

DISCUSSION PAPER SERIES

IZA DP No. 12282

**Bullshitters. Who Are They and What Do
We Know about Their Lives?**

John Jerrim
Phil Parker
Nikki Shure

APRIL 2019

DISCUSSION PAPER SERIES

IZA DP No. 12282

Bullshitters. Who Are They and What Do We Know about Their Lives?

John Jerrim

UCL

Phil Parker

Australian Catholic University

Nikki Shure

UCL and IZA

APRIL 2019

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Bullshitters. Who Are They and What Do We Know about Their Lives?

'Bullshitters' are individuals who claim knowledge or expertise in an area where they actually have little experience or skill. Despite this being a well-known and widespread social phenomenon, relatively few large-scale empirical studies have been conducted into this issue. This paper attempts to fill this gap in the literature by examining teenagers' propensity to claim expertise in three mathematics constructs that do not really exist. Using Programme for International Student Assessment (PISA) data from nine Anglophone countries and over 40,000 young people, we find substantial differences in young people's tendency to bullshit across countries, genders and socio-economic groups. Bullshitters are also found to exhibit high levels of overconfidence and believe they work hard, persevere at tasks, and are popular amongst their peers. Together this provides important new insight into who bullshitters are and the type of survey responses that they provide.

JEL Classification: I24, J16

Keywords: PISA, overclaiming, bullshit

Corresponding author:

Nikki Shure
Department of Social Science
UCL Institute of Education
University College London
20 Bedford Way
London, WC1H 0AL
United Kingdom
E-mail: nikki.shure@ucl.ac.uk

1. Introduction

In his seminal essay-turned-book *On Bullshit*, Frankfurt (2005) defines and discusses the seemingly omnipresent cultural phenomenon of bullshit. He begins by stating that “*One of the most salient features of our culture is that there is so much bullshit. Everyone knows this. Each of us contributes his share*” (Frankfurt, 2005: 1). His book spent weeks on the *New York Times*’ bestsellers list in 2005 and has recently been cited in the post-truth age to better understand Donald Trump (e.g. Jeffries, 2017; Heer, 2018; Yglesias, 2018).

Other philosophers have since expanded on his work, most notably G. A. Cohen in his essay “*Deeper into Bullshit*” (Cohen 2002), but there has been limited large scale empirical research into this issue. We fill this important gap in the literature by providing new cross-national evidence on who is more likely to bullshit and how these individuals view their abilities and social status. This is an important first step in better understanding the seemingly ubiquitous phenomenon of bullshit.

We make use of an influential cross-national education survey administered every three years by the Organisation for Economic Cooperation and Development (OECD), namely the Programme for International Student Assessment (PISA). This data is commonly used by the OECD and education researchers to benchmark education systems or the performance of specific subgroups of pupils (e.g. Anderson et al., 2007; Jerrim and Choi, 2014; Parker et al., 2018), but has never been used to compare participants across countries in terms of their proclivity to bullshit. This paper fills this important gap in the literature.

Previous academic work on bullshit has been limited and mostly theoretical. Black (1983) edited a collection of essays on “*humbug*”, the predecessor of bullshit, which he defines as “*deceptive misrepresentation, short of lying, especially by pretentious word or deed, of somebody's own thoughts, feelings or attitudes*” (Black, 1983: 23). Frankfurt (2005) is the first theoretical treatment of the concept of “*bullshit*” and he situates it in terms of previous philosophical traditions. A crucial aspect of bullshitting in Frankfurt’s work is the fact that bullshitters have no concern for the truth, which is different than a purposeful lie (Frankfurt, 2005: 54). Cohen responds to Frankfurt’s essay and focuses on a slightly different definition of bullshit where “*the character of the process that produces bullshit is immaterial*” (Cohen, 2002: 2).

Petrocelli (2018) is one of the few studies to explore bullshitting empirically. He looks at the “*antecedents of bullshit*”, namely: topic knowledge, the obligation to provide an opinion

hypothesis (i.e. individuals are more likely to bullshit when they feel social pressure to provide a response) and the “ease of passing bullshit hypothesis” (i.e. people are more willing to bullshit when believe they will get away with it). He finds that participants are more likely to bullshit when there is pressure to provide an opinion, irrespective of their actual level of knowledge. Petrocelli also concludes that individuals are more likely to bullshit when they believe they can get away with it, and less likely to bullshit when they know they will be held accountable for the responses they provide (Petrocelli, 2018). His work uses smaller sample sizes than our work ($N \approx 500$) and does not answer the question of who bullshitters are and how they view their abilities or social standing.

Pennycook et al. (2015) is the only other empirical study focused on bullshit. They present experiment participants with “pseudo-profound bullshit” - vacuous statements constructed out of buzzwords - to ascertain when they can differentiate bullshit from meaningful statements and create a Bullshit Receptivity (BSR) scale. Their results point to the idea that some people may be more receptive towards pseudo-profound bullshit, especially if they have a more intuitive cognitive style or believe in the supernatural (Pennycook et al., 2015). Their study focuses on ability to detect bullshit and the mechanisms behind why some people cannot detect bullshit, rather than proclivity to bullshit, which is the focus of this paper.

In psychology, there has been a related literature on overconfidence and overclaiming. More and Healy (2008) provide a thorough overview of existing studies on overconfidence and distinguish between “overestimation”, “overplacement”, and “overprecision” as three distinct types of overconfidence. Overestimation occurs when individuals rate their ability as higher than it is actually observed to be, overplacement occurs when individuals rate themselves relatively higher than their actual position in a distribution, and overprecision occurs when individuals assign narrow confidence intervals to an incorrect answer, indicating overconfidence in their ability to answer questions correctly (More and Healy, 2008). The type of questions we use to construct our bullshit scale are closely related to overestimation and overprecision since the individuals need to not only identify whether or not they are familiar with a mathematical concept, but also assess their degree of familiarity.

Similar to how we define bullshit, overclaiming occurs when individuals assert that they have knowledge of a concept that does not exist. In one of the first studies on overclaiming, Philips and Clancy (1972) create an index of overclaiming based on how often individuals report consuming a series of new books, television programmes, and movies, all of which were not

real products. They use this index to explore the role of social desirability in survey responses. Stanovich and Cunningham (1992) also construct a scale of overclaiming using foils, fake concepts mixed into a list of real concepts, and signal-detection logic for authors and magazines to examine author familiarity. In both of these studies, however, the focus is not on the actual overclaiming index. Randall and Fernandes (1991) also construct an overclaiming index, but use it as a control variable in their analysis of self-reported ethical conduct.

Paulhus, Harms, Bruce, and Lysy (2003) focus more directly on overclaiming. They construct an overclaiming index using a set of items, of which one-fifth are non-existent, and employ a signal-detection formula to measure overclaiming and actual knowledge. They find that overclaiming is an operationalisation of self-enhancement and that narcissists are more likely to overclaim than non-narcissists (Paulhus et al., 2003). Atir, Rosenzweig, and Dunning (2015) find that people who perceive their expertise in various domains favourably are more likely to overclaim. Pennycock and Rand (2018) find that overclaimers perceive fake news to be more accurate. Similar to Atir et al. (2015), we find that young people who score higher on our bullshit index also have high levels of confidence in their mathematics self-efficacy and problem-solving skills.

We contribute to the existing literature on the related issues of bullshitting, overconfidence and overclaiming in three important ways. First, we use a large sample of 40,550 young people from nine Anglophone countries to examine bullshit, which enables us to dig deeper into the differences between subgroups (e.g. boys versus girls, advantaged vs. disadvantaged young people). Second, we provide the first internationally comparable evidence on bullshitting. We use confirmatory factor analysis to construct our scale and test for three hierarchical levels of measurement invariance (configural, metric and scalar). This allows us to compare average scores on our bullshit scale across countries in a robust and meaningful way. Finally, we also examine the relationship between bullshitting and various other psychological traits, including overconfidence, self-perceptions of popularity amongst peers and their reported levels of perseverance. Unlike many previous studies, we are able to investigate differences between bullshitters and non-bullshitters conditional upon a range of potential confounding characteristics (including a high-quality measure of educational achievement) providing stronger evidence that bullshitting really is independently related to these important psychological traits.

Our findings support the view that young men are, on average, bigger bullshitters than young women, and that socio-economically advantaged teenagers are more likely to be bullshitters than their disadvantaged peers. There is also important cross-national variation, with young people in North America more likely to make exaggerated claims about their knowledge and abilities than those from Europe. Finally, we illustrate how bullshitters display overconfidence in their skills, and are more likely to report that they work hard when challenged and are popular at school than other young people.

The paper now proceeds as follows. Section 2 provides an overview of the Programme for International Student Assessment (PISA) 2012 data and our empirical methodology. This is accompanied by Appendix A, where we discuss how we test for measurement invariance of the latent bullshit scale across groups. Results are then presented in section 3, with discussion and conclusions following in section 4.

2. Data

Throughout this paper we focus upon Programme for International Student Assessment (PISA) data collected in 2012. Although around 70 countries participated, we focus upon nine countries where English is the most commonly spoken language in order to minimise concerns about translation of survey items and hence comparability.¹ A multi-stage survey design was used, with schools first divided into a series of strata and then randomly sampled with probability proportional to size. From within each school, a sample of around 30 15-year-olds were randomly selected to participate. A total of 2,689 schools and 62,969 pupils took part in the study from across our nine Anglophone countries, reflecting official response rates of around 80 percent. In all nine countries, the sample was therefore fully compliant with the strict standards set by the OECD. Final student weights are applied throughout our analysis, with each country being given equal weight. Likewise, standard errors are clustered at the school level in order to take the complex PISA survey design into account.

PISA is primarily designed to measure the mathematics, science and reading skills of 15-year-olds across countries via a two-hour achievement test. However, participants also complete a 30-minute questionnaire that gathers information on young people's demographic background and their knowledge, attitudes and experience of subjects they study at school. Mathematics was the focus of PISA 2012, with most test and questionnaire items centred around this subject.

¹ Our identification of bullshitters relies upon participants' responses to some 'fake' questions, as shall be discussed below. We are concerned about how well these fake constructs translate to languages outside of English, and hence focus upon the nine Anglophone countries included within the sample.

Another somewhat unusual feature of PISA in 2012 was that young people were randomly assigned to complete one out of three different versions of the background questionnaire. Throughout this paper, we restrict our analysis to the random sub-sample of 40,550² young people from Anglophone countries who completed either form A or form C, which included the following question:

‘Thinking about mathematical concepts: how familiar are you with the following terms?’

A list of 16 items were then given to students, who were asked to indicate their knowledge of that particular mathematics concept on a five-point scale (ranging from *‘never heard of it’* to *‘know it well, understand the concept’*). These constructs were:

1. Exponential function
2. Divisor
3. Quadratic function
4. Proper number
5. Linear equation
6. Vectors
7. Complex number
8. Rational number
9. Radicals
10. Subjunctive scaling
11. Polygon
12. Declarative fraction
13. Congruent figure
14. Cosine
15. Arithmetic mean
16. Probability

Critically, of these 16 constructs, three of them (items 4, 10 and 12) are fake; students are asked about their familiarity with some mathematics concepts that do not exist. We use participants responses to these three items to form our ‘bullshit’ scale. This is done via estimation of a Confirmatory Factor Analysis (CFA) model, with the three fake items treated as observed

² Sample sizes by country are as follows: Australia 9,246; Canada 13,901; England 2,685; Ireland 3,267; Northern Ireland 1,430; New Zealand 2,762; Scotland 1,901; USA 3,193; Wales 2,165. Although sample sizes differ, senate weights are applied when data is pooled across countries to ensure each nation receives equal weight.

indicators of the latent bullshit construct. These MGCFA models are fitted using Mplus (Muthén and Muthén, 1998-2017), with the final student weights applied and standard errors clustered at the school level. A WLSMV estimator with THETA parameterisation³ was used to account for the ordered categorical nature of the questions (Muthén et al. 2015).

As our aim is to compare average scores on this scale, it is important that we investigate whether the latent construct is consistently understood and measured in the same way across demographic groups. We therefore follow standard practise and test for three hierarchical levels of measurement invariance (configural, metric and scalar) within each country according to the following demographic characteristics:

- Gender
- Socio-economic status
- Mathematics achievement quartile
- Immigrant status

Further details around this methodology and the measurement invariance results can be found in Appendix A. Bullshit scale scores are then derived from these MGCFA models, with average scores then compared across groups where full or partial scalar measurement invariance holds. The bullshit scale has been standardised to mean zero and standard deviation one within each country, so that all differences between groups can be interpreted in terms of an effect size.

Measures of self-efficacy

Within our analysis, we consider whether young people who score highly on the bullshit scale also display a series of other psychological characteristics, the first of which is overconfidence as measured by their mathematics self-efficacy. Specifically, as part of the PISA background questionnaire, participants were asked how confident they are in being able to complete the following eight tasks, according to a four-point scale (ranging from very confident to not confident at all):

Task 1. Using a train timetable to work out how long it would take to get from one place to another.

³ WLSMV is an estimator which is suitable for categorical variables. It performs a probit regression using a robust weighted least squares estimator with a diagonal weight matrix. THETA parameterisation allows the residual variances of the latent trait to be parameters in the model, while excluding scale factors (Muthén & Muthén, 1998-2017).

Task 2. Calculating how much cheaper a TV would be after a 30% discount.

Task 3. Calculating how many square metres of tiles you need to cover a floor.

Task 4. Understanding graphs presented in newspapers.

Task 5. Solving an equation like $3x+5 = 17$.

Task 6. Finding the actual distance between two places on a map with a 1:10,000 scale.

Task 7. Solving an equation like $2(x+3) = (x+3)(x-3)$.

Task 8. Calculating the petrol consumption rate of a car.

Throughout our analysis we dichotomise teenagers' responses, so that we compare the percentage of young people who said they are confident/very confident to the percentage who said they were not confident/not confident at all. The survey organisers have also created a mathematics 'self-efficacy' scale, combining young people's responses to these eight items into a single continuous index. We standardise this scale so that the mean in each country is zero and the standard deviation one. Further details on how we use such measures within our analyses is provided in section 3 below.

Self-belief in problem-solving abilities

Students were asked to indicate how well they believe the following five statements describes them and their problem-solving ability:

1. I can handle a lot of information.
2. I am quick to understand things.
3. I seek explanations for things.
4. I can easily link facts together.
5. I like to solve complex problems.

Throughout our analysis, we consider responses to these five items, focusing upon whether respondents said that the statement was very much/mostly like me (coded one) or if they indicated that it was somewhat/not much/not like me at all (coded zero). A total scale score has also been derived by the survey organisers, which we again standardise to mean zero and standard deviation one.

Self-reported popularity at school

To capture teenagers' self-reported views on their popularity at school, they were asked '*thinking about your school, to what extent do you agree with the following statements*':

1. I feel like an outsider (or left out of things) at school.
2. I make friends easily at school.
3. I feel like I belong at school.
4. I feel awkward and out of place in my school.
5. Other students seem to like me.
6. I feel lonely at school.
7. I feel happy at school.
8. Things are ideal in my school.
9. I am satisfied with my school.

Responses were to be given on a four-point scale, with our analysis of individual questions combining the strongly agree/agree categories and the disagree/strongly disagree categories into a binary scale.

Self-reported measures of perseverance

A series of five items were used in the background questionnaire to capture teenagers' self-reported perseverance with challenging tasks. Specifically, they were asked '*how well does each of the following statements below describe you*' with responses given on a five-point scale (very much like me, mostly like me, somewhat like me, not much like me, not at all like me):

1. When confronted with a problem, I give up easily.
2. I put off difficult problems.
3. I remain interested in the tasks that I start.
4. I continue working on tasks until everything is perfect.
5. When confronted with a problem, I do more than what is expected of me.

We again recode responses to these questions into a binary format, with very/mostly like me coded as one and zero otherwise. An overall scale combining information from all five items has also been derived and standardised to mean zero and standard deviation one.

Problem-solving approaches

As part of the background questionnaire, two hypothetical scenarios were set out to students, who were then asked how they would respond. The first scenario asked:

‘Suppose that you have been sending text messages from your mobile phone for several weeks. Today, however, you can’t send text messages. You want to try and solve the problem. What would you do?’

1. I press every button possible to find out what is wrong.
2. I think about what might have caused the problem and what I can do to solve it.
3. I read the manual.
4. I ask a friend for help.

Students were asked whether they would (a) definitely do this; (b) probably do this; (c) probably not do this or (d) would definitely do this, for each of the four statements above. We combine options (a) with (b) and (c) with (d), allowing us to compare young people who said they would probably/definitely use each strategy versus those who would definitely/probably would not.

The second scenario followed a similar structure, with participants asked:

‘Suppose that you are planning a trip to the zoo with your brother. You don’t know which route to take to get there. What would you do?’

1. I read the zoo brochure to see if it says how to get there.
2. I study a map and work out the best route.
3. I leave it to my brother to worry about how to get there.
4. I know roughly where it is, so I suggest we just start driving.

Participants were provided the same four response options (a to d) as per scenario 1, which we also convert into a binary format as described above.

Analytic models

After comparing average bullshit scale scores across demographic groups, and across countries, we investigate the self-reported self-efficacy, problem-solving skills, perseverance and popularity of bullshitters. To begin, we divide participants into four approximately equal groups (quartiles) based upon their scores on the bullshit scale. Those in the bottom quartile are then labelled ‘non-bullshitters’ (i.e. those young people who overwhelmingly said that they had not heard of the fake mathematics constructs) with the top quartile defined as the bullshitters (i.e. young people who claimed expertise in the fake constructs). Then, for these two groups, we compare how they responded to each of the self-efficacy, problem-solving, popularity and perseverance questions (and overall scale scores) described above.

A limitation with such summary statistics is that there could be confounding characteristics driving the results. For instance, with respect to self-efficacy, it will be particularly important to consider whether bullshitters are much more likely to believe that they can complete each of the eight mathematics tasks than non-bullshitters *after* conditioning upon their actual measured academic ability. In other words, do teenagers who bullshit about their mathematics knowledge also display overconfidence in their mathematics skills? For each of the outcome measures described in section 2, we therefore estimate the following OLS regression model within each country⁴:

$$O_{ij} = \alpha + \beta.BS + \gamma.A + \delta.SES + \tau.D + u_j + \varepsilon_{ij}$$

Where:

O_{ij} = The outcome variable of interest (e.g. teenagers' self-efficacy).

BS = A set of dummy variables reflecting quartiles of the bullshit scale.

A = Teenagers' academic achievement in mathematics, reading, science and problem solving, as measured by the PISA test⁵.

SES = Teenagers' socio-economic status, as measured by the PISA Economic, Social and Cultural Status (ESCS) index.

D = A vector of controls for teenagers' demographic characteristics (e.g. gender and immigrant status).

u_j = School fixed effects.

i = Student i.

j = School j.

The parameters of interest from these models are the estimated β coefficients. These will reveal differences between the bullshitter and non-bullshitter groups, conditional upon their gender, socio-economic status, mathematics, reading, science and problem-solving skills and attendance within the same school. Such conditional associations will help reveal whether bullshitters provide different answers to the self-efficacy, perseverance, popularity and

⁴ As we have dichotomised participants' responses, this is equivalent to estimating a linear probability model for each item, along with a standard OLS model for the overall scale score.

⁵ We include controls for all PISA plausible values in each subject area.

problem-solving questions, compared to non-bullshitters of the same demographic background, of equal academic ability and within the same school.

3. Results

Who are the bullshitters?

Table 1 considers how average scores on the bullshit scale differ between demographic groups. There is an important difference between genders; boys are much more likely to be bullshitters than girls. This holds true across all nine countries, with all differences statistically significant and equivalent to an effect size of between 0.4 and 0.5 standard deviations in most countries. It is also notable how the gender gap in bullshitting is significantly weaker in North America (0.25 in the United States and 0.34 in Canada) than it is in Europe (e.g. gender gaps of between 0.4 and 0.5 are observed for England, Ireland, Scotland and Wales). Consequently, Table 1 provides strong and consistent evidence that teenage boys are bigger bullshitters than teenage girls.

<< Table 1 >>

A similar difference is found with respect to socio-economic status; young people from more advantaged socio-economic backgrounds have higher average bullshit scores than their less advantaged peers. The magnitude of the difference is again not trivial and varies somewhat across countries. For instance, the difference in average bullshit scores between the top and bottom socio-economic quartile stands above 0.6 standard deviations in Scotland and New Zealand, but below 0.3 in England, Canada and the United States. Nevertheless, in all nine countries, the difference is statistically significant at the five percent level. These results therefore provide strong evidence that young people from more affluent backgrounds are more likely to be bullshitters than young people from disadvantaged backgrounds.

The final difference considered in Table 1 is between immigrant and native groups. In most countries, immigrants having significantly higher scores than young people who are country natives. This is particularly pronounced in European Anglophone countries, where immigrants typically score around 0.35 standard deviations higher on the bullshit scale than young people who were born in the country. The association is typically slightly weaker outside of Europe, with there actually being no difference between immigrants and natives in the United States. Hence, although we find a general pattern of immigrants being bigger bullshitters than natives, the strength of this association seems to vary quite substantially between countries (and, thus, characteristics and home locations of the immigrant groups).

Finally, in additional analysis, we have also estimated the within versus between school variation of the bullshit scale within each country. Our motivation was to establish whether bullshitters tend to cluster together within the same school, or if bullshitters are fairly equally distributed across schools. We find that the ICCs tend to be very low; in most countries less than three percent of the variance in the bullshit scale occurs between schools. This perhaps helps to explain why everyone knows a bullshitter; these individuals seem to be relatively evenly spread across schools (and thus peer groups).

Are teenagers in some countries bigger bullshitters than others?

Table 2 provides our comparison of the bullshit scale across eight of the nine Anglophone countries.⁶ The top panel provides the average standardised scale score, while the bottom panel provides t-statistics for pairwise comparisons across countries. Green shading with an asterisk highlights where differences across countries are statistically significant at the five percent level.

<< Table 2 >>

Three broad clusters of countries seem to have emerged. At the top of the rankings are the two North American countries of the United States and Canada. With average scale scores of 0.25 and 0.3, these two countries have significantly higher bullshit scores than any other country. The next three countries (Australia, New Zealand and England) are in the middle of the rankings. Teenagers in these countries exaggerate less about their prowess, on average, than young people in Canada and the United States – by a magnitude equivalent to an effect size of around 0.1. However, they are also significantly bigger bullshitters than young people from Ireland, Northern Ireland and Scotland who form the final group. The average bullshit scale score in these countries ranges between approximately -0.26 (Ireland and Northern Ireland) and -0.43 (Scotland) which is significantly lower than every other country. Moreover, the difference between these countries and North American is sizeable; equivalent to an effect size greater than 0.5. Consequently, despite speaking the same language, and with a closely shared culture and history, we find important variation across Anglophone countries in teenagers' propensity to bullshit.

⁶ Wales has been excluded from this comparison; see Appendix A for further details.

A psychological profile of bullshitters

Table 3 now turns to how bullshitters responded to other items included in the PISA background questionnaire. These results are based upon the pooled sample including young people from across the Anglophone countries. The top panel refers to their self-confidence in completing the eight mathematics tasks described in section 2, while the bottom panel illustrates how they view their problem-solving abilities. Figures refer to the percentage of young people who believe they could complete the task relatively easily or who believe that they have each specific problem-solving skill. The final rows provide the average score for bullshitters and non-bullshitters on the self-efficacy and problem-solving scales. These can be interpreted in terms of an effect size. The raw difference in estimates between these groups are then reported, along with the regression model estimates that control for demographic background, prior academic achievement and school fixed-effects.

<< Table 3 >>

Starting with the results for self-efficacy, there are substantial and statistically significant differences between the high and low bullshit groups on the eight questions asked. For instance, whereas just 40 percent of non-bullshitters were confident that they could work out the petrol consumption of a car (task eight), two-thirds of the bullshitter group claimed that they could do this. Moreover, a sizeable difference can still be observed in the regression model results, illustrating how bullshitters express much higher levels of self-confidence in their skills than non-bullshitters, even when they are of equal academic ability. Specifically, the difference in the average self-efficacy scale score is approaching 0.5 standard deviations; a large and statistically significant effect. Together, these results illustrate how young people who tend to bullshit are also likely to express overconfidence in their skills.

The lower panel of Table 3 confirms these results. When asked about their problem-solving skills, bullshitters are around 20 percentage points more likely to say that they ‘*can handle a lot of information*’, ‘*can easily link facts together*’, ‘*are quick to understand things*’ and ‘*like to solve complex problems*’. Although controlling for achievement, demographics and school characteristics can explain some of the difference between the high and low bullshit groups, significant differences remain; we continue to observe an effect size difference of around one-quarter of a standard deviation, even after such characteristics have been controlled. This again demonstrates the overconfidence expressed by bullshitters.

In Table 4 we report results for young people's self-reported perseverance. Bullshitters are much less likely to say that they give up easily when faced with a difficult problem (eight versus 17 percent) and that they are put off by difficult problems (15 versus 27 percent). Yet they are more likely to say that they exceed expectations when faced with a difficult problem (45 versus 28 percent). In other words, bullshitters claim to have particularly high levels of perseverance when faced with challenging tasks. The difference in the average perseverance scale score between the high and low bullshit groups is 0.50 standard deviations (0.41 once controls have been added) representing a sizeable and statistically significant effect. Table 4 therefore illustrates how bullshitters claim to persevere more with hard-to-solve problems than other groups, independent of a range of background characteristics.

<< Table 4 >>

How do bullshitters claim that they solve problems? Table 5 provides some insight into this issue by summarising how they said they would solve two routine tasks (see section 2 for further details). Interestingly, the most pronounced and statistically significant results are with respect to the most 'socially desirable' (or the most 'obviously sensible') strategy. For instance, if their mobile phone stops sending text messages, bullshitters are somewhat less likely to say that they would press all the buttons to find out what is wrong (49 versus 56 percent) but much more likely to say that they would consult the instruction manual (41 versus 29 percent). Likewise, if they do not know the route to their destination, bullshitters are much more likely to say that they would consult a map than other groups (71 versus 55 percent). Although we do not know what strategy these young people would actually use, Table 5 nevertheless provides some indication that bullshitters are much more likely to say that they would take the most obviously sensible approach.

<< Table 5 >>

Finally, do bullshitters believe that they are popular at school? Table 6 provides some suggestion that this may be the case. The average 'school well-being' scale score is around 0.2 standard deviations higher for bullshitters, and stays at this level even after achievement, demographic and school controls have been added. There is a particularly notable difference in response to the question '*things are ideal at my school*', to which 75 percent of bullshitters agree (compared to 64 percent of non-bullshitters). Therefore, although the evidence is perhaps weaker than for the previous topics considered, we nevertheless find some evidence that

bullshitters are particularly likely to believe that they are popular at school (and certainly believe they are no less popular than their non-bullshitting peers).

<< Table 6 >>

Investigations of possible alternative explanations

There are two primary threats to the validity of our interpretation of the results above. The first alternative explanation is that, rather than capturing young people's propensity to bullshit, the three fake constructs provide evidence of a careless or extreme response style. For instance, some respondents may not be taking the questionnaire seriously, and are simply ticking the top category for every question. A second possibility is that young people's responses are reflecting social desirability bias; that they are providing responses that they believe will be viewed as positively by others (e.g. that they know various mathematics concepts, that they work hard at school etc). Both of these possibilities could lead to a spurious correlation between our bullshit index and the various other psychological traits investigated in the previous sub-section. Similarly, if children with certain characteristics (e.g. boys, immigrants, young people in particular countries) are more likely to provide careless or socially desirable responses, then this could explain why we observe differences between demographic groups.

To explore this possibility further, we investigate how our bullshit index is related to young people's responses to two other questions in the PISA background questionnaire: (a) test motivation and (b) truancy at school. Specifically, children were asked to provide the amount of effort they put into the PISA study using a zero to ten scale and how many times they were absent from school over the last two weeks. If respondents are indeed providing high responses consistently across questions – either due to carelessness, response style or social desirability – then we should observe a strong correlation between our bullshit index and young people's self-reported truancy and test motivation. These results are presented in Table 7.

<< Table 7 >>

We find no evidence that the bullshit index is related to young people's test motivation; the Pearson correlation is -0.03 while Table 7 highlights how the average test effort reported was 7.6 within each bullshit quartile. Similarly, panel (b) of Table 7 illustrates how the bullshit scale is not associated with self-reported truancy from school. Specifically, around 80 percent of young people said they were not absent from school at any point during the last two weeks, regardless of how they responded to the questions which form our bullshit scale.

Together, this provides us with reassurance that the correlations observed in the previous subsection are unlikely to be driven by social desirability bias or other forms of careless/extreme response.

4. Conclusions

Bullshitting is a well-known social phenomenon. It can be summarised as a situation where an individual claims to have knowledge, experience or expertise in some matter, when really they do not. The label “bullshitter” is then assigned to someone who makes such claims on a regular basis; i.e. a person who consistently exaggerates their prowess and/or frequently tells untruths. Although this concept is well-known in everyday life – we all probably know a bullshitter – very little academic research has been conducted into this issue. What, for instance, are the demographic characteristics of bullshitters? Is it a masculine or a feminine trait, and is it something that varies between socio-economic groups? Do young people in some countries tend to bullshit more than those in others? And what other psychological characteristics do bullshitters display; do they display overconfidence, have a tendency to provide socially desirable answers or have an inflated sense of popularity amongst their peers?

This paper has attempted to explore such issues using large-scale, nationally representative data. Focusing upon 15-year-olds from across nine Anglophone countries, we have investigated the characteristics of young people who claim to have knowledge and expertise in three mathematics concepts which are fake. Having derived and established the comparability of our bullshit scale via measurement invariance procedures, we go on to find that young men are more likely to bullshit than young women, and that bullshitting is somewhat more prevalent amongst those from more advantaged socioeconomic backgrounds. Compared to other countries, young people in North America are found to be bigger bullshitters than young people in England, Australia and New Zealand, while those in Ireland and Scotland are the least likely to exaggerate their mathematical knowledge and abilities. Strong evidence also emerges that bullshitters also display overconfidence in their academic prowess and problem-solving skills, while also reporting higher levels of perseverance when faced with challenges and providing more socially desirable responses than more truthful groups.

There are of course limitations to this study, and many issues on the topic of bullshitting that remain unexplored. First, the PISA data analysed are cross-sectional rather than longitudinal. We therefore do not know whether bullshitting is a stable trait that can be consistently observed for an individual over time, or if it is something that changes with age (and the factors

associated with such change). Likewise, the implications of being a bullshitter remain unclear. Although this concept often has negative connotations, being able to bullshit convincingly may be useful in certain situations (e.g. job interviews, negotiations, grant applications). Yet the social and labour market outcomes of bullshitters remains unknown and is thus a key issue in need of further research.

Second, our analysis has only considered the propensity to bullshit in a single area (knowledge of mathematics concepts). Future work should consider the overlap between bullshitting with respect to different areas of life – such as young people’s knowledge/experience of drug taking or of their sexual experiences (for instance). This will help us to identify those individuals who consistently lie about multiple aspects of their life. Finally, it is important we recognise that our bullshit scale was based upon three specific items. Ideally, future research should try to include a greater number of fake constructs in order to maximise precision of the bullshit scale.

Despite these limitations, we believe this paper has started to open an important new area of social science research. Bullshitting is a widely recognised social ‘skill’ which is likely to have an impact upon a person’s life. We have established how some groups are clearly more likely to bullshit (and be caught bullshitting) than others, and that these individuals tend to display certain other psychological traits (most notably a striking overconfidence in their own abilities). It is critical that a developmental perspective is now taken with respect to bullshitting so that we can understand what leads individuals to develop such habits, and whether it turns out to be associated with better or worse social and labour market outcomes.

References

- Anderson, J. O., Lin, H. S., Treagust, D. F., Ross, S. P., & Yore, L. D. (2007). Using large-scale assessment datasets for research in science and mathematics education: Programme for International Student Assessment (PISA). *International Journal of Science and Mathematics Education*, 5(4), 591-614.
- Atir, S., Rosenzweig, E., & Dunning, D. (2015). When knowledge knows no bounds: Self-perceived expertise predicts claims of impossible knowledge. *Psychological Science*, 26(8), 1295-1303.
- Black, M. (1985). *The prevalence of Humbug and other essays*. Cornell University Press.
- Chen, F. (2007). Sensitivity of goodness of fit indexes to lack measurement invariance. *Structural Equation Modeling*, 14, 464-504.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233-255.
- Cohen, G. A. (2002). Deeper into bullshit. *Contours of agency: Essays on themes from Harry Frankfurt*, 321-339.
- Frankfurt, H. G. (2005). *On bullshit*. Princeton University Press.
- Heer, J. (2018, March 15). Worse than a Liar. *The New Republic*. Retrieved from <https://newrepublic.com/article/147504/worse-liar-trump-lies-trudeau>
- Jeffries, S. (2017, May 22). 'Bullshit is a greater enemy than lies' – lessons from three new books on the post-truth era. *The Guardian*. Retrieved from <https://www.theguardian.com/us-news/2017/may/22/post-truth-era-trump-brexit-lies-books>
- Jerrim, J. & Choi, A. (2014). The mathematics skills of school children: How does the UK compare to the high performing East Asian nations? *Journal of Education Policy*, 29, 349–376.
- Kenny, D.; Kaniskan, B. and McCoach, D. (2015). 'The Performance of RMSEA in Models With Small Degrees of Freedom.' *Sociological Methods & Research* 44(3): 486-507.
- Möbius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2011). *Managing self-confidence: Theory and experimental evidence* (No. w17014). National Bureau of Economic Research.
- Muthén, B., Muthén, L., & Asparouhov, T. (2015). *Estimator choices with categorical variables*. Retrieved from <https://www.statmodel.com/download/EstimatorChoices.pdf>
- Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide*. Eight Edition. Los Angeles, CA: Muthén & Muthén.

- Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much?. *The Quarterly Journal of Economics*, 122(3), 1067-1101.
- Parker, P. D., Marsh, H. W., Jerrim, J. P., Guo, J., & Dicke, T. (2018). Inequity and Excellence in Academic Performance: Evidence From 27 Countries. *American Educational Research Journal*, 0002831218760213.
- Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of personality and social psychology*, 84(4), 890.
- Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision making*.
- Pennycook, G. & Rand, D. G. (2018). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking, mimeo.
- Petrocelli, J. V. (2018). Antecedents of bullshitting. *Journal of Experimental Social Psychology*, 76, 249-258.
- Putnick, D. and Bornstein, M. (2016). 'Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research.' *Developmental Review* 41: 71-90.
- Sarsons, H., & Xu, G. (2016). Confidence Men? Gender and Confidence: Evidence among Top Economists, mimeo.
- Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition*, 20(1), 51-68.
- Yglesias, M. (2017, May 30). The Bullshitter-in-Chief. *Vox*. Retrieved from <https://www.vox.com/policy-and-politics/2017/5/30/15631710/trump-bullshit>
- Zieger, L.; Jerrim, J. and Sims, S. (2018). 'Comparing teachers' job satisfaction across countries. A multiple-pairwise measurement invariance approach.' *Forthcoming*.

Table 1. The association between demographic characteristics and average scores on the bullshit scale

(a) Gender

	Girls	Boys	Gap (effect size)	SE
England	-0.23	0.24	0.48*	0.04
Ireland	-0.23	0.23	0.46*	0.04
Scotland	-0.23	0.21	0.44*	0.05
Australia	-0.21	0.21	0.42*	0.02
Wales	-0.21	0.21	0.42*	0.05
New Zealand	-0.20	0.20	0.40*	0.04
Northern Ireland	-0.18	0.17	0.35*	0.05
Canada	-0.17	0.17	0.34*	0.02
USA	-0.13	0.13	0.25*	0.04

(b) Socio-economic status

	Low SES	Q2	Q3	High SES	Gap (effect size)	SE
Scotland	-0.36	0.08	0.09	0.30	0.65*	0.06
New Zealand	-0.29	0.03	0.09	0.33	0.62*	0.06
Ireland	-0.21	0.07	-0.02	0.23	0.44*	0.06
Australia	-0.18	-0.12	0.02	0.25	0.42*	0.03
Wales	-0.17	-0.03	0.04	0.19	0.36*	0.06
England	-0.12	-0.09	0.02	0.17	0.29*	0.06
Canada	-0.13	-0.07	-0.05	0.15	0.28*	0.04
USA	-0.09	0.02	-0.04	0.11	0.20*	0.06

(c) Immigrant group

	Natives	Immigrants	Gap (effect size)	SE
Northern Ireland	-0.02	0.64	0.66*	0.15
Ireland	-0.04	0.34	0.38*	0.08
England	-0.05	0.32	0.37*	0.07
Wales	-0.01	0.35	0.36*	0.11
New Zealand	-0.09	0.26	0.36*	0.05
Scotland	-0.03	0.32	0.36*	0.08
Canada	-0.05	0.12	0.17*	0.03
Australia	-0.04	0.13	0.16*	0.03
USA	0.00	0.01	0.01	0.06

Notes: The bullshit scale has been standardised within each country to mean zero and standard deviation one. The gap refers to the difference between groups in terms of an effect size. SE refers to the standard error of the gap. Northern Ireland excluded from socio-economic status results due to factor scores not able to be calculated. * indicates statistical significance at the five percent level.

Table 2. International comparison of average bullshit scores across Anglophone countries.

(a) Average bullshit scale scores across

Country	Mean	Standard error
Canada	0.298	0.014
USA	0.252	0.023
Australia	0.179	0.010
New Zealand	0.135	0.022
England	0.093	0.021
Ireland	-0.255	0.019
Northern Ireland	-0.265	0.027
Scotland	-0.432	0.025

(b) T-statistics for pairwise country comparisons

	Canada	USA	Australia	New Zealand	England	Ireland	Northern Ireland	Scotland
Canada	-	-	-	-	-	-	-	-
USA	<i>1.72</i>	-	-	-	-	-	-	-
Australia	6.81*	2.93*	-	-	-	-	-	-
New Zealand	6.27*	3.73*	<i>1.83</i>	-	-	-	-	-
England	8.25*	5.22*	3.75*	<i>1.39</i>	-	-	-	-
Ireland	23.23*	17.08*	19.77*	13.32*	12.31*	-	-	-
Northern Ireland	18.26*	14.53*	15.09*	11.35*	10.40*	<i>0.29</i>	-	-
Scotland	25.71*	20.43*	22.75*	17.13*	16.29*	5.65*	4.53*	-

Notes: Average bullshit scale scores have been standardised to mean zero and standard deviation one across the eight Anglophone countries. Wales has been excluded based upon measurement invariance tests. Green shaded cells in panel b indicate where difference across countries is statistically significant (absolute value of the t-statistic is greater than 1.96). Red shaded cells with italic font illustrates where cross-country differences are not statistically significant (absolute value of the t-statistic is less than 1.96).

Table 3. Bullshitters' views of their abilities

	Unconditional			Regression results	
	Non-bullshitters	Bullshitters	Difference	Difference	SE
<u>Self-efficacy</u>					
Believe can could complete Task 1	80%	88%	7%	5%*	1.3%
Believe can could complete Task 2	74%	86%	12%	10%*	1.3%
Believe can could complete Task 3	58%	80%	22%	17%*	1.4%
Believe can could complete Task 4	79%	90%	11%	9%*	1.2%
Believe can could complete Task 5	79%	90%	11%	6%*	1.2%
Believe can could complete Task 6	39%	67%	28%	19%*	1.6%
Believe can could complete Task 7	63%	81%	18%	13%*	1.5%
Believe can could complete Task 8	40%	68%	28%	19%*	1.5%
Scale score (standardised)	-0.41	0.26	0.67	0.48*	0.03
<u>Views of problem-solving ability</u>					
Can handle a lot of information	43%	62%	18%	14%*	2%
Quick to understand things	45%	62%	17%	11%*	2%
Seek explanations for things	61%	66%	5%	2%	2%
Easily link facts together	49%	65%	17%	13%*	2%
Like to solve complex problems	24%	47%	23%	18%*	2%
Scale score (standardised)	-0.36	0.14	0.49	0.35*	0.03

Notes: Figures refer to percent of young people who agree or strongly agree. Non bullshitters refers to young people in the bottom quarter of the derived bullshit scale score distribution, while bullshitters are defined as the top quartile. Regression estimates refers to the difference between bullshitters and non-bullshitters controlling for gender, socio-economic status, immigrant status, PISA reading, maths and science scores and school fixed effects. The 'scale score' row refers to results based upon continuous index combining data across all items. This has been standardised to mean 0 and standard deviation 1, and can therefore be interpreted in terms of an effect size. * indicates that the difference between bullshitters and non-bullshitters is statistically significant at the five percent level. A full list of the self-efficacy tasks can be found in section 2.

Table 4. Bullshitters' views of their perseverance

	Unconditional (% agree)			Regression results	
	Non-bullshitters	Bullshitters	Difference	Difference	SE
When confronted with a problem, I give up easily	17%	8%	-9%	-5%*	1.2%
I put off difficult problem	27%	15%	-12%	-9%*	1.3%
I remain interested in the tasks that I start	46%	60%	14%	14%*	1.6%
I continue working on tasks until everything is perfect.	42%	55%	13%	10%*	1.7%
When confronted with a problem, I do more than what is expected of me	28%	45%	17%	15%*	1.7%
Scale score (standardised)	-0.37	0.13	0.50	0.41	0.03

Notes: See notes to Table 3. * indicates statistically significant difference between bullshitters and non-bullshitters at the five percent level.

Table 5. The bullshitter approach to problem solving

	Unconditional (% agree)			Regression results	
	Non-bullshitters	Bullshitters	Difference	Difference	SE
<u>Task 1.</u>					
I press every button possible to find out what is wrong	56%	49%	-7%	-3%	1.8%
I think about what might have caused the problem and what I can do to solve it	85%	90%	5%	3%*	1.1%
I read the manual	29%	41%	12%	10%*	1.7%
I ask a friend for help	79%	75%	-4%	-2%	1.4%
<u>Task 2.</u>					
I read the zoo brochure to see if it says how to get there	75%	73%	-2%	-1%	1.5%
I study a map and work out the best route	55%	71%	16%	9%*	1.6%
I leave it to my brother to worry about how to get there	34%	27%	-7%	-5%*	1.6%
I know roughly where it is, so I suggest we just start driving	64%	60%	-4%	1%	1.7%

Notes: See notes to Table 3. In task 1, participants were asked ‘*Suppose that you have been sending text messages from your mobile phone for several weeks. Today, however, you can’t send text messages. You want to try to solve the problem. Which of the following would you do?*’ In task 2, participants were asked ‘*Suppose that you are planning a trip to the zoo with your brother. You don’t know which route to take to get there. Which of the following would you do?*’ Figures refer to the percent of young people who said they would either ‘definitely’ or ‘probably’ use this problem-solving strategy. * indicates statistically significant difference between bullshitters and non-bullshitters at the five percent level.

Table 6. Do Bullshitters believe they are popular?

	Unconditional (% agree)			Regression results	
	Non-bullshitters	Bullshitters	Difference	Difference	SE
Left out of things at school	13%	13%	0%	-2%	1.2%
make friends easily at school	85%	89%	4%	5%*	1.3%
feel like I belong at school.	74%	81%	7%	7%*	1.5%
feel awkward/ out of place in my school	15%	16%	1%	1%	1.4%
Other students seem to like me	92%	93%	1%	2%	1.0%
feel lonely at school.	10%	10%	1%	0%	1.1%
feel happy at school.	77%	84%	7%	8%*	1.4%
Things are ideal in my school.	64%	75%	10%	11%*	1.7%
I am satisfied with my school.	76%	83%	8%	9%*	1.5%
Scale score (standardised)	-0.11	0.07	0.18	0.20*	0.04

Notes: See notes to Table 3. * indicates statistically significant difference between bullshitters and non-bullshitters at the five percent level.

Table 7. The association between the bullshit scale and young people’s test motivation and truancy from school

(a) Test motivation

	Average	Standard error
Bottom quartile	7.60	0.03
Second quartile	7.63	0.03
Third quartile	7.63	0.03
Top quartile	7.64	0.03

(b) Truancy from school

	Bullshit index			
	Bottom quartile	Second quartile	Third quartile	Top quartile
None	80%	80%	81%	80%
One or two times	16%	16%	16%	16%
Three or four times	3%	3%	2%	3%
Five or more times	1%	1%	1%	1%
Total	100%	100%	100%	100%

Notes: Figures in panel (a) refer to the average amount of effort children say that they put into the PISA test out of 10. Panel (b) provides column percentages; it refers to the number of times young people said that they skipped school for a whole day over the last two weeks.

Appendix A. Measurement invariance tests

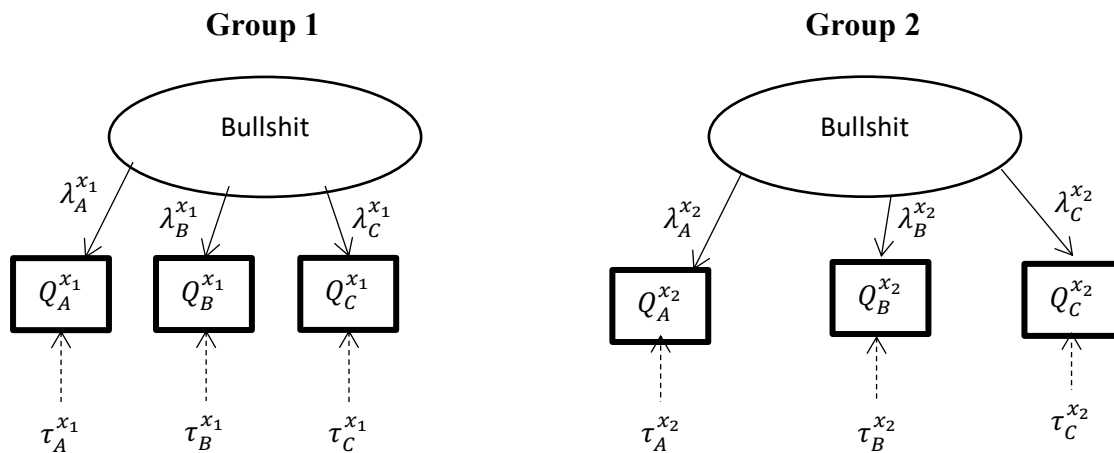
Measurement invariance methods

Teenagers' responses to the three fake mathematics concepts are used to derive the bullshit scale. This is done via estimation of a Confirmatory Factor Analysis (CFA) model, with the three fake items treated as observed indicators of the latent bullshit construct. As our aim is to compare average scores on this scale, it is important that we investigate whether the latent construct is consistently understood and measured in the same way across demographic groups. We therefore follow standard practise and test for three hierarchical levels of measurement invariance (configural, metric and scalar) within each country according to the following demographic characteristics:

- Gender
- Socio-economic status
- Mathematics achievement quartile
- Immigrant status

The intuition behind this approach, with reference to the bullshit scale, is presented in Figure 1. Ovals depict the unobserved latent construct we are trying to measure, while rectangles refer to young people's observed responses to the three fake mathematics items. Specifically, Q_w^x represents a single question w in country x . Factor loadings are represented by λ_w^x , and quantify the strength of the relationship between the latent bullshit trait and question w , in country x . On the other hand, τ_w^x is the 'threshold' and is essentially equivalent to the constant term in a regression model.

Figure 1. A hypothetical example of the MGCFA model to test invariance of the bullshit scale across groups.



The factor loadings (λ_w^x) and thresholds (τ_w^x) are the main properties of the bullshit scale, and the key parameters used to test for ‘measurement invariance’ (i.e. comparability of the bullshit scale) across groups. Basically, measurement invariance involves putting ever more constraints upon the factor loadings (λ_w^x) and thresholds (τ_w^x), to test whether three hierarchical levels of invariance hold. These are:

- Configural invariance (level 1). This requires the same set of questions to be associated with the latent trait across all groups. With respect to Figure 1, if the loadings λ_A^1, λ_B^1 and λ_C^1 are all unequal to zero in group 1 (e.g. boys), they should also be unequal to zero in group 2 (e.g. girls).
- Metric invariance (level 2). This assumes that the factor loadings (λ) are equal across groups. In terms of Figure 1, this means that $\lambda_A^{x_1} = \lambda_A^{x_2}, \lambda_B^{x_1} = \lambda_B^{x_2}$ and $\lambda_C^{x_1} = \lambda_C^{x_2}$. If this level of invariance is established, then it is widely accepted that one can use the bullshit scale as an independent variable in a regression model, and that the estimated parameters can be fairly compared across groups.
- Scalar invariance (level 3). This level of invariance is required for average bullshit scale scores to be fairly compared across groups. This imposes the additional assumption that the thresholds (τ) we wish to compare are also equal. Again returning to Figure 1, we now also need $\tau_A^{x_1} = \tau_A^{x_2}, \tau_B^{x_1} = \tau_B^{x_2}$ and $\tau_C^{x_1} = \tau_C^{x_2}$.

Following standard practise in the literature (Putnick and Bornstein 2016), we determine which level of measurement invariance holds through the use of ‘fit indices’. Although a simple

χ^2 test is sometimes used to decide between such nested model, this is highly sensitive to sample size (Chen, 2007; Cheung and Rensvold, 2002). Given our large sample sizes (several thousand observations in all countries) we focus upon three alternative fit indices instead. These are the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI) and the Root Mean Square Error of approximation. The intuition behind such indices is that they help determine how much worse the MGCFA model fits the data when additional equality constraints are imposed between groups. Zieger, Jerrim and Sims (2018) provide further details. We use the following widely used rules of thumb when deciding which level of measurement invariance holds according to these indices (Chen 2007):

- Configural. $CFI \geq 0.95$, $TLI \geq 0.95$.
- Metric. Decrease in the CFI and TLI versus configural model of less than 0.01.
- Scalar. Decrease in the CFI and TLI versus metric model of less than 0.01. Increase in the RMSEA less than 0.01.

Note that we only use the RMSEA to distinguish between metric and scalar invariance holds. This is due to the configural and metric models having low degrees of freedom, with the RMSEA known to be problematic in such instances (Kenny, Kaniskan and McCoach 2015).

As we shall discuss below, for some comparisons in some countries, full scalar invariance does not seem to hold. Having inspected the parameter estimates, this seems to be mainly driven by the ‘proper number’ item functioning differently across some groups (potentially due to this being confused with the actual mathematical concept of a ‘real number’). In such instances we relax the assumption of full scalar invariance across groups and assume only partial scalar invariance holds instead. Specifically, the ‘proper number’ thresholds are allowed to differ across groups and an alternative bullshit scale scores then produced. Comparisons across groups are then made using these alternative scores to ensure that changing our assumptions regarding full versus partial invariance does not lead to substantially different results.

These MGCFA models are fitted using Mplus (Muthén and Muthén, 1998-2017), with the final student weights applied and standard errors clustered at the school level. A WLSMV estimator with THETA parameterization⁷ was used to account for the ordered categorical nature of the

⁷ WLSMV is an estimator which is suitable for categorical variables. It performs a probit regression using a robust weighted least squares estimator with a diagonal weight matrix. THETA parameterization allows the residual variances of the latent trait to be parameters in the model, while excluding scale factors (Muthén & Muthén, 1998-2017).

questions (Muthén et al. 2015). Bullshit scale scores are then derived from these models, with average scores then compared across groups where scalar measurement invariance approximately holds. The bullshit scale has been standardised to mean zero and standard deviation within each country, so that all differences between groups can be interpreted in terms of an effect size.

Gender

Appendix Table A1 presents results from our measurement invariance tests for the bullshit scale across genders. All three levels of measurement invariance (configural, metric and scalar) hold according to our model fit criteria within each Anglophone country. We consequently conclude that average bullshit scale scores can be legitimately compared between males and females within all nine Anglophone countries.

Socio-economic status

Appendix Table A2 presents the measurement invariance test results with respect to socio-economic status. With the exception of Northern Ireland, configural and metric invariance of the scale holds in every country according to both the CFI and TLI criteria (Northern Ireland passes the former for metric invariance but fails according to the TLI). However, the picture is more mixed for full scalar invariance. Specifically, Australia and New Zealand fail to meet the full-scalar invariance criteria on two out of the three indices, while England fails the RMSEA while Scotland and Northern Ireland fails the CFI. However, once we release the ‘proper number’ item parameters, partial scalar measurement invariance is met in all countries.

Appendix Table A3 illustrates the difference between the full (top panel) and partial (bottom panel) scalar results for comparisons across socio-economic groups. In all countries, we find a non-trivial and statistically socio-economic gap in average bullshit scores regardless of whether the full or partial invariance scale scores are used. However, the magnitude of the gap is smaller based upon the partial invariance results. This is particularly the case in countries like Australia, Ireland, Scotland and New Zealand, where the difference in average bullshit scores between the top and bottom socio-economic group falls by around 0.2. Nevertheless, our key substantive finding that higher socio-economic status pupils are bigger bullshitters than their low socio-economic peers clearly continues to hold.

Mathematics achievement

Appendix Table A4 presents the measurement invariance test results with respect to mathematics. Configural and metric invariance is found to hold for every country according to

both the CFI and TLI criteria. The same is not true, however, with respect to scalar invariance where most countries fail based upon the CFI and RMSEA. Consequently, the full scalar measurement invariance results for mathematics achievement quartiles appear to be particularly problematic. Further inspection of the data and item parameters revealed that higher achieving young people were disproportionately likely to say that they had heard of ‘proper numbers’. Consequently, once we release the threshold parameters for this item under the partial scalar invariance model, all three fit indices take on much more reasonable values.

However, Appendix Table A5 reveals that there are huge discrepancies between the full and partial scalar results with respect to differences between mathematics achievement group. In Ireland, for instance, the difference in the bullshit index between young people in the top and bottom mathematics achievement quartile is 0.87 according to the full scalar invariance results. Yet this drops to 0.11 standard deviations and becomes statistically insignificant for the partial scalar results. A similar difference between the full and partial scalar results holds in most countries, with large declines in the estimated differences between the different achievement groups. This indicates that most of the difference in the bullshit scale scores between high and low ability pupils is being driven by differences in responses to the ‘proper number’ variable. We consequently conclude from Table A5 that differences in average bullshit scores across mathematics achievement groups are not particularly robust, and therefore do not include these results within the paper.

Immigrant status

Appendix Table A6 turns to the measurement invariance results with respect to immigrant status. The fit indices reveal that all countries meet the configural and metric criteria, with most also satisfying the full scalar invariance test as well. The potential exceptions are England, New Zealand and Wales, where the RMSEA criteria for full scalar invariance is not satisfied (though the CFI and TLI criteria is). Moreover, for all nine countries, partial scalar invariance clearly holds.

Appendix Table A7 illustrates that our key conclusions remain intact regardless of whether the full or partial scalar invariance scale scores are used. Specifically, we consistently find immigrants to be bigger bullshitters than country natives in each country other than the United States. Moreover, for most countries, the magnitude of the difference between immigrants and natives remains reasonably consistent between the full and partial scalar scale scores.

Country

The criteria we use for testing measurement invariance across countries is slightly different to the above. Specifically, as noted by Rutkowski and Svetina (2014), the CFI and RMSEA cut-offs typically used to judge whether measurement invariance holds are based upon an assumption of a small number of groups. This is typically violated in cross-national research, when the number of groups is large. Consequently, Rutkowski and Svetina (2014) suggested using slightly more liberal cut-off values for metric invariance tests when more than a handful of groups are being compared. With respect to this paper, they suggested that the change in the CFI between the configural and metric models should be less than 0.02 (rather than the usual 0.01). We follow this advice when testing for measurement invariance across countries.

We began our measurement invariance tests for comparability across countries by including all nine countries within our MGCFA models. However, this revealed a significant problem with Wales, where there was evidence of a very poorly fitting model. Wales has therefore been excluded from our cross-national comparisons, and all invariance tests based upon the remaining eight countries.

Results from these measurement invariance tests are presented in Appendix Table A8. Configural, metric and scalar invariance is met according to the CFI and TLI criteria. Likewise, the change in the RMSEA between the metric and scalar models is sufficiently low to also indicate that full scalar measurement invariance has been met. Nevertheless, to illustrate the robustness of our results, we also produce partial scalar scores where the ‘proper number’ thresholds have been released across three countries (Australia, England and the United States) which contributed the greatest change to the chi-squared statistic between the metric and scalar models.

Cross-country comparisons of average bullshit scores based upon the partial and full scalar invariance models can be found in Appendix Tables 9 and 10. They both provide a similar ranking of countries and have the same pattern of statistically significant differences. Although there are some modest changes in terms of the magnitude of the differences between countries, overall these results suggest that our conclusions regarding cross-country comparisons of bullshit scores are robust.

Table A1. Measurement invariance tests of the bullshit scale by gender

	CFI			TLI			RMSEA		
	Config	Metric	Scalar	Config	Metric	Scalar	Config	Metric	Scalar
Australia	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.007
Canada	1.000	1.000	0.999	1.000	0.999	0.999	0.000	0.018	0.012
England	1.000	0.999	0.999	1.000	0.998	0.999	0.000	0.029	0.019
Ireland	1.000	1.000	0.999	1.000	1.001	1.000	0.000	0.000	0.009
New Zealand	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000
Northern Ireland	1.000	1.000	1.000	1.000	1.004	1.000	0.000	0.000	0.003
Scotland	1.000	1.000	1.000	1.000	1.002	1.001	0.000	0.000	0.000
USA	1.000	1.000	0.998	1.000	0.999	0.999	0.000	0.018	0.027
Wales	1.000	0.998	0.995	1.000	0.993	0.997	0.000	0.052	0.034