Achievement and behaviour in undergraduate mathematics: personality is a better predictor than gender

Lara Alcock[a], Nina Attridge[b], Steven Kenny[a] and Matthew Inglis[a]

[a]*Loughborough University, UK;* [b]*University of Bath, UK*

We investigated two factors that predict students' achievement and behaviour in undergraduate mathematics: gender and personality. We found that gender predicted students' achievement and behaviour when considered in isolation, but ceased to be predictive when personality profiles were taken into account. Furthermore, personality accounted for significantly more variance in undergraduates' achievement and behaviour than did gender, but the converse was not the case. We therefore argue that personality provides the more productive lens through which to understand the behaviour of undergraduate mathematics students. We relate this finding to recent research emphasising gender differences in mathematics education, and suggest that researchers wishing to promote equity in participation at and beyond the undergraduate level should consider shifting their focus to individual differences in personality.

Keywords: gender; personality; undergraduate

Introduction

For decades, researchers have been concerned that mathematics is experienced differently by women and men (e.g. Becker 1995; Boaler 1997; Fenema 1974; Walkerdine 1989). Originally these concerns centred around a 'gender gap' in mathematical achievement: female students were believed to perform less well than male students, especially on advanced problems (e.g. Benbow, & Stanley 1983). However, recent research has demonstrated that this gap has ceased to exist, at least in developed countries (e.g. Else-Quest, Hyde, & Linn 2010; Guiso, Monte, Sapienza, & Zingales 2008; Hyde & Mertz 2009;

Spelke 2005). The primary concern now is the existence of a gender difference in post-compulsory mathematics participation, particularly given that this difference cannot be explained by earlier differences in performance (Noyes 2009). In the UK, 38% of students studying mathematics in higher education in 2008/9 were female (Higher Education Statistics Agency 2011), but only 18% of UK-based mathematics academics and 3% of full professors were female (Hobbs & Kooman 2006).

Theorists have accounted for gender differences in mathematics-related choices and behaviour in a number of ways. Some accounts involve direct causal claims about the lack of effective female role models (e.g. Beilock, Gunderson, Ramirez, & Levine 2010), gendered preferences for different teaching approaches (e.g. Rodd & Bartholomew 2006; Solomon, Lawson & Croft 2011), and even the gendered nature of mathematical knowledge itself (e.g. Burton 1995; Ernest 1991; Walshaw 2001). Other accounts involve indirect causal arguments based on the ideas that common cultural discourses associate mathematics with masculinity, making it more difficult for women to position themselves as successful mathematicians (e.g. Mendick 2006; Solomon 2012), and that pre-existing gender stereotypes differentially influence men and women via implicit social cognitions (e.g. Nosek & Smyth 2011) and stereotype-threat effects (e.g. Good, Aronson, & Jayne 2007; Spencer, Steele, & Quinn 1999).

Here we focus upon the extent to which gender predicts achievement and study behaviour in undergraduate mathematics, the stage immediately prior to the large drop-off in participation. We provide evidence that although, on average, men and women *do* behave differently in this environment, these differences would be better understood as consequences of personality rather than gender. We then consider the possibility that direct causal claims about a gendered experience might have the opposite of their intended effect: that discussing differences in terms of gender might perpetuate stereotypes, and so contribute indirectly to

women opting out of mathematics. We conclude by arguing that, since personality factors are better predictors, this alternative focus offers a way forward for researchers seeking to promote gender equity in mathematics.

Perspectives on gender in mathematics

Researchers working on gender issues in mathematics share the goal of ensuring that no group of students is systematically excluded or deterred from further study; they seek to understand what causes women to opt out of mathematics, with a view to influencing policy and practice. Approaches to this goal, however, vary considerably.

Some researchers have suggested that the disparity in higher-level mathematics participation arises because women and men experience mathematics differently. Some have theorised that mathematics itself is biased against women, and that the discipline must be radically reformed if equity is to be achieved (e.g. Ernest 1991; Walshaw 2001). Ernest (1991), for example, suggested that:

> the neutral view of mathematics […] represents the aggressive masculine half of human
> nature, which has rejected the receptive and compassionate feminine half. […] Success at
> dehumanized male mathematics may diminish our humanity, our ability to care, relate and
> feel (p. 279).

Others have focused on teaching, arguing that typical mathematical pedagogy is discouraging for women (e.g. Burton 1995). Boaler (1997b), for example, reported that in her study "the boys enjoyed individualised work" (p. 297) and that "[understanding] was central to the reasoning of the girls" (p. 296). These claims have been repeated recently, albeit with caveats concerning the importance of opportunities for students of both genders: Boaler, Altendorff, and Kent (2011) (summarising Boaler, 2002) commented on the procedural nature of some mathematics teaching and stated that:

> The boys also suffered from the lack of opportunities to develop conceptual understanding, but this did not concern them as much […]. The girls wanted more and their desire to question, probe and understand concepts was more intense and resulted in disaffection when opportunities for understanding were not given (p. 472).

At the college level, studies in other disciplines have not found gender differences in the degree to which students adopt a deep approach to learning (e.g. Duff, Boyle, Dunleavy, & Ferguson 2004; Wilson, Smart, & Watson 1996), but our study considers the possibility that women and men in undergraduate mathematics might have different learning goals.

Boaler et al. (2011) tempered their claims by expressing caution about balance between promoting equity and treating certain groups as homogeneous, stating that "…we are sensitive to the dangers of concluding that any approach is needed for all girls or boys…" (p. 479). This caution is in keeping with the arguments of researchers who have focused on ways in which common discourses might indirectly affect students' academic choices. Mendick (2006), for instance, took a poststructural approach: she argued that "[o]ppositional discourses about maths as objective not subjective, rational not emotional, and so on, tie it to masculinity" (p. 24) and contended that positioning oneself as mathematical is therefore more difficult for women than for men. Researchers in this tradition have studied ways that societal discourses constrain descriptions of the self and personal choices, and reported ways in which individuals describe their experiences. The focus of such reports is the detail of individual stories (e.g. Solomon 2012), but their overall point is emphatically supported by large-scale empirical studies, as discussed next.

Gender stereotypes, implicit social cognitions and stereotype threat

Despite the evidence on women's mathematical performance, there remains a widespread stereotype that mathematics is a masculine discipline in which women struggle (e.g. Nosek & Smith 2011; Swim 1994). This has systematic, measurable, negative effects upon women's

mathematical achievement and engagement, as shown in the literature on *stereotype threat*. Stereotype threat is said to be operating if individuals experience anxiety when potentially conforming to a negative stereotype, and numerous studies have investigated the effects of gender stereotypes on mathematics performance (see Maloney, Schaeffer, & Beilock 2013, for a recent review). Spencer et al. (1999), for example, found that women performed more poorly than men if told that a mathematics test had previously shown systematic gender differences, but performed as well when told that the test showed no gender differences. Similarly, making undergraduates' gender salient when asking them to tackle mathematics problems – by manipulating the gender balance of their fellow test-takers (Inzlicht & Ben-Zeev 2000) or by asking them to solve problems while wearing a swimming costume (Fredrickson, Roberts, Noll, Quinn, & Twenge 1998) – reduces women's achievement but not men's. In a large-scale study of the implicit social cognitions of over 5,000 adult participants, Nosek and Smith (2011) found that "[f]or women, stronger implicit math=male stereotypes predicted greater negativity toward math, less participation, weaker self-ascribed ability, and worse math achievement" (p.1125); the results for men were weakly in the opposite direction. Moreover, the effect for women was consistent with claims about societal discourses: measures of implicit stereotypes – obtained via reaction-time data from tasks involving implicit associations of mathematics with male and arts with female – revealed stronger relationships than their explicit counterparts, indicating that women may be affected by societal stereotypes even if they consciously reject these.

Such findings suggest that if we wish to reduce inequity in mathematics education, we should work to reduce the saliency of such stereotypes. Consequently, if we wish to claim that women are discouraged either by mathematics or by associated pedagogies, we should ensure that these phenomena are not better explained by other factors. In this paper, rather than assuming that gender is a key variable, we ask whether gender gives us the best

explanation of differences that have been linked to gender in previous research on undergraduate mathematicians' experiences.

Gender and undergraduate mathematics education

Rodd and Bartholomew (2006) followed cohorts of undergraduates through mathematics degree programmes at two highly-regarded UK universities. They noticed a gender difference in attendance rates, and suggested that women have "far higher" attendance (p. 43). This suggestion is consistent with Inglis, Palipana, Trenholm, and Ward's (2011) study of a first year calculus module, in which women mathematics students were found to attend significantly more lectures than men.[1]

Rodd and Bartholomew also suggested that women participate differently in the classroom. They observed that the women in their cohorts (around a third of the total) were "invisible", and suggested that women were less likely to contribute or be noticed in class as a result of attempting to maintain a feminine identity in a masculine discipline. To support this suggestion empirically, they recalled classroom incidents including:

> the woman's answer that was not heard by the man who was giving the lecture; [and] a man
>
> who was described by other students as the 'best student in the year' though in fact the best
>
> result was achieved by a woman who remained silent when hearing the conversation (p. 37).

Such observations can again be related to larger-scale studies. Solomon et al. (2011) found that women reported less positive relationships with their tutors than men, but that four other factors, two of which related to "gendered roles in the learning context" (positive attitudes towards group work and a willingness to ask questions in class) showed no gender differences. These larger-scale results suggest that observed classroom incidents are striking, but may not reflect systematic gender differences in preferred pedagogy or contribution. Indeed, Rodd and Bartholomew noted that not all of the men in their cohorts contributed to classroom discourse (p. 43) and that not all of the women failed to contribute (p. 44).

Solomon et al. (2011) did, however, suggest that women's negative experiences might be moderated by the availability of a Mathematics Learning Support Centre (MLSC) offering dedicated study space in which students could access drop-in support. A majority of UK universities now offer such facilities, as do many non-UK universities (e.g. Perkin & Croft 2004). Solomon et al. suggested that women found MLSC facilities particularly valuable because of the opportunities they offer for collaborative work, although their study did not directly investigate gender differences in the take-up of mathematics support.

Finally, research has investigated possible relationships between gender and use of Virtual Learning Environments (VLEs). Results here are unclear. Some have found that men are more willing than women to engage in dialogue via bulletin boards (Barrett & Lally 1999), whereas others have found the reverse (Jackson, Ervin, Gardner, & Schmitt 2001). In a calculus course, Inglis, Palipana, Trenholm, and Ward (2011) found that male undergraduates more often watched recordings of lectures online than females. However, they noted that the majority of the women in their sample were studying mathematics, whereas most men were studying engineering, so the difference may have been due to discipline rather than gender.

In sum, it is clear that there is scope for clarifying the extent to which gender predicts study behaviours in undergraduate mathematics.

Personality factors

In contemporary social psychology, many of the phenomena discussed above would be considered to be in the domain of personality, not gender. Perhaps students who attend more lectures are more conscientious; perhaps students who enjoy the collegiate atmosphere of MLSCs are more agreeable. This applies to achievement as well as behaviour: Conscientiousness is predictive of academic performance (Duff et al. 2004). Our

investigation of personality factors is thus consistent with the approach used in that field. In particular, we use the most common operationalisation of the construct of personality, which is referred to as the 'Big Five' model. Researchers broadly agree that there are five main dimensions onto which a person's personality can be mapped (for a review see John, Naumann, and Soto 2008):

- *Conscientiousness:* a tendency to show self-discipline and to regulate impulsive behaviours (sample self-report item: "I get chores done right away").

- *Extraversion:* a tendency to be enthusiastic and attention-seeking, especially in social situations ("I talk to a lot of different people at parties").

- *Agreeableness:* a tendency to value getting along with others; to believe that people are essentially honest and trustworthy ("I sympathise with others' feelings").

- *Neuroticism:* a tendency to be anxious and depressed, and to interpret small complications as large difficult problems ("I have frequent mood swings").

- *Openness to Experience:* a tendency to be imaginative and curious; to value abstract thinking and immerse oneself in difficult intellectual tasks ("I have a vivid imagination").

Scores on Big Five factors follow normal distributions and show considerable between-individual variation (e.g. John et al. 2008). Big Five profiles are associated with, among other things, life satisfaction (Boyce, Wood, & Powdthavee 2013), perceived well-being (Hayes & Joseph 2003), job performance (Barrick & Mount 1991), willingness to share money (Paunonen & Ashton 2001), alcohol and tobacco consumption (Paunonen & Ashton 2001), and frequency of group conversations in day-to-day life (Mehl et al. 2006).

For our purposes, it is important to note that there are small but reliable gender differences in scores on the Big Five factors. For instance, Schmitt, Realo, Voracek, and Allik (2008) obtained Big Five profiles of 17,637 participants from 55 nations, finding systematic differences between men's and women's self-report ratings. The sizes of the gender differences varied by culture. In the UK, where we conducted our study, gender differences were found for Neuroticism (women were more neurotic, d=0.55) and Agreeableness (women were more agreeable, d=0.29). Gender differences for the other factors were very small (d≈±0.10). These findings make it plausible that some of the apparent gender differences in undergraduate mathematics might be better understood as personality differences.

Measuring gender and personality

In the discussion above, we have not distinguished between biological sex and gender. Gender, it has been argued, is a social construct; beliefs about individual and societal gender roles both influence and are constructed by our social interactions. This conceptualisation is reflected in large-scale empirical work: there exist many gender-related measures assessing constructs such as adherence to cultural norms, sexism, and feminist identity (Smiler & Epstein 2010). One well-known measure, the Bem Sex Role Inventory, has been used to investigate whether identification with gender roles predicts mathematical achievement in a way that sex does not (Santos, Ursini, Ramirez, & Sanchez 2006). However, like many gender-related measures, this inventory is subject to serious criticism regarding a lack of clarity about its aims and inadequate psychometrics (Hoffman & Borders 2001). Indeed, some researchers have concluded that it is better understood as a measure of the extent to which individuals have dominant, assertive dispositions or expressive, nurturant tendencies (Hoffman & Borders 2001; Smiler & Epstein 2010), meaning that it could be better

understood as a personality measure. Because of this, and because the majority of the research cited above is concerned with effects that differ for women and men, we did not use such measures.

Work on personality is subject to similar debates. A measure such as the Big Five reifies personality, assuming that it is meaningfully measurable. This, again, contrasts with poststructural research, which typically rejects the notion that identity is fixed, and studies *identity work* done as individuals position themselves relative to societal discourses. A poststructuralist researcher might, therefore, be inclined to dismiss personality measures as theoretically misguided, or at least as incommensurable with understandings derived from poststructural analyses. We address two particular concerns associated with this position, and then explain what we believe the personality measurement approach can offer.

One concern is that responses to a personality measure might not capture intrinsic characteristics; they might instead reflect the fact that participants feel compelled to respond in a way that is socially acceptable – or, more insidiously, that participants unknowingly respond in certain ways because patterns of discourse constrain their positions without their knowledge. This is a problem for interpreting the philosophical meaning of any result involving personality measures: should we believe that individuals with different personality profiles behave in different ways because they 'have' different personalities that lead to different behaviours, or because they are differentially vulnerable to societal discourses which lead them to respond in different ways to a personality measure *and* to behave differently in other interactions? This is not a problem from the perspective of empirical research, which accepts the necessity of operationalising theoretical constructs and of evaluating the validity and reliability of the resulting measures. In this case, there is no question that people who respond to a personality measure in a given way *for whatever underlying reason* behave predictably in relation to other measures too. To an empiricist this,

coupled with the sound psychometrics of the test, is sufficient to make the measure useful for scientific progress.

A second concern is that responses to personality measures might not be meaningful because identities are continually co-constructed through interactions in a discourse community, so responses are subject to change in a way that renders a one-time measure meaningless. This is clearly a potential problem, although on this theoretical point there is a smaller gap than one might think between the poststructural approach and recent scientific study of identity-related phenomena. Poststructuralists do not believe that identities change arbitrarily, but rather that individuals must do *identity work* to maintain or develop aspects of their identities. Similarly, psychological research treats individuals not as having fixed beliefs and characteristics, but as being predictable to themselves based on their past behaviour, while also being subject to adjustments prompted by fleeting situational factors (e.g. Ariely & Norton 2008). Thus neither approach contends that personalities might be so variable as to render measures inherently meaningless.

There remains, however, a serious potential problem: personality factors could be well correlated with other measures at a given point in time but dramatically variable in the long term. This would limit their utility in predicting things like study behaviours across the duration of, say, a degree programme. However, Big Five factors are stable across time, at least in adulthood (McCrae & Costa 1990), and (with the exception of Neuroticism) they vary considerably less within individuals than they do between individuals. This does not mean that they are permanently fixed: Boyce, Wood, and Powdthavee (2013) found that over a four-year interval, personality did change to a degree comparable with economic factors such as household income. This stability with the possibility of moderate long-term change is potentially positive for those wishing to promote equity. If personality factors are better predictors than gender for behaviour in undergraduate mathematics, this gives us a way to

analyse education without treating women and men as homogeneous groups and without therefore perpetuating gender stereotypes. Furthermore, if personality factors are stable but malleable, this might allow us to help students become more successful by encouraging productive study behaviours while remaining sensitive to differences in individual needs. We return to this argument after presenting our empirical findings.

Method

Participants

The participants were second-year undergraduates studying degree courses administered by the School of Mathematics at Loughborough University (the study was approved by the university's ethics committee). Approximately half were studying for mathematics degrees and half for joint honours degrees (combining mathematics with, for example, finance or computer science). This institution and department are well respected but not elite – the entry grades for mathematics degrees are high, but comparable with those at numerous institutions around the UK. Students on mathematics-based programmes learn a range of subjects from both pure and applied mathematics, and have approximately 18 contact hours per week split over six subjects per semester in a combination of large lectures, smaller lab and problems classes, and small-group tutorials; again these curricular and pedagogical arrangements are typical for UK mathematics degrees. Finally, around 40% of the department's undergraduates in a typical year are women. The fact that our participants were all from one institution might affect our results in the sense that their experiences and behaviours might be somewhat more similar to each other than they would be to those of students at other institutions. However, with respect to pre-university mathematical preparation, experience of undergraduate mathematics teaching, and experience of the self as a man or woman in relation to the

numbers in a cohort (HESA 2011), we believe that these participants may be considered typical of undergraduate mathematics students in the UK.

In the first week of their second year, all 203 students in the cohort were invited to attend a voluntary session where they completed (in silence) a series of self-report questions about their personality and their approaches to learning. A total of 89 students (34 women, 55 men) attended, so our sample had a gender balance very similar to the UK national figure for mathematics (HESA 2011).

Measures[2]

We treat Gender and Personality as independent variables.

Gender was deduced from the 'Preferred Title' entry (Mr/Mrs/Miss/Ms) in each participant's record in the university student database. We term this variable 'Gender' rather than 'sex' because a) this entry was provided by the participant and so involved self-identification with a gendered form of address, and b) the term 'gender' tends to be used broadly in mathematics education research (e.g. both Boaler, Altendorff, and Kent (2011) and Rodd and Bartholomew (2006) use the term to describe results from observational studies in which they classified participants as male or female based on appearance). Naturally, it is to be expected that a participant's response would coincide with their biological sex in the vast majority of cases, so readers who prefer to reserve the use of the term 'gender' for more nuanced analyses of social interactions might prefer to think of this variable as 'sex' throughout.

Big Five Personality factors were measured using the Mini-IPIP scales (Donnellan et al. 2006). These scales use four items per factor, with a mixture of positively-worded items ("I get chores done right away") and negatively-worded items ("I often forget to put things back in their proper place"). Participants are asked to report how well each item describes them, using a five-point scale from "strongly disagree" to "strongly agree". Responses are reverse-

scored if necessary, then summed to give five measures, each on a 4-to-20 scale. The Mini-IPIP shows convergent, discriminate and criterion-related validity and good test-retest reliability (Donnellan et al. 2006).

We treat achievement, approaches to learning, and study behaviours (MLSC and VLE use) as dependent variables.

Academic Achievement was assessed by obtaining students' grades from their first year mathematics modules. Because single-honours students studied more mathematics than joint-honours students, we calculated students' grand mean marks on their mathematics modules only.

Approaches to Learning were measured using Biggs's Revised Two Factor Study Process Questionnaire (R-SPQ-2F, Biggs, Kember, & Leung 2001). This comprises Deep Approach to Learning and Surface Approach to Learning scales, each consisting of five strategy items (about what the student does) and five motive items (about what the student is aiming to achieve). For example, "My aim is to pass the course while doing as little work as possible" is a Surface Motive item whereas "I find that at times studying gives me a feeling of deep personal satisfaction" is a Deep Motive item. Participants use five-point Likert scales to report the extent to which each item describes them, and these ratings are combined to yield scores on 5-to-25 scales; each scale has been shown to be internally reliable (Biggs et al. 2001).

Mathematics Support Uptake was measured by recording each time a participant entered the Mathematics Learning Support Centre during the academic year (attendees were required to swipe their University ID card – this was routinely enforced by a receptionist).

VLE Resource Use was measured by interrogating server logs to calculate the total number of times each student accessed a VLE resource during the academic year. Because not all the sampled students studied the same modules, and because lecturers varied in the resources that

they provided online, we restricted our count to the four core modules studied by every participant.

Representativeness of the sample

To check that our sample was representative of the cohort, we first assessed whether there was a relationship between students' genders and attendance at the data-collection session. There was no significant relationship: 41% of men attended, compared with 49% of women, $p=.373$ (Fisher's Exact Test).

We then compared the module marks of students who did and did not participate, using an ANOVA with two between-subjects factors: Participation (yes/no) and Gender (male/female).[3] We found significant main effects of both factors: those who participated achieved marks approximately 8% higher than those who did not, 66% versus 58%, $F(1,199)=15.373$, $p<.001$, $\eta_p^2=.072$; the gender effect is discussed later. There was no Gender × Participation interaction, $F<1$, $\eta_p^2=.001$, so it was not the case that high(/low)-achieving women participated more than high(/low)-achieving men. Thus, although our participants were to some extent self-selected, there is no reason to believe that this self-selection systematically skewed the data. This, together with issues of representativeness discussed under *Participants*, means that we can be reasonably confident about using our data to make inferences about gender-related issues in the achievement and behaviour of undergraduate mathematics students.

Results

We report our main results in three sections. First, we show that there were significant differences in the personalities of women and men in our sample.[4] Secondly, we analyse Academic Achievement, Approaches to Learning, MLSC use and VLE-resource use with

respect to Gender, then with respect to Gender while controlling for Personality. Finally, we compare the amount of variance accounted for by the Gender and Personality predictors.

Personality differences

We first compared men's and women's self-report ratings for each of the Big Five factors, subjecting participants' ratings for Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism to a one-way MANOVA with Gender as the factor. The influence of Gender was significant, $F(5,79)=5.018$, $p<.001$, $\eta_p^2=.241$. Considering each factor separately revealed significant differences between men and women's mean ratings for Agreeableness, $F(1,83)=7.606$, $p=.007$, Conscientiousness, $F(1,83)=7.089$, $p=.009$, and Neuroticism, $F(1,83)=6.091$, $p=.016$; women rated higher on all three, with effect sizes of $d=0.694$, 0.551, and 0.570 respectively. We note that these effect sizes are somewhat larger than those found in the general UK population by Schmitt et al. (2008). Determining whether these differences are important (one could hypothesise, for example, that undergraduate mathematics attracts students with particular personality profiles) would be a difficult but valuable task. It would require determining the personality profiles of a representative sample of the UK population, and practicalities mean that neither our study nor Schmitt et al.'s adopted truly representative sampling techniques. For our purposes in this paper, however, these differences establish the potential for personality-based rather than gender-based accounts for differences in achievement and behaviour in undergraduate mathematics.

Predictors of achievement and behaviour

We next investigated whether Gender was a significant predictor of academic achievement. In our sample, as in Rodd and Bartholomew's (2006) and Stout et al.'s (2011) studies, women received higher mean marks than men: 69% versus 63%, $t(87)=2.556$, $p=.012$,

d=0.548. We then investigated whether this Gender difference was independent of personality differences by conducting a one-way ANCOVA with Academic Achievement as the dependent variable, Gender as the factor, and each of the Big Five factor scores as a covariate. We found that, having controlled for personality differences, Gender ceased to be a significant predictor of Academic Achievement, $F(1,78)$=1.660, p=.201. Instead, Conscientiousness was a significant predictor, $F(1,78)$=3.981, p=.049, η_p^2=.049, and Agreeableness was a borderline significant predictor, $F(1,78)$=3.074, p=.083, η_p^2=.038.

We then investigated whether Gender was a significant predictor of Approaches to Learning, using each individual's Deep Motive, Deep Strategy, Surface Motive, and Surface Strategy scores. We subjected the data to a one-way MANOVA with four dependent variables (Deep Motive, Deep Strategy, Surface Motive, Surface Strategy) and one independent variable (Gender). Contrary to some earlier research (Wilson et al. 1996) but in line with other findings (Boaler, Altendorff, & Kent 2011), Gender was a borderline significant predictor, $F(4,83)$=2.387, p=.058, η_p^2=.103. In particular, women had significantly higher ratings than men on the Deep Motive subscale, 17.0 versus 15.4, $F(1,86)$=6.121, p=.015, η_p^2=.067. However, when we conducted a one-way MANCOVA including the Big Five as covariates, the borderline significant effect for Gender disappeared, $F(4,75)$=1.443, p=.228, η_p^2=.071. Instead, Conscientiousness, $F(4,75)$=3.788, p=.007, η_p^2=.168 and Openness, $F(4,75)$=3.862, p=.007, η_p^2=.171, were significant predictors of students' R-SPQ-2F profiles. Conscientiousness was negatively correlated with the Surface Strategy subscale, r=-.217, p=.041, and positively correlated with the Deep Motive, r=.246, p=.020, and Deep Strategy subscales, r=.333, p=.002. Openness was negatively correlated with the Surface Strategy, r=-.402, p<.001, and Surface Motive subscales, r=-.245, p=.022, and positively correlated with the Deep Strategy subscale, r=.220, p=.040.

We then investigated whether Gender was a significant predictor of how often students visited the MLSC. Because MLSC entries did not follow a normal distribution (many students never used the MLSC), we adopted a non-parametric approach, comparing the median number of times men and women entered the MLSC using a Mann Whitney U test. In line with claims about women valuing MLSC provision (Solomon, Lawson, & Croft 2011), Gender was a significant predictor, $U=624$, $p=.007$: the median number of MLSC entries was 4.5 for women and 1.0 for men (means were 9.71 and 5.64 respectively). We then conducted an ordinal logistic regression with MLSC entries as the dependent variable, Gender as the independent variable, and the Big Five factors as covariates, finding that Gender ceased to be a significant predictor of MLSC entries, Wald $\chi^2(1)=0.916$, $p=.339$. Instead, Openness was a significant predictor, Wald $\chi^2(1)=4.976$, $p=.026$, and Conscientiousness was a borderline significant predictor, Wald $\chi^2(1)=3.460$, $p=.063$. Students with lower Openness ratings and students with higher Conscientiousness ratings visited the MLSC more often.

Finally, we investigated whether Gender was a significant predictor of VLE resource use. We found a borderline significant difference between the number of accesses for men, 326.4, and women, 387.3, $t(84)=1.763$, $p=.081$. However, when we conducted a one-way ANCOVA with Gender as the factor and the Big Five factors as covariates, the Gender effect disappeared, $F(1,75)=.008$, $p=.929$. Instead, we found that Conscientiousness was a significant predictor of the number of times resources were accessed, $F(1,75)=4.842$, $p=.031$, $\eta_p^2=.061$, and that Openness was a borderline significant predictor, $F(1,75)=3.566$, $p=.063$, $\eta_p^2=.045$. More conscientious students accessed resources more frequently, and students more open to experience accessed resources less frequently.

To summarise, when considered in isolation, Gender was found to predict mathematical achievement, approaches to learning, mathematics support uptake, and VLE

resource use. However, all of these relationships ceased to be significant when students' personality profiles were taken into account.

Variance explained: Personality versus Gender

The above analyses demonstrate that the effects of Personality factors and Gender are confounded in the context of undergraduate mathematics education. What seem at first to be gender issues can, in fact, be understood as personality issues. Nevertheless, it might be that gender could be used as a proxy for personality: if students' gender allowed one to predict their behaviour and achievement with nearly as much accuracy as their personality profiles, gender might be a reasonable pragmatic alternative to a personality scale.

To assess this possibility, we ran six sets of hierarchical regression models, two for each of achievement, MLSC use and VLE use (ordinal logistic regressions in the case of MLSC use). We used the first set of models for each measure to determine whether the Big Five factors predicted any variance over and above that accounted for by Gender. To achieve this we entered Gender in the first block, followed in blocks 2-6 by each of the Big Five factors. In order to subject our hypothesis to the most stringent examination, we entered the Big Five factors in reverse order of predictive power (as determined by a one-block regression).[5] In the second set of models, we did the reverse: to determine whether Gender predicted any variance in our outcome measures over and above that which could be accounted for by Personality, we entered each of the Big Five factors in the first five blocks, followed by Gender in the last.

Table 1 shows the resultant models for mathematical achievement (a similar pattern was observed for MLSC and VLE use, so we omit details). When added to the first model, both Agreeableness and Conscientiousness accounted for significant variance over and above that accounted for by Gender and the first three Big Five factors (although this was only

borderline in the case of Agreeableness). Conversely, Gender did not account for extra variance when it was added to the second model after the Big Five.

To summarise, Personality had predictive power over and above Gender, in fact around twice as much in the case of mathematical achievement: using Gender alone to predict achievement resulted in an R^2 of .075 (Model 1 in Table 1a), whereas using the Big Five alone resulted in an R^2 of .154 (Model 5 in Table 1b). But the reverse was not the case: Gender did not have any predictive power over and above Personality.

Summary

We used four measures of undergraduate mathematics students' achievement and behaviour: academic achievement, approaches to learning, MLSC use, and VLE resource use. When considered in isolation, Gender was associated with each of these measures. However, all of these relationships ceased to be predictive when students' personality profiles were taken into account. Personality factors explained more variance in undergraduate mathematicians' behaviour and achievement, and gender did not explain significant variance beyond that explained by personality.

These are important findings because many researchers have claimed that men and women have different pedagogical preferences, and that therefore mathematical experiences are gendered (e.g., Becker 1995; Boaler 1997; Rodd & Bartholomew 2006; Solomon et al. 2011). Our findings suggest that this claim is valid, but also that it is misleading because it directs attention to a variable that does not provide the most predictive power. It seems likely, in fact, that the gender differences observed by earlier researchers arise from a confound between gender and personality, and that personality factors would provide a more nuanced and accurate view of the complex social world of undergraduate mathematical study.

To conclude this section, we observe that neither personality nor gender explained a large proportion of the variance in undergraduates' achievement (or their MLSC or VLE use) – R^2 values reported in both sets of regression models were relatively low. This is not surprising: many factors influence that influence mathematics achievement (e.g. working memory, general intelligence, teaching quality) would not be expected to vary systematically with personality or gender. Nevertheless, these low R^2 values confirm that there are large individual differences among undergraduate mathematics students, so it is important to avoid inappropriate generalisations or stereotyping based on small between-group differences.

General Discussion

In this final section we discuss the social implications of our findings. We begin with two negative consequences that may result from an inappropriate focus on gender: the potential 'invisibility' of certain groups and the perpetuation of harmful stereotypes.

Earlier mathematics education research has described women as a disadvantaged group: Rodd and Bartholomew (2006) suggested that women undergraduates are 'invisible' in the classroom, and Solomon et al. (2011) argued that this problem could be ameliorated somewhat by mathematics support. Such claims relate to questions of equity: higher education institutions are responsible for all students, so it is important to know which resources are used and appreciated by which groups and to consider additional provision if a group is underserved. However, our findings suggest that a focus on gender obscures the needs of different students – men and women with 'atypical' personality profiles – and that Rodd and Bartholomew could have been observing a phenomenon that is better understood in terms of personality differences. If institutions and individual lecturers are to successfully engage all their students, they need to engage men with high Neuroticism ratings and women who are disagreeable. Seeing the classroom solely through the lens of gender obscures

within-gender variability, potentially making 'invisible' substantial numbers of students of both genders.

Secondly, recall that the widespread stereotype that mathematics is a masculine discipline can act via stereotype threat and implicit social cognitions to negatively influence women's mathematical achievement and participation (e.g. Maloney, Schaeffer, & Beilock 2013). We argue that such findings should give education researchers pause for thought. Some have written about men and women as though they constitute homogeneous groups, and as though membership of the group of females interacts with mathematics or pedagogy to directly discourage women from further mathematical studies. Boaler (1997), for example, claimed that "*the boys* enjoyed individualised work" (p. 297, our emphasis) and that "[understanding] was central to the reasoning of *the girls*" (p. 296, our emphasis), seemingly implying that there was little or no within-gender variance in her sample. Other researchers have proposed that mathematical knowledge itself is biased against women (e.g. Ernest 1991). In fact, women and men show substantial within-group variance in personality profiles, and we have shown that this accounts for many apparent between-group differences in their behaviour and success in undergraduate mathematics.

Of course, it remains the case that gender might influence participation via effects that were not revealed by our study. We used only a short (though standard and widely-used) personality measure, and we investigated only four outcome measures; other methods might reveal relationships between gender and other variables that cannot be accounted for by personality factors. Perhaps, for instance, factors influencing participation operate only at earlier educational levels or in other educational systems, meaning that it would be valuable to determine whether similar results occur in school-level mathematics or in other cultural settings. Or perhaps gender is a superior predictor for other aspects of experience in undergraduate mathematics: although women and men behave and achieve similarly in

relation to our measures, they might, for instance, have systematically different emotional reactions to their day-to-day study experiences. But such differences have yet to be established empirically, and our results are not consistent with the existence of a scientific reason for treating gender as a privileged variable of interest in undergraduate mathematics. Thus, we suggest that researchers investigating gender in this and in broader contexts should begin considering personality factors as covariates.

In such educational research, it is important to remember that many studies on gender effects – including ours – are inherently correlational in nature, and so cannot be used to make direct claims about causality. While it certainly seems plausible that being more conscientious would cause better achievement, our results do not prove this – we use the term 'predictors' only in the technical sense appropriate to the reported analyses. This, of course, is also true for research that employs alternative methods: because gender cannot be randomly assigned, it is risky under any circumstances to imply that gender is a causal factor in student behaviours and experiences. The particular risk for theorists in mathematics education is that promoting the idea of gender differences might perpetuate unhelpful stereotypes, and thus indirectly suppress women's mathematical progression. Others have expressed concern about this possibility: Mendick (2006), for instance, acknowledged that discussions of gender differences could be self-perpetuating, but argued that:

> [s]topping feminist research on gender differences is not going to end discussion about them; what it will do is restrict even further the range of stories through which we can make sense of them (p. 102).

Providing alternative narratives is clearly a sensible aim, but we suggest that it would need to be enacted very thoroughly in order to avoid, for instance, transmitting a simplistic view of gender in mathematics to prospective teachers via teacher education programmes. Female teachers' mathematics anxiety levels have been shown to correlate with poor performance in

their female (but not male) pupils (Beilock et al. 2010), so it is important that the nuances of such a message are properly understood.

In our view, there is no doubt that a focus on gender has served, historically, to draw attention to important inequities in mathematics education. But perhaps a change of focus is both scientifically and socially appropriate at this stage. We suggest that one way forward is exemplified by our approach in this paper: if we can identify and investigate alternative factors that provide more accurate predictions of students' behaviours and choices, this provides an opportunity to drive the discussion in a new direction. This is especially promising if these alternative factors provide more realistic opportunities for intervention. People do not ordinarily change their gender, but they might over time change their personality characteristics (Boyce, Wood, & Powdthavee 2013), and they can certainly be guided to change their beliefs about those characteristics. Studies have shown that people often hold implicit beliefs that personality and intelligence are fixed, and that teaching students that these qualities may be developed can help them to be resilient in the face of academic challenges, can positively affect willingness to engage with difficult material, and can lead to improved performance (Yeager & Dweck 2012). We suggest that related interventions might be particularly useful in mathematics, because students entering university often conceive of mathematics as a set of rules and procedures to be learned for reproduction rather than as a coherent theory to be understood by engaging with difficult problems (Crawford, Gordon, Nicholas, & Prosser 1998). For students whose personalities tend to be associated with less productive beliefs and behaviours, perhaps interventions focused on the possibility of learning and change would encourage development towards deeper approaches to learning.

To summarise, we do not claim that gender is irrelevant in undergraduate mathematics education. Clearly it is relevant. But this is not because women as a group

perform more poorly in mathematics – they do not. Nor is it because women as a group engage in different ways with learning resources – *individual* women engage differently in a way that is systematically related to personality factors, but so do individual men. Rather, gender is important because society still holds stereotypes that run counter to the evidence about mathematical potential and achievement: women know that society thinks of mathematics as a male domain, and this knowledge can have a negative impact on their performance and participation. Thus we argue that if we wish to increase female participation in post-compulsory mathematics, we should not speak of women as a homogeneous group, and we should not claim that they are disadvantaged by mathematics as a subject. Rather, we should continue to publicise accurate statistics about achievement, and we should continue to develop our understanding of the complex social world of mathematical study by endeavouring to identify the most accurate predictors of behaviour and achievement.

Notes

1. A re-analysis of their data indicated that on average women mathematics students attended 62% of their lectures and men attended 49%, $t(206.8)=3.484$, $p=.001$, $d=0.459$.

2. Our dataset is available at http://dx.doi.org/10.6084/m9.figshare.865640.

3. For seven participants we had missing data for one or two variables (e.g., because a participant did not complete every questionnaire item). A total of 8 values (0.69% of the dataset) were missing; such cases were dealt with using pairwise deletion.

4. Prior to conducting the regression and ANOVA analyses we checked our predictors for collinearity. As would be expected (the five personality variables are, by definition, orthogonal), there was no evidence of this: all VIF values were well within the acceptable range (all < 2).

5. This ensured minimal ΔR^2 values for the most predictive of the personality variables.

References

Ariely, D., & Norton, M. I. (2008). How actions create – not just reveal – preferences. *Trends in Cognitive Sciences*, *12*, 13-16.

Barrett, E. & Lally, V. (1999). Gender differences in an on-line learning environment. *Journal of Computer Assisted Learning, 15*, 48-60.

Barrick, M. R., & Mount, M., (1991). The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology, 44*, 1-26.

Becker, J. (1995). Women's ways of knowing in mathematics. In G. Kaiser & P. Rogers (Eds.) *Equity in mathematics education: Influences of feminism and culture* (pp. 164-174). London: Falmer.

Beilock, S. L., Gunderson, E. A., Ramirez, G. & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences*, *107*, 1060-1063.

Benbow, C. P., & Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: More facts. *Science, 222*, 1029–1031.

Biggs, J., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor Study Process Questionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 77,* 133-149.

Boaler, J. (1997). Reclaiming school mathematics: The girls fight back. *Gender and Education, 9,* 285-305.

Boaler, J. (2002). *Experiencing school mathematics: Traditional and reform approaches to teaching and their impact on student learning.* Lawrence Erlbaum, Mahwah, NJ.

Boaler, J., Altendorff, L., & Kent, G. (2011). Mathematics and science inequalities in the United Kingdom: When elitism, sexism and culture collide. *Oxford Review of Education, 37,* 457-484.

Boyce, C.J., Wood, A.M., & Powdthavee, N. (2013). Is personality fixed? Personality changes as much as "variable" economic factors and more strongly predicts changes to life satisfaction. *Social Indicators Research, 111*, 287-305.

Burton, L. (1995). Moving towards a feminist epistemology of mathematics. *Educational Studies in Mathematics, 28*, 275-291.

Crawford, K., Gordon, S., Nicholas, J., & Prosser, M. (1998). Qualitatively different experiences of learning mathematics at university. *Learning and Instruction, 8*, 455-468.

Donnellan, M.B., Oswald, F.L., Baird, B.M., & Lucas, R.E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment, 18,* 192-203.

Duff, A., Boyle, E., Dunleavy, K., & Ferguson, J. (2004). The relationship between personality, approach to learning and academic performance. *Personality and Individual Differences, 36*, 1907-1920.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136,* 103-127.

Ernest, P. (1991). *The philosophy of mathematics education*. New York: Falmer Press.

Fenema, E. (1974). Mathematics learning and the sexes: A review. *Journal for Research in Mathematics Education, 5,* 126-139.

Fredrickson, B. L., Roberts, T.-A., Noll, S. M., Quinn, D. M. & Twenge, J. M. (1998). That swimsuit becomes you: Sex differences in self-objectification, restrained eating, and math performance. *Journal of Personality and Social Psychology, 75*, 269-284.

Good, C., Aronson, J., & Jayne, A. H. (2007). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology, 29*, 17–28.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science*, *320*, 1164-1165.

Hayes, N., & Joseph, S. (2003). Big 5 correlates of three measures of subjective well-being. *Personality and Individual Differences, 34*, 723–727.

HESA (2011). *Statistics - Students and qualifiers at UK HE institutions.* www.hesa.ac.uk [Accessed 1/11/11].

Hobbs, C. & Kooman, E. (2006). *Statistics on Women in Mathematics*. Available online: http://womenandmath.files.wordpress.com/2007/09/statisticswomen.pdf [accessed 26/10/11].

Hoffman, R. M., & Borders, L. D. (2001). Twenty-five years after the Bem sex-role inventory: A reassessment and new issues regarding classification variability. *Measurement and Evaluation in Counseling and Development, 34*, 39–55.

Hyde, J.A., & Mertz, J.E. (2009). Gender, culture and mathematics performance. *Proceedings of the National Academy of Sciences*, *106*, 8801-8807.

Inglis, M., Palipana, A., Trenholm, S., & Ward, J. (2011). Individual differences in students' use of optional learning resources. *Journal of Computer Assisted Learning, 27,* 490-502.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11*, 365−371.

Jackson, L. A., Ervin, K. S., Gardner, P. D. & Schmitt, N. (2001). Gender and the internet: women communicating and men searching. *Sex Roles, 44*, 363–379.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W.

Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114–158). New York: Guilford.

Maloney, E. A., Schaeffer, M. W., & Beilock, S. L. (2013). Mathematics anxiety and stereotype threat: Shared mechanisms, negative consequences and promising interventions. *Research in Mathematics Education, 15*, 115–128.

McCrae, R. R. & Costa, P. T. (1990). *Personality in adulthood*. New York: The Guildford Press.

Mendick, H. (2006). *Masculinities in Mathematics*. Maidenhead: Open University Press.

Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal, 48*, 1125–1156.

Noyes, A. (2009). Exploring social patterns of participation in university-entrance level mathematics in England. *Research in Mathematics Education, 11*, 167-183.

Paunonen, S. V., & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology, 81*, 524–539.

Perkin G. & Croft T. (2004) Mathematics support centres – the extent of current provision. *MSOR Connections,* 4, 14–18.

Rodd, M. & Bartholomew, H. (2006). Invisible and special: Young women's experiences as undergraduate mathematics students. *Gender and Education, 18,* 35-50.

Santos, D., Ursini, S., Ramirez, M. P., & Sanchez, G. (2006). Mathematics achievement: Sex differences vs. gender differences. In J. Novotná, H. Moraová, M. Krátká, M., & N. Stehlíkovía (Eds.), *Proceedings of the 30th Conference of the International Group for the Psychology of Mathematics Education*, (Vol. 5, pp 41–48). Prague, Czech Republic: PME.

Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology, 94*, 168–182.

Smiler, A.P., & Epstein, M. (2010). Measuring gender: Options and issues. In J.C. Chrisler & D.R. McCreary (Eds.), *Handbook of gender research in psychology* (pp.133-158). New York: Springer.

Solomon, Y. (2012). Finding a voice? Narrating the female self in mathematics. *Educational Studies in Mathematics*, *80*, 171-183.

Solomon, Y., Lawson, D., & Croft, T. (2011). Dealing with 'fragile identities': resistance and refiguring in women mathematics students. *Gender and Education, 23,* 565-583.

Spelke, E.S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist*, 60, 950-958.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35,* 4-28.

Stout, J. G., Dasgupta, N., Hunsinger, M. & McManus, M. A. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology and mathematics (STEM). *Journal of Personality and Social Psychology, 100,* 255-270.

Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology, 66*, 21-36.

Walkerdine, V. (1989). *Counting girls out*. London: Virago.

Walshaw, M. (2001). A Foucauldian gaze on gender research: What do you do when confronted with the tunnel at the end of the light? *Journal for Research in Mathematics Education, 32*, 471–492.

Wilson, K. L., Smart, R. M., & Watson, R. J. (1996). Gender differences in approaches to learning in first year psychology students. *British Journal of Educational Psychology, 66,* 59-71.

Yeager, D. S., & Dweck, C. S. (2012). Mindsets that promote resilience: When students believe that personal characteristics can be developed. *Educational Psychologist*, *47*, 302-314.

Table 1. Hierarchical regression models with achievement as the dependent variable. Model (a) demonstrates that the Big Five explain significant variance over that explained by Gender. Model (b) demonstrates that Gender does not account for significantly more variance than that accounted for by the Big 5. * $p<.05$, † $p<.1$.

Model (a)	Predictors of Achievement	R^2	Change in R^2
1	Gender	.075	.075*
2	Gender, Extraversion	.084	.009
3	Gender, Extraversion, Neuroticism	.089	.005
4	Gender, Extraversion, Neuroticism, Openness	.089	.000
5	Gender, Extraversion, Neuroticism, Openness, Agreeableness	.129	.040†
6	Gender, Extraversion, Neuroticism, Openness, Agreeableness, Conscientiousness	.171	.042*

Model (b)	Predictors of Achievement	R^2	Change in R^2
1	Extraversion	.003	.003
2	Extraversion, Neuroticism	.025	.022
3	Extraversion, Neuroticism, Openness	.026	.001
4	Extraversion, Neuroticism, Openness, Agreeableness	.094	.067*
5	Extraversion, Neuroticism, Openness, Agreeableness, Conscientiousness	.154	.067*
6	Extraversion, Neuroticism, Openness, Agreeableness, Conscientiousness, Gender	.171	.018