry. *J. Cogn. Psychol.* **25**, 220–228 (2013).

50.   Hegarty, M. Spatial Thinking in Undergraduate Science Education. *Spat. Cogn. Comput.* **14**, 142–167 (2014).

51.   Oswald, T. M. *et al.* Clinical and Cognitive Characteristics Associated with Mathematics Problem Solving in Adolescents with Autism Spectrum Disorder. *Autism Res. Off. J. Int. Soc. Autism Res.* **9**, 480–490 (2016).

52.   Weiss, L. G., Saklofske, D. H., Coalson, D. L. & Raiford, S. E. CHAPTER 3 - Theoretical, Empirical and Clinical Foundations of the WAIS-IV Index Scores. in *WAIS-IV Clinical Use and Interpretation* (eds. Weiss, L. G., Saklofske, D. H., Coalson, D. L. & Raiford, S. E.) 61–94 (Academic Press, 2010). doi:10.1016/B978-0-12-375035-8.10003-5.

53.   Day, J., Stang, J. B., Holmes, N. G., Kumar, D. & Bonn, D. A. Gender gaps and gendered action in a first-year physics laboratory. *Phys. Rev. Phys. Educ. Res.* **12**, 020104 (2016).

54.   Kelly, A. M. Social cognitive perspective of gender disparities in undergraduate physics. *Phys. Rev. Phys. Educ. Res.* **12**, 020116 (2016).

55.   Lorenzo, M., Crouch, C. H. & Mazur, E. Reducing the gender gap in the physics classroom. *Am. J. Phys.* **74**,

118–122 (2006).

56.   Madsen, A., McKagan, S. B. & Sayre, E. C. Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap? *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **9**, 020121 (2013).

57.   McCullough, L. Gender, Context, and Physics Assessment. *J. Int. Womens Stud.* **5**, 20–30 (2013).

58.   Nissen, J. M. & Shemwell, J. T. Gender, experience, and self-efficacy in introductory physics. *Phys. Rev. Phys. Educ. Res.* **12**, 020105 (2016).

59.   Eddy, S. L. & Brownell, S. E. Beneath the numbers: A review of gender disparities in undergraduate education across science, technology, engineering, and math disciplines. *Phys. Rev. Phys. Educ. Res.* **12**, 020106 (2016).

60.   Gewin, V. Gender divide in physics spans globe. *Nature* **473**, 547–547 (2011).

61.   *Status and Trends on the Education of Racial and Ethnic Groups, Indicator 26: STEM Degrees.* (2019).

62.   Freeman, S. *et al.* Active learning increases student performance in science, engineering, and mathematics. *Proc. Natl. Acad. Sci.* **111**, 8410 (2014).

63.   Hake, R. R. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *Am. J. Phys.* **66**, 64–74 (1998).

64.   Pollock, S. J., Finkelstein, N. D. & Kost, L. E. Reducing the gender gap in the physics classroom: How sufficient is interactive engagement? *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **3**, 010107 (2007).

65.   Fischer, F. U., Wolf, D., Scheurich, A. & Fellgiebel, A. Association of Structural Global Brain Network Properties with Intelligence in Normal Aging. *PLoS ONE* **9**, e86258 (2014).

66.   Heuvel, M. P., Stam, C. J., Kahn, R. S. & Hulshoff Pol, H. E. Efficiency of functional brain networks and intellectual performance. *J. Neurosci. Off. J. Soc. Neurosci.* **29**, 7619–24 (2009).

67.   Hilger, K., Ekman, M., Fiebach, C. J. & Basten, U. Efficient hubs in the intelligent brain: Nodal efficiency of hub regions in the salience network is associated with general intelligence. *Intelligence* **60**, 10–25 (2017).

68.   Langer, N. *et al.* Functional brain network efficiency predicts intelligence. *Hum. Brain Mapp.* **33**, 1393–1406 (2012).

69.   Li, Y. *et al.* Brain Anatomical Network and Intelligence. *PLoS Comput. Biol.* **5**, e1000395 (2009).

70.   Jung, R. E. & Haier, R. J. The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behav. Brain Sci.* **30**, 135–154; discussion 154-187 (2007).

71.   Neubauer, A. C. & Fink, A. Intelligence and neural efficiency: Measures of brain activation versus measures of functional connectivity in the brain. *Intelligence* **37**, 223–229 (2009).

72.   Santarnecchi, E., Galli, G., Polizzotto, N. R., Rossi, A. & Rossi, S. Efficiency of weak brain connections support general cognitive functioning. *Hum. Brain Mapp.* **35**, 4566–4582 (2014).

73.   Hearne, L. J., Mattingley, J. B. & Cocchi, L. Functional brain networks related to individual differences in human intelligence at rest. *Sci. Rep.* **6**, 32328 (2016).

74.   Dubois, J., Galdi, P., Paul, L. K. & Adolphs, R. A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **373**, (2018).

75.   Hilger, K., Ekman, M., Fiebach, C. J. & Basten, U. Intelligence is associated with the modular structure of intrinsic brain networks. *Sci. Rep.* **7**, 16088 (2017).

76.   Laird, A. R. *et al.* Behavioral interpretations of intrinsic connectivity networks. *J. Cogn. Neurosci.* **23**, 4022–4037 (2011).

77.   Smith, S. M. *et al.* Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 13040–5 (2009).

78.   Greene, A. S., Gao, S., Scheinost, D. & Constable, R. T. Task-induced brain state manipulation improves prediction of individual traits. *Nat. Commun.* **9**, 2807 (2018).

79.   Hestenes, D., Wells, M. & Swackhamer, G. Force concept inventory. *Phys. Teach.* **30**, 141–158 (1992).

80.   Bartley, J. E. *et al.* Brain activity links performance in science reasoning with conceptual approach. *Npj Sci. Learn.* **4**, 1–8 (2019).

81.   Šidák, Z. Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *J. Am. Stat. Assoc.* **62**, 626–633 (1967).

82.   Docktor, J. & Heller, K. Gender Differences in Both Force Concept Inventory and Introductory Physics Performance. *AIP Conf. Proc.* **1064**, 15–18 (2008).

83.   Neisser, U. The concept of intelligence. *Intelligence* **3**, 217–227 (1979).

84.   Weinberg, R. A. Intelligence and IQ: Landmark issues and great debates. *Am. Psychol.* **44**, 98–104 (1989).

85.   Howe, M. J. A. *IQ in Question: The Truth about Intelligence*. (SAGE, 1997).

86.   Hunt, E. The Concept and Utility of Intelligence. in *Intelligence, Genes, and Success: Scientists Respond to The Bell Curve* (eds. Devlin, B., Fienberg, S. E., Resnick, D. P. & Roeder, K.) 157–176 (Springer New York, 1997). doi:10.1007/978-1-4612-0669-9_7.

87.   Richardson, K. What IQ Tests Test. *Theory Psychol.* **12**, 283–314 (2002).

88.   Schönemann, P. H. Do IQ tests really measure intelligence? *Behav. Brain Sci.* **6**, 311–313 (1983).

89.   Sternberg, R. J. What Should Intelligence Tests Test? Implications of a Triarchic Theory of Intelligence for Intelligence Testing. *Educ. Res.* **13**, 5–15 (1984).

90.   Holdnack, J. A., Drozdick, L. W., Iverson, G. L. & Chelune, G. J. Chapter 6 - Serial Assessment with WAIS–IV and WMS–IV. in *WAIS-IV, WMS-IV, and ACS* (eds. Holdnack, J. A., Drozdick, L. W., Weiss, L. G. & Iverson, G. L.) 279–329 (Academic Press, 2013). doi:10.1016/B978-0-12-386934-0.00006-7.

91.   Cullum, C. M. & Larrabee, G. J. CHAPTER 6 - WAIS-IV Use in Neuropsychological Assessment. in *WAIS-IV Clinical Use and Interpretation* (eds. Weiss, L. G., Saklofske, D. H., Coalson, D. L. & Raiford, S. E.) 167–187 (Academic Press, 2010). doi:10.1016/B978-0-12-375035-8.10006-0.

92.   Estevis, E., Basso, M. R. & Combs, D. Effects of Practice on the Wechsler Adult Intelligence Scale-IV Across 3- and 6-Month Intervals. *Clin. Neuropsychol.* **26**, 239–254 (2012).

93.   Beilock, S. L. & DeCaro, M. S. From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *J. Exp. Psychol. Learn. Mem. Cogn.* **33**, 983–998 (2007).

94.   Piacente, T. Verbal comprehension of university students. *Orientación Soc.* **no. 12**, (2012).

95.   Helwig, R., Rozek-tedesco, M. A., Tindal, G., Heath, B. & Almond, P. J. Reading as an Access to Mathematics Problem Solving on Multiple-Choice Tests for Sixth-Grade Students. *J. Educ. Res.* **93**, 113–125 (1999).

96.   Simmons, F. R., Willis, C. & Adams, A.-M. Different components of working memory have different relationships with different mathematical skills. *J. Exp. Child Psychol.* **111**, 139–155 (2012).

97.   Chen, W.-C. & Whitehead, R. Understanding physics in relation to working memory. *Res. Sci. Technol. Educ.* **27**, 151–160 (2009).

98.   Staver, J. R. & Jacks, T. The influence of cognitive reasoning level, cognitive restructuring ability, disembedding ability, working memory capacity, and prior knowledge on students' performance on balancing equations by inspection. *J. Res. Sci. Teach.* **25**, 763–775 (1988).

99.   Cassels, J. R. T. & Johnstone, A. H. The effect of language on student performance on multiple choice tests in chemistry. *J. Chem. Educ.* **61**, 613 (1984).

100.   Daneman, M. & Hannon, B. Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT. *J. Exp. Psychol. Gen.* **130**, 208–223 (2001).

101.   Andreassen, R. & Bråten, I. Examining the prediction of reading comprehension on different multiple-choice tests. *J. Res. Read.* **33**, 263–283 (2010).

102.   Halloun, I. A. & Hestenes, D. Common sense concepts about motion. *Am. J. Phys.* **53**, 1056–1065 (1985).

103.   Neidorf, T., Arora, A., Erberber, E., Tsokodayi, Y. & Mai, T. Review of Research into Misconceptions and Misunderstandings in Physics and Mathematics. in *Student Misconceptions and Errors in Physics and Mathematics: Exploring Data from TIMSS and TIMSS Advanced* (eds. Neidorf, T., Arora, A., Erberber, E., Tsokodayi, Y. & Mai, T.) 11–20 (Springer International Publishing, 2020). doi:10.1007/978-3-030-30188-0_2.

104.   Eaton, P., Vavruska, K. & Willoughby, S. Exploring the preinstruction and postinstruction non-Newtonian world views as measured by the Force Concept Inventory. *Phys. Rev. Phys. Educ. Res.* **15**, 010123 (2019).

105.   Buffler, A., Lubben, F., Ibrahim, B. & Pillay, S. A model-based framework for understanding the role of visualization in physics education. in *Proceedings of the 16th Annual Meeting of the Southern African Association for Research in Mathematics, Science and Technology Education, Maseru, Lesotho* 435–441 (2008).

106.   Kozhevnikov, M., Hegarty, M. & Mayer, R. E. Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers. *Cogn. Instr.* **20**, 47–77 (2002).

107.   Kozhevnikov, M. & Thornton, R. Real-Time Data Display, Spatial Visualization Ability, and Learning Force and Motion Concepts. *J. Sci. Educ. Technol.* **15**, 111 (2006).

108.   Madsen, A. M., Larson, A. M., Loschky, L. C. & Rebello, N. S. Differences in visual attention between those who correctly and incorrectly answer physics problems. *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **8**, 010122 (2012).

109.   Pallrand, G. J. & Seeber, F. Spatial ability and achievement in introductory physics. *J. Res. Sci. Teach.* **21**,

507–516 (1984).

110. Poldrack, R. A. Is "efficiency" a useful concept in cognitive neuroscience? *Dev. Cogn. Neurosci.* **11**, 12–17 (2015).

111. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* **52**, 1059–1069 (2010).

112. Charles, M. & Bradley, K. Indulging Our Gendered Selves? Sex Segregation by Field of Study in 44 Countries. *Am. J. Sociol.* **114**, 924–976 (2009).

113. Baird, C. L. Male-dominated stem disciplines: How do we make them more attractive to women? *IEEE Instrum. Meas. Mag.* **21**, 4–14 (2018).

114. Richman, L. S., vanDellen, M. & Wood, W. How Women Cope: Being a Numerical Minority in a Male-Dominated Profession. *J. Soc. Issues* **67**, 492–509 (2011).

115. Barone, C. Some Things Never Change: Gender Segregation in Higher Education across Eight Nations and Three Decades. *Sociol. Educ.* **84**, 157–176 (2011).

116. Nosek, B. A., Banaji, M. R. & Greenwald, A. G. Math = male, me = female, therefore math not = me. *J. Pers. Soc. Psychol.* **83**, 44–59 (2002).

117. Spencer, S. J., Steele, C. M. & Quinn, D. M. Stereotype Threat and Women's Math Performance. *J. Exp. Soc. Psychol.* **35**, 4–28 (1999).

118. Keller, J. & Dauenheimer, D. Stereotype Threat in the Classroom: Dejection Mediates the Disrupting Threat Effect on Women's Math Performance. *Pers. Soc. Psychol. Bull.* **29**, 371–381 (2003).

119. Keller, J. Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *Br. J. Educ. Psychol.* **77**, 323–338 (2007).

120. Zhao, W. *et al.* Individual Differences in Cognitive Performance Are Better Predicted by Global Rather Than Localized BOLD Activity Patterns Across the Cortex. *Cereb. Cortex* doi:10.1093/cercor/bhaa290.

121. Dunst, B., Benedek, M., Koschutnig, K., Jauk, E. & Neubauer, A. C. Sex differences in the IQ-white matter microstructure relationship: A DTI study. *Brain Cogn.* **91**, 71–78 (2014).

122. Haier, R. J., Jung, R. E., Yeo, R. A., Head, K. & Alkire, M. T. The neuroanatomy of general intelligence: sex

matters. *NeuroImage* **25**, 320–327 (2005).

123. Jiang, R. *et al.* Gender Differences in Connectome-based Predictions of Individualized Intelligence Quotient and Sub-domain Scores. *Cereb. Cortex* (2019) doi:10.1093/cercor/bhz134.

124. Ryman, S. G. *et al.* Fronto-Parietal gray matter and white matter efficiency differentially predict intelligence in males and females. *Hum. Brain Mapp.* **37**, 4006–4016 (2016).

125. Satterthwaite, T. D. *et al.* Linked Sex Differences in Cognition and Functional Connectivity in Youth. *Cereb. Cortex* **25**, 2383–2394 (2015).

126. Schmithorst, V. J. Developmental sex differences in the relation of neuroanatomical connectivity to intelligence. *Intelligence* **37**, 164–173 (2009).

127. Schmithorst, V. J. & Holland, S. K. Sex differences in the development of neuroanatomical functional connectivity underlying intelligence found using Bayesian connectivity analysis. *NeuroImage* **35**, 406–419 (2007).

128. Wu, K. *et al.* Topological Organization of Functional Brain Networks in Healthy Children: Differences in Relation to Age, Sex, and Intelligence. *PLOS ONE* **8**, e55347 (2013).

129. Troyanskaya, O. *et al.* Missing value estimation methods for DNA microarrays. *Bioinformatics* **17**, 520–525 (2001).

130. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *NeuroImage* **62**, 782–790 (2012).

131. Andersson, J. L. R., Jenkinson, M. & Smith, S. *Non-linear registration, aka spatial normalization*.

132. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* **17**, 825–841 (2002).

133. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* **20**, 45–57 (2001).

134. Siegel, J. S. *et al.* Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. *Hum. Brain Mapp.* **35**, 1981–1996 (2014).

135. Shen, X., Tokoglu, F., Papademetris, X. & Constable, R. T. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage* **82**, 403–415 (2013).

136. Craddock, R. C., James, G. A., Holtzheimer, P. E., Hu, X. P. & Mayberg, H. S. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum. Brain Mapp.* **33**, 1914–1928 (2012).

137. Power, J. D. *et al.* Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage* **84**, (2014).

138. Abraham, A. *et al.* Machine learning for neuroimaging with scikit-learn. *Front. Neuroinformatics* **8**, 14 (2014).

139. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

140. Lynall, M.-E. *et al.* Functional Connectivity and Brain Networks in Schizophrenia. *J. Neurosci.* **30**, 9477–9487 (2010).

141. Ginestet, C. E. & Simmons, A. Statistical parametric network analysis of functional connectivity dynamics during a working memory task. *NeuroImage* **55**, 688–704 (2011).

142. Gong, G. *et al.* Mapping Anatomical Connectivity Patterns of Human Cerebral Cortex Using In Vivo Diffusion Tensor Imaging Tractography. *Cereb. Cortex* **19**, 524–536 (2009).

143. Hayasaka, S. & Laurienti, P. J. Comparison of characteristics between region-and voxel-based network analyses in resting-state fMRI data. *NeuroImage* **50**, 499–508 (2010).

144. He, Y., Chen, Z. J. & Evans, A. C. Small-World Anatomical Networks in the Human Brain Revealed by Cortical Thickness from MRI. *Cereb. Cortex* **17**, 2407–2419 (2007).

145. Bassett, D. S., Meyer-Lindenberg, A., Achard, S., Duke, T. & Bullmore, E. Adaptive reconfiguration of fractal small-world human brain functional networks. *Proc. Natl. Acad. Sci.* **103**, 19518 (2006).

146. Bassett, D. S., Nelson, B. G., Mueller, B. A., Camchong, J. & Lim, K. O. Altered resting state complexity in schizophrenia. *NeuroImage* **59**, 2196–2207 (2012).

147. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).

148. McKinney, W. Data Structures for Statistical Computing in Python. in *Proceedings of the 9th Python in Science Conference* (eds. Walt, S. van der & Millman, J.) 56–61 (2010). doi:10.25080/Majora-92bf1922-00a.

149. Millman, K. J. & Aivazis, M. Python for Scientists and Engineers. *Comput. Sci. Eng.* **13**, 9–12 (2011).

150. Oliphant, T. E. Python for Scientific Computing. *Comput. Sci. Eng.* **9**, 10–20 (2007).

151. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

152. Perez, F. & Granger, B. E. IPython: A System for Interactive Scientific Computing. *Comput. Sci. Eng.* **9**, 21–29

(2007).

153. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

154. Anderson, M. J. & Robinson, J. Permutation Tests for Linear Models. *Aust. N. Z. J. Stat.* **43**, 75–88 (2001).

155. Freedman, D. & Lane, D. A Nonstochastic Interpretation of Reported Significance Levels. *J. Bus. Econ. Stat.* **1**, 292 (1983).

156. Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M. & Nichols, T. E. Permutation inference for the general linear model. *NeuroImage* **92**, 381–397 (2014).

157. Rosseel, Y. lavaan: An R Package for Structural Equation Modeling. *J. Stat. Softw.* **48**, 1–36 (2012).