Stereotype threat:

The role of question type in female maths performance

Lucy Charlotte Davies

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Institute of Psychological Sciences

October 2013

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Dr. Russell Hutter. Without his support, encouragement, and expertise I would have been lost. I would like to thank my supervisors Prof. Mark Conner and Prof. Constantine Sedikides for their guidance and manuscript feedback. Thank you also to Dr. Chantelle Wood for her advice and technical assistance. I would like to thank Brian Wilkinson and St. Aidans School for their collaboration, and thank you to the Leeds Institute of Psychological Sciences for the funding which allowed me to undertake this research. Finally, a special thanks to my friends and family for their endless support and encouragement.

Abstract

Contemporary educational environments rely heavily on performance-based evaluations. Indeed, maths performance at GCSE has become a matter of national debate.Stereotype threat (ST) is often cited as an explanation for gender differences in maths performance, yet little research has focused on differential ST susceptibility across question types. Furthermore, despite its high relevance to education, the ST phenomenon has received little attention in real examination settings. This research builds on Jamieson and Harkins (2007) *mere effort* approach to investigate ST in both lab and field studies. Mere effort theory proposes that ST motivates test-takers to disprove an active negative stereotype. When responding to *solve* type questions based on prepotent (i.e., well learned) knowledge, activation of a negative stereotype can motivate test-takers and improve performance. However, *comparison* type questions (requiring logic or estimation), often result in performance decreases, because test-takers seek to disprove the negative stereotype leading to a failure in inhibiting prepotent (i.e. solve) information. Findings from Study 1 supported the mere effort perspective; threatened females maths performance was dependent on question type. Study 2 showed that the effects transferred to educational setting during an undergraduate statistics practise exam. In Study 3, female and male secondary school pupils were tested in a GCSE maths practise exam environment. The interactive effects of ST and question type were replicated in females' maths performance, whereas males' maths performance was augmented under ST irrespective of question type. The focus moved to mere effort's ST processes in Studies 4 and 5. A moderating role of *inhibitory ability* as a ST protective mechanism was found in Study 4. However this was not specific to

comparison question performance, and thus suggested that the overproduction of prepotent responses is not the main processes driving ST effects. In Study 5, ST seemingly increased test-takers *performance motivation* (i.e., the motivation to perform well and undermine the stereotype), influencing their question type *preference* for solve versus comparison questions. The present research attests to the important role of maths question type in determining ST effects. However, the motivated application of prepotent responses as an explanatory mechanism is questioned and discussed with reference to the alternative *working memory* (WM) (Schmader & Johns, 2003) ST explanation.

Contents

List of Tables

List of Figures

Chapter 1

Thesis overview

“Girls like to have methods clearly defined and explained and to be shown how to use them. Girls’ success can be attributed to their ability to follow rules rather than ‘real understanding’.”

- Leicestershire Primary Team (2005) ‘Girls’ achievement in mathematics’

1.1 Investigating maths question type and stereotype threat

As the above quote illustrates, there is widely held belief that girls and boys have different approaches to mathematical problems. Girls favour a more rule-based structure of learnt formulas and equations, whereas boys prefer a more unstructured approach (Gallagher et al., 2000; Royer, Tronsky, Chan, Jackson, & Marchant, 1999). Furthermore, the quote also seemingly implies a pernicious maths-gender stereotype operates in our society. The stereotype that ‘women are poorer at mathematics’ may be *threatening* to performance; and undermines females’ mathematical ability in test situations (Nosek et al., 2009; Picho, Rodriguez, & Finnie, 2013). This may help explain the gender-gap in maths performance. Indeed, recent UK A-level mathematical exam results revealed 18% of male candidates received A* grades (a slight increase on 2012), in comparison to only 14.8% of female candidates; a 1.8 percentage point fall from 2012 (Adams, 2013).

A plethora of stereotype threat (ST) research has focused on how the maths-gender stereotype affects female maths performance based on question *difficulty* (e.g., O'Brien & Crandall, 2003; Steele, Spencer, & Aronson, 2002). However, it is unclear how ST effects may interact with the *type* of maths question. Specifically, does variation in how a question can be answered differentially impact maths performance under ST? A recent ST explanation that focuses on differences in question type is Jamieson and Harkins' (2007) *mere effort* account. The current research uses mere effort to investigate how maths question type may differentially affect females' maths performance in response to the maths-gender ST.

1.2 Mere effort

Are female test-takers more susceptible to reduced performance following ST on some types of maths questions more than others? Mere effort (Jamieson & Harkins, 2007) proposes that the motivation to disprove the negative stereotype potentiates whatever response is prepotent (i.e., most likely to be produced) on a task. Threatened test-takers' performance is determined by whether the potentiated prepotent response is the correct approach for the *type* of question encountered (Jamieson & Harkins, 2007, 2009). That is, if the question can be correctly worked out using the prepotent response. In the context of maths, female test-takers' prepotent response is to apply a *solve* approach using learnt formula and knowledge (Jamieson & Harkins, 2009; 2012). Thus, in response to ST, females become motivated to disprove the maths-gender stereotype, resulting in the overproduction and application of the prepotent solve response. This facilitates performance on *solve* questions based on prepotent learnt knowledge, but debilitates performance on *comparison* questions (that require an approach based on logic and estimation). The present research uses solve and comparison questions to investigate how differences in question type, based on the prepotent response, affect ST effects on maths performance.

1.3 Thesis aims and structure

The thesis is structured as follows: First, Chapter 2 reviews the maths-gender ST literature and outlines the mere effort account of ST, and then tackles some of the issues that remain to be resolved. Second, in Chapter 3, the outcomes of mere effort are replicated and extended. Third, the applicability of mere effort effects to educational settings are tested in Chapter 4. Fourth, Chapters 5 and 6 are devoted to testing the processes hypothetically underpinning mere effort. Finally, the current findings are summarised and discussed with reference to previous work, and future research directions are suggested in Chapter 7. This is outlined in greater detail below.

In Chapter 2 a review of the maths-gender ST literature is presented and the mechanisms underlying ST outlined. Specifically, the motivation-based mere effort account is focused upon, and an explanation of how question type interacts with ST is detailed. Evidence for mere effort effects in the maths-performance domain are discussed, alongside research investigating gender differences in mathematical problem solving. Additionally, literature pertaining to motivational and inhibitory processes is detailed and linked to the proposed mechanisms underlying mere effort. In particular, (via inhibition literature) links between mere effort and the alternative working memory (WM) explanation of ST (Schmader & Johns, 2003) is explored. Finally, issues surrounding ST replicability are discussed, as well as the importance of ST field research. From the review three key research aims are identified: (a) to test the maths-gender ST on the outcome of females' maths performance, based on maths question type (i.e., solve vs. comparison); (b) to test the mere effort account in educational settings (i.e., university and school examinations); and (c) to test mere effort's prepotent responses and motivational ST processes.

Chapter 3 addresses the first aim to investigate Jamieson and Harkins' (2009) mere effort account in the maths performance domain. Study 1 tests the hypothesis that in response to the maths-gender ST, females' maths performance will be augmented on solve questions and debilitated on comparison questions (c.f. controls). Study 1 is one of the few studies investigating ST effects based on question type (Beilock, Rydell, & McConnell, 2007; Jamieson & Harkins, 2009, 2012) and uses GCSE rather than Graduate Record Examination (GRE) maths questions. Maths performance is measured using both (a) unadjusted maths performance that is used to mark real examinations such as GCSEs, and (b) the adjusted percentage of problems solved, as is typical of ST research (e.g., Jamieson & Harkins, 2009). Thus, Study 1 tests to see if the interactive effects of question type and ST are generalizable to different maths tests using real test marking formats.

Next, in Chapter 4 the relevance of ST to educational equality (Huguet & Regner, 2007) and the deficit of ST research in educational settings (Wicherts, Dolan, & Hessen, 2005) is addressed. Studies 2 and 3 investigate whether the interactive effects of ST and question type are replicated during university and secondary school examination conditions respectively. Furthermore, Study 3 includes a male cohort to investigate how ST may affect males' maths performance and is therefore relevant to the maths-gender performance gap (Stoet & Geary, 2012).

In Chapters 5 and 6, the focus turns to mere effort's mechanisms to address the final research aim. Specifically, research is presented that investigates the processes hypothetically driving mere effort; *prepotent response inhibition* and *motivation*, respectively. In Chapter 5, Study 4 examines the potential moderating role of inhibitory ability. It is hypothesised that, under ST, inhibitory ability will protect maths performance on comparison questions. High inhibitors will be more able to suppress the incorrect solve approach and apply the correct comparison approach. As the prepotent solve response does not need to be inhibited for solve questions, performance on solve questions should be unaffected by inhibitory ability.

The role of performance motivation in response to the maths-gender ST forms the basis for Chapter 6. Previous mere effort research has only measured performance motivation using task performance itself (Jamieson & Harkins, 2007; McFall, Jamieson, & Harkins, 2009). This creates difficulty distinguishing ST processes from ST effects. Study 5 adapts an existing independent motivation measure from Forbes and Schmader (2010) (which crucially remains maths orientated). This allows a test of performance motivation using a maths question type *preference* task (administered following the main performance measure). Specifically, if ST inherently motivates and leads test-takers to apply a prepotent solve approach, participants should demonstrate greater preference to select solve questions (i.e., where the prepotent response can be successfully applied) than comparison questions.

Finally, a discussion and interpretation of the overall findings is presented in Chapter 7. The main findings illustrate the importance of question type in determining how the maths-gender ST impacts female maths performance, and that the ST effects (based on maths question type) are generalizable to females' maths performance in educational setting. ST is also shown to enhance males' maths performance (i.e., stereotype lift), irrespective of question type, that may contribute to the maths-gender performance gap. Furthermore, the discussion highlights difficulties in differentiating the present findings as evidence for the mere effort versus the WM accounts of ST. In particular, although ST seemingly heightens participants' motivation, the ability to inhibit the prepotent response protected maths performance *overall*. This was regardless of whether the question could be correctly answered using the prepotent response. An integrated approach of ST is outlined, that provides a more unified perspective of the mere effort and WM accounts to explain the ST phenomenon. The limitations of the present research are discussed and future research ideas identified.

1.3.1 Summary

To summarise, using the mere effort account, the present research examines how different *types* of maths questions may affect maths performance in response to the maths-gender ST. This is also tested in field work. In addition, mere effort's proposed ST mechanisms (i.e., prepotent response inhibition and performance motivation) are investigated. The implications of the findings and discussion of the WM account of ST, as both an alternative and potential complimentary explanation, are addressed. This facilitates a better understanding of how differences in question type may affect the relationship between the maths-gender ST and female maths performance. It also informs those involved in educational settings as to some of the issues impeding performance in test situations.

Chapter 2

Stereotype threat: The role of question type in female maths performance

---

2.1 Aims of the literature review

Boys have outperformed girls at GCSE (General Certificate of Secondary Education) maths for the third year in a row (Shepherd, 2011). This has widely been accredited to the decision to drop coursework in the subject, increasing the importance of maths examinations. As noted by a former chief examiner, girls often feel less confident in high-stakes testing than boys. Consequently, singular assessments may have a considerable gender impact on results and be discriminatory against girls (Sellgren, 2013). With significant increases in students' selection of maths and science courses at GCSE and A-level (Department of Education Report, 2010), and universities favouring top grades in these traditional subjects (Shepherd, 2011), it is paramount to ensure all students are able to perform to their full ability in examination conditions.

The identification of factors that facilitate or hinder performance is essential. An instance of debilitated performance that has come to the fore in recent years is the stereotype threat effect (ST). ST is the situational phenomenon that results in reduced performance following exposure to a salient negative self-relevant stereotype associated with the task (Steele, 1997; Steele & Aronson, 1995). In particular, ST occurs when people feel that their performance will be *evaluated* in light of a negative stereotype, and a *fear* that they may confirm the negative stereotype (Brodish & Devine, 2009). Indeed, in situations where the stereotype may apply, a malaise confronts individuals sensing that evaluation will result from anything they do or from any personal qualities that fit the stereotype (Spencer, Steele, & Quinn, 1999; Steele & Aronson, 1995). Thus, in test situations, ST may hinder the ability of stigmatised individuals through concerns that their performance will be associated with the stereotype.

In this chapter, evidence for a maths-gender ST will be explored alongside an examination of how ST affects maths performance. The current ST literature will be discussed and reviewed to identify gaps in the research and areas for development. Specifically, the focus will be on the mere effort account (Jamieson & Harkins, 2007) of ST. This is because mere effort, unlike the dominant cognitive models of ST (e.g., Schmader, Johns, & Forbes, 2008), focuses on how differences in the *type* of question may interact with ST effects. Mere effort's *motivation* and *prepotent response* ST processes can potentially explain how a variation in question type may differentially affect maths performance under ST. Therefore, motivational and prepotent response inhibitory mechanisms will be outlined. This will ultimately inform the studies to be conducted as part of the current research.

2.2 The maths-gender stereotype threat

Researchers have argued that ST is intrinsic to performance environments where negative stereotypes may apply (e.g., Steele, 1997; Steele & Aronson, 1995). Indeed, in accordance with earlier reports of boys outperforming girls in GCSE maths examinations (Shepherd, 2011), research suggests that there exists a pernicious maths-gender stereotype that may undermine females' mathematical ability in test situations (Nosek et al., 2009; Picho, Rodriguez, & Finnie, 2013). The stereotype 'men are better at mathematics' or 'women are poorer at mathematics' triggers female test-takers' concern that their performance may be evaluated by or conform to the negative stereotype (Shapiro & Neuberg, 2008). This concern disrupts and undermines their mathematical performance (Schmader et al., 2008). Furthermore, research has proposed that ST is more likely to significantly affect individuals who have a high self-investment in the stereotype domain (Aronson, Lustina, Good, Keough, & Steele, 1999; Martens, Johns, Greenberg, & Schimel, 2006; Steele, Spencer, & Aronson, 2002), and are high-achieving members of the stereotyped group (Steele, 1997). Consequently, even female students who enjoy mathematics and *believe* that boys and girls perform equally can still be susceptible to the maths-gender ST.

Research has also indicated that the social transmission of gender-related maths attitudes may serve to perpetuate gender-stereotypical roles and reinforce the ST (Eccles & Jacobs, 1986; Eccles, Jacobs, & Harold, 1990; Jacobs & Eccles, 1992). For example, parent and teacher expectancies for children's maths competence are often gender-biased, and play a critical role in children's maths performance, maths course-taking, and pursuit of maths-related career paths (Beilock, Gunderson, Ramirez, & Levine, 2010; Gunderson, Ramirez, Levine, & Beilock, 2012). Indeed, the gender gap in maths performance is typically not observed until middle school (Hyde, Fennema, & Lamon, 1990; Maccoby & Jacklin, 1974). This indicates that gender differences in maths performance are the result of a strong pattern of socialisation to maths success or failure, rather than gender differences in innate ability (Schwartz & Hanson, 1992). Negative stereotypes are therefore deeply ingrained in our society. It is essential to investigate the underperformance of stigmatised groups in performance settings (such as females in maths examinations) in order to understand how the ST effect operates (Forbes & Schmader, 2010).

2.2.1 Testing the maths-gender stereotype threat

The ST effect has been tested in female sample groups via a range of experimental manipulations (e.g., explicitly stating that men out-perform women on the test; Brown & Pinel, 2003; Ford, Ferguson, Brooks, & Hagadone, 2004; Keller, 2002; Keller & Dauenheimer, 2003; Quinn & Spencer, 2001; Spencer et al., 1999). Other common maths-gender ST manipulations include: informing women that their maths performance will be compared to the maths performance of men (Rosenthal & Crisp 2009; Schmader, 2002); making women the gender-minority in the maths test environment (Ben-Zeev et al., 2005; Schmader & Johns, 2003; Sekaquaptewa & Thompson, 2003); informing women that an upcoming test is diagnostic of mathematical ability (Cadinu, Maass, Lombardo, & Frigerio, 2006; Schmader & Johns, 2003); and making gender-identity salient (i.e., asking participants to indicate their gender on a questionnaire (Schmader, 2002; Schmader & Johns, 2003).

The implications of ST are disconcerting: the experimental conditions analogous to many 'real-world' situations that women routinely encounter in standardised ability tests (e.g., scholastic examinations such as GCSEs, or employment selection contexts) indicate that females' maths potential may be significantly impeded (Eccles et al., 1990; Hyde et al., 1990; Steele & Davies, 2003). Indeed, research has suggested that women are less likely to persist in stereotyped domains, creating a gender gap in science, technology, engineering and maths (STEM) fields (Gunderson et al., 2012; Major, Spencer, Schmader, Wolfe, & Crocker, 1998). It is therefore paramount to address ST in the educational system. This will help to alleviate ST's detrimental effects on performance, subject selection and persistence in the stereotyped domain, and will ultimately help to ensure equal opportunity for all students.

2.2.2 Stereotype threat performance effects

A variation exists in the degree to which ST affects performers. The performance decrements in response to ST have been well documented (Brown & Day, 2006; Cole, Matheson, & Anisman, 2007; Good, Aronson, & Harder, 2008; Good, Aronson, & Inzlicht, 2003; Neuville & Croizet, 2007). Indeed, in a recent meta-analysis synthesising 17 years of ST research, Picho et al. (2013) found that, on average, ST debilitated female test-takers' maths performance compared to their non-threatened counterparts. Alternatively, a reactive effect (i.e., a performance increase) has sometimes been observed when the ST is encountered.

2.2.2.1 Reactivity

One reactive effect is *stereotype lift*; a tangible increase in performance when participants make downward comparisons with outgroups considered stereotypically poorer at the task (Chalabaev, Stone, Sarrazin, & Croizet, 2008; Walton & Cohen, 2003). In other words, participants do not necessarily believe they are good at a specific task, but believe that they are relatively better than other participants at the task. For example, male participants performed better on a maths test when they were made aware of the negative female maths stereotype (Walton & Cohen, 2003). A similar, although distinct, performance enhancing effect has been observed when an alternative *positive* self-relevant identity is emphasised to participants. This is a *stereotype boost* effect, whereby in a performance domain, an individual may be negatively stereotyped by one identity and positively stereotyped by another. For example, Asian American women demonstrated improved maths performance when their ethnic identity was made salient (i.e., stereotype boost effect), and poorer performance when their gender identity was made salient (i.e., ST effect). The observed performance effects were consistent with the respective group stereotypes (i.e., "Asians are good at maths," "females are poor at maths") (Shih, Ambady, Richeson, & Fujita, 2002; Shih, Pittinsky, & Ambady, 1999).

There are, however, occasions when reactivity arises following exposure to ST that are *not* triggered by either downward comparisons or emphasis on alternative identities (e.g., Jamieson & Harkins, 2009). The present research aims to build upon current findings in order to more fully understand the maths-gender ST and its differential performance effects (i.e., performance enhancement or debilitation). Clearly, ST has a substantial influence on performance; however the mechanisms driving the effects and how they interact are still unclear and disputed (Jamieson & Harkins, 2007; Schmader et al., 2008).

2.2.3 <u>Stereotype threat mechanisms</u>

In order to understand the full impact of ST on maths performance, researchers must first investigate the underlying ST mechanism(s) driving ST effects (McFall, Jamieson, & Harkins, 2009). Research has documented a number of seemingly competing explanations, implicating: reduced *working memory* (WM) *capacity* (Beilock, Rydell, & McConnell, 2007; Bonnot & Croizet, 2007; Schmader et al., 2008); *anxiety* (Bosson, Haymovitz, & Pinel, 2004; Spencer et al., 1999);

increased *arousal* (Ben-Zeev et al., 2005; Murphy, Steele, & Gross, 2007; O'Brien & Crandall, 2003); *expectancy* (Cadinu, Maas, Frigerio, Impagliazzo, & Latinotti, 2003); *withdrawal of effort* (Stone, 2002; Stone, Perry, & Darley, 1997); *prevention focus* (Keller & Dauenheimer, 2003; Seibt & Forster, 2004); and *mere effort* (Jamieson & Harkins, 2007, 2009, 2012).

Indeed, it is likely that ST is mediated by multiple processes (Steele et al., 2002). For instance, Schmader et al. (2008) developed a ST process model that integrated research on stress arousal, vigilance, WM, and self-regulation. However, while the motivational processes associated with ST were described, the model focused upon how performance is harmed by the disruption to WM resources required for the task. In contrast, Jamieson and Harkins's (2007) mere effort account focuses on how high levels of motivation in response to ST drive ST effects. The heightened levels of motivation lead to ST's differential performance effects (i.e., reduction and increases in performance) based on the *type* of question encountered. Thus, in order to investigate how a variation in maths question type interacts with ST effects, the current research will capatalise on mere effort. The following section details and outlines the mere effort explanation of ST, before more closely linking it to maths performance.

2.3 The mere effort account

Jamieson and Harkins's (2007) research attests to the importance of the motivational component of participants' responses when encountering ST in performance domains. According to the mere effort explanation of ST, when a negative stereotype is associated with performance, individuals actively set out to perform well and undermine the stereotype (Harkins, 2006; Jamieson & Harkins, 2007). This goal potentiates a previously well-learned, *prepotent* response (Jamieson & Harkins, 2007; McFall et al., 2009). Specifically, the prepotent response is a habitual response tendency that is most likely to be produced in a given situation (or context) (Grandjean & Collette, 2011). Performance under ST is therefore dependent on whether the prepotent response is the correct approach or not to answer the question. Thus, while previous research has argued that ST tends to facilitate performance on *simple* maths problems but debilitates performance on *complex* ones (O'Brien & Crandall, 2003; Steele et al., 2002). In contrast, mere effort argues that it is *problem type* rather than difficulty per se that affects performance following ST (Jamieson & Harkins, 2009). That is, if the *type* of question can be correctly answered using the potentiated prepotent response. Therefore, according to mere effort, differences in how people engage with the type of problem encountered are a key part in understanding ST effects. To recap, the motivation to disprove the activated stereotype enhances the prepotent response (see Figure 2.1); if the prepotent response is correct, performance is facilitated (i.e., reactivity); if not, performance is inhibited (i.e., the ST effect)

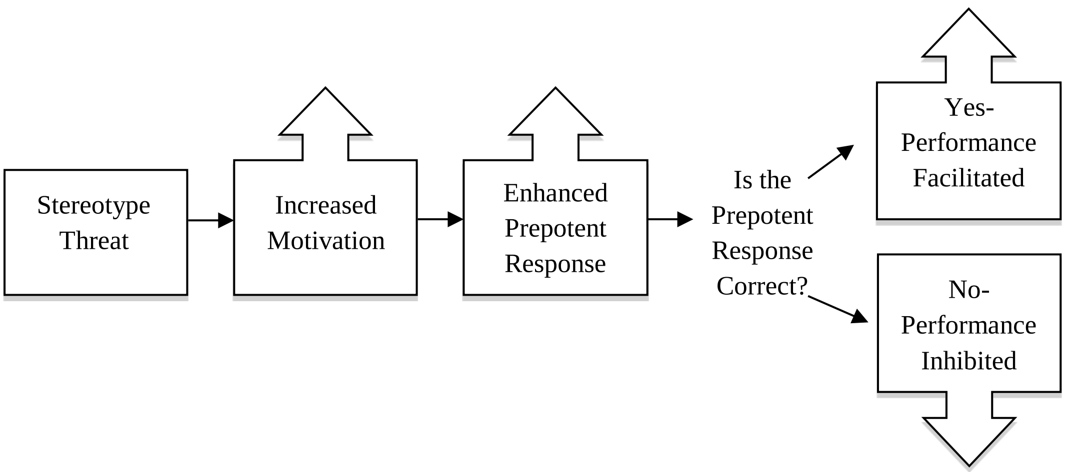


Figure 2.1. The mere effort account of stereotype threat on performance.

The mere effort explanation is a derivative of drive theory (Zanjonc, 1965), in which dominant responses occur as a function of increased arousal (i.e., drive, motivation) (Ben-Zeev et al., 2005; Cottrell, 1972; Jamieson & Harkins, 2009; O'Brien & Crandall, 2003; Zajonc, 1965). Physiological arousal facilitates dominant responses and inhibits non-dominant ones (Hull, 1943). This process operates via a bottom up-up mechanism, whereby motivation increases the likelihood of generating the prepotent response, rather than impairing top-down control (Jamieson & Harkins, 2011). In other words, motivation is the catalyst in the response process; it builds the response up from base level rather than breaking it down. Harkins and his colleagues also proposed that, as well as potentiating prepotent responses, the newfound motivation directs effort towards correcting inaccurate responses (Jamieson & Harkins, 2007; McFall et al., 2009). Correction, however, can only occur if perceivers recognise their response as inaccurate, have knowledge of the correct response, and are in a position to implement the response (Jamieson & Harkins, 2007).

Mere effort therefore provides a cogent account for how ST may facilitate or debilitate performance based on question type (Jamieson & Harkins, 2007). The following section will outline mere effort's predictions as specific to the maths-gender performance domain and mathematical problem solving (Jamieson & Harkins, 2009, 2012).

2.3.1 Mathematical problem solving

Mathematical problem solving has been defined as a process that involves several dynamic activities including: understanding the problem, making a plan, carrying out the plan, and revision (Willson, Fernandez, & Hadaway, 1993).

Similarly, research has supported the notion that a maths problem solver must both correctly *interpret* the problem and correctly *execute* the problem (Montague, 2006; Royer & Garofoli, 2005). In other words, representing and understanding the problem is the basis for finding a successful solution to the problem. A base of mathematical knowledge is needed, which is then organised into a specific set of applications and heuristics (i.e., strategies and techniques) (Willson et al., 1993). During the problem solving process, an individual might apply a number of different strategies (i.e., solution rubric, logical mathematical reasoning, trial and error, etc.) in order to correctly answer the question (Gallagher et al., 2000).

2.3.1.1 Question (problem) type

The prepotent (i.e., dominant) strategy of working out the correct answer to quantitative maths problems is to apply the *solve* approach (Jamieson & Harkins, 2009, 2012). In this approach test-takers apply *learnt* formulas and equations to compute an answer, in contrast to the *comparison* approach, in which logic, estimation, or intuition is applied (Gallagher & De Lisi, 1994; Gallagher et al., 2000; Quinn & Spencer, 2001) (see Figure 2.2). In particular, females demonstrate a stronger preference for the solve approach than males do, indicating that males and females have different problem solving patterns (Gallagher et al., 2000; Royer, Tronsky, Chan, Jackson, & Marchant, 1999; Tartre, 1990). Indeed, Gallagher et al. (2000) suggested that males tended to be more flexible than females in applying solution strategies. Females tended to adhere to classroom-learned procedures when solving maths problems more than their male counterparts. This suggests that women are less likely to use shortcuts and estimation techniques for solving unfamiliar and complex problems (Gallagher, 1998). The gender differences in problem solving strategies have been linked to a range of different variables, such as learners' psychological characteristics (Meyer, Turner, & Spencer, 1997), teachers' beliefs and instructions (Carr, Jessup, & Fuller, 1999), learning styles (Kimball, 1989; Schwartz & Hanson, 1992), and classroom structure (Pearson & West, 1991; for a review see Zhu, 2007).

Solve Type:

Work out $3\frac{3}{4} \div 1\frac{2}{3}$

Comparison Type:

(a) Oscar has a spinner with eight sections.
Four of the sections are Red, two are Green, one is Blue and one is Yellow.
He spins the spinner 200 times.
His results are shown in the table.

Colour	Red	Green	Blue	Yellow
Frequency	105	48	22	25

Do the results suggest that the spinner is fair?
Explain your answer.

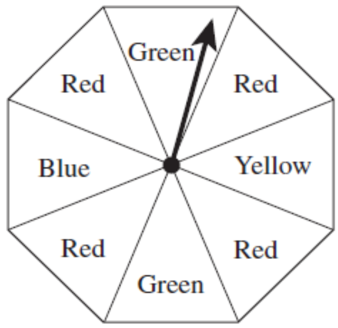


Figure 2.2. Example of solve and comparison type maths questions.

Females' preference for the solve approach in mathematical problem solving supports the idea that the solve approach is their prepotent (i.e., dominant) approach. According to mere effort, this solve response will be potentiated (i.e., more strongly used) under ST. Therefore, when investigating ST on maths performance, it is important to examine question type as ST may influence females' selection of problem solving strategies (Quinn & Spencer, 2001).

2.3.1.2 Stereotype threat and question type

The mere effort account has been tested using maths question type. Jamieson and Harkins (2009) indexed performance using two types of maths problems that differ in the most efficient approach to answer the question: *solve problems* (e.g., solve an equation) and *comparison problems* (e.g., logic and estimation). Consistent with mere effort's predictions; threatened female participants performed better than controls on solve problems, but less well than controls on comparison problems. The threatening stereotype that 'women are bad at maths' increased female test-takers' motivation to perform well in order to disprove the stereotype, subsequently enhancing their prepotent response to apply the solve approach (see Figure 2.3). The prepotent solve approach is correct and facilitated performance for solve questions, but is incorrect and debilitated performance for comparison questions (i.e., where the comparison approach is correct).

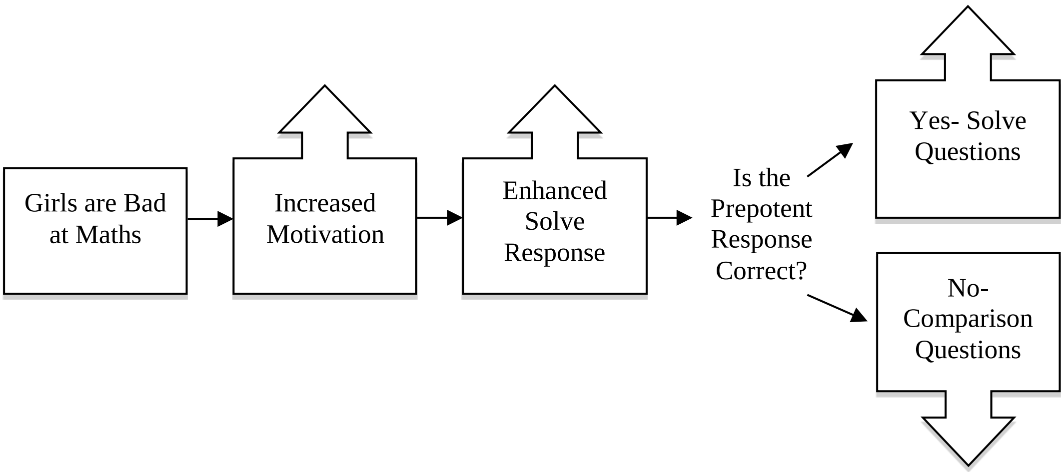


Figure 2.3. The mere effort account of the maths-gender stereotype threat on maths performance dependent upon question type.

The effect of debilitated performance on comparison problems is greater than the effect of facilitated performance on solve problems (Jamieson & Harkins, 2009). In other words, under ST, performance is much worse on comparison problems than it is better on solve problems. This occurs because the prepotent response of solving equations (i.e., relevant formulas and operations) is female test-takers preferred maths approach (Gallagher et al., 2000; Royer et al., 1999). The solve approach is generally known and applied by all females test-takers, but females subject to ST are more motivated to apply this approach than non-threatened females. Thus, for solve questions, the enhanced performance under ST is limited to females' increased effort to disprove the stereotype within the restricted examination time. In contrast, for comparison questions, performance is more strongly debilitated by ST. The potentiated solve approach serves to handicap test-takers comparison mathematical ability: Threatened female test-takers must both recognise that the solve approach is wrong and adopt the correct comparison approach. Furthermore, the more motivated threatened test-takers are, the stronger the solve approach will be potentiated, creating greater decrements to comparison performance. Therefore, in all, threatened females perform better on solve problems, more poorly on comparison problems, and more poorly overall than their non-threatened counterparts (Jamieson & Harkins, 2009).

This has considerable implications for maths revision techniques; revising learnt maths solutions and thereby strengthening the solve approach may aid answering solve examination questions. However, this may be detrimental to questions that require a comparison approach to find the correct answer. It may not always be the case that 'practice makes perfect', and that conversely, students best efforts may actually be harmful to their performance. Consequently, although research has suggested that performance can be improved by breaking the link between stereotypes and performance (Johns, Schmader, & Martens, 2005); efforts should be made to address how to assist stigmatised individuals in focusing their motivation more effectively during task performance.

The mere effort account thus is defined by and differs from other accounts of ST. Mere effort argues that negative effects on cognitive capacities (e.g., processing interference; Schmader & Johns, 2003) or withdrawal of effort (e.g., test-anxious students appear to become less motivated in evaluative contexts; Hancock & Dawson, 2001) are not the primary mechanisms underpinning ST effects. In contrast, the perspective argues that situations involving ST increase the *motivation to perform well* in order to disconfirm the negative self/group affiliated stereotype (Jamieson & Harkins, 2007; O'Brien & Crandall, 2003). However, these efforts may be misdirected because ST potentiates the prepotent response that may not always be contextually correct. In other words, awareness of a negative stereotype can fuel the motivation to disprove it, but the motivation may sometimes, erroneously, be applied to the incorrect approach. In the next section mere effort's ST mechanisms (i.e., motivation and prepotent responses) will be more closely examined, as will methods of measurement.

2.4 Mere effort processes

Process-oriented ST research requires new methods to be adapted to move from documenting ST effects to measuring the underlying mechanisms (Jamieson & Harkins, 2011). It is important to test mere effort's ST mechanisms independent from ST effects to clarify *how* ST operates, and to consider whether other competing explanations (e.g., WM; Schmader & Johns, 2003) are also relevant (see Schmader et al., 2008).

2.4.1 Motivation

It is clear that motivation is a key component of mere effort. ST motivates individuals to place their efforts in one direction (i.e., the prepotent response) even if this is not the correct approach to take. However, mere effort does not make predictions for effort or motivation *independent of* task performance (Jamieson & Harkins, 2007). The motivation outlined is specifically *performance motivation*, intrinsic to the task itself and therefore can not be disassociated or measured separately (Jamieson & Harkins, 2011). In other words, the process and outcome variables are one and the same. Therefore, unlike other forms of motivation that can be more explicitly measured (Tapia & Marsh, 2004); Jamison and Harkins (2007, 2011) argue that performance motivation can only be measured using task performance.

However, using maths performance as evidence for *both* ST effects and ST mechanisms creates difficulties when interpreting findings for *how* ST operates. The present research will use a novel process-focused designed study to measure performance motivation separately from task performance. For example, Forbes and Schmader (2010) implemented a maths motivation task using a question choice design (maths vs. verbal questions). Preference to answer maths questions over the verbal questions provided an index of maths motivation. The present research will adapt Forbes and Schmader’s (2010) study design to incorporate question type (solve vs. comparison questions). Thus, *question type preference* under ST will provide another potential indicator of performance motivation. Specifically, under

ST, it is expected that motivated individuals will have a preference to select solve questions (c.f. comparison questions) as the prepotent solve approach will have been activated.

To summarise, mere effort's performance motivation will be tested using the traditional index of maths score, and the novel separate index of question type preference. Motivation following threat may initiate the mere effort ST process (see Figure 2.1) but what other mechanisms might be involved? One potential candidate that may play a moderating role is the ability to *inhibit* potentiated prepotent responses.

2.4.2 Inhibitory ability and prepotent responses

Inhibition is an executive function that controls an individual's capacity to be able to block out cognitive interference (von Hippel, Silver, & Lynch, 2000). Stated otherwise, inhibitory ability works to keep irrelevant information from entering the focus of attention and suppresses automatic, prepotent responses that are inappropriate for the task at hand (Friedman et al., 2008; Hasher, Quig, & May, 1997). In the context of the threatening stereotype 'women are bad at maths', inhibitory ability may therefore serve to suppress the prepotent response generated (i.e., the solve response). This enables other approaches (i.e., the comparison response) to be considered and applied. Indeed, Carr and Steele (2009) proposed that the experience of ST may induce a fixed way of thinking; ST interferes with the ability to inhibit old strategies in order to develop more successful ones for problem solving.

However, rather than being a single unitary construct, inhibition-related processes are a family of functions that can be clustered into several distinct categories (Friedman & Miyake, 2004; Hasher, Lustig, & Zacks, 2007). For example, Hasher and colleagues proposed three functions of inhibition: access, deletion, and restraint. Specifically, the *restraint* function is most closely associated with the mere effort account. The restraint function of inhibition suppresses automatic, prepotent responses to allow for other, more contextually appropriate, responses to be considered and applied (Hasher et al., 2007). Similarly, Friedman and Miyake (2004) defined the distinct inhibition function of *prepotent response inhibition*; the ability to deliberately suppress dominant, automatic, or prepotent responses. Taken together, this suggests that individuals with higher levels of prepotent response inhibition (or restraint) will be more able to suppress prepotent solve responses potentiated by ST.

The Stroop task (Macleod, 1991; Stroop, 1935) is a classical paradigm commonly used to assess prepotent response inhibition (e.g., Dao-Castellana et al., 1998; Davidson, Zacks, & Williams, 2003; Friedman & Miyake, 2004; Mutter, Naylor, & Patterson, 2005). The task is to inhibit a dominant habitual response (i.e., reading colour words) to apply a different novel requirement (i.e., naming the colour the words are printed in). For example, in a typical Stroop task, participants are presented with one of three trial types and asked to name the ink colour of the stimuli. Trials consist of congruent (e.g., the word “RED” printed in red ink, respond red), incongruent (e.g., the word “RED” printed in blue ink, respond blue), or neutral stimuli (e.g., “XXX” printed in red ink, respond red). The longer durations to complete the incongruous trials compared (i.e., minus) to the congruous trails can indicate an inability to inhibit prepotent but contextually inappropriate responses (Stroop, 1935). In other words, individuals that perform poorly (i.e., more slowly between conditions) on the Stroop have poor response inhibition. The present research will implement a Stroop task to assess if a variation in prepotent response inhibition moderates ST effects on maths performance dependent on question type. Specifically, does the ability to suppress incorrect prepotent responses generated under ST enable other approaches to be applied? The ability to inhibit the solve approach would potentially protect threatened test-takers' maths performance for comparison questions.

There is a gap in ST literature investigating the potential role of motivated inhibitory ability on performance under ST. Furthermore, inhibitory ability would provide a clear link between mere effort (Jamieson & Harkins, 2007) and WM (Schmader & Johns, 2003) explanations of ST. Specifically, inhibition is a component of WM. Research has proposed that ST lowers performance by reducing WM capacity (Beilock et al., 2007; Rydell, McConnell, & Beilock, 2009; Schmader & Johns, 2003), thus interfering with the ability to undertake the task at hand. Those with higher (vs. lower) WM ability may be better equipped to cope with ST (Regner et al., 2010; Schmader et al., 2008), as research has shown taxing WM resources increases the difficulty of inhibiting prepotent responses (Grandjean & Collette, 2011). Taken together, it follows that individuals with higher inhibitory ability (i.e., greater WM ability) may self-protect (Sedikides, 2012; Sedikides & Green, 2009) by being more able to suppress incorrect prepotent response tendencies. Testing inhibitory ability (indexed by Stroop) offers a unique opportunity to assess an individual differences factor that should interact with threat and question type.

Hence, the role of question type is a key determinant of maths performance under ST. The experience of ST may motivate test-takers to perform well and undermine the stereotype, potentiating prepotent responses. Performance is determined by whether the prepotent response is the correct approach or not (i.e., if it needs to be inhibited). The present research will investigate ST effects based on question type, and also will aim to directly test mere effort's ST mechanisms (i.e., motivation and prepotent responses) in process-focused studies. Furthermore, as ST is based on real world phenomena, it is important not only to test ST effects and mechanisms in the lab, but also to replicate these effects in the field.

2.5 Replicating stereotype threat in lab and field research

The importance of replication has come to the fore in recent years following the scientific fraud of a prolific social-cognitive psychology researcher (Levelt Commission, 2012). Indeed, the robustness of ST has been questioned (Ganley et al., 2013), and there have been suggestions of publication bias towards significant ST findings (Stoet & Geary, 2012). Thus, in order to provide reliable replication the same maths task will be used to test ST across the lab studies in the current work. This will enable a clearer interpretation of the potential interactive effects of question type and ST.

Furthermore, despite its high relevance to education, at present the maths-gender ST has received little attention in real exam or school settings (Huguet & Regner, 2007; Wei, 2012; Wicherts, Dolan, & Hessen, 2005). As previously discussed, the mere effort account provides a potential explanation for the observed gender differences in maths examination performance such as GCSEs (see Section 2.1). The lab offers a controlled environment to test the account as an explanation. However, it is important to also replicate these effects in the field. Indeed, in a recent replication and extension of Anderson, Lindsay, and Bushman's (1999) novel work measuring the external validity of laboratory research, Mitchell's (2012) meta-analysis revealed that although overall psychology lab studies usually replicate real-world (r = .71), Social psychology needed the most improvement (r = .53). In particular, of the different Social psychology topics, lab studies of gender differences were least likely to translate to real world, which may be due to the small effect sizes often found in these studies (Mitchell, 2012). Consequently, it is paramount that the current research tests the mere effort account of ST in the field, in order to establish to what degree ST effects based on maths question type generalise beyond the laboratory (Sackett, Schmitt, Ellingson, & Kabin, 2001).

2.6 Summary

ST is often cited as an explanation for gender differences in maths performance (Nosek et al., 2009; Picho et al., 2013), yet little research has focused on differential ST susceptibility across maths question types. Furthermore, despite its high relevance to education, the ST phenomenon has received little attention in real examination settings. This research will build on Jamieson and Harkins's (2007) mere effort account to investigate ST in both lab and field studies. Mere effort proposes that ST motivates test-takers to disprove an active negative stereotype. When responding to solve type questions based on prepotent learnt knowledge, activation of a negative stereotype can motivate test-takers and improve performance. However, comparison type questions (requiring logic or estimation), often result in performance decrements, because test-takers seek to disprove the negative stereotype leading to a failure in inhibiting prepotent (i.e. solve) information. In conjunction with testing ST effects based on question type, the current review has highlighted the need for process-orientated research to specifically test mere effort's ST mechanisms (i.e., prepotent responses and motivation). The motivated application of prepotent responses as an explanatory ST process will be tested and discussed.

2.6.1 Research aims

The literature review forms the basis for three key aims, which are identified below, accompanied by a brief outline of how each chapter will address these aims:

(a) To test outcomes of the maths-gender ST on female maths performance, based on maths question type (i.e., solve vs. comparison);

(b) To test the mere effort account in educational settings (i.e., university and school examinations);

(c) To test mere effort's prepotent responses and motivational ST processes.

Chapter 3 seeks to provide constructive replication of Jamieson and Harkins's (2009) mere effort account in the maths performance domain. Study 1 will test the differential ST performance effects (i.e., facilitation vs. debilitation) based on the type of maths question encountered (i.e., solve vs. comparison) in response to ST. Two GCSE maths tests (solve vs. comparison), consisting of questions pretested for difficulty (Pilot study 1a), will be used to test females' maths performance in a between-subjects design. ST effects will be measured by indexing performance using unadjusted and adjusted maths test scores.

In Chapter 4, Studies 2 and 3 will investigate mere effort account in the field for the first time, to test if the interactive effects of ST and question type on maths performance transfer beyond the laboratory. ST is frequently cited as a determinant of educational inequality, yet there is a deficit in ST research conducted in real educational settings (Huguet & Regner, 2007). Study 2 will test female undergraduate students during a university statistics mock exam. Study 3 will test male and female secondary school students in a GCSE mock exam.

The focus turns to mere effort's ST mechanisms in Chapter 5. Specifically, Study 4 will implement a Stroop task to investigate the potential moderating role of *prepotent response inhibition* on ST effects, based on the type of maths question encountered. As in Study 1, ST effects will be measured by performance on the maths test(s) (solve vs. comparison). Study 4 will test the hypothesis that threatened female participants who have higher levels of inhibitory ability (measured by Stroop performance) will be more able to inhibit the potentiated prepotent solve response, compared to their lower inhibitory ability counterparts. That is, under ST, inhibitory ability will protect performance on comparison questions, as it will enable high inhibitors to suppress the incorrect solve approach and apply the correct comparison approach. Performance on solve questions should be unaffected by inhibitory ability as the prepotent solve response does not need to be inhibited. Chapter 5 will discuss the implications of the findings with reference to the alternative WM explanation of ST (Beilock et al., 2007).

Chapter 6 will aim to directly test *performance motivation* in response to the negative maths-gender stereotype. Previous research has only measured performance motivation using task performance itself (Jamieson & Harkins, 2007; McFall et al., 2009). This can create difficulties interpreting the ST processes from ST effects. Study 5 will employ a separate process-orientated task, to test performance motivation using maths question type *preference* under ST. Solve and comparison question will be pretested for difficulty (Pilot study 5a). It is hypothesised that ST will lead to a greater selection of solve versus comparison questions.

Chapter 3

Question type and stereotype threat

Mere effort theory proposes that the maths-gender stereotype threat (ST) has differential effects on female maths performance depending on the type of maths question encountered. ST motivates test-takers to disprove an active negative stereotype. This facilitates a prepotent solve response, augmenting performance for solve type questions (e.g., equations), but reducing performance for comparison type questions (e.g., estimations). Study 1 replicated and extended Jamieson and Harkins's (2009, 2012) findings. Question type was tested in a between-subjects design using General Certificate of Secondary Education (GCSE) solve versus comparison maths tests. Following ST, solve questions resulted in performance facilitation, whereas comparison questions resulted in performance reduction (for both unadjusted and adjusted scores). This finding supports the notion that question type is key to understanding the outcomes of ST.

3.1 Introduction

It has been argued that stereotype threat (ST) is “likely to be mediated in multiple ways – cognitively, affectively and motivationally” (Steele, Spencer, & Aronson, 2002, p. 397). The current research focuses predominantly on a motivational contribution to ST performance effects – namely, Jamieson and Harkin’s (2007) mere effort account. Jamieson and Harkin’s (2007, 2009, 2011) motivation-based mere effort model discusses how efforts to disprove negative stereotypes paradoxically harms performance to confirm the ST.

As outlined in Chapter 2 (Section 2.3), the mere effort account proposes that ST motivates individuals to want to perform well. This motivation potentiates a prepotent response (i.e., dominant or most likely), that if correct facilitates performance and if incorrect debilitates performance. Thus, performance is determined by whether the prepotent response is correct or not to answer the *type* of question (Jamieson & Harkins, 2007). These predictions have been supported using the maths-gender stereotype ‘women are poor at maths’ (Jamieson & Harkins, 2009, 2012). Following exposure to ST, female participants’ performance was increased for maths problems where the prepotent tendency to solve them was correct (e.g., equations). However, this approach was not correct for comparison problems (e.g., probability) and resulted in debilitated performance.

3.1.1 <u>The importance of question type</u>

The mere effort account therefore highlights an important difference from other ST research (e.g., Ganley et al., 2013; O’Brien & Crandall, 2003; Steele et al., 2002). The account proposes that the *type* of maths problem rather than the level of difficulty is key to understanding ST performance effects (Jamieson & Harkins, 2007, 2009; O' Brien & Crandall, 2003). This may explain the inconsistencies in previous ST research and weak effects when analysing maths performance (e.g., Stoet & Geary, 2012). For example, a recent large-scale study conducted by Ganley et al. (2013) found no evidence that female participants' maths performance was impeded by ST. Ganley et al. (2013) argued that too much emphasis is placed on ST as an explanation for female maths underperformance to the detriment of other key factors that may be involved (e.g., mathematics anxiety, mathematics interest; Ceci & Williams, 2010). The authors conducted three experiments: two with young adolescents and a third with children, younger adolescents, and older adolescents. Despite girls overall underperformance compared to boys, there were no ST effects observed (girls underperformed in both stereotype and control conditions).

However, in all of their studies, Ganley et al. (2013) used *fairly difficult* maths assessments informed by previous ST research (Neuville & Croizet, 2007; Nguyen & Ryan, 2008; O'Brien & Crandall, 2003; Spencer et al., 1999; Steele, 1997). Furthermore, they conducted follow-up analyses using only relatively difficult items (i.e., those with less than 50% correct). This type of analyses fails to control for potential differences in *how* individuals engage with the *type* of maths problem that may play a key role in determining performance (Jamieson & Harkins, 2007). Specifically, if the type of maths question can be answered correctly using the prepotent response that may subsequently facilitate or debilitate performance under ST. The results of ST studies, such as Ganley et al.'s (2013), that have not examined or controlled for the potential confounding effects of maths question type are hard to interpret. It may be that problem type rather than problem difficulty per se, impedes performance under ST.

Support for the importance of problem type rather than difficulty has been documented (Jamieson, 2009, Exp 2). Jamieson (2009) tested the maths-gender ST and manipulated orthogonally maths problem type (solve vs. comparison) and problem difficulty (test average of 75% for easy vs. 50% for difficult). If question difficulty is a factor determining ST, then performance should be facilitated regardless of question type on the easy maths test and debilitated on the hard one (O'Brien & Crandall, 2003). In contrast, Jamieson (2009) found support for question type as a driving factor in ST effects. Regardless of question difficulty, the experience of ST debilitated performance on comparison problems and facilitated performance on solve problems. That is, under ST, performance did not differ as a function of difficulty level, but instead depended on whether the prepotent response was correct or not. Similarly, Jamieson and Harkins (2009) tested maths solve and comparison questions that had a mean overall accuracy of 50% for each type (comparison range = 38% to 60%, solve range = 42% to 63%). The questions were taken from the quantitative section of the Graduate Record Examinations (GRE) and were in multiple choice question (MCQ) formats. The authors found that ST effects still occurred as a function of question type when controlling for question difficulty. In light of the findings, further investigation is needed to contribute to the limited body of question type ST research.

3.1.2 Aims of Study 1

The aim of Study 1 is to determine if the maths-gender ST impacts females' maths performance based on question type. In accord with earlier research (Jamieson & Harkins, 2009, 2012), performance will be measured using two different types of maths questions: solve and comparison, while question difficulty will be controlled for. Study 1 will therefore seek to replicate of Jamieson and Harkins's (2009) findings, and will establish a basis for the current research to further investigate question type and ST. At present, the investigation of question type in ST literature is limited (e.g., Beilock et al., 2007; Jamieson & Harkins, 2009, 2012). Furthermore, Study 1 will extend upon the original findings in several ways. First, it uses a different version and format of maths examination questions: General Certificate of Education (GCSE) rather than GRE, and full answers rather than MCQ. This will help establish the generalizability of ST effects based on question type across different standardised maths examinations. Second, it implements a between-subjects design to reduce interference across question types; to allow a clear picture of the differential ST effects on performance based on question type. Finally, analyses of (1) actual unadjusted maths test scores, as well as (2) the adjusted maths score of total percentage of problems solved (Jamieson & Harkins, 2009) will serve as the dependent measures. This enables the examination of ST effects on maths performance typical to both real standardised maths tests (e.g., in educational settings) that use unadjusted scores, and ST research that use adjusted scores (e.g., Gimmig, Huguet, Caverni, & Cury, 2006; Schmader & Johns, 2003).

3.2 Pilot study 1a

A pilot study was conducted to create both a solve and comparison question type maths test that closely resembled a GCSE test for the main Study 1 (see Appendix B). A GCSE is a compulsory academic qualification for core subjects taken by students (14-16 years) in England and Wales. A maths test was administered consisting of 18 questions (*solve* = 9, *comparison* = 9) all set at the

GCSE (higher tier) level and taken from non-calculator examinations selected from an online academic source (www.aqa.org.uk). Thirty female participants aged 18-23 years (M = 19.67, SD = 1.67) were allotted 70 minutes to complete the test. This was analogous to the time per question allocated in GCSE examinations. To control for mathematical ability, all participants had a GCSE grade of C or above. They identified as British Caucasian and spoke English as their first language. To create the two tests differing in question type (i.e., solve vs. comparison), while of equal difficulty, five questions each worth three marks were selected across each question type that elicited similar overall scores. Criteria specified by Jamieson (2009) were used to determine question type (see Appendix A). From the thirty participants, the overall scores of the questions selected for each maths test did not differ in difficulty across the solve (M = 7.40, SD = 2.90) versus comparison (M = 7.03, SD = 2.89) questions, $t(29) = .59$, $p = .56$. Therefore, the finalised versions of the solve and comparison maths tests consisted of five 3-mark questions equally balanced for difficulty, with 18 minutes test completion time allowed (see Appendix C).

3.3 Study 1

Study 1 tested the prediction that participants experiencing ST will be motivated to undermine the active stereotype which in turn will enhance activation of the prepotent solve response. Participants subject to ST are hypothesised to perform better on the solve question maths test (i.e., where the prepotent response is correct) and worse on comparison question maths test (i.e., where the prepotent response is incorrect) than their non-threatened counterparts. When ST is not activated (control condition), there will be no differences in performance across question type.

3.3.1 Method

3.3.1.1 Participants and design

A power calculation conducted using the computer software GPower (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that a sample size of 128 would be sufficient to detect a significant interaction effect with a power of .80 and an alpha of .05. One hundred and sixty female University of Leeds undergraduates (*age range* = 18-23 years, *M* = 19.32, *SD* = 1.04) participated. All participants had achieved a maths GCSE of grade C previously, and identified as British Caucasian, with English as their first language. Participants were randomly assigned to the conditions of a 2 (task diagnosticity: high, low) x 2 (question type: comparison, solve) between-subjects design.

3.3.1.2 Materials

Task diagnosticity was manipulated by instructing participants that "Previous research has shown gender differences on this test" (i.e., high diagnosticity) or that "Previous research has shown no gender differences on this test" (i.e., low diagnosticity). This manipulation (adapted from Steele & Aronson, 1995) successfully induces or removes the maths-gender ST, respectively (Brown & Pinel, 2003; Jamieson & Harkins, 2007; Spencer et al., 1999; see Nguyen & Ryan, 2008, for a review). Demographic data was collected via the computer program E-prime, using questions that included participants: age, gender, ethnicity, first language, nationality, GCSE and any additional maths qualifications. Lastly, a two-item ST manipulation check was also included, taken from Jamieson and Harkins (2007, 2009): "To what extent are there gender differences in performance on this task?" (1 = *no gender differences*, 11 = *gender differences*); "Who do you believe performs better in this task?" (1 = *males perform better*, 6 = *males and females perform the same,* 11 = *females perform better*).

3.3.1.3 Procedure

The Female experimenter escorted participants into the lab one at a time and informed them that they would be involved in a series of short tasks. Following allocation to experimental conditions, participants were asked to carefully read the maths test instructions on the front cover (that included the manipulation, see Section 3.3.1.2) before signalling to the experimenter that they were ready to begin. All participants completed as many as possible of five 3-mark questions on the pre-tested non-calculator mathematics pen-and-paper test(s) (see Appendix C). A test time constraint of 18 minutes was implemented, resulting in approximately one minute allowed per mark (analogous to the time allocated in GCSE examinations). A ruler, pencil, and pen were also provided to simulate an exam environment. Upon completion, participants completed several measures (outlined in Section 3.3.1.2).

3.3.1.4 Dependent measures

The dependent measures were maths score (out of 15) and total percentage of problems solved (maths score divided by 3 times the number of questions attempted, as each question is worth 3 marks, and then multiplied by 100).

3.3.2 <u>Results</u>

3.3.2.1 Manipulation checks

Responses to the manipulation check items were analysed using independent samples t-tests. Here and throughout the thesis adjusted dfs are applied when Levene's test for equality of variances is violated, indicating that the homogeneity of variance across the samples cannot be assumed. The first item ("To what extent are there gender differences in performance on this task?"), led participants in the high diagnosticity condition to report that gender differences existed to a greater extent (M = 5.58, SD = 2.58) than participants in the low diagnosticity condition (M = 3.95, SD = 2.67), $t(158) = 3.93$, $p < .001$. Similarly, in regards to the second item ("Who do you believe performs better in this task?"), participants in high diagnosticity condition reported greater expectancies that males would perform better than females on the task (M = 4.81, SD = 1.81) in comparison to participants in the low diagnosticity condition (M = 5.34, SD = 1.35), $t(154.88) = -2.09$, $p = .02$. In all, participants in the high diagnosticity condition were aware of, and expected task performance to reflect the maths-gender stereotype. The task diagnosticity manipulation was effective.

3.3.2.2 Maths performance

Maths test performance data was analysed using a 2 (task diagnosticity) x 2 (question type) between-subjects Analysis of Variance (ANOVA). The task diagnosticity, $F(1, 156) = .001$, $p = .50$ and question type, $F(1, 156) = 1.59$, $p = .20$ main effects were not significant.

Crucially, however, the interaction was significant, $F(1, 156) = 4.83$, $p = .03$, $\eta p^2 = .03$ (Figure 3.1). In the case of high diagnosticity, participants in the comparison question type condition (M = 7.12, SD = 2.71), underperformed relative to participants in the solve question type condition (M = 8.91, SD = 3.65), $t(79.27) = 2.59$, $p = .01$. This pattern did not emerge in the low diagnosticity condition: Participants in the comparison question type condition (*M* = 8.26, *SD* = 3.09) performing analogously to those in the solve question type condition (*M* = 7.77, *SD* = 3.53), *t*(72) = -.63, *p* = .53. The interaction was viewed from another angle. For comparison questions, participants underperformed in the high diagnosticity (*M* = 7.12, *SD* = 2.71) relative to the low diagnosticity (*M* = 8.26, *SD* = 3.09) condition, *t*(79) = -1.76, *p* = .04. However, for solve questions, participants tended to perform better in the high diagnosticity (*M* = 8.91, *SD* = 3.65) relative to the low diagnosticity (*M* = 7.77, *SD* = 3.53) condition at a level approaching significance, *t*(77) = 1.40, *p* = .08. These results are supportive of the hypotheses.

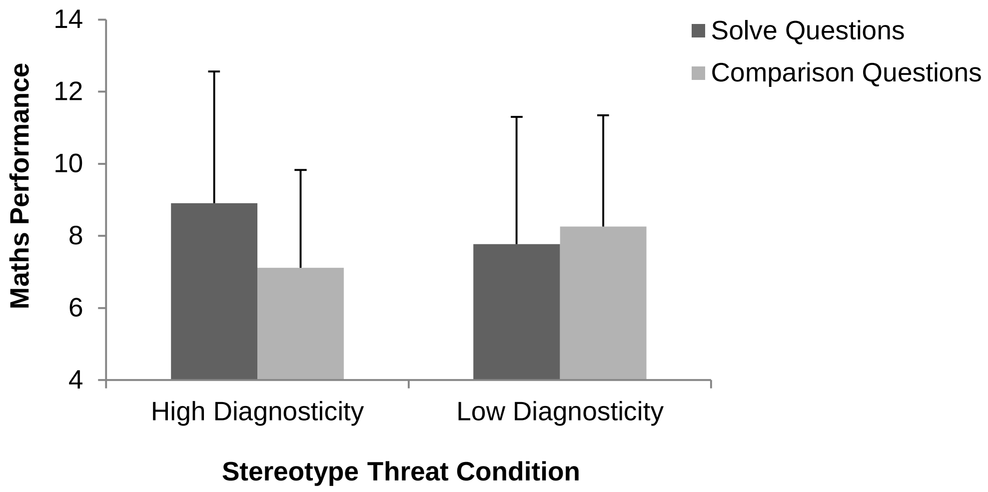


Figure 3.1. Mean maths score on the maths task as a function of question type and diagnosticity in Study 1. The error bars represent the standard deviations.

3.3.2.3 Percentage of problems solved

Significant effects were also found using the total percentage of problems solved. This was calculated by dividing the total maths score by the number of questions attempted (that has been multiplied by 3 as each question is worth 3 marks) and then multiplying the score by 100. Similarly to the maths performance data, the data was subjected to a 2 (task diagnosticity) x 2 (question type) between-subjects ANOVA. No main effects were obtained for either task diagnosticity, $F(1, 156) = .001$, $p = .49$ or question type, $F(1, 156) = .81$, $p = .37$. Instead, there was a significant interaction, $F(1, 156) = 4.95$, $p = .03$, $\eta p^2 = .03$ (Figure 3.2). As hypothesised, in the case of high diagnosticity, participants in the comparison question type condition ($M = 49.66\%$, $SD = 18.39\%$) solved a smaller percentage of the questions they attempted relative to participants in the solve question type condition ($M = 60.30\%$, $SD = 23.33\%$), $t(81.14) = 2.36$, $p = .01$. This effect did not emerge in the case of low diagnosticity: Participants in the comparison question type condition ($M = 57.32\%$, $SD = 19.93\%$) performing analogously to participants in the solve question type condition ($M = 52.81\%$, $SD = 23.96\%$), $t(72) = -.88$, $p = .14$. In addition, as hypothesised, for comparison questions, participants solved a smaller percentage of the questions they attempted in the high diagnosticity ($M = 49.66\%$, $SD = 18.39\%$) relative to those in the low diagnosticity ($M = 57.32\%$, $SD = 19.93\%$) condition, $t(79) = -1.80$, $p = .04$. However, for solve questions, participants tended to solve a greater percentage of the questions they attempted in the high diagnosticity ($M = 60.30\%$, $SD = 23.33\%$) relative to the low diagnosticity ($M = 52.81\%$, $SD = 23.96\%$) condition, at a level approaching significance, $t(77) = 1.40$, $p = .08$. The results are consistent with the maths performance data and are in line with the hypotheses.

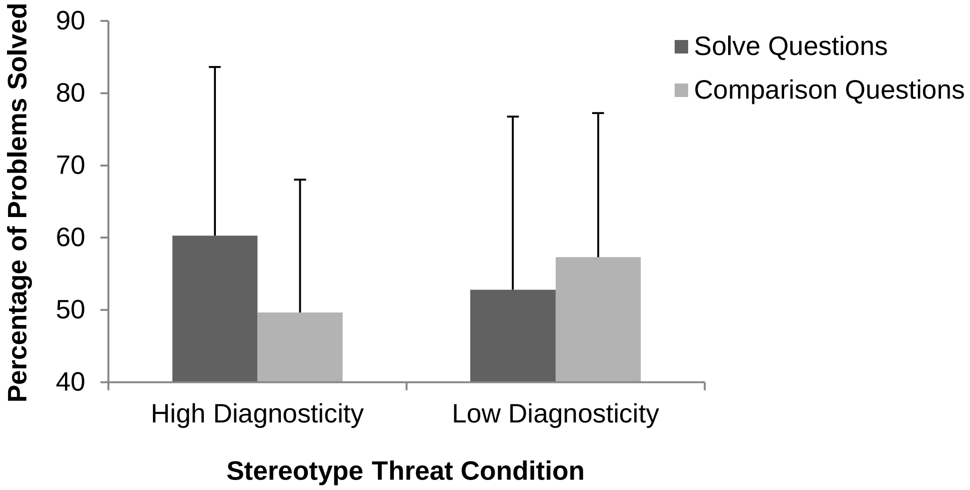


Figure 3.2. Mean percentage of problems solved on the maths task as a function of question type and diagnosticity in Study 1. The error bars represent the standard deviations.

3.4 Discussion

Study 1 set out to examine the potential interactive effects of the maths-gender ST and question type on females' maths test performance. Consistent with Jamieson and Harkins (2009, 2012), female test-takers experiencing ST performed better when responding to maths *solve* questions than *comparison* questions. Whereas, in the control condition, participants performed equally well across maths question types. The interactive effects were also replicated using Jamieson and Harkins's (2009) adjusted maths score of total percentage of problems solved, traditionally used to measure ST effects (e.g., Gimmig et al., 2006; Schmader & Johns, 2003). Females subject to ST solved a greater percentage of solve problems correctly than their non-threatened counterparts. In contrast, for comparison problems, females under threat solved a smaller number of these problems correctly than controls.

Importantly, these interactive effects on performance were observed when maths question *difficulty* was controlled for across maths question type (solve vs. comparison). Thus, the ST effects on maths performance in Study 1 are the result of differences in the *type* of question encountered. This is contrary to some ST literature that has focused on maths question difficulty (e.g., Ganley et al., 2013; O'Brien & Crandall, 2003; Steele et al., 2002). Consequently, Study 1 findings indicate that differences in question type could have potentially confounded ST findings that solely focused on question difficulty; such as those by Ganley et al. (2013), where ST effects were not observed. The maths-gender ST may have led participants to perform better on solve type questions and worse on comparison type questions, which served to confound any ST effects.

3.4.1 <u>Differences in maths question type</u>

Specifically, in Study 1, the differences in maths question type were determined by their correct solution approach (see Appendix A). The solve type questions required the application of well-learned formulas (i.e., the solve approach), whereas comparison type questions required a more reasoned, logical approach (i.e., the comparison approach). As the solve approach is the prepotent (or preferred) method of maths problem solving by female test-takers (e.g., Gallagher et al., 2000; Royer, Tronsky, Chan, Jackson, & Marchant, 1999), the mere effort account argues that this approach is the one that is potentiated by the motivation to disprove the maths-gender stereotype (Jamieson & Harkins, 2009). Thus, maths performance was determined by whether the question could be answered correctly using the prepotent *solve* response. ST facilitated performance when the prepotent response was correct (i.e., on solve questions), but debilitated performance when the prepotent response was incorrect (i.e., on comparison questions) (Jamieson & Harkins, 2009, 2012).

This interactive effect has only been previously demonstrated with quantitative GRE problems (Jamieson & Harkins, 2009, 2012). Indeed, at present maths-gender ST research focuses predominantly on GRE maths questions conducted in the United States (see Picho, Rodriguez, & Finnie, 2013). Furthermore, it is the first time the effect has been shown using an unadjusted maths score, a typical measure of score for standardised maths exams such as GCSEs. Study 1 therefore served as a replication and extension of Jamieson and Harkins's (2009, 2012) findings to demonstrate the robustness of ST effects based on maths question type (categorised by their most efficient solution strategy) across different standardised maths tests and scoring methods. Replication is a particularly important issue in ST research: Picho et al.'s (2013) recent meta-analysis of 17 years of maths-gender ST research highlighted the deficit in constructive replication, with the literature focusing too much on the breadth rather than depth of research. Indeed, the importance of replication, particularly translating laboratory findings to the field will be investigated next in Chapter 4 (Sackett, Schmitt, Ellingson, & Kavin, 2001). ST is derived from a real-world phenomenon and it is therefore important to investigate whether the effects of question type in Study 1 are generalizable to real-world maths performance.

3.4.2 Summary

Study 1 replicated and extended the mere effort account of ST by demonstrating that; ST differentially affects performance dependent on maths question type, and the effects are transferable across different types and formats of maths questions. Thus Study 1 provides a clear foundation for the current ST research to investigate how a variation in question type interacts with ST effects. The interactive effects of ST and question type are next tested in the field, using studies set in a university and a secondary school under examination conditions.

Chapter 4

Testing stereotype threat and question type in educational settings

---

Stereotype threat (ST) is frequently cited as a determinant of educational equality. However, despite its high relevance to education, the ST phenomenon has received little attention in educational settings. Studies 2 and 3 investigated the role of question type on maths performance following ST, to test if the interactive effects (observed in laboratory Study 1) were replicated for real-world maths performance. Study 2 tested female psychology undergraduates during a statistics mock exam and found that ST debilitated performance on comparison questions and overall maths performance. However, no evidence was found for ST facilitating performance on solve questions. In Study 3, both male and female secondary school students were tested during a General Certificate of Education (GCSE) maths mock exam. As expected, following ST, female participants' maths performance was augmented for solve questions and debilitated for comparison questions. Male participants' maths performance was lifted under ST conditions, regardless of maths question type. The implications for mere effort as an explanation for ST effects in real-world maths performance are discussed.

4.1 Introduction

The importance of replicability in psychology is more important than ever following replication crisis (see Asendorpf et al., 2013), particularly in replicating experimental effects in the field (Sackett, Schmitt, Ellingson, & Kabin, 2001; Shen et al., 2011). The external validity of laboratory findings have been investigated by meta-analyses of laboratory and field studies to assess the impact of research settings on results within a particular area of research (e.g., Avolio, Reichard, Hannah, Walumbwa, & Chan, 2009). For example, Mitchell's (2012) meta-analysis of 217 lab-field comparisons from 82 meta-analyses found that the external validity of laboratory research differed considerably by psychological subfield, research topic, and effect size. Mitchell (2012) posited despite the usual replicability of psychological lab studies in real-world research (r = .71), Social psychology needed the most improvement (r = .53), with studies of gender differences least likely to translate to the real-world.

Likewise, Picho, Rodriguez, and Finnie's (2013) large-scale meta-analysis of maths-gender ST research indicated that the literature is "plagued with insufficient replication" (p. 326). It is therefore important to conduct field studies to demonstrate that causal relations observed in the laboratory hold in the field (e.g., Behrman & Davey, 2001; Levitt & List, 2007), as well as provide constructive replication (Picho et al., 2013). This will help dispel external validity concerns stemming from the use of laboratory experimental designs (Aguinis & Lawal, 2013; Brutus, Gill, & Duniewicz, 2010), and accentuate the cumulative and incremental nature of progress in psychological science (Shen et al., 2011).

4.1.1 Stereotype threat field research

Replication in the field is particularly important for stereotype threat (ST) in maths performance, as ST is frequently cited as a determinant of educational equality. However, despite its high relevance to education, the ST phenomenon has received little attention in real exam or school settings (Huguet & Regner, 2007; Wei, 2012; Wicherts, Dolan, & Hessen, 2005). Sackett et al. (2001) argued that research has yet to demonstrate whether and to what degree ST convincingly generalizes beyond the laboratory and cautioned over interpreting current ST findings. At present only a few studies have tested the maths-gender ST in test settings high in ecological validity (e.g., classrooms, group settings) (Croizet & Claire, 1998; Huguet & Regner, 2007; Keller, 2002; Keller & Dauenheimer, 2003; Walsh, Hickey, & Duffy, 1999). Of these ST studies, the results have been mixed. In some cases, ST effects were generally small or non-existent (Cullen, Hardison, & Sackett, 2004; Cullen, Waters, & Sackett, 2006; Ganley et al., 2013; Stricker & Ward, 2004) despite the large and representative samples used.

For example, Stricker and Ward (2004) conducted two field studies to evaluate the effects of inquiring about ethnicity and gender on test performance (see Steele & Aronson, 1995). These researchers used two standardised academic ability tests in actual test administrations, and were unable to replicate strong ST effects on test performance for minority and female groups. In contrast, maths-gender ST effects *were* found in a number of other field studies (Good, Aronson, & Inzlicht, 2003; Huguet & Regner, 2007; Keller, 2007; Keller & Dauenheimer, 2003; Miyake et al., 2010; Ramirez & Beilock, 2011; Wei, 2012). Studies by Keller (2007) and Keller and Dauenheimer (2003) indicated that adolescent girls' maths performance at secondary school was influenced by ST in the classroom setting. Similarly, Huguet and Regner (2007) found that middle school girls performance on tasks they believed measured mathematical skill (e.g., geometry) was affected by ST in quasi-ordinary classroom circumstances (i.e., as close to normal classroom conditions).

It is difficult to draw a clear overall picture as each study differs in its particular focus. For instance, Keller (2007) investigated ST with maths-identification, whereas Miyake et al. (2010) focused on values affirmation (i.e., one’s important values). Differences in ST manipulations also factor: Huguet and Renger (2007) questioned Keller and Dauenheimer’s (2003) explicit activation of the ST (i.e., stating that men had outperformed women in past research) versus more ordinary school circumstances (e.g., characterising a task as diagnostic of ability). Furthermore, emphasis is often placed on *question difficulty* (Ganley et al., 2013; Huguet & Regner, 2007; Keller, 2007), as is typical of ST lab experiments (Neuville & Croizet, 2007; Nguyen & Ryan, 2008; O'Brien & Crandall, 2003; Spencer, Steele, & Quinn, 1999; Steele, 1997). However, as shown earlier in the current research (see Study 1), the *type* of question also has a significant role on performance under ST. As previously discussed (see Chapter 2, Section 2.3), when a negative stereotype is associated with performance, individuals are seemingly motivated to perform well and actively set out to disprove the stereotype, potentiating prepotent responses (Harkins, 2006; Jamieson & Harkins, 2007; McFall, Jamieson, & Harkins, 2009). Performance is therefore dependent on whether the prepotent response is contextually appropriate, or correct. In the context of maths, females’ prepotent response is to apply a *solve* approach (applying learned formula), indicated by their stronger preference for the solve approach over a comparison approach (using estimation and intuition) than males (Gallagher et al., 2000; Royer, Tronsky, Chan, Jackson, & Marchant, 1999; Tartre, 1990). Therefore, arguably, it is question type rather than difficulty which determines performance under ST.

4.1.2 Aims of Study 2 and 3

Studies 2 and 3 will test the hypothesis that (as found in Study 1), under ST female participants will perform worse on comparison questions relative to solve questions (c.f. controls) in educational settings. So far no research has investigated the potential interactive effects of ST and question type in the field. Studies 2 and 3 will therefore aim to replicate the ST effects based on question type observed in lab Study 1 in the field, using (a) undergraduate students' performance on a university mock statistics exam, and (b) secondary school pupils' performance on the General Certificate of Education (GCSE) maths exam.

4.2 Study 2

Study 2 tested the mere effort account of ST during a first year psychology undergraduate mock statistics exam at a UK University during 2011-12. The Research Skills 1 (RS1) practice examination has been administered for the last five years as part of the students' Research Skills 1 module at the University of Leeds. The module is a core component of psychology undergraduate curriculum, designed to educate students in statistical research methods and analyses. As it is a pre-existing test, questions were categorised as being either solve or comparison in nature; with those that could not be categorised not included in the analysis (N= 14). Due to the greater number of solve (N = 18) relative to comparison (N = 8) questions, the *proportion* of correct scores for solve and comparison question were analysed. The exam is statistically rather than specifically maths based, and uses a multiple choice question (MCQ) format. However, as previous maths-gender ST literature has demonstrated (e.g., Huguet & Regner, 2007), tests that are presumed to be diagnostic of maths ability still produce ST effects. Therefore, as statistics are part of the UK maths curriculum, and included in GSCE, AS and A-level syllabus' (see www.aqa.org.uk/subjects/mathematics), it is likely the students will associate the exam as being diagnostic of their maths ability. To enable parity with Study 1 (and previous ST field research: Keller & Dauenheimer, 2003) an explicit ST manipulation was implemented (see Section 4.2.1.2). Maths ability was controlled for as it is a course requirement for students studying psychology at the University of Leeds to have a maths GCSE of at least a B grade.

It was hypothesised that female participants subject to ST would perform better on solve questions (prepotent response correct) and worse on comparison questions (prepotent response incorrect) compared to their non-threatened counterparts. To the author's knowledge, this is the first field study investigating ST using question type in a statistics exam.

4.2.1 Method

4.2.1.1 Participants and design

Two hundred and ten female subjects were recruited via an opportunity sample during their first year psychology RS1 mock exam on two separate occasions in 2011 and 2012 (*age range* = 18 – 21 years, *M* = 18.32, *SD* = .58). All participants identified as British Caucasian, with English as their first language. Participants were allocated to a 2 (task diagnosticity: high, low) x 2 (question type: comparison, solve) mixed design, with repeated measures on the second factor.

4.2.1.2 Materials

The RS1 statistics mock exam consisted of a 40-item pen and paper MCQ (see Appendix D). This exam paper has been used for the past five years as the RS1 course practice exam. Each question had a choice of one of four possible answers, with each correct answer worth 1-mark. Prior to the exam, questions were reviewed and categorised as being either solve or comparison (see Figure 4.1). Questions that could not be categorised (i.e., could be answered using combination of solve and comparison approaches) were not included in the analyses (N = 14). Task diagnosticity was elicited by adapting the manipulation from Study 1. In the task instructions participants were informed that “In previous years in the RS1 exam we have found that women are less competent at statistics compared to men” (i.e. high diagnosticity) or that “In previous years in the RS1 exam we have found no differences in statistical ability across men and women” (i.e. low diagnosticity). The same manipulation check and additional measure questions from Study 1 were used but in pen and paper format (see Appendix D).

Solve Type:

2. What is the modal score in the following list of test scores: 55, 55, 59, 65, 65, 65, 70?

   a. 62.5
   b. 55
   c. 60
   d. 65

Comparison

32. Which of the following statements is correct?

   a. An extremely small p-value indicates that the actual data differs markedly from that expected if the null hypothesis were true.
   b. The p-value measures the probability of making a Type II error.
   c. The larger the p-value, the stronger the evidence against the null hypothesis
   d. A large p-value indicates that the data is inconsistent with the alternative hypothesis.

Figure 4.1. Example of the Research Skills 1 exam solve and comparison type questions for Study 2.

4.2.1.3 Procedure

The Female examiner placed an exam paper on each desk before participants entered the examination room. Participants were invited to sit at a desk and were instructed examination conditions applied (i.e., no talking or conferring, and seated separately). As was the customary practice in the exam, all students had pocket calculators at their disposal. To ensure that participants were randomly assigned to study conditions, the exam papers were distributed in random order and the examiner had no influence on which desk the participants chose to sit at. Participants were allocated to one of two conditions: high diagnosticity versus low diagnosticity. Participants were instructed to turn over their paper and to carefully read the examination instructions before they began. The ST manipulation was included in the instructions (see Section 4.2.1.2). All participants were given 1.5 hours to complete as many as possible of the 40 statistics MCQ questions (see Appendix D). Upon completion participants were instructed to raise their hand, and were given the manipulation check to complete before turning overleaf to read the study debrief. Participants were free to ask the examiner any further questions at the end of the exam and thanked for their time. The time constraints and examination setting were analogous to the conditions of the RS1 exam participants complete as part of their undergraduate degree course and thus replicated a real university examination.

4.2.1.4 Dependent Measures

The RS1 exam proportion of correct scores for solve (N = 18) and comparison questions (N = 8) were the main dependent measures. This was calculated by dividing participants' score for each question type by the maximum score for each question type.

4.2.2 <u>Results</u>

4.2.2.1 Manipulation checks

The data for the 2011 and 2012 RS1 mock exams were pooled and all analyses that follow are derived from this. Responses to the manipulation check items were analysed using independent samples t-tests. The first item ("To what extent are there gender differences in performance on this task?"), led participants in the high diagnosticity condition to report that gender differences existed to a greater extent (M = 5.45, SD = 2.76) than participants in the low diagnosticity condition (M = 3.53, SD = 2.63), $t(208) = 5.07$, $p < .001$. The second item was incorrectly printed for the test in 2011 and was discarded. The 2012 cohort responded to the second manipulation check item ("Who do you believe performs better in this task?") equally across the high diagnosticity (M = 5.21, SD = 1.66) and low diagnosticity (M = 5.44, SD = 2.03) conditions, $t(75) = -.46$, $p = .32$. Thus, participants in the high diagnosticity ST condition were aware of but did not necessarily believe that task performance would reflect the negative group stereotype. The manipulation was partially effective.

4.2.2.2 Exam performance

Participants' maths test performance was subjected to a 2 (task diagnosticity) x 2 (question type) mixed Analysis of Variance (ANOVA), with repeated measures on the second factor. Maths performance was indexed by the *proportion* of correct scores for solve and comparison questions (calculated by dividing participants' score for each question type by the maximum score for each question type). In contrast to Study 1, significant main effects were obtained for task diagnosticity, $F(1, 208) = 4.92$, $p = .01$, $\eta p^2 = .02$, and question type, $F(1, 208) = 78.33$, $p < .001$, $\eta p^2 = .27$. Participants performed significantly worse in the high diagnosticity ($M = .48$, $SE = .01$) compared to the low diagnosticity ($M = .52$, $SE = .02$) conditions, and worse on the comparison questions ($M = .44$, $SE = .01$) than the solve questions ($M = .56$, $SE = .01$) respectively. Importantly, the interaction was also significant, $F(1, 208) = 13.89$, $p < .001$, $\eta p^2 = .06$ (see Figure 4.2).

A paired samples t-test was used to unpack the interaction. As expected, in the case of high diagnosticity, participants answering comparison questions ($M = .39$, $SD = .19$), underperformed relative to when answering solve questions ($M = .56$, $SD = .16$), $t(122) = 9.87$, $p < .001$, $d = -.97$. However, this pattern also emerged in the low diagnosticity condition; participants answering comparison questions ($M = .49$, $SD = .19$) underperformed relative to when answering solve questions ($M = .56$, $SD = .18$), $t(86) = 3.30$, $p = .001$, $d = -.38$. The interaction was then analysed using an independent samples t-test to compare each question type across high versus low diagnosticity. As hypothesised, for comparison questions, participants underperformed in the high diagnosticity ($M = .39$, $SD = .19$) relative to the low diagnosticity ($M = .49$, $SD = .19$) condition, $t(208) = -3.66$, $p < .001$, $d = -.53$. However, for solve questions, participants performed similarly in the high diagnosticity condition ($M = .56$, $SD = .16$) relative to the low diagnosticity ($M = .56$, $SD = .18$) condition, $t(208) = .11$, $p = .46$. These results are in part consistent with the hypotheses.

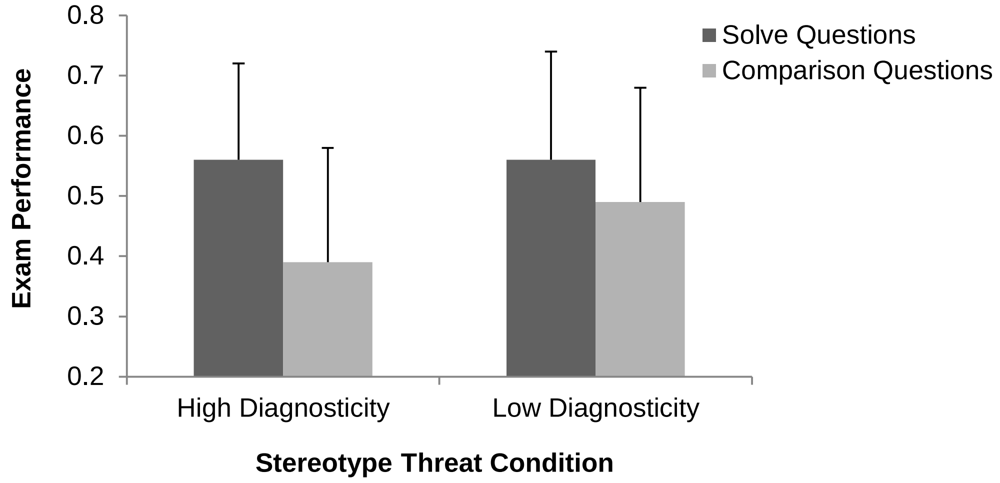


Figure 4.2. Mean proportion of problems solved on the Research Skills 1 exam as a function of question type and diagnosticity in Study 2. The error bars represent the standard deviations.

4.2.3 Study 2 summary

The findings from Study 2 provide compelling new evidence for the effects of ST under real exam conditions: the experience of ST differentially impacted female maths performance based on the type of maths question. Crucially, and in support of the hypotheses, participants performed significantly worse following ST on comparison questions (c.f. controls participants). However, performance was not augmented under ST on solve questions; participants performing analogously across ST conditions. This may have occurred because (as discussed in Chapter 2, Section 2.3.1.2), the effect of performance debilitation on comparison problems is greater than the effect of facilitation on solve problems (Jamieson & Harkins, 2009). Indeed, it was the performance detriment to the comparison questions under ST that drove the harmful significant effect of ST on overall exam performance. However, as performance was worse for comparison questions than solve questions across both ST and control conditions, this indicates that the comparison questions were overall harder than the solve questions. This highlights a practical difficulty with conducting ST field research, as the pre-existing exam did not allow for questions to be matched for difficulty. The interactive effects of question type and ST were next investigated in a secondary school population in Study 3.

4.3 Study 3

The aim of Study 3 was to investigate the effects of the maths-gender ST on maths performance dependent on question type in a secondary school setting. Understanding the experience of ST in a school maths exam environment is fundamental in efforts to reduce inequalities in education (Huguet & Regner, 2007; Wei, 2012; Wicherts et al., 2005). Indeed, recent UK educational statistics reveal that for maths and additional maths GCSEs in 2011 and 2012, a higher cumulative percentage of boys that took the exam achieved a greater number of top grades than girls (GCSE Results, 2012). Similarly, the participating secondary school in the current study has reported gender differences between the number of A*'s achieved: In 2011, 32% of boys compared to 21% of girls, and in 2012, 29% of boys compared to 14% of girls achieved the top maths grade. Thus, it is important to understand how ST may interact with the type of maths question encountered. Indeed, as reported by the school head of mathematics, the maths department had observed that girls often underperformed when faced with unstructured (i.e., comparison) maths questions (B. Wilkinson, personal communication, December 12, 2012).

The present study provides an opportunity to test a GCSE target age sample (14-16 years) to see if the findings from Study 1 are reproduced during real GCSE maths examination conditions. The GCSE solve and comparison maths tests from Study 1 will be combined into a single maths test. Therefore Study 3's maths test will comprise of both solve and comparison questions (see Appendix E) that is typical of a real GCSE exam (i.e. tested within subjects). Study 3 will also include *male* participants. As Stoet and Geary (2012) have argued, it is important to include a male control group in order to draw clear conclusions about how ST may lead to gender differences in performance.

In accordance with Studies 1 and 2, threatened females are predicted to perform more poorly on comparison problems and better on solve problems than females not subject to ST (Jamieson & Harkins, 2009). Furthermore, as ST is proposed to harm performance on comparison questions more than it improves performance on solve questions; female maths performance under ST should overall be worse than their non-threatened counterparts. Male maths performance should not differ as a function of ST.

4.3.1 Method

4.3.1.1 Participants and design

One hundred and ninety one secondary school pupils at St. Aidans School in North Yorkshire, UK, participated (*female* = 94, *male* = 97), ranging in age between 14-16 years (*M* = 14.79, *SD* = .56), the target age for GCSEs. All had identified as British Caucasian, with English as their first language. Participants were assigned to a 2 (task diagnosticity: high, low) x 2 (gender: male, female) x 2 (question type: comparison, solve) mixed design, with repeated measures on the third factor.

4.3.1.2 Materials

The maths test was a combination of the solve and comparison maths tests from Study 1 (see Appendix E). The maths test therefore comprised of ten GCSE maths questions (*solve* = 5, *comparison* = 5) evenly matched for difficulty. The test also copied GCSE test formatting (including the test cover and formula sheet) to closely resemble a real GCSE maths test paper. As in Study 1, the same ST manipulation was included in the test instructions, alongside demographic questions (e.g., gender, age, ethnicity, and nationality). Participants in the high diagnosticity condition were informed that “Previous research has shown gender differences on this test”, whereas in the low diagnosticity condition participants were informed that “Previous research has shown no gender differences on this test”. A pen and paper version of the 2-item manipulation check from Study 1 was also included (see Appendix E).

4.3.1.3 Procedure

The test was administered in the school examination hall during a maths lesson period. The Female examiner, accompanied by 3 male school maths teachers, set out the examination hall with a maths paper on each individual desk. As per typical GCSE examination procedure, participants were invited into the exam hall and to sit at a desk. Examination conditions were enforced (i.e., no talking or conferring, only stationary permitted on their desks, no calculators). To ensure that participants were randomly assigned to experimental conditions, the exam papers were distributed in random order and the experimenter had no influence on which desk the participants chose to sit. Participants were allocated to one of two conditions: high diagnosticity versus low diagnosticity. Participants were instructed to fill out the information on the front of the maths paper (age, gender, nationality, and ethnicity) and to carefully read the test instructions (including the ST manipulation). Subsequently, they were allowed 35 minutes to complete as many as possible of the ten 3-mark questions on the pre-tested, non-calculator mathematical pen and paper test (see Appendix E). The time restriction allowed approximately one minute per mark as per real GCSE examinations. After the allotted time, participants were each given the 2-item pen and paper manipulation check to fill in (see Appendix E). The examiner then collected the test papers and manipulations checks, before debriefing the participants as a group and thanking them for their time.

4.3.1.4 Dependent measures

The dependent measures were maths score for each question type (out of 15) and total percentage of problems solved. This was calculated by dividing participants' maths score by the number of questions they had attempted (that has been multiplied by 3 as each question is worth 3 marks) and then multiplied by 100.

4.3.2 Results

4.3.2.1 Manipulation checks

Participants in the high diagnosticity condition (*M* = 5.45, *SD* = 2.94) reported that gender differences existed on the test to a greater extent than participants in the low diagnosticity condition (*M* = 4.50, *SD* = 2.51), $t(189) = 2.41$, $p = .01$. However there were no differences in reported beliefs as to who would perform better on the test (male vs. female) between participants in the high diagnosticity (*M* = 5.18, *SD* = 1.56) relative to the low diagnosticity condition (*M* =

5.41, *SD* = 1.64), *t*(189) = -.99, *p* = .16. In all, although participants in the high diagnosticity condition were aware of the negative stereotype they did not expect this to reflect in task performance. The manipulation was partially effective.

4.3.2.2 Exam performance

Participants' maths test performance was subjected to a 2 (task diagnosticity) x 2 (gender) x 2 (question type) mixed ANOVA, with repeated measures on the third factor. There was no significant main effect of gender, $F(1, 187) = .52$, $p = .47$, and a marginal significant main effect was obtained for task diagnosticity, $F(1, 187) = 3.13$, $p = .08$, $\eta p^2 = .02$. There was also a significant main effect of question type, $p = .02$, $\eta p^2 = .03$. No significant interaction emerged for gender x task diagnosticity, $F(1, 187) = .62$, $p = .43$. However, a significant interaction was observed for question type x gender, $F(1, 187) = 4.76$, $p = .03$, $\eta p^2 = .03$. There was also a significant interaction for question type x task diagnosticity, $F(1, 187) = 5.28$, $p = .02$, $\eta p^2 = .03$. Importantly, and of most interest, the 3-way interaction was significant, $F(1, 187) = 16.64$, $p < .001$, $\eta p^2 = .08$ (see Figure 4.3). The 3-way interaction was next decomposed by diagnosticity x question type across gender.

4.3.2.3 Male exam performance

In contrast to the hypothesis, for male participants, there was a significant main effect for task diagnosticity, $F(1, 95) = 2.93$, $p = .05$, $\eta p^2 = .03$. In line with ST lift research (e.g., Chalabaev, Stone, Sarrazin, & Croizet, 2008; Walton & Cohen, 2003), male participants maths performance was augmented in the high diagnosticity ($M = 6.99$, $SE = .45$) relative to the low diagnosticity ($M = 5.91$, $SE = .44$) condition. There was no significant main effect of question type, $F(1, 95) = .02$, $p = .90$, and no significant diagnosticity x question type interaction, $F(1, 95) =$

1.56, $p = .21$. Therefore, male maths performance was lifted by the maths-gender ST, but this was not affected by the type of maths question (solve vs. comparison) encountered.

4.3.2.4 Female exam performance

In contrast, for female participants, there was a significant main effect of question type; female performance was greater on the solve ($M = 7.33$, $SE = .37$) than comparison ($M = 6.18$, $SE = .29$) questions, $F(1, 92) = 10.56$, $p = .002$, $\eta p^2 = .10$. No main effect of task diagnosticity was observed, $F(1, 92) = .55$, $p = .23$. Importantly, as in Studies 1 and 2, the diagnosticity x question type was significant, $F(1, 92) = 20.71$, $p < .001$, $\eta p^2 = .18$.

In the high diagnosticity group, female participants performed significantly worse on comparison questions ($M = 5.58$, $SD = 2.91$) relative to solve questions ($M = 8.35$, $SD = 3.68$), $t(39) = 4.41$, $p < .001$. As expected, this pattern did not emerge in the low diagnosticity condition, with no significant differences for female performance on comparison questions ($M = 6.78$, $SD = 2.71$) relative to solve questions ($M = 6.31$, $SD = 3.40$), $t(53) = -1.17$, $p = .12$. Furthermore, as hypothesised, for comparison questions, female participants underperformed in the high diagnosticity ($M = 5.58$, $SD = 2.91$) relative to the low diagnosticity ($M = 6.78$, $SD = 2.71$) condition, $t(92) = -2.06$, $p = .02$. In contrast, for solve questions, female participants performance was augmented in the high diagnosticity ($M = 8.35$, $SD = 3.68$), relative to the low diagnosticity ($M = 6.31$, $SD = 3.40$) condition, $t(92) = 2.77$, $p = .001$. The results therefore support the hypotheses.

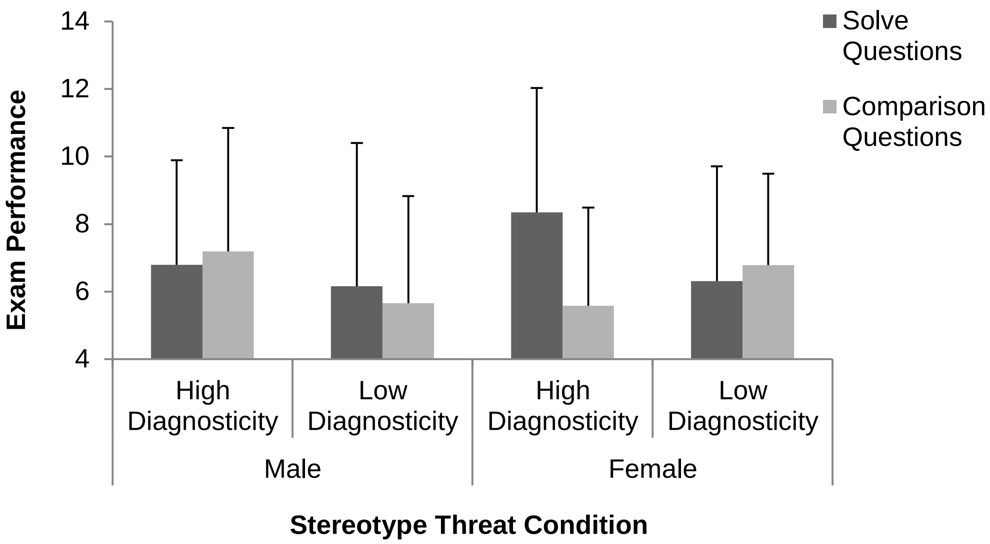


Figure 4.3. Mean maths score on the maths exam as a function of question type, diagnosticity and gender in Study 3. The error bars represent the standard deviations.

4.3.2.5 Percentage of problems solved

A similar pattern of results were found for the percentage of problems solved. As in Study 1, this was calculated by dividing participants total maths score by the number of questions they had attempted (that has been multiplied by 3 as each question was worth 3 marks) and then multiplying the score by 100. The data was subjected to a 2 (task diagnosticity) x 2 (gender) x 2 (question type) mixed ANOVA, with repeated measures on the third factor. No significant main effect was obtained for gender $F(1, 187) = .11$, $p = .74$. A significant main effect was found for diagnosticity, $F(1, 187) = 3.45$, $p = .03$, $\eta p^2 = .02$, and a marginal significant effect of question type was observed, $F(1, 187) = 3.13$, $p = .08$, $\eta p^2 = .02$. As with the maths performance data, no significant interaction emerged for gender x task diagnosticity, $F(1, 187) = .22$, $p = .64$. A significant interaction was observed again for question type x gender, $F(1, 187) = 5.40$, $p = .02$, $\eta p^2 = .03$, and the question type x task diagnosticity interaction was also significant, $F(1, 187) = 4.50$, $p = .04$, $\eta p^2 = .02$. Crucially, the 3-way interaction was significant again, $F(1, 187) = 14.63$, $p < .001$, $\eta p^2 = .07$ (see Figure 4.4). The 3-way interaction was next decomposed by diagnosticity x question type across gender.

4.3.2.6 Male percentage of problems solved

As with the maths performance data and in contrast to the hypothesis, for male participants, there were a marginally significant main effect for task diagnosticity, $F(1, 95) = 2.45$, $p = .06$, $\eta p^2 = .03$. Male participants answered a greater number of maths questions correctly that they attempted in the high diagnosticity (M = 49.40%, SE = 2.94%) relative to the low diagnosticity (M = 43.00%, SE = 2.85%) condition. There was no significant main effect of question type, $F(1, 95) = .14$, $p = .71$, and no significant diagnosticity x question type interaction, $F(1, 95) = 1.35$, $p = .25$. Therefore, as with the maths performance data, male participants answered a greater percentage of maths questions they attempted correctly under ST conditions, but this was not affected by maths question type.

4.3.2.7 Female percentage of problems solved

The female results were also generally consistent with the female exam performance data. A significant main effect of question type was observed; female participants solved a greater percentage of questions they attempted for solve (M = 50.91%, SE = 2.46%) than comparison (M = 43.56%, SE = 1.88%) questions, $F(1, 92) = 9.13$, $p = .003$, $\eta p^2 = .09$. No main effect of task diagnosticity, $F(1, 92) = 1.09$, $p = .15$, was observed. The diagnosticity x question type interaction was significant, $F(1, 92) = 19.29$, $p < .001$, $\eta p^2 = .17$.

In the high diagnosticity condition, female participants solved a smaller percentage of the comparison questions they attempted (*M* = 40.31%, *SD* = 18.90%) relative to solve questions (*M* = 57.79%, *SD* = 25.34%), $t(39) = 4.19$, $p < .001$. As expected, this pattern did not emerge in the low diagnosticity condition, with no significant differences for the percentage of problems solved for comparison questions (*M* = 46.82%, *SD* = 17.80%) relative to solve questions (*M* = 43.59%, *SD* = 22.12%), $t(53) = -1.23$, $p = .11$. When viewed across question type, as expected, for comparison questions, female participants underperformed on the questions they attempted in the high diagnosticity (*M* = 40.31%, *SD* = 18.90%), relative to the low diagnosticity (*M* = 46.82%, *SD* = 17.80%) condition, $t(92) = -1.74$, $p = .04$. In contrast, for solve questions, female participants answered significantly more questions they attempted correctly in the high diagnosticity (*M* = 57.79%, *SD* = 25.34%) relative to the low diagnosticity (*M* = 43.59%, *SD* = 22.12%) condition, $t(92) = 2.89$, $p = .003$.

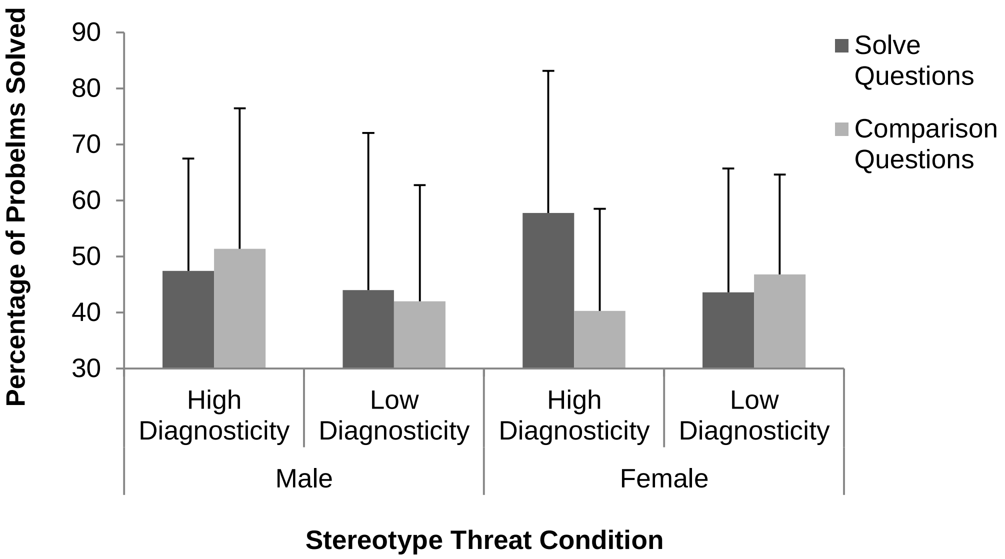


Figure 4.4. Mean percentage of problems correctly solved on the maths exam as a function of question type, diagnosticity and gender in Study 3. The error bars represent the standard deviations.

4.3.3 Study 3 summary

Study 3 attests to the real-world applicability of ST effects based on question type. ST effects were observed in a school environment, using real GCSE maths questions in typical examination conditions on the target GCSE population (14-16 year old students). Consistent with previous mere effort laboratory research (Jamieson & Harkins, 2009: Exp 1) there was a significant interaction between diagnosticity, question type and gender. As hypothesised, female maths performance under ST was dependent on the *type* of maths question encountered. Specifically, females subject to ST performed better on *solve* questions (i.e., prepotent response correct) and worse on *comparison* questions (i.e., prepotent response incorrect) than their non-threatened counterparts. However, ST did not reduce female maths performance overall. Study 3 findings also revealed that males' maths performance was lifted under ST, irrespective of question type. This fits with previous ST lift research that has shown tangible increases in male maths performance when they were made aware of the negative female maths stereotype (Walton & Cohen, 2003). Male participants are able to make downward comparisons with females stereotypically poorer at the task that 'lifts' their maths performance. Therefore, under ST, the lift effect for male overall maths performance, occurring simultaneously with female debilitated performance on comparison questions, may exacerbate gender differences in maths performance.

4.4 Discussion

An interactive effect of question type and ST was observed outside the lab in high-ecological educational test settings for the first time. In Study 2, the RS1 mock exam was completed by psychology undergraduate students in examination conditions. Similarly, the GCSE maths test in Study 3 was tested on the target GCSE secondary school population in real examination conditions. The pattern of findings generally supported the mere effort explanation of the maths-gender ST (Jamieson & Harkins, 2009, 2012), to show that ST differentially impacts female maths performance depending on the type of maths question encountered. The importance of replicating the interactive effects of ST and maths question type (observed in Study 1) in real examination settings (Huguet & Regner, 2007; Wei, 2012; Wicherts et al., 2005) make the present findings a key contribution to understanding ST in education. These contributions include:

First, Studies 2 and 3 demonstrate that ST harms female maths performance on comparison questions in real educational test settings. In both the undergraduate and secondary school samples, female performance on comparison questions was significantly decreased under ST relative to controls. Study 3 also found strong effects for females' enhanced performance on solve questions under ST. Female test-takers subject to threat outperformed their non-threatened counterparts on solve questions. Replicating the laboratory findings in high ecological settings (i.e. university and secondary school exams) provides strong evidence that ST is relevant to educational practice. Indeed, the differential effects of question type on performance under ST were shown using a measure of *unadjusted* maths score (i.e., actual maths performance) using repeated measures on question type (i.e., participants completed both question types) that is typical to real examination marking and test formats. Furthermore, female performance on comparison questions was debilitated under ST during real test based situations but without the real additional threatening consequences (i.e., test performance did not count towards real academic grades). This attests to the power of ST in natural educational environments (Keller and Dauenheimer, 2003).

Second, whereas Study 2 findings demonstrated that ST negatively impacted female maths performance overall, Study 3 did not find ST effects for overall maths performance. That is, in Study 3, threatened female test-takers' solve performance was facilitated to the same extent that it was harmed for comparison performance. The facilitated solve performance protected their overall maths performance under ST. This suggests that the differential effects of different maths question types can 'cancel out' any ST effects and, as previously suggested in Chapter 3 (see Section 3.1.1), may potentially confound ST results that do not control for question type. It is interesting that, despite the general consistency between the lab and field results, in Study 3 the ST performance facilitation of the maths GCSE solve questions was stronger in the field. This highlights the complex interplay between ST and maths performance in real-world exam environments, and suggests that mere effort's motivational account may not be able to explain overall ST effects on maths performance alone.

Third, Study 3 findings revealed that male secondary school participants' maths performance was improved under ST, regardless of maths question type. This suggests that, in response to the maths-gender stereotype, males' maths performance was lifted by the downward social comparison they could make with females (Walton & Cohen, 2003). Thus, in contrast to female maths performance, ST influenced male maths performance irrespective of maths question type. This illustrates differences in how male and female maths performance is affected by ST to potentially widen the maths-gender performance gap. Furthermore, these findings are consistent with the idea that it is motivation to *disprove* the ST that leads individuals to rely on the prepotent response. As female test-takers are the social group stigmatised by the maths-gender ST, whereas males can make a downward social comparison, only females' prepotent response is activated by the ST. Thus, only females' maths performance is affected by whether the question can be correctly answered using the solve response.

4.4.1 Explicit stereotype threat manipulation

The explicit manipulation used in the present work leads to some reduced applicability of ST in educational settings because this is not what students are typically exposed to. Previous research has contested the validity of using explicit ST measures. For example, Huguet and Regner (2007) criticised Keller and Dauenheimer (2003) for informing participants that the maths test produced (or did not produce) gender differences. Huguet and Regner (2007) used quasi-ordinary classroom circumstances to manipulate ST by altering the gender composition of the groups of test-takers. Indeed, other research has investigated the potential influence of coed versus single sex learning environments as a ST manipulation (Kessels & Hannover, 2008; Picho & Stephens, 2012). However, in a recent meta-analysis, Picho et al. (2013) showed that ST was not moderated by the nature of testing environment or sex composition of the participants: Females' performance was unaffected by test settings that were homogeneous or where they formed the majority. Thus, the implementation of an explicit ST manipulation (as in laboratory Study 1) enabled a clearer indication of ST effects.

4.4.3 Summary

The findings from Studies 2 and 3 demonstrate the interactive effects of maths question and ST for the first time in field research. The laboratory effects from Study 1 were generally replicated in female test-takers' real-world maths performance, during an undergraduate statistics mock exam (Study 2) and a secondary school GCSE maths mock exam (Study 3). Furthermore, Study 3 found evidence for female test-takers' augmented *solve* performance following ST, implicating the role of heightened motivation in response to ST. Study 3 findings also revealed that the maths-gender ST lifted male overall maths performance. Study 4 will return to the lab (Chapter 5) to investigate the mechanisms underlying the mere effort account of ST. Specifically, Study 4 will test for the overproduction of the prepotent solve response (activated by the maths-gender ST) via the ability to *inhibit* the prepotent response.

Chapter 5

The role of inhibitory ability in stereotype threat

Study 4 investigated the overproduction of prepotent responses in response to the maths-gender stereotype. Mere effort argues that the motivation to disprove the negative stereotype potentiates the solve response. Maths performance is therefore dependent on whether type of maths question can be answered correctly using the potentiated solve approach: If not (i.e. when faced with comparison questions), then solve responses must be inhibited. This suggests a potential moderating role of inhibitory ability. Threatened test-takers with higher inhibitory ability were predicted to be more able to inhibit the incorrect solve response and apply the correct comparison approach when answering comparison questions. Inhibitory ability would therefore help test-takers to overcome detrimental ST performance effects. Performance on solve questions was predicted to remain unaffected as the solve response does not need to be inhibited. However, while higher levels of inhibitory ability did protect overall maths performance following ST, this was not dependent on question type. The implications of these findings are discussed with reference to alternative working memory (WM) model of ST (e.g., Beilock, Rydell, & McConnell, 2007).

5.1 Introduction

Individual differences in inhibition-related functions in normal adults have been proposed as underlying variations in: Working memory (WM) (De Beni, Palladino, & Cornoldi, 1998); problem solving (Passolunghi, Cornoldi, & De Liberto, 1999); and general cognitive ability (Dempster & Corkhill, 1999). However, rather than being a single unitary construct, research suggests that inhibition-related processes are a family of functions that can be clustered into several distinct categories (Dempster, 1993; Friedman & Miyake, 2004; Nigg, 2000). For example, Friedman and Miyake (2004) tested three functions; (1) prepotent response inhibition - the ability to suppress dominant, automatic, or prepotent responses; (2) resistance to distractor interference - the ability to ignore or resolve interference from task-irrelevant information in the external environment; and, (3) resistance to proactive interference - the ability to block task-irrelevant information from memory that was once relevant. Of these, prepotent response inhibition is most straightforwardly associated with the mere effort account of stereotype threat (ST) (Jamieson & Harkins, 2007), in the active suppression of dominant responses potentiated by ST.

5.1.1 <u>Inhibiting the prepotent response</u>

In the context of the ST ‘women are bad at maths’, the inhibitory mechanism *prepotent response inhibition* may therefore serve to suppress the prepotent response generated (i.e., to apply the solve response), in order for other approaches (i.e., the comparison response) to be considered. Indeed, Carr and Steele (2009) proposed that the experience of ST may induce a perseverant way of thinking. ST interferes with test-takers’ ability to replace old strategies with more successful ones for problem solving. The problem solver must have the ability to inhibit the previous response and develop a new response. Carr and Steele's (2009) ST activation of inflexible perseverance shares similarities with Jamieson and Harkins's (2009) potentiation of the prepotent response. Specifically, maths test-takers experiencing ST use the dominant problem solving strategy of trying to 'solve' all questions, instead of generating and using other strategies (such as reasoning, logic or estimation) that are more efficient for some questions. Therefore, the *ability to inhibit* the prepotent solve approach when it is not required (i.e., for comparison questions), could enable test-takers to use the comparison approach and may potentially improve maths performance under ST.

5.1.2 Testing for the inhibition of prepotent response

The inhibition of prepotent responses is akin to Hasher, Lustig, and Zacks (2007) inhibitory function of *restraint* that suppresses automatic, prepotent responses to enable the use of other, more-contextually appropriate responses. This suggests that the prepotent response inhibition function associated with the mere effort account is fairly straightforward to test. Typically, the Stroop task (Stroop, 1935) is used as a test of prepotent response inhibition. Participants name the colour in which colour words and neutral words are printed, ignoring the dominant tendency to read the words. Theoretically, it follows that participants who are more able to suppress their prepotent reading response for the Stroop task should perform better under ST on comparison questions, as they should also be more able to suppress the incorrect prepotent solve response (potentiated by the ST). The current study will therefore use a Stroop task (Stroop, 1935) to assess the potential role of inhibitory ability in moderating ST effects based on question type.

5.1.3 Working memory and stereotype threat

Inhibition and performance following ST can be related to mere effort (Jamieson & Harkins, 2007), but also potentially the WM perspective of ST (Schmader & Johns, 2003). This is because inhibitory ability is a component of WM: Inhibitory functions regulate and control the contents of WM to help efficiently manage the cognitive system (Hasher & Zacks, 1988). WM can be conceptualised as a short-term memory system involved in the control, regulation, and active maintenance of a limited amount of information required for task goals (Miyake & Shah, 1999). Inhibition operates in service of task goals by hindering goal-irrelevant information that becomes active in parallel with goal-relevant information (Hasher & Zacks, 1988). Individual differences in WM may therefore dictate the amount of goal-directed attention that is available for task-relevant information, while simultaneously inhibiting irrelevant information (Barrett, Tugade, & Engle, 2004). Therefore, some individuals are better at inhibiting task irrelevant information than others.

ST research has investigated how individual differences in WM influence performance on WM intensive tasks such as mathematical problem solving (Beilock, Rydell, & McConnell, 2007). The core of the WM explanation of ST is that the negative stereotype harms performance by disrupting WM resources needed to perform certain types of maths problems (Beilock et al., 2007; Rydell, McConnell, & Beilock, 2009; Schmader & Johns, 2003). Research has suggested that high-WM individuals may be better equipped to cope with ST than low-WM individuals (Regner et al., 2010; Schmader, Johns, & Forbes, 2008). There seems little doubt that WM is implicated in maths problems, but whether this is via threat disrupting WM or some other process is a contentious issue. Therefore, ST effects can potentially be interpreted in terms of the WM disruption or other approaches, for example mere effort.

5.1.4 <u>Aims of Study 4</u>

The main aim of Study 4 is to investigate the potential role of inhibitory ability moderating ST effects based on question type. This will help determine if ST does operate via mere effort's overproduction of prepotent responses. Specifically, if ST does potentiate prepotent responses, then the greater aptitude to suppress the prepotent (solve) response should protect threatened females' maths performance specifically for *comparison* questions. Theoretically, high inhibitors should be more able to inhibit the solve response and apply the correct comparison approach so that they can perform to their full mathematical ability. Performance for the solve questions should be unaffected by inhibitory ability as the prepotent solve response does not need to be inhibited. Furthermore, as it is the experience of ST that potentiates prepotent responses, this pattern should not emerge in the control condition.

5.2 Study 4

Study 4 tested the prediction that inhibitory ability would moderate ST on maths task performance. Inhibitory ability was measured using Stroop task performance. Following ST, participants who have higher levels of inhibitory ability are hypothesised to be more able to inhibit the prepotent solve response. This will enable high inhibitors to apply the correct comparison approach and perform to their full ability on comparison questions. Conversely, participants who have lower levels of inhibitory ability will be less capable of inhibiting the incorrect solve response. Therefore, low inhibitors will be unable to apply the correct comparison approach, resulting in debilitated performance on comparison questions. Performance on solve questions should not differ as a function of inhibitory ability as the prepotent solve response does not need to be inhibited. Control participants' performance should be unaffected.

5.2.1 Method

5.2.1.1 Participants and design

One hundred and sixty five female University of Leeds undergraduates ranging in age between 18-25 years (*M* = 20.50, *SD* = 1.61) were tested. All had achieved a maths General Certificate of Education (GCSE) of grade C and identified as British Caucasian, with English as their first language. Participants were assigned to a 2 (task diagnosticity: high, low) x 2 (question type: comparison, solve) between-subjects design. Inhibitory ability was also assessed as a continuous variable. The design was implemented using a moderated regression (Baron & Kenny, 1986). Therefore, standardised scores were used in the main analyses.

5.2.1.2 Materials

The same maths tests (see Appendix C), ST manipulation, manipulation checks, and additional questions were used as Study 1 (see Chapter 3, Section 3.2.1.2), while also including a Stoop task. The Stroop task involved 10 practice trials and 48 experimental trials (16 of each trial type; congruent, incongruent, and neutral). On each trial participants were presented with a fixation asterisk (+) in the center of the screen for 1000 milliseconds (ms), followed by the presentation of a target colour word (i.e., blue, red, green, yellow) or hash key (i.e., ####), in either congruent (baseline) or incongruent (interference) coloured font. The hash key represented the neutral condition. The words were printed in size 24 Arial Font. The ‘Z’ response key denoted the correct response for words printed in red or green font, whereas the ‘M’ key denoted the correct response for words printed in blue or yellow font. Stimuli remained on screen until participants responded and the next trial began after an inter-trial interval of 1000ms.

5.2.1.3 Procedure

The Female experimenter escorted participants into the lab individually. Participants first completed the Stroop task on-screen (see Section 5.2.1.2), lasting approximately 5 minutes. Participants were instructed that colour words (red, blue, yellow, green) and hash strings (####) would be presented on the screen in one of the following colours: red, blue, yellow, or green. They were asked to press the button corresponding to the ink colour of the word as quickly as they could, whilst ignoring the word itself. They were asked to respond as quickly and as accurately as possible, and that the first ten trials were to practice before moving on to the real experimental trials. Second, after completing the Stroop task and following a 5-minute break, participants were given the maths test and asked to read carefully the front cover. The ST manipulation was included in the test instructions. Subsequently, participants were allowed 18 minutes to complete as many as possible of the five 3-mark questions on the non-calculator mathematical pen and paper test(s) (comparison vs. solve question type), in accord with Study 1 (see Appendix C). On completion, participants responded to the manipulation check and additional questions.

5.2.1.4 Dependent measures

Maths score (out of 15) and the total percentage of problems correct out of the total number of problems attempted were the main measures. Response time differences between median congruent and incongruent accurate responses on the Stroop task provided an index of inhibitory ability.

5.2.2 Results

5.2.2.1 Calculating inhibitory ability

Stroop task trials on which participants made incorrect responses (errors) were removed (7.9%). Median reaction times to incongruent and congruent trials for each participant were calculated, and a difference score was computed by subtracting the median reaction time on congruent trials from the median reaction time on incongruent trials. Median trial reaction times were used because they are less susceptible to outliers than mean reaction times (see Whelan, 2008). Lower scores indicated greater inhibitory ability.

5.2.2.2 Manipulation checks

In regard to the first manipulation check item, participants in the high diagnosticity condition reported that gender differences existed to a greater extent (M = 5.70, SD = 2.97) than participants in the low diagnosticity condition (M = 3.04, SD = 2.59), $t(163) = 6.07$, $p < .001$. For the second manipulation check item, participants in high diagnosticity condition (M = 4.95, SD = 2.11) did not differ in their beliefs of who would perform better on the task (male vs. female) from participants in the low diagnosticity condition (M = 4.93, SD = 1.99), $t(163) = .04$, $p = .48$. In all, participants in the high diagnosticity condition were aware of, but did not necessarily believe, that task performance would reflect the negative group stereotype. The manipulation was partially effective.

5.2.2.3 Maths performance[1]

The predictors of task diagnosticity (high diagnosticity vs. low diagnosticity), question type (solve vs. comparison), and inhibitory ability were measured in the regression analysis. No main effect emerged for task diagnosticity, β = .01, p = .44, whereas there was a significant main effect of question type, β = -.14, p = .04. Participants performed better on the solve than comparison questions. There was also a significant main effect of inhibitory ability, β = -.19, p = .001: High inhibitors performed better on the maths test than their low inhibitor counterparts. As hypothesised, the task diagnosticity x question type interaction was significant, β = .18, p = .02. A significant diagnosticity x inhibitory ability interaction was also observed, β = .25, p = .02, whereas there was no significant question type x inhibitory ability interaction, β = -.17, p = .13. Furthermore, in contrast to the hypothesis, no interaction was found for task diagnosticity x question type x inhibitory ability, β = .11, p = .32, ΔR-squared = .19. This

---

[1] A 2 (task diagnosticity) x 2 (question type) between-subjects Analysis of Covariance (ANCOVA), with inhibitory ability as a covariate, was also undertaken to control for the potential confounding role of inhibitory ability on maths performance. No significant main effects were obtained for task diagnosticity, $F(1, 160) = .02$, $p = .44$, or question type, $F(1, 160) = 2.48$, $p = .12$. The covariate, inhibitory ability, was significantly related to maths performance, $F(1, 160) = 6.10$, $p = .01$, $\eta p^2 = .04$, suggesting that inhibitory ability has a separate effect on maths performance to the interaction. Importantly, when controlling for the covariate effect of inhibitory ability the interaction was still significant, $F(1, 160) = 4.60$, $p = .03$, $\eta p^2 = .03$.

indicates that the task diagnosticity x question type interaction was not moderated by inhibitory ability.

The task diagnosticity x question type interaction was unpacked by conducting simple regressions separately for each diagnostic condition (see Figure 5.1). In the low diagnosticity condition there was no effect of question type on maths performance, $\beta = .05$, $p = .67$, whereas, in the high diagnosticity condition there was a significant effect of question type on maths performance, $\beta = -.28$, $p = .004$. Participants subject to ST performed significantly better for the solve relative to the comparison questions. Simple regressions were next conducted separately for each question type. As hypothesised, for comparison questions, participants in the high diagnosticity condition underperformed relative to the low diagnosticity condition, $\beta = -.24$, $p = .01$. However, for solve questions, no significant differences in maths performance were observed across the diagnostic conditions, $\beta = -.12$, $p = .14$.

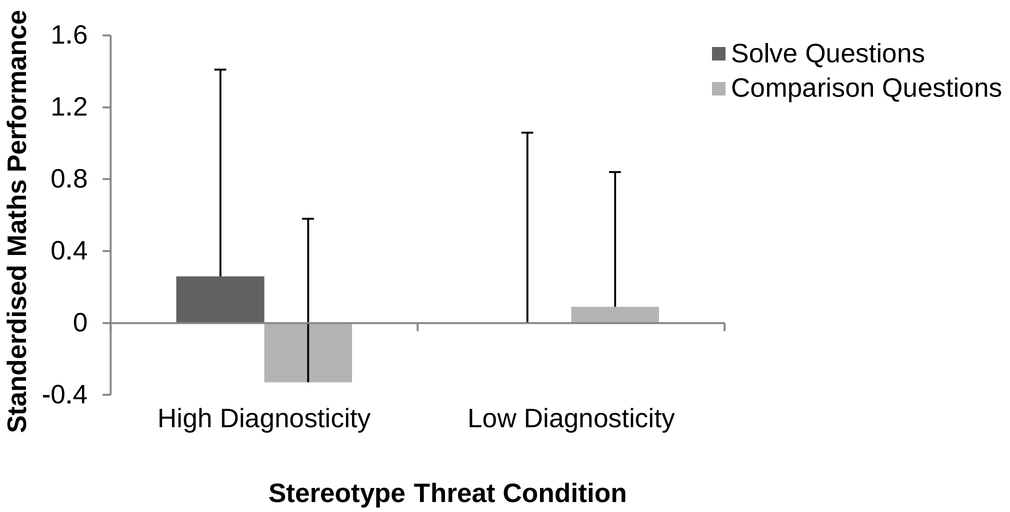


Figure 5.1. Standardised mean maths score on the maths task as a function of question type and diagnosticity in Study 4. The error bars represent the standardised standard deviations.

The task diagnosticity x inhibitory ability interaction was also explored. In the low diagnosticity condition there was no effect of inhibitory ability on maths performance, $\beta = .07$, $p = .27$. In contrast, in the high diagnosticity condition, there was a significant effect of inhibitory ability on maths performance, $\beta = -.30$ $p = .002$. High diagnosticity resulted in low inhibitors underperforming (c.f. high inhibitors).

5.2.2.4 Percentage of problems solved[2]

A similar pattern of results were found for the percentage of problems solved. This was calculated by dividing participants' total maths score by 3 times the number of questions they had attempted (as each question is worth 3 marks) and multiplying it by 100. No main effect emerged for task diagnosticity, $\beta = .02$, $p = .42$. A marginal significant main effect was observed for question type, $\beta = -.11$, $p = .07$: Participants answered a greater percentage of questions they attempted correctly for solve questions than comparison questions. There was also significant main effect of inhibitory ability, $\beta = -.16$, $p = .02$. High inhibitors answered a greater percentage of questions they attempted correctly than their low inhibitor counterparts. In terms of the interactions, as hypothesised, the task diagnosticity x question type interaction was significant, $\beta = .18$, $p = .01$. The task diagnosticity x

---

[2] The percentage of problems solved data was subjected to a 2 (task diagnosticity) x 2 (question type) between-subjects ANCOVA, with inhibitory ability as a covariate. No main effects were obtained for either task diagnosticity, $F(1, 160) = .04$, $p = .42$, or question type, $F(1, 160) = 1.64$, $p = .20$. However, a significant main effect of inhibitory ability on maths performance was found, $F(1, 160) = 4.11$, $p = .02$, $\eta p^2 = .03$. Participants with higher levels of inhibitory ability solved a greater percentage of questions they attempted correctly than those with lower inhibitory ability. A marginal interaction was also observed, $F(1, 160) = 3.50$, $p = .06$, $\eta p^2 = .02$.

inhibitory ability interaction was also significant again, $\beta = .23$, $p = .03$. In contrast to the maths performance data, the question type x inhibitory ability interaction was also significant, $\beta = .27$, $p = .02$. Importantly, no interaction was found for task diagnosticity x question type x inhibitory ability, $\beta = .14$, $p = .19$, ΔR-squared = .10. This suggests that the task diagnosticity x question type interaction for percentage of problems solved was not moderated by inhibitory ability.

The significant interactions were unpacked using simple regressions. For the task diagnosticity x question type interaction (see Figure 5.2), analysis revealed that there was no effect of question type on the percentage of questions correctly solved in the low diagnosticity condition, $\beta = .05$, $p = .33$. As expected, in contrast, in the high diagnosticity condition participants answered correctly a significantly greater percentage of questions they attempted for the solve relative to the comparison questions, $\beta = -.24$, $p = .01$. Furthermore, as hypothesised, for comparison questions participants solved a smaller percentage of the questions they attempted in the high diagnosticity relative to those in the low diagnosticity condition, $\beta = .22$, $p = .02$. However, for solve questions, there were no significant differences in the percentage of questions correctly answered that were attempted across the diagnostic conditions, $\beta = -.10$, $p = .19$.

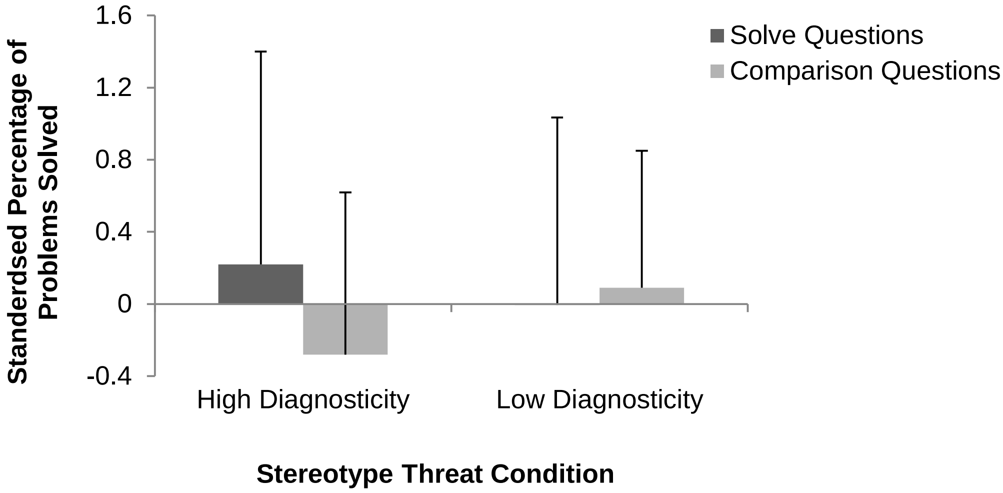


Figure 5.2. Standardised mean percentage of problems solved on the maths task as a function of question type and diagnosticity in Study 4. The error bars represent the standardised standard deviations.

For the task diagnosticity x inhibitory ability interaction, in the low diagnosticity condition there was no effect of inhibitory ability on the percentage of questions attempted answered correctly, β = .80, p = .25. In comparison, in the high diagnosticity condition performance, participants with higher inhibitory ability answered a greater percentage of questions they attempted correctly than their lower inhibitory ability counterparts, β = -.26, p = .01. The question type x inhibitory interaction decomposition revealed a marginally significant effect of inhibitory ability on the percentage of questions correctly answered for solve questions, β = -.16, p = .07. High inhibitors answered a greater percentage of questions they attempted correctly than low inhibitors. Similarly, for comparison questions, high inhibitors also answered a significantly greater percentage of questions they attempted correctly than low inhibitors, β = -.18, p = .05.

5.3 Discussion

Study 4 investigated the potential role of inhibitory ability moderating ST effects based on question type. Specifically, do high levels of inhibitory ability protect female maths performance under ST for *comparison* questions? Study 4 findings illustrated that while inhibitory ability did protect threatened test-takers' maths performance, this was irrespective of the *type* of maths question encountered. Inhibitory ability moderated ST effects on maths performance overall: Low inhibitors subject to ST performed worse on the maths test relative to their high inhibitor counterparts. This suggests that, contrary to the hypothesis, inhibitory ability did not protect threatened females' maths performance specifically on comparison questions by inhibiting the incorrect potentiated solve response.

5.3.1 Inhibition of the prepotent response

Study 4 did not find support for mere effort's overproduction of the prepotent response. The ability to inhibit the prepotent response did not protect maths performance specifically for *comparison* questions, that require the prepotent solve approach to be inhibited for other approaches (such as the comparison approach) to be applied. Instead, high levels of inhibitory ability protected maths performance under ST for *both* solve and comparison maths questions, irrespective of whether the prepotent approach was correct or not. This indicates that solve questions also require inhibitory resources. Indeed, research has proposed that *all* mathematical cognition involves central executive resources of the WM (DeStefano & LeFevre, 2004), and inhibitory ability plays an essential role in the efficient operation of WM (Hasher & Zacks, 1988). Thus, while not consistent with the mere effort account, the finding that inhibitory ability was positively related to overall maths performance under ST may be explained by the WM perspective.

5.3.2 Working memory as alternative explanation

Researchers investigating WM as an explanation for ST effects have explored the relationship between maths question presentation, WM resources and ST (Beilock et al., 2007). Beilock et al. (2007) proposed that ST harms performance by interfering with WM resources needed to compute certain types of maths problems. The authors argued that maths problems can be differentiated in terms of the specific demands they make on WM resources. Specifically, well-practised maths problems are less reliant on WM resources relative to questions that require novel problem solutions. Thus, under ST, novel maths problems suffered greater decrements to performance than well-practised questions. Indeed, Beilock et al. found that ST effects were alleviated by the practice of susceptible maths problems. Practise resulted in the correct solution being stored in the long-term memory rather than requiring WM-intensive computations. In other words, under ST, performance on maths problems was improved by making the problem solution less reliant on WM (Beilock et al., 2007).

In the context of Study 4 findings, solve questions require well-practised (i.e., learnt knowledge) solutions that are arguably less reliant on WM resources than comparison questions, that require a more reasoned novel WM-intensive approach. Therefore, in line with the current study findings, one would expect individuals experiencing ST to perform worse on comparison than solve questions, as WM resources required to answer comparison questions will be undermined by the ST. Furthermore, as previously discussed (see Section 5.1.3) inhibitory ability helps control and regulate WM resources: Higher levels of inhibitory ability are associated with greater WM capacity (Kane & Engle, 2003; Kane, Bleckley, Conway, & Engle, 2001). Thus, in the current study, higher levels of inhibitory ability may be indicative of higher levels of WM, and subsequently helped protect performance under ST. High inhibitors may have had more WM resources (c.f. low inhibitors) enabling them to cope simultaneously with both the negative stereotype and maths question computations.

In contrast to these findings, Beilock and Carr (2005) found that high WM individuals' maths performance was more harmed under high-pressure performance situations than their low WM counterparts. Furthermore, the WM account does not provide an explanation for female test-takers' augmented solve maths performance (c.f. controls) in Studies 1 and 3. This implicates the potential role of the motivated component of ST, as hypothesised by mere effort (Jamieson & Harkins, 2009).

5.3.4 Summary

Study 4 tested mere effort's prepotent response mechanism in response to ST. It was hypothesised that the ability to inhibit the prepotent solve response (indicated by Stroop task performance) would protect performance specifically for comparison questions (that require a comparison approach). Findings supported the moderating role of inhibitory ability on maths performance under ST; however this was irrespective of maths question type. Inhibitory ability protected threatened females' maths performance for both solve and comparison questions. The findings can be interpreted using the WM account as an alternative explanation. However, as with the mere effort approach, continued process-orientated research is required to establish *how* question type differentially affects ST effects. Chapter 6 will investigate mere effort's *performance motivation* mechanism, to test if ST motivates test-takers to perform well, affecting their maths question type *preference.*

Chapter 6

The role of motivation in stereotype threat

Study 5 tested the notion that performance motivation (the motivation to disprove the negative stereotype) is implicated in ST. Previous mere effort research measured performance motivation using task performance itself (e.g., Jamieson & Harkins, 2007). That is the outcome (performance) and the process (motivation to disprove the stereotype), are both measured via the dependent variable. This leads to difficulty in distinguishing ST processes from ST effects. Study 5 adapted a maths motivation task (Forbes & Schmader, 2010) to measure motivation in response to ST independently of maths performance. Performance motivation was indexed by maths question type (solve vs. comparison) preference. ST led female participants to select a greater number of solve versus comparison maths questions in line with mere effort.

6.1 Introduction

Performance motivation (i.e., the motivation to perform well) is the initial mechanism that drives stereotype threat (ST) effects according to the mere effort perspective (Jamieson & Harkins, 2007; McFall, Jamieson, & Harkins, 2009). Jamieson and Harkins (2011) propose that performance motivation is intrinsically linked to performance on any given task: it therefore cannot be measured independently. However, this conceptualisation renders performance motivation potentially directly untestable and thus unfalsifiable. For example, Jamieson and Harkins (2007) propose that, in response to ST, individuals are motivated to perform well (i.e., performance motivation), activating whatever response is prepotent for the task. If the prepotent response is correct performance will be facilitated and if prepotent response is incorrect performance will be debilitated (c.f. controls). Consequently, if ST induced performance facilitation and inhibition effects are observed, this is taken as support for the motivational mechanism underlying the effects. In other words, task performance provides evidence for *both* the ST effect and ST mechanism. However, as previously discussed in Chapter 5, other explanations of ST, such as the working memory (WM) account (Beilock, Rydell, & McConnell, 2007) may provide an alternative mechanism to explain these effects (see Chapter 5, Section 5.3.2). Demonstrating ST effects are therefore insufficient as evidence for *how* ST operates. It is essential to design tests that specifically investigate mere effort's ST mechanisms (independent of ST effects) in order to gain a clearer view.

6.1.1 Mere effort and motivation

Jamieson and Harkins (2011) propose that the experience of ST is inherently motivating. Thus, experimental tasks must be related to the negative stereotype to elicit ST effects. One must care (i.e., be motivated) that one's performance could reflect badly on the self (Inzlicht & Ben-Zeev, 2000) and/or one's group (Wout, Danso, Jackson, & Spencer, 2008). In their 2007 study, Jamieson and Harkins reported finding evidence for performance motivation using task performance on an antisaccade task (i.e. the task and process measure were one and the same). The antisaccade task required participants to inhibit their prepotent tendency (to look at a presented cue), and instead look at the opposite side of the display. To elicit ST the task was described as a measure of "visospatial capacity" that was diagnostic of maths ability and had produced gender differences. Subsequently, threatened female test-takers looked in the wrong (prepotent) direction more often than controls, but overall performed better than controls when given time to implement correction and launch corrective saccades. Therefore, the motivation to perform well under ST (i.e., performance motivation) led participants to respond as quickly as possible, even when required to first inhibit the incorrect prepotent response (Jamieson & Harkins, 2007). This test of performance motivation is elegant; although it indexes motivation by task performance (i.e., the speed of accurate responses), the application and correction of prepotent responses were also directly measured. This enabled a clearer interpretation of test-takers' response processes under ST.

6.1.2 Testing performance motivation in maths

One potential measure of performance motivation on maths tests is to test participants *question preference* under ST. Specifically, does ST influence participants' choice of maths question? Question preference is closely associated with task performance but is not measured by task performance itself. It follows that if ST motivates participants to perform well through applying the prepotent solve response, threatened test-takers, when presented with a choice, may demonstrate a stronger preference to answer solve questions. In other words, under ST, test-takers will choose to answer more solve than comparison questions relative to their non-threatened counterparts. Similarly, Forbes and Schmader (2010) implemented a maths motivation task using question choice. The task comprised of 30 maths and 30 verbal remote association problems presented to participants on a series of choice screens. On each screen, participants were asked which type of problem they would like to work on: a maths problem (e.g., "Solve for x: 20 x 16 x 19 x 7") or a remote associates problem (e.g., "Find a fourth word that somehow relates to the following three words: athlete's, web, rabbit"). Maths motivation was indexed by the total length of time spent working on or looking at the maths problems over the course of an allocated 10-minute period. Therefore, the more maths questions selected, the more time spent working on these questions, indicating a greater level of motivation. Question choice, or preference, offers a unique opportunity to test performance motivation separately from maths performance for the first time. This will enable the motivational mechanism of the mere effort account to be tested by comparing question preference under ST for solve versus comparison questions.

6.1.3 Aims of Study 5

Study 5 aims to test the motivational component of the mere effort account. Question preference will be used as an index of performance motivation. Using questions that differ in type (i.e., solve vs. comparison) but are equal in difficulty, participants question choice under ST will give a potential measure of their performance motivation. That is, if ST is inherently motivating and leads test-takers to apply a prepotent solve approach, participants should demonstrate a greater preference to answer solve questions (i.e., prepotent response correct) than comparison questions (prepotent response incorrect). Participants will not be required to answer the questions following their choice; the key measure is the question selection itself. Maths performance will be measured separately on a maths test (as in Studies 1 & 4). Therefore, ST effects and the ST mechanism will be tested separately.

6.2 Pilot study 5a

A pilot study was undertaken to create a maths question selection task. A maths paper was administered consisting of 18 maths questions (*solve* = 9, *comparison* = 9) all set at the GCSE (higher tier) level, and taken from non-calculator papers selected from an online academic source (www.aqa.org.uk). Thirty female student participants (*age range* = 18-24 years, M = 19.79, SD = 1.72) were tested, all with a GCSE grade of C or above to control for mathematical ability. Five questions for each question type (i.e., solve vs. comparison) each worth 3 marks and elicited similar overall scores were selected. The overall total scores of the questions selected for the solve (M = 50.6, SD = 15.08) versus comparison questions (M = 46.6, SD = 15.69), $t(8)$ = .41, p = .69 did not differ significantly. Therefore the finalised question selection task consisted of ten 3-mark questions (5 solve vs. 5 comparison) equally balanced for difficulty. In order to ensure question order did not confound question selection choice, questions were ordered alternatively solve and comparison and were counterbalanced (i.e., there were five variations of question order) (see Appendix F).

6.3 Study 5

Study 5 tested the prediction that ST would influence test-takers' maths question preference. That is, under ST, participants will select to answer more solve than comparison questions. Control participants question preference will be unaffected.

6.3.1 Method

6.3.1.1 Participants and design

One hundred and three female University of Leeds undergraduates ranging in age between 18-22 years (M = 19.36, SD = 1.07) were tested. All had achieved a maths GCSE of grade C and identified as British Caucasian, with English as their first language. Participants were assigned to a 2 (task diagnosticity: high, low) x 2 (question type selected: comparison, solve) mixed design, with repeated measures on the second factor.

6.3.1.2 Materials

The same maths task, ST manipulation, manipulation checks, and additional questions were used as in lab Studies 1 and 4 (see Chapter 3, Section 3.3.1.2). A 10-question maths selection task was also included (see Pilot study 5a).

6.3.1.3 Procedure

The Female experimenter escorted participants into the lab individually. In accord with Study 1, participants first were given the pen and paper non-calculator maths test(s) (solve vs. comparison question type) (see Appendix C). Participants were asked to carefully read the instructions on the front cover that included the ST manipulation. They were then allocated 18 minutes to complete as many as possible of the five 3-mark questions. On completion, participants were given the question selection task (see Appendix F), and instructed to choose a total of any 5 questions to ostensibly answer from a choice of 10 (5 of each question type), by marking a cross in a box next to each question to be answered. Questions were labelled as either “Type A” or “Type B” and participants were informed in the task instructions that “Psychologists have identified two different types of maths questions labelled here as either type A or type B.” Therefore, participants were made aware that there were *differences* between question types but *not* what these differences were.[1] Finally, after making their selection, participants completed manipulation checks and additional measure questions, and were thanked for their time.

6.3.1.4 Dependent measures

The number of solve versus comparison questions selected (from 5 out of 10) on the maths question selection task was the main measure.

[1] Type A were solve questions and Type B were comparison questions.

6.3.2 Results[2]

6.3.2.1 Manipulation checks

Participants responded to the first manipulation check item that gender differences existed on the test to a greater extent in the high diagnosticity (M = 6.11, SD = 2.56) relative to the low diagnosticity (M = 2.89, SD = 2.64) condition, $t(101) = 6.24$, $p < .001$. On the second manipulation check item, there was a marginally significant trend for participants in the high diagnosticity condition to report that they believed males (vs. females) would perform better on the test (M = 4.75, SD = 2.18) in comparison to participants in the low diagnosticity condition (M = 5.26, SD = 1.29), $t(93.25) = -1.46$, $p = .07$. Similarly to Studies 2, 3, and 4, the manipulation was only partially effective.

6.3.2.2 Maths question selection order effects

A 5 (question selection task) x 2 (question type selected) mixed Analysis of Variance (ANOVA), with repeated measures on the second factor, was conducted to check for the potential effects of question order on question selection. As expected, no significant interaction between question type selected x question selection task, $F(1, 98) = .19$, $p = .94$, $\eta p^2 = .01$, was observed. The counterbalancing of question order across the question selection task was successful.

---

[2] Maths performance was analysed but is not reported. The focus of the study is maths question type selection under stereotype threat.

6.3.2.3 Maths question selection

Participants' maths question selection data was then subjected to the main 2 (task diagnosticity) x 2 (question type selected) mixed Analysis of Covariance (ANCOVA), with repeated measures for the second factor. Question type (i.e., which type of maths test participants completed before the question selection task) was the covariate, to rule out the potential confound of the type of maths questions answered influencing the type of questions later selected. There were no significant main effects for question type selected, $F(1, 100) = .55$, $p = .46$, or task diagnosticity, $F(1, 100) = .00$, $p = .50$. As expected, the covariate, question type, was not significant, $F(1, 100) = .00$, $p = 1.00$, and therefore did not affect participants question selection choice. Crucially, as hypothesised, there was a significant interaction between task diagnosticity x question type selected, $F(1, 100) = 4.11$, $p = .05$, $\eta p^2 = .04$ (see Figure 6.1).

The significant interaction was decomposed using t-tests. As expected, in the case of high diagnosticity, participants chose to answer solve questions ($M = 3.18$, $SD = 1.00$) significantly more than comparison questions ($M = 1.82$, $SD = 1.00$), $t(56) = 5.01$, $p < .001$. In contrast, in the case of low diagnosticity, there were no differences in question choice between solve ($M = 2.76$, $SD = 1.16$) and comparison ($M = 2.24$, $SD = 1.16$) questions, $t(45) = 1.53$, $p = .13$. Participants in the high diagnosticity condition chose to answer significantly more solve questions ($M = 3.18$, $SD = 1.00$) relative to participants in the low diagnosticity condition ($M = 2.76$, $SD = 1.16$), $t(101) = 1.95$, $p = .03$. Participants in the high diagnosticity condition therefore also chose to answer significantly fewer comparison questions ($M = 1.82$, $SD = 1.00$) than participants in the low diagnosticity condition ($M =$

2.24, *SD* = 1.16), *t*(101) = -1.95, *p* = .03. The results are supportive of the hypothesis.

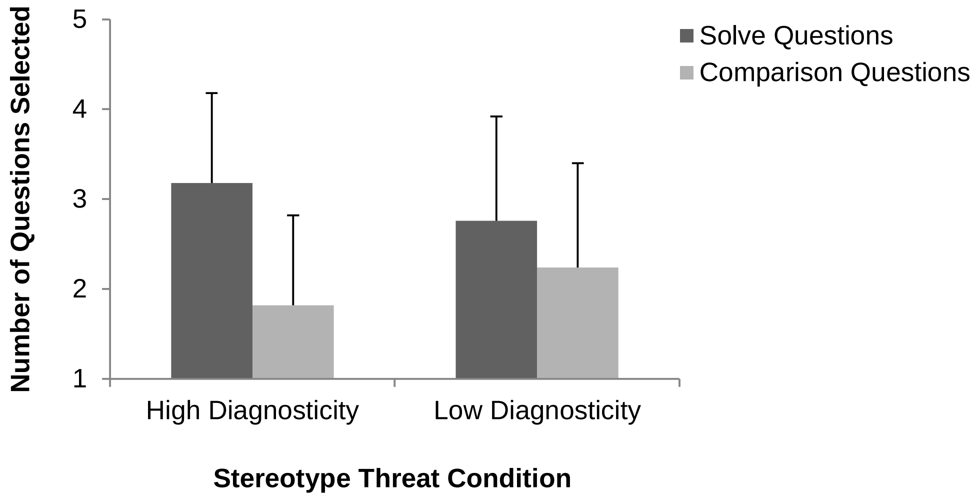


Figure 6.1. Mean number of questions selected for each question type on the maths question selection task as a function of diagnosticity in Study 5. The error bars represent the standard deviations.

6.3 Discussion

The hypothesis that ST affects female test-takers' maths question type preference was tested. Study 5 demonstrates a greater preference to answer solve questions than comparison questions following ST. The maths-gender ST therefore led participants to choose more solve questions. As expected, no differences in question type preference were observed in the control condition. Importantly, maths type question selection was not affected by the type of maths test (i.e., solve vs. comparison) completed beforehand.

The present study findings therefore support the motivational component of the mere effort account (Jamieson & Harkins, 2007, 2009). Performance motivation was measured using the novel method of question type preference and was tested independently of maths performance. Thus, the mechanism was tested separately from the effect using a task that was specifically process-orientated yet appeared to participants as maths related. As hypothesised, those experiencing threat were motivated to perform well and undermine the negative stereotype that subsequently influenced their question type preference.

However, difficulties arise in distinguishing how performance motivation influences question type selection under ST. Specifically, why does performance motivation lead participants to choose more solve than comparison questions? According to mere effort, under ST, performance motivation potentiated the prepotent solve response (Jamieson & Harkins, 2009, 2012), that subsequently led participants to choose to answer questions where the solve approach is applicable (i.e., solve questions). Another potential explanation for current study findings is that performance motivation led threatened test-takers to select more solve questions because these questions arguably rely less on WM resources that may have been disrupted by ST.

It is therefore difficult to infer how performance motivation drives ST effects and how it interacts with other cognitive and affective ST processes (Steele, Spencer, & Aronson, 2002). In Study 4, when testing the prepotent response mechanism of ST, no support was found for prepotent response inhibition moderating performance on comparison questions. Inhibition instead moderated performance for both question types, irrespective of whether the prepotent response was the correct approach to answer the question. This implicates the role of other ST explanations, such as the WM account (Beilock, et al., 2007; Schmader & Johns, 2003). Clearly, further process-orientated research is needed to continue to design new methods to test ST processes to understand how ST operates. Combined with the previous studies, the current research has investigated the mere effort account of ST; testing both its proposed ST effects and ST mechanisms using question type. The implications of the research and a more in-depth discussion will be undertaken in the general discussion in Chapter 7.

6.3.1 Summary

The findings from Study 5 suggest that the maths-gender ST not only impacts female test-takers' maths performance but also their maths question preference. Following exposure to the ST, female participants were more inclined to choose to answer solve relative to comparison questions. The observed effects are consistent with performance motivation as a ST mechanism. This is in accord with the mere effort perspective that ST seemingly motivates test-takers to perform well to disprove the negative stereotype. However, determining precisely *why* performance motivation led participants to choose more solve questions should be addressed by future research. According to mere effort, performance motivation potentiated the prepotent solve response resulting in the preference to answer solve questions. However, the preference for solve questions may also be interpreted using the WM explanation; test-takers choose to answer more solve (vs. comparison) questions as these rely less on WM resources that may be impaired by ST. Chapter 7 will further discuss and interpret the findings.

Chapter 7

General discussion

7.1 The maths-gender stereotype threat revisited

In Chapter 2, three main research aims to investigate the maths-gender stereotype threat (ST) ‘women are poorer at mathematics’ were identified:

(a) To test outcomes of the maths-gender ST on female maths performance, based on maths question type (i.e., solve vs. comparison);
(b) To test the mere effort account in educational settings (i.e., university and school examinations);
(c) To test mere effort’s prepotent responses and motivational ST processes.

In Chapters 1 and 2, ST was explored as an explanation for gender differences in maths performance. Specifically, the pernicious maths-gender stereotype may undermine females’ maths performance in examinations, resulting in tangible performance decrements (i.e., ST effects) (e.g., Brown & Day, 2006; Neuville & Croizet, 2007). Alternatively, performance increases have also been documented following exposure to ST (e.g., Chalabaev, Stone, Sarrazin, & Croizet, 2008; Shih, Ambady, Richeson, & Fujita, 2002). An understanding of how ST affects maths performance is paramount in ensuring all students are able to perform in examination conditions to their full ability. The mere effort account (Jamieson & Harkins, 2007) was reviewed as an explanation for differential performance effects (i.e., performance enhancement or debilitation) in ST. Mere effort focuses on the motivation to disprove the negative stereotype potentiating whatever response is prepotent (i.e., most likely to be produced) on a task. Threatened test-takers' performance is therefore determined by whether the prepotent response is correct or not for the type of question encountered at task (Jamieson & Harkins, 2007, 2009).

In the context of maths, female test-takers' prepotent response is to apply the *solve* approach using learnt formula and knowledge (Jamieson & Harkins, 2009). Thus, when responding to *solve type* questions based on prepotent learnt knowledge, activation of a negative stereotype may motivate test-takers and improve performance. However, *comparison type* questions (requiring logic or estimation), may result in performance decreases, because test-takers seek to undermine the negative stereotype leading to a failure in inhibiting prepotent (i.e. solve) information. Testing the differential ST performance effects (i.e., facilitation vs. debilitation) based on the type of maths question encountered (i.e., solve vs. comparison) in response to ST was investigated in Chapter 3.

Chapter 2 highlighted the necessity to test the maths-gender ST in the field. ST is frequently cited as a determinant to educational equality, yet there is a deficit in research replicating ST effects in real exam or school settings (Huguet & Regner, 2007). Accordingly, Chapter 4 investigated the mere effort account of ST in educational settings. The potential interactive effects of ST and question type were tested during a university statistics mock exam (Study 2) and a secondary school maths mock exam (Study 3). Study 3 also included *male* participants to provide clearer evidence for how ST leads to gender differences in maths performance

(Stoet & Geary, 2012). That is, if male maths performance is also affected by the maths-gender ST.

Chapter 2 also identified the need for new research methodologies to more directly identify the mechanisms implicated in ST, particularly in the mere effort explanation. This issue was tackled in Chapters 5 and 6, which focused on testing *prepotent response inhibition* and *motivational* ST processes, respectively. Specifically, Chapter 5 investigated whether the ability to inhibit the prepotent solve response (indexed by Stroop task performance) protected performance on comparison questions under ST. In Chapter 6, a new method to measure performance motivation was devised, using maths question type preference for solve versus comparison questions.

7.1.1 <u>Summary of research undertaken</u>

This research set out to investigate how a variation in the type of question encountered at test may influence ST performance effects. This was investigated using Jamieson and Harkins's (2007) mere effort account of the maths-gender ST as an explanation for female maths underperformance. A substantial body of research has focused on differences in maths question *difficulty* in relation to ST effects (O'Brien & Crandall, 2003; Steele, Spencer, & Aronson, 2002), yet little research has investigated differential ST susceptibility across maths question *type*. It was hypothesised that, under ST, female maths performance would be facilitated on *solve type* questions, as the potentiated prepotent solve response (i.e., to apply learnt formula) is the correct approach to answer these questions. In contrast, it was hypothesised that performance would be debilitated on *comparison type* questions that instead require the comparison approach (i.e., reasoning, estimation). This research next examined if the interactive effect of ST and question type were replicable in the field. Two studies were conducted in a university and a secondary school under examination conditions. Finally, the mechanisms potentially underlying mere effort (i.e., prepotent responses and motivation) were explored. In terms of prepotent responses, it was hypothesised that the superior ability to inhibit the prepotent solve response (potentiated by ST) would protect maths performance for comparison questions. That is, inhibitory ability would enable test-takers to inhibit the solve approach and apply the correct comparison approach required to answer comparison questions. The motivation to disprove the negative ST was also hypothesised to affect maths question type *preference*, leading to selection of more solve versus comparison questions. This was expected because the motivation to disprove the stereotype leads to the overproduction of the prepotent response, attracting participants to answer more questions where the prepotent response can be applied (i.e., solve questions).

In Chapter 3, Study 1 addressed aim (a) ‘to test outcomes of the maths-gender ST on female maths performance, based on maths question type (i.e., solve vs. comparison)’. Study 1 tested female participants using two maths tests equal in difficulty (see Pilot study 1a), but differing in question type (i.e., solve vs. comparison) in a between-subjects design. Task diagnosticity was manipulated by instructing participants that the test had (or had not) been shown to produce gender differences (adapted from Steele & Aronson, 1995). Study 1 replicated and extended Jamieson and Harkins’s (2009) findings to show that performance on the solve question type maths test was facilitated, whereas performance on the comparison question type maths test was reduced (c.f. controls), following ST. The interactive effect was shown for the first time using UK standardised secondary school General Certificate of Education (GCSE) maths questions. Performance was indexed performance using both unadjusted maths score typically used to grade real maths examinations (such as GCSEs), and the adjusted percentage of problems solved widely used in ST research (e.g., Gimmig, Huguet, Caverni, & Cury, 2006; Schmader & Johns, 2003).

Studies 2 and 3 in Chapter 4 tested aim (b) ‘to test the mere effort account in educational settings (i.e., university and school examinations)’. The role of question type on maths performance under ST was tested in the field, to see if the interactive effects (as found in laboratory Study 1) were replicated during real-world maths performance. Study 2 tested female psychology undergraduates during a statistics mock exam and found that ST debilitated performance on comparison questions and overall maths performance. However, no evidence was found for ST facilitating performance on solve questions. In Study 3, both male and female secondary school students aged 14-16 years (the target GCSE age) were tested during a GCSE maths mock exam. As expected, in Study 3, female maths performance was augmented for solve questions and debilitated for comparison questions. In addition, consistent with ST lift research (Walton & Cohen, 2003), male maths performance was also improved under ST, irrespective of question type. In all, Studies 2 and 3 findings generally demonstrated that question type and ST interact to specifically affect female maths test performance in real test scenarios.

Aim (c) ‘to test mere effort’s prepotent responses and motivational ST processes’ was addressed in Chapters 5 and 6. Process-focused measures were designed and incorporated in Studies 4 and 5 to directly test potential prepotent response inhibition and motivational ST mechanisms respectively. In Chapter 5, Study 4 implemented a Stroop task to investigate the potential moderating role of inhibitory ability on ST effects, dependent on the type of maths question encountered. The greater ability to inhibit the prepotent solve response was hypothesised to enable test-takers to apply the correct comparison approach. This would protect comparison question performance when encountering ST. Solve question performance would be unaffected as the prepotent solve response would not need to be inhibited. However, in contrast to this premise, inhibitory ability moderated female maths performance under ST overall, rather than specifically for comparison questions. The findings suggest that alternative explanations of ST, such as working memory (WM) (Schmader & Johns, 2003) may be useful.

Study 5, in Chapter 6, adapted a maths motivation task (Forbes & Schmader, 2010) to innovatively measure performance motivation in response to ST. Previous research has only measured performance motivation using task performance itself (Jamieson & Harkin, 2007; McFall, Jamieson, & Harkins, 2009), which can create difficulty in distinguishing ST processes from ST effects. In Study 5 a maths question selection task was employed, that indexed performance motivation using maths question type preference. Questions were pretested and equal in difficulty across maths question type (see Pilot study 5a). In line with predictions, ST led female test-takers to select a greater number of solve versus comparison maths questions.

7.2 Contributions of the present research

In this section, the contributions of the present research are detailed and linked to the literature reviewed in the previous chapters. Evidence for the outcomes and processes implicated in the mere effort account will be discussed.

7.2.1 Understanding stereotype threat effects

The current research found reliable evidence of ST effects on females' maths performance. Specifically, important contributions were made in understanding differential ST performance effects based on maths question type (i.e., research aim a), and replicating these effects in educational settings (i.e., research aim b).

Indeed, in accord with research aim a, previous ST research has largely focused on maths question difficulty rather than question type (Neuville & Croizet, 2007; Nguyen & Ryan, 2008; O'Brien & Crandall, 2003; Spencer, Steele, & Quinn, 1999; Steele, 1997). For example, as discussed in Chapter 3, a recent large-scale ST study conducted by Ganley et al. (2013) focused on the use of *fairly difficult* mathematic assessments and conducted follow-up analyses only on relatively difficult questions (i.e., those with less than 50% correct) (see Section 3.1). However, not controlling for maths question type can potentially confound ST findings; as shown in Study 1, the maths-gender ST led participants to perform better on solve type questions and worse on comparison type questions compared to their non-threatened counterparts. Question difficulty was controlled for across the maths question types. Thus, the observed ST effects on maths performance in Study 1 resulted from differences in the *type* of question encountered. Furthermore, these performance increases and decreases under ST largely 'cancelled out' any overall ST performance effects: ST only affected maths performance when including question type within the analysis. The null results from ST research, such as Ganley et al. (2013), in which maths question type was not manipulated are therefore difficult to clearly interpret.

Hence, Study 1 findings confirm the importance of question type in determining how ST affects performance: Female test-takers are more susceptible to the negative effects of ST on comparison type questions and bolstered on solve type questions. The effects serve to replicate and extend Jamieson and Harkins's (2009, 2012) findings, to show the interactive effects of ST and question type are generalizable across different standardised maths tests and marking formats. Indeed, a recent meta-analysis of 17 years of maths-gender ST research revealed that the maths-gender ST is predominantly tested using Graduate Record Examination (GRE) maths questions, and highlighted a deficit in constructive replication (see Picho, Rodriguez, & Finnie, 2013).

7.2.1.1 Stereotype threat in the field

In line with research aim b, ST effects based on question type were replicated in both university and secondary school examination settings (Studies 2 and 3). This clearly demonstrates the real-world applicability of the maths-gender ST impeding female maths performance, and contributes to the limited body of ST research conducted in real educational environments (Huguet & Regner, 2007; Wei, 2012; Wicherts, Dolan, & Hessen, 2005). Indeed, mere effort's maths performance effects had only previously been tested in the laboratory. Moreover, of the ST field research, emphasis is often placed on maths question difficulty rather than question type (e.g., Ganley et al., 2013; Huguet & Regner, 2007; Keller, 2007). Thus, Studies 2 and 3 address a significant gap in current maths-gender ST field research to demonstrate that question type interacts with ST in educational settings.

Study 2 tested female psychology undergraduate students during their Research Skills 1 (RS1) statistics mock exam; a key part of the University of Leeds psychology undergraduate syllabus. With the exception of the explicit ST manipulation and manipulation checks, the exam and examination conditions were identical to those for the RS1 mock exam. The limitations of the explicit ST manipulation are later discussed in Section 7.4.2. Exam questions were categorised as being either solve or comparison in nature; questions that could not be categorised were not included in the analysis (see Chapter 4, Section 4.2.1.2). Crucially, replicating Study 1 findings, ST differentially impacted exam performance based on question type. Female test-takers performed significantly worse under ST on the comparison questions (c.f. controls). However, in contrast, female test-takers' performance for solve questions was unaffected by ST. This may have occurred (as discussed in Chapter 2, Section 2.3.1.2), according to mere effort, because ST decreases performance for comparison questions to a greater extent than it improves performance for solve problems (Jamieson & Harkins, 2009). Indeed, the performance detriment to the comparison questions resulted in the negative impact of the maths-gender ST on exam performance. These findings therefore shed new light on how female performance may be disrupted by ST in real-world test situations.

Study 3 findings also contributed to understanding how the maths-gender ST operates in educational settings. Specifically, Study 3 tested a secondary school GCSE target age group (14-16 years) of *both* male and female students during a GCSE maths mock exam. This enabled a closer examination of how ST effects may contribute to the gender gap in mathematics (i.e., how ST may differentially impact male and female maths performance). There has been controversy surrounding ST's role in the maths-gender performance gap[1] because there is a lack of field studies testing both male and female participants (Stoet & Geary, 2012). Similarly to Study 2, typical examination conditions were enforced (see Chapter 4, Section 4.3.1.3), so that the exam environment replicated that of a real GCSE examination. Importantly, the same GCSE maths questions from laboratory Studies 1 and 4 were included (combined in one maths test; see Appendix E) in order to directly test the interactive ST and question type laboratory effects in a real-world exam environment.

Study 3's findings revealed that GCSE maths exam performance is susceptible to the maths-gender ST. Consistent with previous mere effort laboratory research (Jamieson & Harkins, 2009: Exp 1) there was a significant interaction between diagnosticity, question type and gender. As expected (and as found in the present research's Study 1), ST differentially affected female maths performance based on question type: Performance was enhanced on solve questions and debilitated on comparison questions (c.f. controls). However, in contrast to the traditional concept of ST impeding female maths performance (Steele, 1997), ST did not reduce maths performance overall. This suggests that the role of other factors that may attenuate female maths performance beyond the scope of the mere effort explanation. Furthermore, male maths performance was *improved* under ST (c.f. controls), regardless of the type of maths question encountered. ST may affect male and female performance via different mechanisms. Specifically, female maths

[1] For example, recent UK educational statistics show that for maths and additional maths GCSEs in 2011 and 2012, a greater cumulative percentage of boys that took the exam achieved a greater number of top grades than girls (GCSE Results, 2012).

performance may have been affected (as proposed by mere effort) by the heightened motivation and overproduction of the prepotent response under ST; that led to differential performance effects based on question type. Whereas male maths performance seems to have been lifted by the downward social comparison they could make with females (Walton & Cohen, 2003), resulting in overall enhanced maths performance under ST. The findings therefore illustrate how ST can manipulate both male and female test-takers' maths performance.

Overall, the pattern of findings in Studies 2 and 3 demonstrate how a variation in question type can affect ST effects in educational settings. Study 3 also suggests that it is exclusively female performance that is susceptible to ST effects based on question type, as male maths performance was unaffected by the type of maths question encountered (male maths performance was lifted under ST irrespective of question type). Collectively, the present research adds to the growing body of evidence for the role of question type in determining female maths performance under ST (e.g., Beilock, Rydell, McConnell, 2007; Jamieson & Harkins, 2009, 2012).

7.2.2 Understanding stereotype threat processes

In order to understand *how* ST and question type interact, it is important to test the ST processes underlying the ST performance effects. For instance, although the present findings are supportive of mere effort's performance outcomes, they are also generally consistent with the WM account (Schmader & Johns, 2003) as an alternative explanation. It is therefore erroneous to suggest that the observed ST effects based on question type are evidence exclusively for the mere effort explanation.

As previously discussed in Chapter 5 (see Section 5.3.2), WM proposes that ST operates by disrupting WM memory resources required to answer certain types of questions (Beilock et al., 2007). Specifically, maths questions that require problem-solving approaches more heavily based on WM resources (e.g., the novel comparison approach) suffer greater performance decrements under ST than questions that rely on more long-term memory based approaches (e.g., the well-learned solve approach). Therefore, as found in the present research, performance on comparison questions would be harmed under ST as the negative stereotype interferes with test-takers' WM resources needed to answer comparison questions.

Thus, the contribution of the present research findings as evidence for the mere effort account requires the review of Studies 4 and 5 (which tested mere effort's ST mechanisms). That is, while Studies 1, 2, and 3 tested for mere effort's ST performance outcomes, Studies 4 and 5 investigated mere effort's ST processes. It was essential that methods were designed and implemented to explicitly test mere effort's ST processes independently from ST effects. This is important because it has previously been argued that mere effort is task specific (Jamieson & Harkins, 2007), which presents a problem in distinguishing ST processes from ST effects. Studies 4 and 5 process-focused approach enabled the closer examination of mere effort ST processes. These included; the overproduction of (a) *prepotent responses* activated by heightened levels of (b) *performance motivation* to disprove to the stereotype.

7.2.2.1 Prepotent response inhibition

Study 4 investigated the potential role of inhibition in ST effects, specifically; prepotent response inhibition. Prepotent response inhibition refers to the ability to supress contextually specific, dominant response tendencies (i.e., prepotent responses) that are inappropriate for a given task (Friedman et al., 2008; Grandjean & Collette, 2011; Hasher, Quig, & May, 1997). Mere effort proposes that ST operates via the over activation of prepotent responses from the motivation to disprove the negative stereotype. Thus, in the context of the maths-gender ST, if ST does potentiate the solve response, then threatened female test-takers with a greater aptitude to supress prepotent responses may be able to inhibit the solve response. The ability to inhibit the solve response may act as a buffer to protect their performance on comparison questions. Prepotent response inhibition would enable other new approaches (i.e., the comparison response) to be considered and applied for more successful problem solving on these questions (Carr & Steele, 2009).

Study 4 implemented a Stroop task to specifically measure test-takers' prepotent response inhibition ability (e.g., Dao-Castellana et al., 1998; Davidson, Zacks, & Williams, 2003; Friedman & Miyake, 2004; Mutter, Naylor, & Patterson, 2005). Measuring inhibitory ability as a moderator provided a method to directly test mere effort's prepotent response mechanism. Inhibition tasks (such as the Stroop task) are reliant on the same mechanisms that hypothetically underpin mere effort (i.e., the ability to inhibit the dominant habitual response). The Stroop task requires participants to inhibit the dominant prepotent tendency to read the colour of the word and instead identify the colour it is printed in (MacLeod, 1991; Stroop, 1935). Therefore, those with the superior ability to inhibit the prepotent response to read the word on the Stroop task should also be more able to inhibit the prepotent solve approach on the maths task. The ability to inhibit the solve approach would protect comparison question performance under ST, by enabling the correct comparison approach to be applied.

Intriguingly, rather than protecting female maths performance against ST on comparison questions (as the prepotent solve response would not need to be inhibited for solve questions), inhibitory ability moderated female maths performance *overall* (i.e., for both solve and comparison questions). That is, low inhibitory ability was associated with poorer maths performance (c.f. high inhibitory ability) in general following ST. Furthermore, in terms of ST effects, as in Studies 1, 2, and 3, the interactive effects of ST and maths question type were found. Study 4 findings therefore provide evidence for mere effort's ST outcomes, but do not support mere effort's prepotent response activation. The overproduction of prepotent responses following ST was not observed; the ability to suppress prepotent responses protected maths performance irrespective of whether the prepotent approach was correct or not.

The findings of a general inhibitory ability ST mechanism are perhaps more consistent with the WM account of ST. As described at the beginning of this section, WM proposes that maths-gender ST impairs maths performance by reducing WM capacity needed to undertake maths questions (Beilock et al., 2007; Rydell, McConnell, & Beilock, 2009; Schmader & Johns, 2003). Indeed, research has proposed that *all* mathematical cognition involves central executive resources of the WM (DeStefano & LeFevre, 2004). Consequently, those with higher (vs. lower) WM ability may be more able to cope with ST; as they have more resources to cope simultaneously with both the ST and maths problem-solving demands (Regner et al., 2010; Schmader, Johns, & Forbes, 2008). Inhibitory mechanisms are an essential component of WM (Hasher & Zacks, 1988). As discussed in Chapter 5 (see Section 5.1.3), research has demonstrated the positive relationship between WM capacity and inhibitory ability (Grandjean & Collette, 2011; Kane & Engle, 2003; Kane, Bleckley, Conway, & Engle, 2001). Thus, in Study 4, a greater level of inhibitory ability (measured by Stroop) may have been indicative of higher levels of WM that helped protect performance under ST, as there were more WM resources able to attend to both the negative stereotype and maths task requirements. The full implications of Study 4 findings with regards to WM will be discussed further in the following section (see Section 7.3).

7.2.2.2 Performance motivation

The present research also tested another mere effort ST mechanism; performance motivation. In Studies 1 and 3, the improvement in performance on solve type questions implicated the motivational component of the mere effort account. The majority of ST explanations, such as the WM account, focus on negative phenomenological experiences in response to the maths-gender ST (see Schmader et al., 2008). That is, when encountering the negative stereotype, female test-takers may experience feelings of self-doubt (Steele & Aronson, 1995), anxiety (Beilock et al., 2007) and negative emotions (Keller & Dauenheimer, 2003). In contrast, non-stigmatised (i.e., male) test-takers may feel confident and energised by the positive stereotype. The mere effort account differs to argue that stigmatised individuals may also feel energised through motivation to disprove the threatening stereotype (i.e., performance motivation). This would explain inconsistencies in self-reported measures of negative phenomelogical experiences (see Wheeler & Petty, 2001, for a review). In addition, this may contribute to previous research findings that blurring intergroup boundaries (i.e., by focusing on shared gender characteristics) reduces ST effects on maths performance (Rosenthal & Crisp, 2006). That is, reducing intergroup bias and weakening the negative stereotype may relatively lessen the performance motivation to disprove it (that can lead to the potentiation of incorrect responses).

Performance motivation has only previously been measured using task performance itself (Jamieson & Harkins, 2007, 2011). However, as demonstrated by the prepotent response mechanism in Study 4 (see Section 7.2.2.1), evidence for ST effects are insufficient as evidence for ST processes. That is, mere effort's performance effects were found *despite* a lack of support for its ST prepotent response mechanism. Thus, it was important Study 5 designed a task able to index performance motivation *separately* from maths performance but was still closely related to maths performance. The innovative adaption of Forbes and Schmader's (2010) maths motivation task enabled performance motivation to be measured using maths *question type preference* (solve vs. comparison). As expected, Study 5 revealed that ST led female test-takers to select to answer more solve than comparison questions. No differences in question type preference were observed for control participants. Thus, Study 5 findings suggest that the experience of ST is inherently motivating (Jamieson & Harkins, 2011): ST enhances test-takers' performance motivation to disprove the stereotype that drives them to select questions that they will perform well on.

The preference for solve questions arguably indicates that performance motivation led participants to choose questions where the potentiated prepotent solve response can be applied. However, in light of the previous findings (see

Section 7.2.1.1), another possible interpretation of these results is that ST motivated participants to perform well (i.e., mere effort), but that this resulted in the preference for questions that rely less on WM resources (i.e., WM account). Specifically, if ST disrupted WM resources, participants may have chosen to answer more solve questions as these require the recall of learnt formula stored in the long-term memory, rather than comparison questions that need novel WM-intensive solutions (Beilock et al., 2007).

In sum, in terms of understanding ST processes, the present research has found inconsistent evidence for mere effort’s ST mechanisms. First, in regards to prepotent response inhibition, Study 4’s finding that a greater level of inhibitory ability (indicative of greater levels of WM) protected overall maths performance is more supportive of the WM account. The findings suggest that ST operates predominantly through taxing WM resources that are needed to compute maths questions, rather than over activating prepotent responses. Second, Study 5 revealed that ST did heighten test-takers’ performance motivation, affecting their preference to answer more solve questions that would result in facilitated performance under ST. However it is unclear whether this selection was based upon the prepotent response being correct for solve questions, or that solve questions require less WM resources. It is interesting that the present research finds evidence that can be related to both the mere effort and WM accounts of ST. The implications of the present research findings (both ST effects and mechanisms) are discussed in the following section.

7.3 Implications of the present research

Collectively, the present research findings testify to the importance of question type in determining ST effects on maths performance. However, it is still unclear *why* question type and ST interact, and which ST account provides the best explanation. The present research findings were mixed for the processes proposed by mere effort, suggesting that the theory is limited. The findings point towards the potential integration of the mere effort and WM accounts of ST.

7.3.1 Mere effort and working memory: an integrated approach

In contrast to Jamieson and Harkins's (2007) claims that mere effort is incompatible with the WM account of ST, the present research findings reveal that the mere effort account may provide a complimentary theoretical perspective. Indeed, to investigate the interactive effects of ST and question type in the present research, maths questions were categorised as being either solve or comparison in nature using Jamieson and Harkins's (2009) criteria (see Appendix A). Jamieson and Harkins (2009) focus on differences in question type based on whether the question can be efficiently answered using the prepotent response. This is arguably not dissimilar from question type differences based on how much the question relies on WM resources (Beilock et al., 2007). The prepotent solve response is a well-learned dominant maths approach, that may be more likely to be stored in the long-term memory. In contrast, the comparison approach requires novel solution strategies based on estimation and reasoning that are likely to rely more heavily on WM resources. Thus, under ST, the activation of the prepotent response (mere effort) alongside the disruption to WM resources (WM account) may have led to decreased performance on comparison questions. Indeed, the idea that these two seemingly competing ST explanations are compatible has recently been supported using an age-related ST on older adults' memory performance (Mazerolle, Regner, Morisset, Rigalleau, & Huguet, 2012). Mazerolle and colleagues (2012) found that ST reduced older adults' use of WM memory processes and simultaneously strengthened their use of automatic memory processes. Mazerolle et al. (2012) findings therefore implicate the respective roles of *both* executive WM resources and prepotent responses in ST effects. The potential links between the WM and mere effort perspectives are now discussed in more detail through the motivation and inhibitory findings from the current research.

7.3.1.1 The role of motivation

In Schmader et al.'s (2008) integrated ST process model, the researchers acknowledged, in line with mere effort, ST does increase the *motivation* to perform well and combat the negative stereotype. However, despite recognising that this heightened motivation may lead to the reliance on automatic responses, the model predominantly focuses on how the threatened test-takers are motivated to resolve the cognitive imbalance created by ST. As threatened test-takers struggle against the ST, this burdens their executive WM resources (needed to perform on the task). This can explain why, as found in Studies 1, 2, 3, and 4, performance on comparison questions is debilated under ST. Comparison questions arguably require more novel WM-intensive computations that are susceptible to interference from ST. Indeed, it may also explain why the motivation to avoid stereotype confirmation by performing well leads test-takers to select more solve than comparison type questions (Study 5). Solve questions rely less on WM resources (c.f. comparison questions) that may be disrupted by the ST.

However, Schmader et al.'s (2008) model does not specifically explain *how* motivation may lead to performance increases on solve questions (as found in Studies 1 and 3). Mere effort on the other hand does attempt to explain this: The heightened performance motivation strengthens automatic response tendencies and potentiates the overproduction of the prepotent response. Therefore, in response to ST, performance is facilitated or impaired depending on whether the activated prepotent response is the correct approach or not. In the context of the maths-gender ST, this explains why performance may be enhanced for solve questions as participants are more motivated to use the prepotent solve response.

7.3.1.2 The role of inhibition

Despite the potential role of mere effort's prepotent response process, Study 4 illustrated that it is not the main mechanism driving ST effects. One would expect if ST operated chiefly through the overproduction of prepotent responses, then the ability to inhibit the prepotent response would specifically protect performance on questions where the prepotent response needs to be inhibited. In contrast, inhibitory ability moderated maths performance overall under ST. High levels of inhibitory ability protected performance on both solve and comparison questions, regardless of whether the prepotent solve response was correct or not. Inhibitory ability may therefore play a more complex role in the interplay between ST mechanisms, rather than simply inhibiting the prepotent response (as proposed by mere effort). Indeed, as previously discussed (Section 7.2.2.1), there is ample evidence that inhibitory ability is indicative of WM, with research proposing an interactive link between the WM and inhibitory efficiency (Kane et al., 2001; Kane & Engle, 2003; Redick, Calvo, Gay, & Engle, 2011; Roberts, Hager, & Heron, 1994). Thus, higher levels of inhibitory ability found in Study 4 may have indicated higher levels of WM that subsequently protected performance under ST. Performance may have been protected as individuals with a high levels of WM (c.f. low levels of WM) have a higher threshold of WM resources needed for task performance while simultaneously coping with threat (Schmader et al., 2008).

Collectively, the present research supports a more unified perspective of the mere effort and WM accounts of ST. Similarities between the accounts' question type criteria have been identified (i.e., the prepotent solve approach requires less WM resources than the comparison approach) and the potential contribution of mere effort's prepotent responses and motivational mechanisms to the WM model have been discussed. Specifically, while the experience of ST may motivate participants to alleviate threat, taxing their WM resources (Schmader et al., 2008), it may also motivate test-takers to combat the ST and potentiates automatic response tendencies (Jamieson & Harkins, 2007).

7.3.2 Female maths question preference explained

The integration of both the mere effort and WM accounts may also help understand the development of females' preference for solve questions (Gallagher et al., 2000; Royer, Tronsky, Chan, Jackson, & Marchant, 1999). The opening quote in Chapter 1 highlighted how differences in male and female mathematical problem solving are observed from as early an age as primary school. Specifically, girls are believed to favour a more structured rule-based approach, and do not necessarily show 'real understanding' (Leicestershire Primary Team, 2005). A potential explanation, stemming from the current findings, is that ST may shape girls *preference* for the solve approach and create a negative performance cycle.

Indeed, research has suggested that the maths-gender ST may interfere with females' ability to learn mathematical skills and operations (Rydell, Rydell, & Boucher, 2010). Thus, the experience of threat may disrupt their WM resources needed to learn and 'understand' new types of maths solutions. Consequently, this may increase their reliance on formulas and structured approaches that can be stored in the long-term memory. The reliance and preference for the 'solve approach' may ultimately result in this approach becoming their dominant and therefore prepotent response tendency. Thus, females' maths performance may be continually impeded by ST, which both disrupts their ability to learn and use new WM-intensive computations. This serves to increase females' adherence to apply the prepotent solve approach.

7.4 Future research and limitations

This thesis has explored the contributions and implications of the present ST research. However, there remain a number of unexplored areas, limitations and research ideas resulting from the current research findings. This section aims to explore these potential research areas and weaknesses more fully.

7.4.1 <u>Future research</u>

There has been a recent shift in ST research from simply identifying ST effects to testing ST mechanisms (Jamieson & Harkins, 2011). The present research contributes to the growing body of process-orientated ST research; that aims to identify underlying ST mechanisms and synthesise these into a clearer picture of how ST operates. In particular, similarities and cohesion between the mere effort and WM accounts of ST have been explored, to create a more comprehensive perspective of how ST may interact with the type of question encountered. Furthermore, Studies 4 and 5 emphasise the need to design and implement process-focused methodologies to directly test ST mechanisms separately from ST effects. For example, the inclusion of a WM depletion task (vs. no depletion) could help determine *why* ST led to the increased preference for solve questions. Specifically, if ST taxes WM resources needed to compute maths questions, then those in the WM depletion condition would have shown an even grater tendency to select solve questions (that rely less on WM computations).

It is important that research also continues to test and replicate the effects of ST based on question type, and (as in Studies 2 and 3) extend investigation in ST field research. This will facilitate understanding of the robustness of the relationship between question type and ST effects, as well as establishing to what degree these effects generalise beyond the laboratory (Sackett, Schmitt, Ellingson, & Kabin, 2001). Indeed, the ultimate aim of future research should be to develop practical interventions to assist stigmatised individuals to perform to their full maths potential. For example, following Study 4 findings that inhibitory ability protected female maths performance from the detrimental effects of ST; one potential route would be to improve inhibitory ability via WM (as inhibitory ability is a component of WM).

Here developmental research is informative: Karbach and Kray (2009) suggested that the performance debilitating effects of negative stereotype may be alleviated by trained improvements in WM. Karbach and Kray (2009) showed that task-switching training led to selective enhancements in both task-switching performance and inhibitory control (measured using Stroop). This finding has implications for the transferability of trained improvements in WM to increase individuals' inhibitory ability that subsequently may moderate ST effects. If inhibitive ability can be taught, then this skill might be used to curtail unwanted performance deficits prompted by ST. Indeed, the plasticity of inhibition has been shown in young adults, with tangible improvements in Stroop performance (i.e., an index of inhibitory ability) observed following training (Davidson et al., 2003; Dulaney & Rogers, 1994; MacLeod, 1998). Thus, the present research findings provide a basis for future research efforts that may investigate the potential of inhibitory ability training as a ST intervention. This may enable female test-takers to channel their increased performance motivation when encountering the maths-gender ST (as found in Study 5) into performing to their full potential.

7.4.2 Limitations of the present research

The majority of ST studies implement an explicit maths-gender ST manipulation (e.g., Brown & Pinel, 2003; Crisp, Bache, & Maitner, 2009; Keller, 2002; Quinn & Spencer, 2001); and the present research is not an exception. As expected, across all of the present research studies, the first manipulation check item revealed that ST participants reported greater gender differences on the maths test relative to their non-threatened counterparts. However, in all of the present research studies (except Study 1), the second manipulation check item revealed no differences in reported beliefs as to who would perform better on the maths test (male vs. female) across ST conditions. Thus, despite threatened participants being aware of the negative ST, they did not necessarily endorse this view and believe that task performance would reflect the negative group stereotype. The manipulation was therefore only partially effective and so test-takers may not have fully experienced ST. One potential solution would be to implement a stronger manipulation of threat. For example Rosenthal, Quinn, and Seddon (2009), following weak ST effects, repeated an experiment with a more explicit stereotype manipulation. That is, as well as being informed that gender-based performance would be compared, participants were also told that this was because females had been shown to perform worse on the task than males.

Furthermore, in Studies 2 and 3, the explicit ST manipulation potentially has some reduced applicability to real-world scenarios. Indeed, informing test-takers that the maths test had (vs. had not) been shown to produced gender differences is not what examiners would typically tell their students (Huguet & Regner, 2007). Thus, future research could implement a more realistic ST manipulation by altering the gender composition of the test-taker group (e.g., Kessels & Hannover, 2008; Picho & Stephens, 2012). However, as discussed in Chapter 4 (see Section 4.4.1), Picho et al.'s (2013) meta-analysis found weak effects for the nature of the maths test environment on ST effects: Female test-takers did not benefit more from test situations that were homogeneous or testing contexts where they formed the majority. Therefore, despite the limitations of the explicit manipulation, it enabled a clearer indication of ST effects. Future research could potentially investigate how ST is activated in real-world test environments.

In terms of ST processes, task engagement has been implicated in confounding the relationship between WM and ST effects (Beilock & Carr, 2005; Gimmig et al., 2006). Research has indicated that low WM individuals may be less affected by threat because they experience less anxiety due to this threat (Gimmig et al., 2006; Schmader et al., 2008). Thus, in Study 4, despite finding effects consistent with WM efficiency (indicated by inhibitory ability) moderating the affect of the maths-gender ST on performance, a measure of task engagement could also have been included.

7.5 Conclusions

The present research set out to investigate the pernicious maths-gender ST. Specifically, whether ST interacts with the *type* of maths question to affect maths performance. The mere effort approach of ST was focused upon, using *solve* (e.g., equations) and *comparison* (e.g., probability) type questions. First, differences in how question type differentially affected performance under ST (i.e., performance facilitation or debilitation) were addressed. Second, the application of mere effort to educational settings (i.e., university and school examinations) was investigated. Third, the processes potentially driving mere effort were explored (i.e., prepotent responses and performance motivation). The present research revealed that variation in how a question can be answered differentially impacts female maths performance under ST. The interactive effects of question type and ST on female maths performance were also shown to be applicable in educational settings. Additionally, the finding that male maths performance was augmented under ST, irrespective of question type, suggests that ST lift operates via different mechanisms to ST effects. The present research therefore illustrates how the maths-gender ST can alter both male and female test-takers' maths performance and exacerbate the maths-gender performance gap.

Furthermore, in terms of ST processes, the present research suggests a more unified perspective of mere effort and WM accounts to explain the ST phenomenon. Indeed, similarities were identified between mere effort and WM in defining question type. Specifically, whether a question can be answered using the prepotent response (according to mere effort) arguably matched differences in how much the question relies on WM resources. For example, comparison questions that do not use the prepotent solve response also require novel WM intensive computations, whereas solve questions that use the prepotent response do not rely heavily on WM resources.

Mere effort's motivational and prepotent response mechanisms were also related to the WM model of ST. A strength of the mere effort perspective is that it helps to explain how performance motivation can augment performance on solve questions; via the activation of the prepotent solve response. However, the present research findings indicate that the overproduction of prepotent responses do not drive ST effects. The ability to inhibit the prepotent response moderated maths performance under ST *overall*, regardless of whether the prepotent response needed to be inhibited or not (i.e., not specifically to comparison questions). Inhibitory ability may therefore index WM levels that help individuals cope with the maths-gender ST. Individuals with dispositionally higher levels of WM may have more WM resources to simultaneously deal with both the threat and maths task demands. Thus, through the research aims set out in Chapter 2, this thesis has found:

(a) Question type (i.e., solve vs. comparison) differentially affect the outcome of female test-takers' maths performance under the maths-gender ST;

(b) ST effects based on maths question type are generalizable to female maths performance in educational settings;

(c) Male maths performance is lifted by the maths-gender ST, irrespective of question type;

(d) Inhibitory ability moderates the impact of ST on female maths performance (irrespective of maths question type);

(e) ST increases female test-takers' motivation to perform well and influences their maths question preference;

(f) Links between mere effort and WM accounts of ST may provide a complimentary theoretical perspective. The experience of ST may motivate participants to alleviate the threat, taxing their WM resources (Schmader et al., 2008); and may also motivate test-takers to combat the ST, potentiating the prepotent response (Jamieson & Harkins, 2007).

Future research should continue to investigate how maths test composition (i.e., question type) affects female maths performance under ST. Research should also continue to test ST mechanisms using a process-focused approach, with a view to develop and integrate these mechanisms into a more comprehensive understanding of how ST operates. This requires the design and implementation of new research methodologies to specifically test ST mechanisms separately from ST effects. Additionally, due to its high relevance to education, future maths-gender ST research should move towards replicating and testing ST in the field, in real examination settings. Ultimately, such research will enable the development of practical ST interventions, such as inhibition training, that may assist female test-takers to both learn and perform to their full maths ability to help eradicate the maths-gender ST.

References

Adams, R. (2013, August 15). A-level results 2013: fewer students get top grades for second year running. *The Guardian.* Retrieved from http://www.theguardian.com/education/2013/aug/15/a-level-results-2013-students-top-grades

Aguinis, H., & Lawal, S. O. (2013). eLancing: A review and research agenda for bridging the science-practice gap. *Human Resource Management Review, 23*(1), 6-17.

Anderson, C. A., Lindsay, A. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science, 8*(1), 3-9.

Ankney, C. D. (1992). Sex-differences in relative brain size - the mismeasure of woman, too. *Intelligence, 16*(3-4), 329-336.

Aronson, J., Lustina, M. J., Good, C., Keough, K., & Steele, C. M. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*(1), 29-46.

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen J. J. A., Fiedler, K., Fiedler, S., Funder, D. C., Kliegl, R., Nosek, B. A., Perugini, M., Roberts, B. W., Schmitt, M., van Aken, M. A. G., Weber, H., & Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27,* 108-119.

Avolio, B. J., Reichard, R. J., Hannah, S. T., Walumbwa, F. O., & Chan, A. (2009). A meta-analytic review of leadership impact research: Experimental and quasi-experimental studies. *Leadership Quarterly, 20*(5), 764-784.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173-1182.

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin, 130*(4), 553-573.

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior, 25*(5), 475-491.

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 of the United States of America, 107*(5), 1860-1863.

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology-General, 136*(2), 256-276.

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology, 41*(2), 174-181.

Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science, 12*(3), 225-229.

Bonnot, V., & Croizet, J. C. (2007). Stereotype internalization and women's math performance: The role of interference in working memory. *Journal of Experimental Social Psychology, 43*(6), 857-866.

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology, 40*(2), 247-255.

Brodish, A. B., & Devine, P. G. (2009). The role of performance-avoidance goals and worry in mediating the relationship between stereotype threat and performance. *Journal of Experimental Social Psychology, 45*(1), 180-185.

Brown, R. P., & Day, E. A. (2006). The difference isn't black and white: Stereotype threat and the race gap on Raven's Advanced Progressive Matrices. *Journal of Applied Psychology, 91*(4), 979-985.

Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology, 39*(6), 626-633.

Brutus, S., Gill, H., & Duniewicz, K. (2010). State of science in industrial and organizational psychology: A review of self-reported limitations. *Personnel Psychology, 63*(4), 907-936.

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology, 33*(2), 267-285.

Cadinu, M., Maass, A., Lombardo, M., & Frigerio, S. (2006). Stereotype threat: The moderating role of locus of control beliefs. *European Journal of Social Psychology, 36*(2), 183-197.

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science, 16*(7), 572-578.

Carr, M., Jessup, D. L., & Fuller, D. (1999). Gender differences in first-grade mathematics strategy use: Parent and teacher contributions. *Journal for Research in Mathematics Education, 30*(1), 20-46.

Carr, P. B., & Steele, C. M. (2009). Stereotype threat and inflexible perseverance in problem solving. *Journal of Experimental Social Psychology, 45*(4), 853-859.

Ceci, S. J., & Williams, W. M. (2010). Sex differences in math-intensive fields. *Current Directions in Psychological Science, 19*(5), 275-279.

Chalabaev, A., Stone, J., Sarrazin, P., & Croizet, J. C. (2008). Investigating physiological and self-reported mediators of stereotype lift effects on a motor task. *Basic and Applied Social Psychology, 30*(1), 18-26.

Cole, B., Matheson, K., & Anisman, H. (2007). The moderating role of ethnic identity and social support on relations between well-being and academic performance. *Journal of Applied Social Psychology, 37*(3), 592-615.

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review, 8*(2), 331-335.

Conway, A. R. A., & Engle, R. W. (1994). Working-memory and retrieval - a resource-dependent inhibition model. *Journal of Experimental Psychology-General, 123*(4), 354-373.

Conway, A. R. A., Tuholski, S. W., Shisler, R. J., & Engle, R. W. (1999). The effect of memory load on negative priming: An individual differences investigation. *Memory & Cognition, 27*(6), 1042-1050.

Cottrell, N. B. (1972). Social facilitation. In C. G. McClintock (Ed.), *Experimental social psychology* (pp. 185-236). New York: Holt, Rinehart, & Winston.

Crisp, R. J., Bache, L. M., & Maitner, A. T. (2009). Dynamics of social comparison in counter-stereotypic domains: Stereotype boost, not stereotype threat, for women engineering majors. *Social Influence, 4*(3), 171-184.

Croizet, J. C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24*(6), 588-594.

Croizet, J. C., Despres, G., Gauzins, M. E., Huguet, P., Leyens, J. P., & Meot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin*, *30*(6), 721-731.

Cullen, M. J., Hardison, C. M., & Sackett, P. R. (2004). Using SAT-grade and ability- job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology, 89*(2), 220-230.

Cullen, M. J., Waters, S. D., & Sackett, P. R. (2006). Testing stereotype threat theory predictions for math-identified and non-math-identified students by gender. *Human Performance, 19*(4), 421-440.

Dao-Castellana, M. H., Samson, Y., Legault, F., Martinot, J. L., Aubin, H. J., Crouzel, C., Feldman, L., Barrucand, D., Rancurel, G., Feline, A., & Syrota, A. (1998). Frontal dysfunction in neurologically normal chronic alcoholic subjects: metabolic and neuropsychological findings. *Psychological Medicine, 28*(5), 1039-1048.

Davidson, D. J., Zacks, R. T., & Williams, C. C. (2003). Stroop interference, practice, and aging. *Aging Neuropsychology and Cognition, 10*(2), 85-98.

De Beni, R., Palladino, P., Pazzaglia, F., & Cornoldi, C. (1998). Increases in intrusion errors and working memory deficit of poor comprehenders. *Quarterly Journal of Experimental Psychology Section a- Human Experimental Psychology, 51*(2), 305-320.

Dempster, F. N. (1993). Exposing our students to less should help them learn more. *Phi Delta Kappan, 74*(6), 433-437.

Dempster, F. N., & Corkill, A. J. (1999). Individual differences in susceptibility to interference and general cognitive ability. *Acta Psychologica, 101*(2-3), 395-416.

Department of Education. (2010). *Maths and science education: the supply of high achievers at A level.* Retrieved from http://www.education.gov.uk/publications/eOrderingDownload/DFE-RR079.pdf

DeStefano, D., & LeFevre, J. A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16*(3), 353-386.

Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in stroop interference with practice for young and old adults. *Journal of Experimental Psychology-Learning Memory and Cognition, 20*(2), 470-484.

Eccles, J. S., & Jacobs, J. E. (1986). Social forces shape math attitudes and performance. *Signs, 11*(2), 367-380.

Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender-role stereotypes, expectancy effects, and parents socialization of gender differences. *Journal of Social Issues, 46*(2), 183-201.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Forbes, C. E., Schmader, T., & Allen, J. J. B. (2008). The role of devaluing and discounting in performance monitoring: a neurophysiological study of minorities under threat. *Social Cognitive and Affective Neuroscience, 3*(3), 253-261.

Forbes, C. E., & Schmader, T. (2010). Retraining attitudes and stereotypes to affect motivation and cognitive capacity under stereotype threat. *Journal of Personality and Social Psychology, 99*(5), 740-754.

Ford, T. E., Ferguson, M. A., Brooks, J. L., & Hagadone, K. M. (2004). Coping sense of humor reduces effects of stereotype threat on women's math performance. *Personality and Social Psychology Bulletin, 30*(5), 643-653.

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language, 51*(1), 136-158.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology-General, 137*(2), 201-225.

Gallagher, A. M. (1998). Gender and antecedents of performance in mathematics testing. *Teachers College Record, 100*, 297-314.

Gallagher, A. M., & De Lisi, R. (1994). Gender differences in scholastic aptitude test- mathematics problem solving among high-ability students. *Journal of Educational Psychology, 86*, 204-211.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology, 75*(3), 165-190.

Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effects on girls' mathematics performance. *Developmental Psychology.*

GCSE Higher-Tier Mathematics Examinations. (n.d). Retrieved October 18, 2010, from http://www.aqa.org.uk/subjects/mathematics

GCSE Results. (2012). Retrieved from http://docs.google.com/spreadsheet/pub?key=0AoEZjwuqFS2PdEZfSVpFd0UwdExROXlQbHR4d2laUHc&gid=0

Gimmig, D., Huguet, P., Caverni, J. P., & Cury, F. (2006). Choking under pressure and wording memory capacity: When performance pressure reduces fluid intelligence. *Psychonomic Bulletin & Review, 13*(6), 1005-1010.

Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology, 29*(1), 17-28.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24*(6), 645-662.

Grandjean, J., & Collette, F. (2011). Influence of response prepotency strength, general working memory resources, and specific working memory load on the ability to inhibit predominant responses: A comparison of young and elderly participants. *Brain and Cognition, 77*(2), 237-247.

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles, 66*(3-4), 153-166.

Hancock, D. R., & Dawson, R. (2001). Effects of test anxiety and evaluative threat on students' acheivements and motivation. *Journal of Educational Research, 94*, 284-290.

Harkins, S. G. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology, 91*(3), 436-455.

Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In C. J. In A. Conway, M., A. Miyake, A., & Towse, J. (Eds.), *Variation in working memory* (pp. 227-249). New York: Oxford University Press.

Hasher, L., Quig, M. B., & May, C. P. (1997). Inhibitory control over no-longer-relevant information: Adult age differences. *Memory & Cognition, 25*(3), 286-295.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 193-225). New York: Academic Press.

Huguet, P., & Regner, I. (2007). Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology, 99*(3), 545-560.

Hull, C. L. (1943). *Principles of behavior:* New York: Appleton- Century-Crofts.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance - a meta-analysis. *Psychological Bulletin, 107*(2), 139-155.

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*(5), 365-371.

Jacobs, J. E., & Eccles, J. S. (1992). The impact of mothers gender-role stereotypic beliefs on mothers and childrens ability perceptions. *Journal of Personality and Social Psychology, 63*(6), 932-944.

Jamieson, J. P. (2009). *The role of motivation in blatant stereotype threat, subtle stereotype threat, and stereotype priming.* (Doctoral Dissertation, Northeastern University). Retrieved from http://iris.lib.neu.edu/cgi/viewcontent.cgi?article=1009&context=psych_diss&sei-redir=1&referer=http%3A%2F%2F

Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology, 93*(4), 544-564.

Jamieson, J. P., & Harkins, S. G. (2009). The effect of stereotype threat on the solving of quantitative GRE problems: A mere effort interpretation. *Personality and Social Psychology Bulletin, 35*(10), 1301-1314.

Jamieson, J. P., & Harkins, S. G. (2011). The intervening task method: implications for measuring mediation. *Personality and Social Psychology Bulletin, 37*(5), 652-661.

Jamieson, J. P., & Harkins, S. G. (2012). Distinguishing between the effects of stereotype priming and stereotype threat on math performance. *Group Processes & Intergroup Relations, 15*(3), 291-304.

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle - Teaching stereotype threat as a means of improving women's math performance. *Psychological Science, 16*(3), 175-179.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology-General, 130*(2), 169-183.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology-General, 132*(1), 47-70.

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science, 12*, 978-990.

Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles, 4*7(3-4), 193-198.

Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *British Journal of Educational Psychology, 77*, 323-338.

Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin, 29*(3), 371-381.

Kessels, U., & Hannover, B. (2008). When being a girl matters less: Accessibility of gender-related self-knowledge in single-sex and coeducational classes and its impact on students' physics-related self-concept of ability. *British Journal of Educational Psychology, 78*, 273-289.

Kimball, M. M. (1989). A new perspective on womens math achievement. *Psychological Bulletin, 105*(2), 198-214.

Leicestershire Primary Team. (2005). *Girls' achievement in mathematics.* Retrieved from http://www.leics.gov.uk/girls_and_maths.doc

Levelt Committee. (2012). *Flawed science: The fraudulent research practices of social psychologist Diederik Stapel.* Retrieved from https://www.commissielevelt.nl/wp-content/uploads_per_blog/commissielevelt/2013/01/finalreportLevelt1.pdf

Levitt, S. D., & List, J. A. (2007). Viewpoint: On the generalizability of lab behaviour to the field. *Canadian Journal of Economics-Revue Canadienne D Economique, 40*(2), 347-370.

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence, 32*(5), 481-498.

Maccoby, E. E., & Jacklin, C. N. (1974). Myth, reality and shades of gray - What we know and don't know about sex differences. *Psychology Today, 8*(7), 109-112.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect - An intergrative review. *Psychological Bulletin, 109*(2), 163-203.

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition, 26*, 201-211.

Major, B., Spencer, S., Schmader, T., Wolfe, C., & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. *Personality and Social Psychology Bulletin, 24*(1), 34-50.

Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42*(2), 236-243.

Mazerolle, M., Regner, I., Morisset, P., Rigalleau, F., & Huguet, P. (2012). Stereotype threat strengthens automatic recall and undermines controlled processes in older adults. *Psychological Science, 23*(7), 723-727.

McFall, S. R., Jamieson, J. P., & Harkins, S. G. (2009). Testing the mere effort account of the evaluation-performance relationship. *Journal of Personality and Social Psychology, 96*(1), 135-154.

Meyer, D. K., Turner, J. C., & Spencer, C. A. (1997). Challenge in a mathematics classroom: Students' motivation and strategies in project-based learning. *Elementary School Journal, 97*(5), 501-521.

Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science, 7*(2), 109-117.

Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science, 330*(6008), 1234-1237.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.

Montague, M. (2006). Self-regulation strategies for better math performance in middle school. In M. Montague & A. Jitendra (Eds.), *Mathematics* *education for students with earning disabilites* (pp. 89-107). New York: Guilford.

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect the objective and subjective experiences of women in math, science, and engineering settings. *Psychological Science 18*(10), 879-885.

Mutter, S. A., Naylor, J. C., & Patterson, E. R. (2005). The effects of age and task context on Stroop task performance. *Memory & Cognition, 33*(3), 514-530.

Neuville, E., & Croizet, J. C. (2007). Can salience of gender identity impair math performance among 7-8 years old girls ? The moderating role of task difficulty. *European Journal of Psychology of Education, 22*(3), 307-316.

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin, 126*(2), 220-246.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., … Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences of the United States of America, 106*(26), 10593-10597.

Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93*(6), 1314-1334.

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin, 29*(6), 782-789.

Passolunghi, M. C., Cornoldi, C., & De Liberto, S. (1999). Working memory and intrusions of irrelevant information in a group of specific poor problem solvers. *Memory & Cognition, 27*(5), 779-790.

Pearson, J. C., & West, R. (1991). An initial investigation of the effects of gender on student questions in the classroom - Developing a descriptive base. *Communication Education, 40*(1), 22-32.

Picho, K., Rodriguez, A., & Finnie, L. (2013). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: A meta-analysis. *Journal of Social Psychology, 153*(3), 299-333.

Picho, K., & Stephens, J. M. (2012). Culture, context and stereotype threat: A comparative analysis of young ugandan women in coed and single-sex schools. *Journal of Educational Research, 105*(1), 52-63.

Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues, 57*(1), 55-71.

Ramirez, G., & Beilock, S. L. (2011). Writing about testing worries boosts exam performance in the classroom. *Science, 331*(6014), 211-213.

Redick, T. S., Calvo, A., Gay, C. E., & Engle, R. W. (2011). Working memory capacity and go/no-go task performance: selective effects of updating, maintenance, and inhibition. *Journal of Experimental Psychology-Learning Memory and Cognition, 37*(2), 308-324

Regner, I., Smeding, A., Gimmig, D., Thinus-Blanc, C., Monteil, J. M., & Huguet, P. (2010). Individual differences in working memory moderate stereotype-threat effects. *Psychological Science, 21*(11), 1646-1648.

Roberts, R. J., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General, 123*(4), 374-393.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology-General, 126*(3), 211-227.

Rosenthal, H. E. S., & Crisp, R. J. (2006). Reducing stereotype threat by blurring intergroup boundaries. *Personality and Social Psychology Bulletin, 32*(4), 501-511.

Rosenthal, H., Quinn, K., & Seddon, E. (2009). *The key processes of stereotype threat: Stereotype activation and self-categorization.* Paper presented at the Social Psychology Section Annual Conference, University of Sheffield, UK.

Royer, J. M., & Garofoli, L. M. (2005). Cognitive contribution to sex differences. In A. M. Gallagher & J. C. Kaufman (Eds.), G*ender differences in mathematics: An intergrative psychological approach* (pp. 99-120). Cambridge: Cambridge University Press.

Royer, J. M., Tronsky, L. N., Chan, Y., Jackson, S. J., & Marchant, H. (1999). Math-fact retrieval as the cognitive mechanism underlying gender differences in math test performance. *Contemporary Educational Psychology, 24*(3), 181-266.

Rushton, J. P. (1992). Differences in brain size. *Nature, 358*(6387), 532-532.

Rydell, R. J., McConnell, A. R., & Beilock, S. L. (2009). Multiple social identities and stereotype threat: imbalance, accessibility, and working memory. *Journal of Personality and Social Psychology, 96*(5), 949-966.

Rydell, R. J., Rydell, T. R., & Boucher, K. L. (2010). The effect of negative performance stereotypes on learning. *Journal of Personality and Social Psychology, 99*(6), 883-896.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education - Prospects in a post-affirmative-action world. *American Psychologist, 56*(4), 302-318.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*(2), 194-201.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*(3), 440-452.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*(2), 336-356.

Schwartz, W., & Hanson, K. (1992). *Equal mathematics education for female students.* Retrieved from http://www.ericdigests.org/1992-4/equal.htm

Sedikides, C. (2012). Self-protection. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (2nd Ed.), (pp. 327-353). New York, NY: Guilford Press.

Seibt, B., & Forster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology, 87*(1), 38-56.

Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology, 39*(1), 68-74.

Sellgren, K. (2013). GCSE changes to final exams 'will disadvantage girls’. *BBC News.* Retrieved from http://www.bbc.co.uk/news/education-21955004

Shapiro, J. R., & Neuberg, S. L. (2008). When do the stigmatized stigmatize? The ironic effects of being accountable to (perceived) majority group prejudice-expression norms. *Journal of Personality and Social Psychology, 95*(4), 877-898.

Shen, W. N., Kiger, T. B., Davies, S. E., Rasch, R. L., Simon, K. M., & Ones, D. S. (2011). Samples in applied psychology: Over a decade of research in review. *Journal of Applied Psychology, 96*(5), 1055-1064.

Shepherd, J. (2011, August 18). A-level results: boys catching up in maths and sciences. *The Guardian.* Retrieved from http://www.theguardian.com/education/2011/aug/18/a-levels-boys-close-maths-sciences-gap

Shepherd, J. (2011, August 25). GCSE results 2011: girls widen their lead. *The Guardian.* Retrieved from http://www.theguardian.com/education/2011/aug/25/gcse-results-girls-beat-boys

Shih, M., Ambady, N., Richeson, J. A., & Fujita, K. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology, 83*(3), 638-647.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*(1), 80-83.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4-28.

Stapel, D. A., & Lindenberg, S. (2011). Coping with chaos: How disordered contexts promote stereotyping and discrimination. *Science, 332*(6026), 251-253.

Steele, C. M. (1997). A threat in the air - How stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6), 613-629.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test-performance of African-Americans. *Journal of Personality and Social Psychology, 69*(5), 797-811.

Steele, C. M., & Davies, P. G. (2003). Stereotype threat and employment testing: A commentary. *Human Performance, 16*(3), 311-326.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology, 34,* 379-440.

Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement? *Review of General Psychology, 16*(1), 93-102.

Stone, J. (2002). Battling doubt by avoiding practice: The effects of stereotype threat on self-handicapping in white athletes. *Personality and Social Psychology Bulletin, 28*(12), 1667-1678.

Stone, J., Perry, Z. W., & Darley, J. M. (1997). "White men can't jump": Evidence for the perceptual confirmation of racial stereotypes following a basketball game. *Basic and Applied Social Psychology, 19*(3), 291-306.

Stricker, L. J., & Ward, W. C. (2004). Stereotype threat, inquiring about test takers' ethnicity and gender, and standardized test performance. *Journal of Applied Social Psychology, 34*(4), 665-693.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Tapia, M., & Marsh II, G. E. (2004). An instrument to measure mathematics attitudes. *Academic Exchange Quarterly, 8*, 16-21.

Tartre, L. A. (1990). Spatial orientation skill and mathematical problem-solving. *Journal for Research in Mathematics Education, 21*(3), 216-229.

von Hippel, W., Silver, L. A., & Lynch, M. E. (2000). Stereotyping against your will: The role of inhibitory ability in stereotyping and prejudice among the elderly. *Personality and Social Psychology Bulletin*, *26*, 523-532.

Walsh, M., Hickey, C., & Duffy, J. (1999). Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles, 41*(3-4), 219-240.

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*(5), 456-467.

Wei, T. E. (2012). Sticks, stones, words, and broken bones: New field and lab evidence on stereotype threat. *Educational Evaluation and Policy Analysis, 34*(4), 465-488.

Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin*, *127*(6), 797-826.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*, 475-482.

Wicherts, J. M., Dolan, C. V., & Hessen, D. J. (2005). Stereotype threat and group differences in test performance: A question of measurement invariance. *Journal of Personality and Social Psychology, 89*(5), 696-716.

Wilson, J. W., Fernandez, M. L., & Hadaway, N. (1993). *Mathematical problem solving.* Retrieved from http://jwilson.coe.uga.edu/emt725/PSsyn/PSsyn.html

Wout, D., Danso, H., Jackson, J., & Spencer, S. (2008). The many faces of stereotype threat: Group- and self-threat. *Journal of Experimental Social Psychology, 44*(3), 792-799.

Zacks, R. T., & Hasher, L. (1994). Directed ignoring: Inhibition regulation of working memory. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory and language* (pp. 241-264). New York: Academic Press.

Zajonc, R. B. (1965). The requirements and design of a standard group task. *Journal of Experimental Social Psychology, 1*(1), 71-88.

Zhu, Z. (2007). Gender differences in mathematical problem solving patterns: A review of the literature. *International Education Journal, 8*(2), 187-203.

Appendix A

Question type explained

An example and explanation of solve and comparison maths question types.

Solve Type:

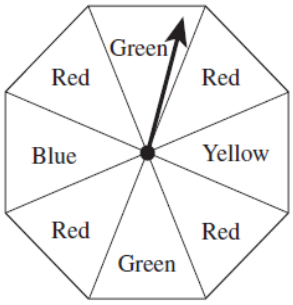
Work out $3\frac{3}{4} \div 1\frac{2}{3}$

For this problem, the test-taker must apply a formula to covert the fraction into 15 over 4 and 5 over 3. The second fraction must then be inverted to 3 over 5. The top and bottom numbers of the fraction must then be multiplied: 15 times 3 and 4 times 5 respectively. This results in the correct answer of 45 over 20 (or simplified equivalent). Thus, solve problems involve the application and computation of equations.

Comparison Type:

(a) Oscar has a spinner with eight sections.
Four of the sections are Red, two are Green, one is Blue and one is Yellow.
He spins the spinner 200 times.
His results are shown in the table.

Colour	Red	Green	Blue	Yellow
Frequency	105	48	22	25



Do the results suggest that the spinner is fair?
Explain your answer.

This problem can be solved using logic and estimation. First, the test-taker must estimate the theoretical probabilities and compare these to the relative experimental probabilities. The test-taker must then logically deduce that these are approximately correct and that the spinner is fair. Thus, comparison problems require a combination of logic and estimation, rather than the application of a learnt formula.

Appendix B

Pilot study 1a:

Maths question types

Maths test

Table B.1.

Question type (solve vs. comparison) for each maths question in Pilot study 1a.

Question Number	Question type
1	Solve
2	Solve
3	Comparison
4	Solve
5	Solve
6	Comparison
7	Comparison
8	Solve
9	Solve
10	Comparison
11	Comparison
12	Solve
13	Solve
14	Comparison
15	Comparison
16	Solve
17	Comparison
18	Comparison

Centre Number						Candidate Number				

For Examiner's Use

MATHEMATICS (SPECIFICATION A)
Higher Tier
Paper 1 Non-calculator

UNIVERSITY OF LEEDS

For this paper you must have:
- mathematical instruments.

You must **not** use a calculator.

Time allowed: 1 hour 10 minutes

Instructions
- Use black ink or black ball-point pen. Draw diagrams in pencil.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- You must answer the questions in the spaces provided. Answers written in margins or on blank pages will not be marked.
- Do all rough work in this book.

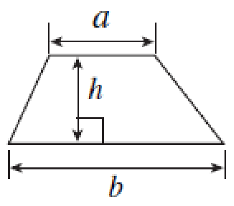
Information
- The maximum mark for this paper is **60**
- The marks for questions are shown in brackets.
- You may ask for more answer paper, graph paper and tracing paper. This must be tagged securely to this answer booklet.

Advice
- In all calculations, show clearly how you work out your answer.

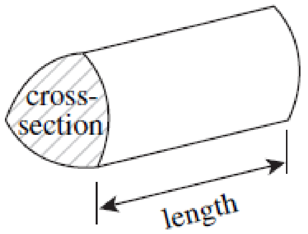
For Examiner's Use	
Pages	Mark
3	
4–5	
6–7	
8–9	
10–11	
12–13	
14–15	
16–17	
18–19	
20–21	
TOTAL	
Examiner's Initials	

Formulae Sheet: Higher Tier

Area of trapezium = $\frac{1}{2}(a+b)h$

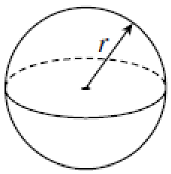


Volume of prism = area of cross-section × length



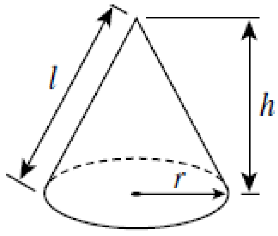
Volume of sphere = $\frac{4}{3}\pi r^3$

Surface area of sphere = $4\pi r^2$



Volume of cone = $\frac{1}{3}\pi r^2 h$

Curved surface area of cone = $\pi r l$

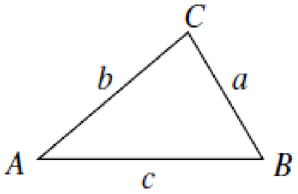


In any triangle ABC

Area of triangle = $\frac{1}{2}ab \sin C$

Sine rule $\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$

Cosine rule $a^2 = b^2 + c^2 - 2bc \cos A$



The Quadratic Equation

The solutions of $ax^2 + bx + c = 0$, where $a \neq 0$, are given by

$$x = \frac{-b \pm \sqrt{(b^2 - 4ac)}}{2a}$$

Areas outside the box will not be scanned for marking

Answer **all** questions in the spaces provided.

1 Work out the value of $\frac{a(3b + 1)}{5}$ when $a = -2$ and $b = 3$

....

....

....

Answer .... *(3 marks)*

2 In a school, there are 200 students in Year 11.
110 of these students are girls.

What percentage of these students are boys?

....

....

....

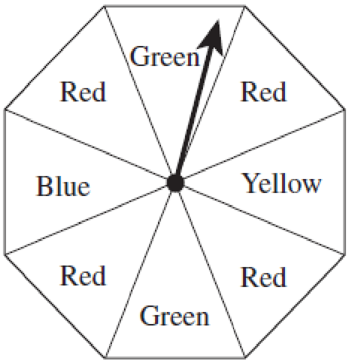
....

Answer .... % *(3 marks)*

Areas outside the box will not be scanned for marking

3 (a) Oscar has a spinner with eight sections.
Four of the sections are Red, two are Green, one is Blue and one is Yellow.
He spins the spinner 200 times.
His results are shown in the table.

Colour	Red	Green	Blue	Yellow
Frequency	105	48	22	25



3 (a) (i) Explain why the relative frequency of Green is 0.24

....

....

(1 mark)

3 (a) (ii) Do the results suggest that the spinner is fair?
Explain your answer.

....

....

....

(2 marks)

4 Work out $1\frac{7}{8} \times 2\frac{2}{5}$

Give your answer in its simplest form.

....

....

....

....

Answer .... *(3 marks)*

5 (a) Show **clearly** that $(n + 1)^2 + (n - 1)^2 \equiv 2n^2 + 2$

....

....

....

....

(3 marks)

Areas outside the box will not be scanned for marking

6 x, y and z represent lengths.

For each expression, put a tick in a box to show whether it represents a length, an area, a volume or none of these.

Expression	Length	Area	Volume	None of these
$3x + y + 2z$				
$x + z^2$				
$x^3 + 5y^2z$				

(3 marks)

7 (a) Solve the inequality $3x + 2 \leqslant 8$

....

....

....

Answer .... *(2 marks)*

7 (b) Write down all the integer values of x satisfying this inequality $-4 \leqslant 2x < 4$

....

....

Answer .... *(2 marks)*

8 $A = 6$ and $B = -7$

Work out the value of $\dfrac{A(B+2)}{3}$

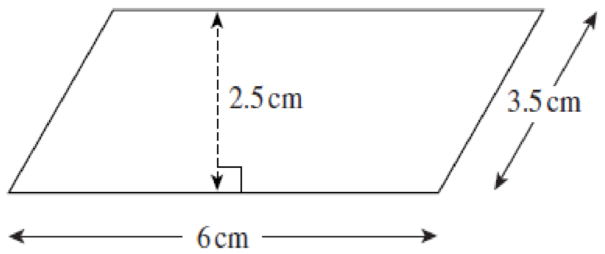
....

....

....

Answer .... *(3 marks)*

9 The diagram shows a parallelogram.



Not drawn accurately

Calculate the area of the parallelogram.
State the units of your answer.

....

....

....

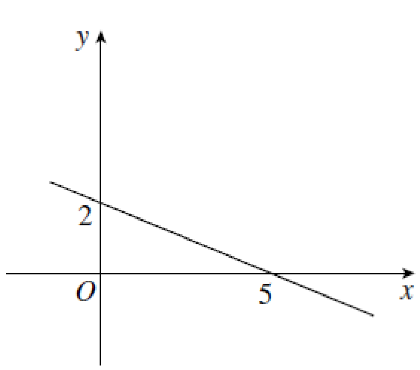
Answer .... *(3 marks)*

10 Each of these equations represents the graph of a straight line.

A: $5y + 10 = 2x$	B: $5x + 2y = 10$	C: $2y + 10 = 5x$	D: $2x + 5y = 10$

The four graphs are shown in the diagrams below.

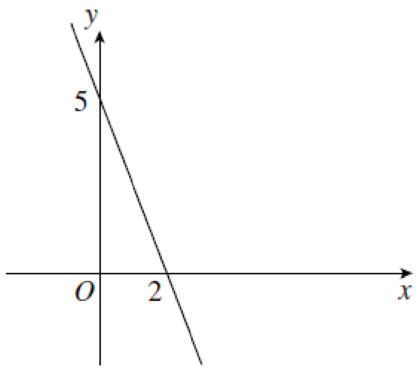
Which equation represents which graph?



....

....

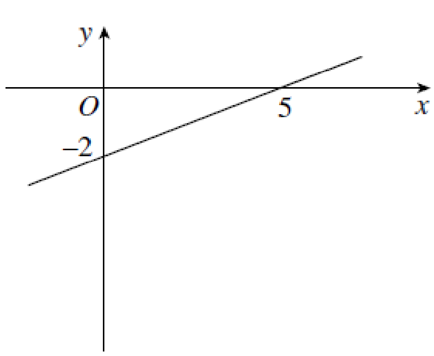
This is equation ....



....

....

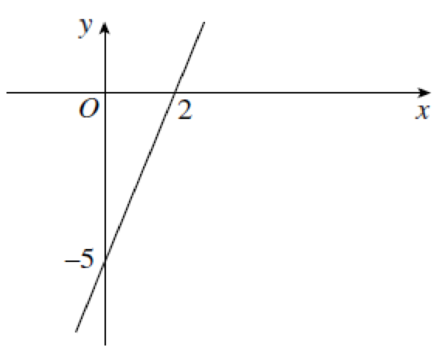
This is equation ....



....

....

This is equation ....



....

....

This is equation ....

(3 marks)

Areas outside the box will not be scanned for marking

11 Grace buys a packet of ten hyacinth bulbs.
They all look the same.

Seven of the bulbs will produce Pink flowers, three will produce Blue flowers.
A bulb is taken at random and planted.
A second bulb is taken at random and planted.

Calculate the probability that the two bulbs will produce **at least one** Blue flower.

....

....

....

....

....

....

Answer .... *(3 marks)*

Areas outside the box will not be scanned for marking

12 Leah, Chloe and Maya share £400 between them.
Leah receives the smallest amount of £90
The ratio of Leah's share to Chloe's share is 2 : 3

Work out how much Maya receives.

....

....

....

....

....

....

Answer £ .... *(3 marks)*

13 Work out $3\frac{3}{4} \div 1\frac{2}{3}$

....

....

....

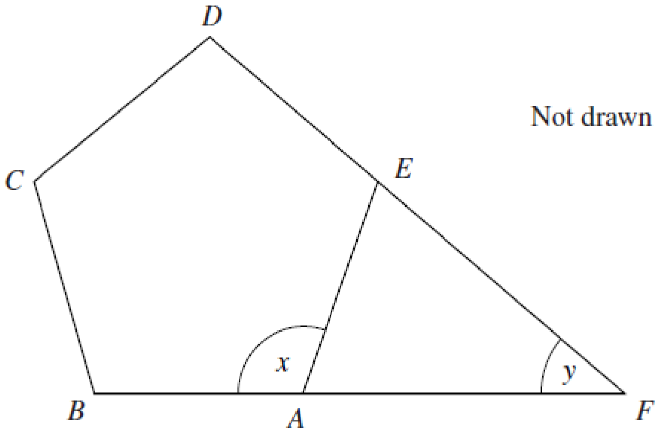
....

....

....

Answer .... *(3 marks)*

14 *ABCDE* is a regular pentagon.
DEF and *BAF* are straight lines.



14 (a) Which **one** of these statements is true?

1 The exterior angle of a regular pentagon is equal to $360° \div 5 = 72°$

2 The interior angle of a regular pentagon is equal to $360° \div 5 = 72°$

3 The exterior angle of a regular pentagon is equal to $360° - 72° = 288°$

4 The interior angle of a regular pentagon is equal to $360° - 72° = 288°$

Answer .... *(1 mark)*

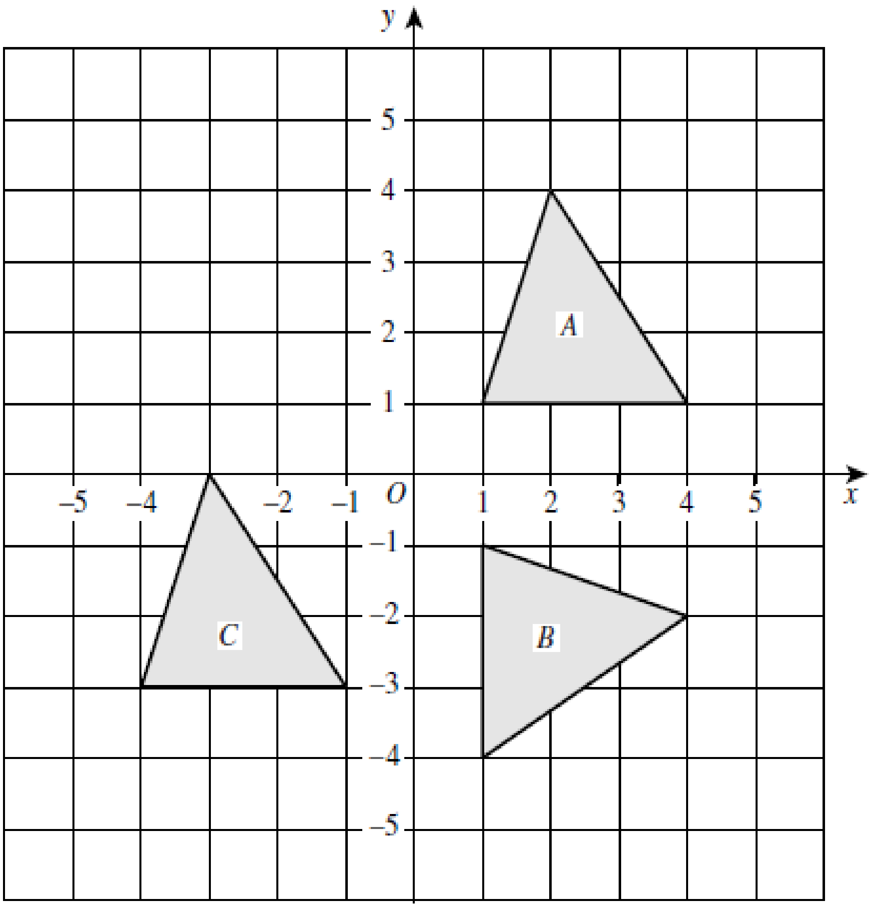
14 (ii) Work out the size of the angle marked *y* on the diagram.

....

....

Answer .... degrees *(2 marks)*

15 Triangles *A*, *B* and *C* are shown on the grid.



Describe fully the **single** transformation that maps triangle *A* onto triangle *B*.

....

....

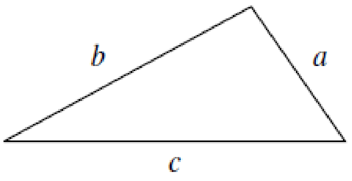
....

(3 marks)

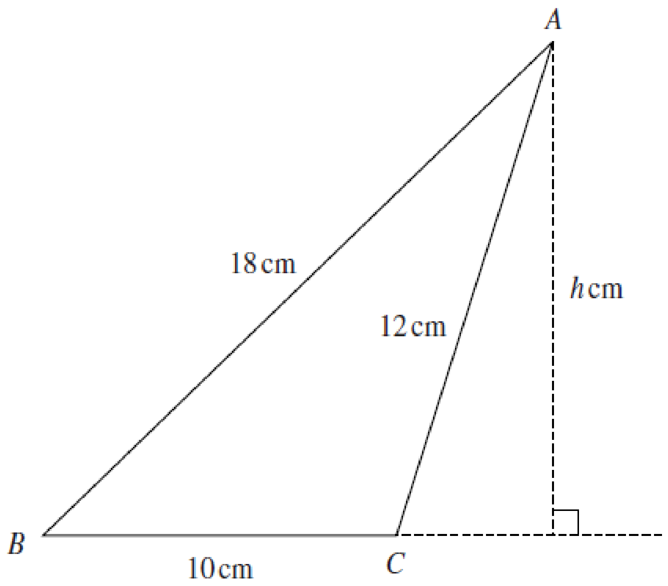
16 Hero's formula for the area of a triangle with sides of length a, b and c is

Area = $\sqrt{s(s-a)(s-b)(s-c)}$

where $s = \frac{1}{2}(a + b + c)$



The diagram shows triangle ABC in which AB = 18 cm, AC = 12 cm and BC = 10 cm.
The perpendicular distance from A to BC is h cm.



Not drawn accurately

Calculate the value of h.
Give your answer in the form $p\sqrt{2}$, where p is an integer.
You **must** show your working.

....

....

....

....

....

....

....

....

Answer .... cm *(6 marks)*

17 A Golf Club has 600 members.
A stratified sample of members is taken, by age group.

The table shows the age grouping of the members.
Some information is given in the table.

Age group	Junior	18 – 39	40 – 59	Senior
Number of members	100			120
Number in sample	20		35	

Complete the table.

....

....

....

....

....

....

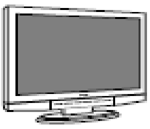
....

....

(3 marks)

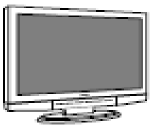
18 George wants to buy a new television.
He sees the same television on special offer at two different stores.

Teleworld

40% off 

Normal price £480

SuperSave

$\frac{1}{3}$ off 

Normal price £420

Which store sells the television more cheaply?
You **must** show your working.

....

....

....

....

....

Answer .... *(5 marks)*

END OF QUESTIONS

Appendix C

Maths tests

(solve vs. comparison)

(high diagnosticity vs. low diagnosticity)

for Studies 1, 4, & 5

Centre Number						Candidate Number				

For Examiner's Use

MATHEMATICS (SPECIFICATION A)
Higher Tier
Paper 1 Non-calculator

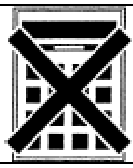
UNIVERSITY OF LEEDS

Note:

- **Previous research has shown gender differences on this test.**

For this paper you must have:
- mathematical instruments.

You must **not** use a calculator.



Time allowed: 18 minutes

Instructions

- Use black ink or black ball-point pen. Draw diagrams in pencil.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- You must answer the questions in the spaces provided. Answers written in margins or on blank pages will not be marked.
- Do all rough work in this book.

Information

- The maximum mark for this paper is **15**
- The marks for questions are shown in brackets.
- You may ask for more answer paper, graph paper and tracing paper. This must be tagged securely to this answer booklet.

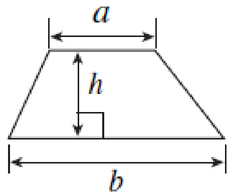
Advice

- In all calculations, show clearly how you work out your answer.

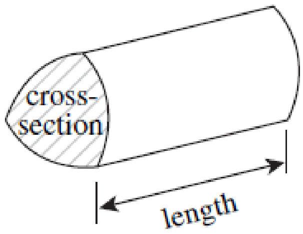
For Examiner's Use	
Pages	Mark
3	
4–5	
6–7	
8–9	
10–11	
12–13	
14–15	
16–17	
18–19	
20–21	
TOTAL	
Examiner's Initials	

Formulae Sheet: Higher Tier

Area of trapezium = $\frac{1}{2}(a+b)h$

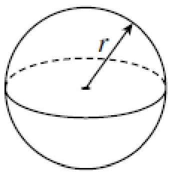


Volume of prism = area of cross-section × length



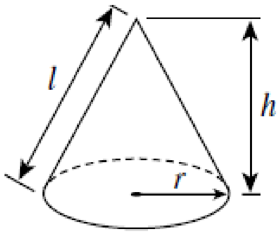
Volume of sphere = $\frac{4}{3}\pi r^3$

Surface area of sphere = $4\pi r^2$



Volume of cone = $\frac{1}{3}\pi r^2 h$

Curved surface area of cone = $\pi r l$

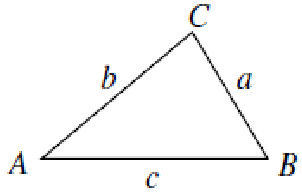


In any triangle ABC

Area of triangle = $\frac{1}{2}ab\sin C$

Sine rule $\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$

Cosine rule $a^2 = b^2 + c^2 - 2bc\cos A$



The Quadratic Equation

The solutions of $ax^2 + bx + c = 0$, where $a \neq 0$, are given by

$$x = \frac{-b \pm \sqrt{(b^2 - 4ac)}}{2a}$$

Answer **all** questions in the spaces provided.

1 Work out $1\frac{7}{8} \times 2\frac{2}{5}$

Give your answer in its simplest form.

....

....

....

....

Answer .... *(3 marks)*

2 Show **clearly** that $(n + 1)^2 + (n - 1)^2 \equiv 2n^2 + 2$

....

....

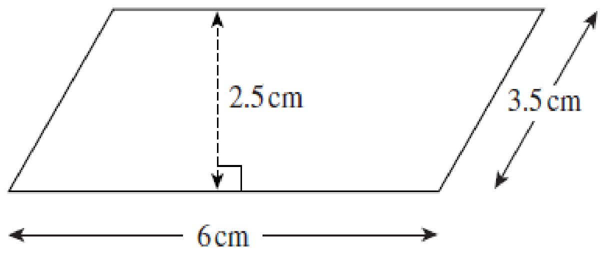
....

....

(3 marks)

Areas outside the box will not be scanned for marking

3 The diagram shows a parallelogram.



Not drawn accurately

Calculate the area of the parallelogram.
State the units of your answer.

....

....

....

Answer .... *(3 marks)*

Areas outside the box will not be scanned for marking

4 Leah, Chloe and Maya share £400 between them.
Leah receives the smallest amount of £90
The ratio of Leah's share to Chloe's share is 2 : 3

Work out how much Maya receives.

....

....

....

....

....

....

Answer £ .... *(3 marks)*

5 Work out $3\frac{3}{4} \div 1\frac{2}{3}$

....

....

....

....

....

....

Answer .... *(3 marks)*

END OF QUESTIONS

Centre Number						Candidate Number				

For Examiner's Use

MATHEMATICS (SPECIFICATION A)
Higher Tier
Paper 1 Non-calculator

UNIVERSITY OF LEEDS

Note:

- **Previous research has shown no gender differences on this test.**

For this paper you must have:

- mathematical instruments.

You must **not** use a calculator.

Time allowed: 18 minutes

Instructions

- Use black ink or black ball-point pen. Draw diagrams in pencil.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- You must answer the questions in the spaces provided. Answers written in margins or on blank pages will not be marked.
- Do all rough work in this book.

Information

- The maximum mark for this paper is **15**
- The marks for questions are shown in brackets.
- You may ask for more answer paper, graph paper and tracing paper. This must be tagged securely to this answer booklet.

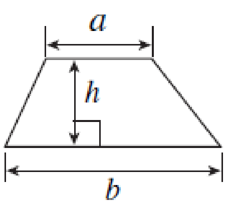
Advice

- In all calculations, show clearly how you work out your answer.

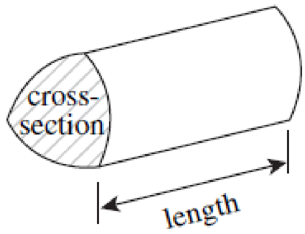
For Examiner's Use	
Pages	Mark
3	
4–5	
6–7	
8–9	
10–11	
12–13	
14–15	
16–17	
18–19	
20–21	
TOTAL	
Examiner's Initials	

Formulae Sheet: Higher Tier

Area of trapezium = $\frac{1}{2}(a+b)h$

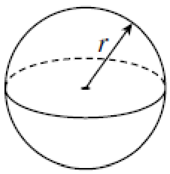


Volume of prism = area of cross-section × length



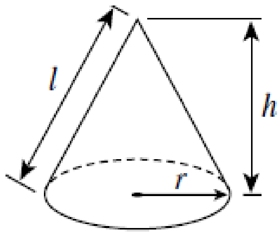
Volume of sphere = $\frac{4}{3}\pi r^3$

Surface area of sphere = $4\pi r^2$



Volume of cone = $\frac{1}{3}\pi r^2 h$

Curved surface area of cone = $\pi r l$

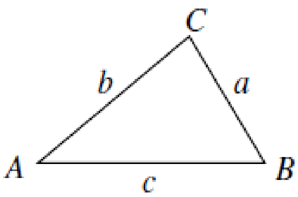


In any triangle ABC

Area of triangle = $\frac{1}{2}ab\sin C$

Sine rule $\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$

Cosine rule $a^2 = b^2 + c^2 - 2bc\cos A$



The Quadratic Equation

The solutions of $ax^2 + bx + c = 0$, where $a \neq 0$, are given by

$$x = \frac{-b \pm \sqrt{(b^2 - 4ac)}}{2a}$$

Answer **all** questions in the spaces provided.

1 Work out $1\frac{7}{8} \times 2\frac{2}{5}$

Give your answer in its simplest form.

....

....

....

....

Answer .... *(3 marks)*

2 Show **clearly** that $(n + 1)^2 + (n - 1)^2 \equiv 2n^2 + 2$

....

....

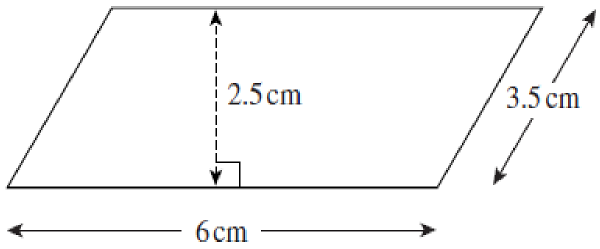
....

....

(3 marks)

Areas outside the box will not be scanned for marking

3 The diagram shows a parallelogram.



Not drawn accurately

Calculate the area of the parallelogram.
State the units of your answer.

....

....

....

Answer .... *(3 marks)*

Areas outside the box will not be scanned for marking

4 Leah, Chloe and Maya share £400 between them.
Leah receives the smallest amount of £90
The ratio of Leah's share to Chloe's share is 2 : 3

Work out how much Maya receives.

....

....

....

....

....

....

Answer £ .... *(3 marks)*

5 Work out $3\frac{3}{4} \div 1\frac{2}{3}$

....

....

....

....

....

....

Answer .... *(3 marks)*

END OF QUESTIONS

Centre Number						Candidate Number				

For Examiner's Use

MATHEMATICS (SPECIFICATION A)
Higher Tier
Paper 1 Non-calculator

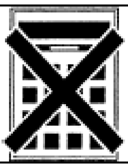
UNIVERSITY OF LEEDS

Note:

- **Previous research has shown gender differences on this test.**

For this paper you must have:
- mathematical instruments.

You must **not** use a calculator.



Time allowed: 18 minutes

Instructions
- Use black ink or black ball-point pen. Draw diagrams in pencil.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- You must answer the questions in the spaces provided. Answers written in margins or on blank pages will not be marked.
- Do all rough work in this book.

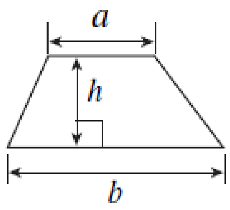
Information
- The maximum mark for this paper 15 .
- The marks for questions are shown in brackets.
- You may ask for more answer paper, graph paper and tracing paper. This must be tagged securely to this answer booklet.

Advice
- In all calculations, show clearly how you work out your answer.

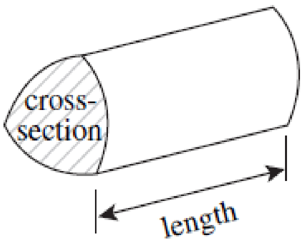
For Examiner's Use	
Pages	Mark
3	
4–5	
6–7	
8–9	
10–11	
12–13	
14–15	
16–17	
18–19	
20–21	
TOTAL	
Examiner's Initials	

Formulae Sheet: Higher Tier

Area of trapezium = $\frac{1}{2}(a+b)h$

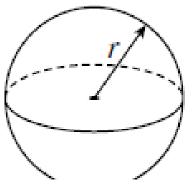


Volume of prism = area of cross-section × length



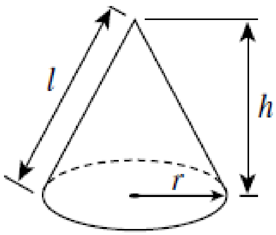
Volume of sphere = $\frac{4}{3}\pi r^3$

Surface area of sphere = $4\pi r^2$



Volume of cone = $\frac{1}{3}\pi r^2 h$

Curved surface area of cone = $\pi r l$

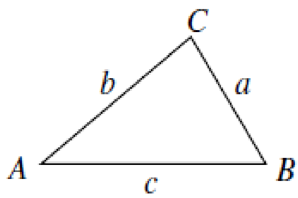


In any triangle ABC

Area of triangle = $\frac{1}{2}ab\sin C$

Sine rule $\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$

Cosine rule $a^2 = b^2 + c^2 - 2bc\cos A$



The Quadratic Equation

The solutions of $ax^2 + bx + c = 0$, where $a \neq 0$, are given by

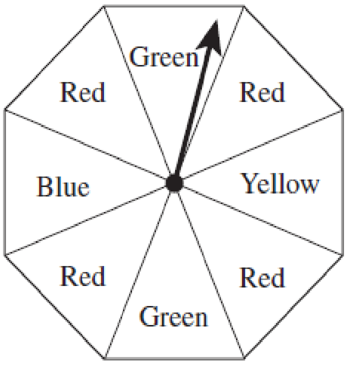
$$x = \frac{-b \pm \sqrt{(b^2 - 4ac)}}{2a}$$

Areas outside the box will not be scanned for marking

Answer **all** questions in the spaces provided.

1 (a) Oscar has a spinner with eight sections.
Four of the sections are Red, two are Green, one is Blue and one is Yellow.
He spins the spinner 200 times.
His results are shown in the table.

Colour	Red	Green	Blue	Yellow
Frequency	105	48	22	25



1 (a) (i) Explain why the relative frequency of Green is 0.24

....

....

(1 mark)

1 (a) (ii) Do the results suggest that the spinner is fair?
Explain your answer.

....

....

....

(2 marks)

2 x, y and z represent lengths.

For each expression, put a tick in a box to show whether it represents a length, an area, a volume or none of these.

Expression	**Length**	**Area**	**Volume**	**None of these**
$3x + y + 2z$				
$x + z^2$				
$x^3 + 5y^2z$				

(3 marks)

3 A Golf Club has 600 members.
A stratified sample of members is taken, by age group.

The table shows the age grouping of the members.
Some information is given in the table.

Age group	Junior	18 – 39	40 – 59	Senior
Number of members	100			120
Number in sample	20		35	

Complete the table.

....

....

....

....

....

....

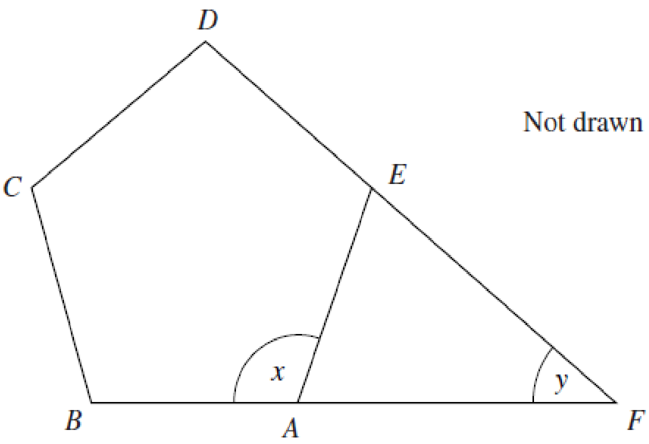
....

....

(3 marks)

Areas outside the box will not be scanned for marking

4 *ABCDE* is a regular pentagon.
DEF and *BAF* are straight lines.



Not drawn accurately

4 (a) Which **one** of these statements is true?

1 The exterior angle of a regular pentagon is equal to $360° \div 5 = 72°$

2 The interior angle of a regular pentagon is equal to $360° \div 5 = 72°$

3 The exterior angle of a regular pentagon is equal to $360° - 72° = 288°$

4 The interior angle of a regular pentagon is equal to $360° - 72° = 288°$

Answer .... *(1 mark)*

4 (ii) Work out the size of the angle marked y on the diagram.

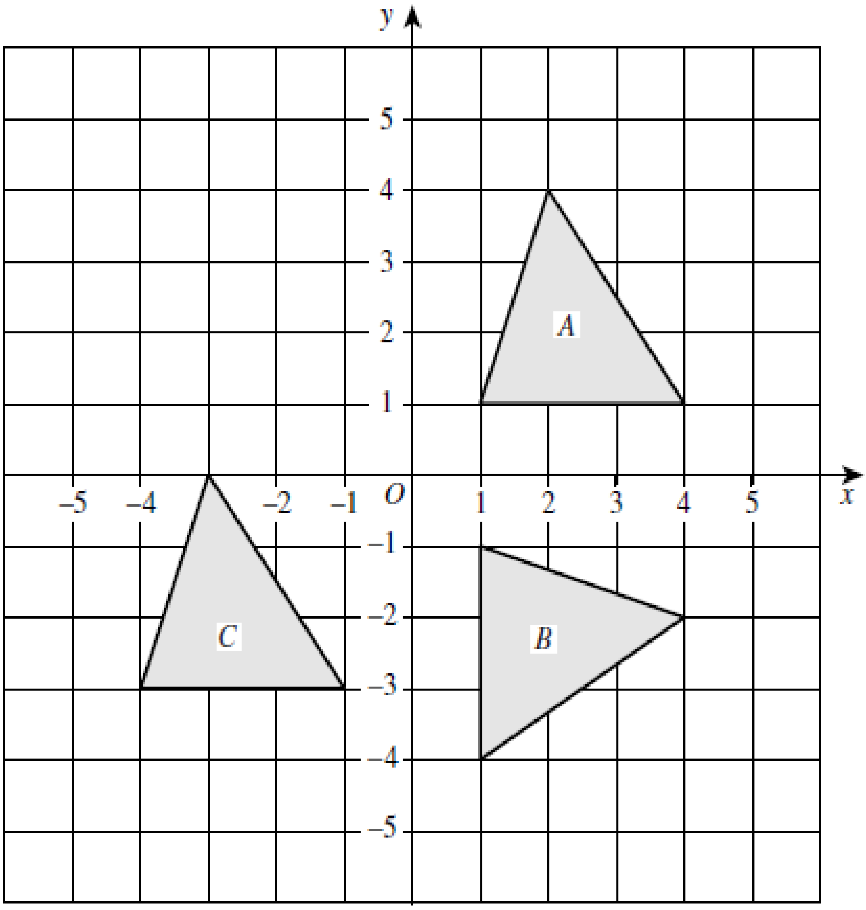
....

....

Answer .... degrees *(2 marks)*

Areas outside the box will not be scanned for marking

5 Triangles A, B and C are shown on the grid.



Describe fully the **single** transformation that maps triangle A onto triangle B.

....

....

....

(3 marks)

END OF QUESTIONS

Centre Number						Candidate Number				

For Examiner's Use

MATHEMATICS (SPECIFICATION A)
Higher Tier
Paper 1 Non-calculator

UNIVERSITY OF LEEDS

Note:
- **Previous research has shown no gender differences on this test.**

For this paper you must have:
- mathematical instruments.

You must **not** use a calculator.

Time allowed: 18 minutes

Instructions
- Use black ink or black ball-point pen. Draw diagrams in pencil.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- You must answer the questions in the spaces provided. Answers written in margins or on blank pages will not be marked.
- Do all rough work in this book.

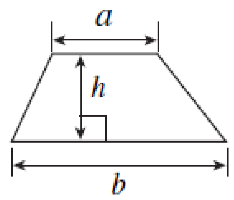
Information
- The maximum mark for this paper is **15**
- The marks for questions are shown in brackets.
- You may ask for more answer paper, graph paper and tracing paper. This must be tagged securely to this answer booklet.

Advice
- In all calculations, show clearly how you work out your answer.

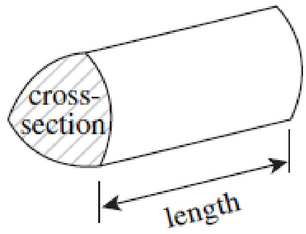
For Examiner's Use	
Pages	Mark
3	
4–5	
6–7	
8–9	
10–11	
12–13	
14–15	
16–17	
18–19	
20–21	
TOTAL	
Examiner's Initials	

Formulae Sheet: Higher Tier

Area of trapezium = $\frac{1}{2}(a+b)h$

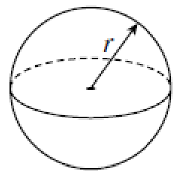


Volume of prism = area of cross-section × length



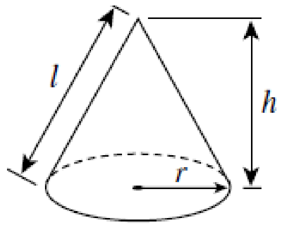
Volume of sphere = $\frac{4}{3}\pi r^3$

Surface area of sphere = $4\pi r^2$



Volume of cone = $\frac{1}{3}\pi r^2 h$

Curved surface area of cone = $\pi r l$

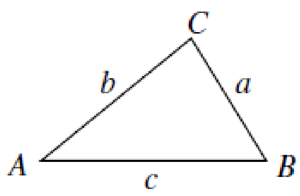


In any triangle ABC

Area of triangle = $\frac{1}{2}ab\sin C$

Sine rule $\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$

Cosine rule $a^2 = b^2 + c^2 - 2bc\cos A$



The Quadratic Equation

The solutions of $ax^2 + bx + c = 0$, where $a \neq 0$, are given by

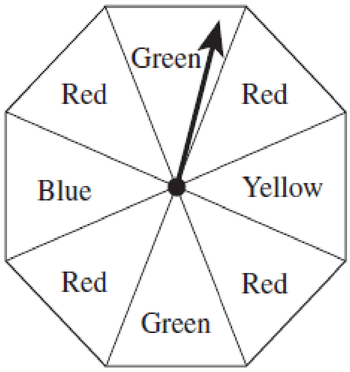
$$x = \frac{-b \pm \sqrt{(b^2 - 4ac)}}{2a}$$

Areas outside the box will not be scanned for marking

Answer **all** questions in the spaces provided.

1 (a) Oscar has a spinner with eight sections.
Four of the sections are Red, two are Green, one is Blue and one is Yellow.
He spins the spinner 200 times.
His results are shown in the table.

Colour	Red	Green	Blue	Yellow
Frequency	105	48	22	25



1 (a) (i) Explain why the relative frequency of Green is 0.24

....

....

(1 mark)

1 (a) (ii) Do the results suggest that the spinner is fair?
Explain your answer.

....

....

....

(2 marks)

2 x, y and z represent lengths.

For each expression, put a tick in a box to show whether it represents a length, an area, a volume or none of these.

Expression	**Length**	**Area**	**Volume**	**None of these**
$3x + y + 2z$				
$x + z^2$				
$x^3 + 5y^2z$				

(3 marks)

3 A Golf Club has 600 members.
A stratified sample of members is taken, by age group.

The table shows the age grouping of the members.
Some information is given in the table.

Age group	Junior	18 – 39	40 – 59	Senior
Number of members	100			120
Number in sample	20		35	

Complete the table.

....

....

....

....

....

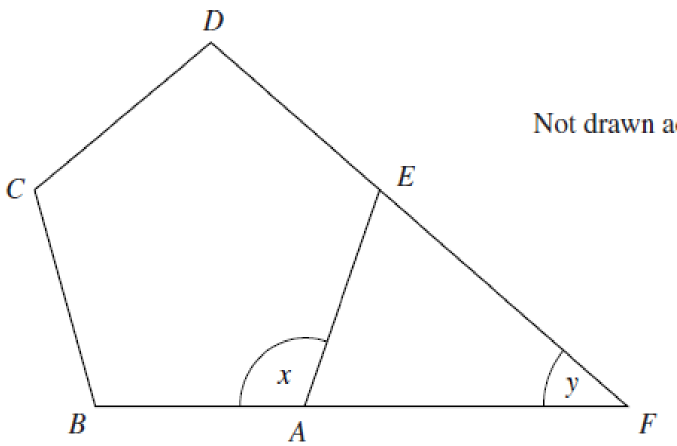
....

....

....

(3 marks)

4 *ABCDE* is a regular pentagon.
DEF and *BAF* are straight lines.



Not drawn accurately

4 (a) Which **one** of these statements is true?

1 The exterior angle of a regular pentagon is equal to $360° \div 5 = 72°$

2 The interior angle of a regular pentagon is equal to $360° \div 5 = 72°$

3 The exterior angle of a regular pentagon is equal to $360° - 72° = 288°$

4 The interior angle of a regular pentagon is equal to $360° - 72° = 288°$

Answer .... *(1 mark)*

4 (ii) Work out the size of the angle marked y on the diagram.

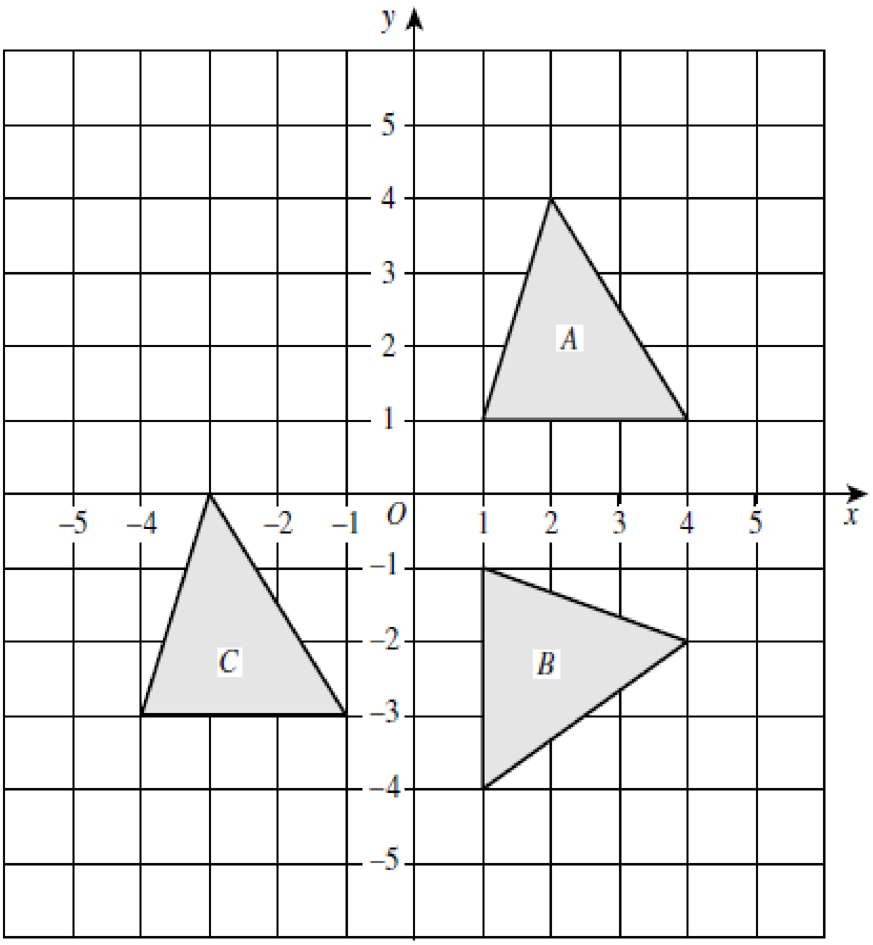
....

....

Answer .... degrees *(2 marks)*

Areas outside the box will not be scanned for marking

5 Triangles *A*, *B* and *C* are shown on the grid.



Describe fully the **single** transformation that maps triangle *A* onto triangle *B*.

....

....

....

(3 marks)

END OF QUESTIONS

Appendix D

Study 2:

Exam question types

Research Skills 1 exam (high diagnosticity vs. low diagnosticity)

Manipulation check

Table D.1.

Question type (solve vs. comparison vs. uncategorised) for each Research Skills 1 exam question in Study 2.

Question Number	Question type
1	Uncategorised
2	Solve
3	Uncategorised
4	Uncategorised
5	Uncategorised
6	Uncategorised
7	Uncategorised
8	Uncategorised
9	Uncategorised
10	Solve
11	Solve
12	Solve
13	Solve
14	Solve
15	Solve
16	Solve
17	Solve
18	Solve
19	Comparison
20	Comparison
21	Solve
22	Solve
23	Solve
24	Solve
25	Uncategorised
26	Uncategorised
27	Uncategorised
28	Uncategorised
29	Comparison
30	Uncategorised
31	Comparison
32	Comparison
33	Comparison
34	Solve
35	Comparison
36	Uncategorised
37	Solve
38	Solve
39	Solve
40	Comparison

PSYC1036 – Research Skills 1 - Mock Examination

This question paper consists of 10 printed pages and 40 Multiple Choice Questions (MCQ).

Formula sheets and statistical tables are provided.

Each question has four possible answers, only one is correct

Student Number:………………………………………………

Age: ....

Gender (Please circle): Male/Female

Ethnicity: ....

Please note: In previous years in the RS1 exam we have found that women are less competent at statistics compared to men.

---

1. A _________ represents the frequency counts in discrete categories for two nominal level variables.

   a. bar chart
   b. histogram
   c. box plot
   d. contingency table

2. What is the modal score in the following list of test scores: 55, 55, 59, 65, 65, 65, 70?

   a. 62.5
   b. 55
   c. 60
   d. 6

3. The mean, mode, and median:

   a. are all measures of central tendency
   b. are never equal
   c. are always equal
   d. both a and b.

4. The Wilcoxon Signed Ranks test is:

   a. a parametric version of the Mann-Whitney U-test
   b. a non-parametric version of the independent groups t-test
   c. a non-parametric test equivalent to the Binomial Sign Test
   d. both b and c

5. What is the Mann-Whitney U-test?

   a. a parametric version of the Wilcoxon Signed Ranks test
   b. non-parametric version of the independent groups t-test
   c. a non-parametric equivalent of the Binomial Sign Test
   d. a non-parametric equivalent of the Chi-square test of independence

6. Which of the following is a Repeated Measures design?

   a. all participants perform in all conditions
   b. each condition is repeated twice
   c. different participants perform in each condition
   d. none of the above

7. Choose the best definition for what is meant by the term ‘dependent variable’:

   a. A variable that is allowed to vary at random
   b. a predictor variable
   c. an outcome variable
   d. a variable that is manipulated by the experimenter

8. The effectiveness of a new pain-killer drug B was compared with that of drug A which had been in use for many years. Randomly selected patients were assigned to two treatments, with either drug A or drug B. What statistic should be used in order to compare the effectiveness of the two drugs?

   a. The Chi-square test for relatedness
   b. The Mann-Whitney U test
   c. The dependent t test
   d. The independent groups t test

9. The value of the degrees of freedom is best defined as:

    a. The critical value of a test statistic
    b. The number of scores free to vary in the calculation of a test statistic
    c. Always equal to N-1
    d. The number of dependent variables

Questions 10-19

Read the following scenario and use the information to help answer the questions that follow.

In a large national company the mean salary for all males in middle management with 3 to 5 years experience is £28,000. The salaries (expressed in thousands of pounds) for a random sample of 10 similarly experienced females in middle management are:

24, 27, 31, 21, 19, 26, 30, 22, 15, 36

10. What is the mean female salary (in thousands of pounds)?

    a. 25.00
    b. 25.10
    c. 28.00
    d. 26.55

11. What is the variance of the sample?

    a. 5.91
    b. 38.77
    c. 6.23
    d. 34.89

12. What is the standard deviation of the sample?

    a. 5.91
    b. 38.77
    c. 6.23
    d. 34.89

13. What is the value of ΣX^2?

    a. 348.9
    b. 63001
    c. 6300.1
    d. 6649

14. What is the estimated variance of the population?

    a. 5.91
    b. 38.77
    c. 6.23
    d. 34.89

15. What is the estimated standard deviation of the population?

    a. 5.91
    b. 38.77
    c. 6.23
    d. 34.89

16. What is the estimated standard error of the mean based on this sample?

    a. 6.23
    b. 3.16
    c. 1.97
    d. -2.9

17. What is the (one sample) *t*-score for this sample?

    a. 1.47
    b. -1.47
    c. 1.97
    d. -2.9

18. Given the original question, which critical value of *t* should be used to assess the significance of the obtained *t*-score at the two-tailed 5% level?

    a. 1.833
    b. 2.306
    c. 2.262
    d. 1.383

19. Based on your answers to questions 10 and 18, which of the following is the most appropriate conclusion to draw?

    a. There is evidence to suggest that the female salaries are significantly different from the male salaries.
    b. There is no evidence to suggest that the female salaries are significantly different from the male salaries.
    c. There is evidence to suggest that the female salaries are significantly lower than the male salaries.
    d. There is insufficient information to draw any conclusions

20. 14 people took part in a word-recall experiment under two conditions, in a quiet room and in a noisy room, using a counterbalanced design. What statistic should be used in order to compare the two conditions?

    a. The Chi-square test for relatedness
    b. The Wilcoxon Signed Ranks test
    c. The dependent t test
    d. The independent t test

21. A group of 8 boys is compared with a group of 6 girls in the number of errors made in a series of problem-solving tasks. What statistic should be used assuming that the samples do not come from a normally distributed population?
    a. The Chi-square test for Goodness-of-fit
    b. The Wilcoxon Signed Ranks test
    c. The Mann Whitney U test
    d. The Binomial Sign test

22. The degrees of freedom for a one-sample t-test with a sample of 9 participants is:

    a. 7
    b. 8
    c. 9
    d. None of these

23. The degrees of freedom for a Chi-square test when there is one independent variable with 7 levels is:

    a. 5
    b. 6
    c. 7
    d. None of these

24. In a χ^2 test for a contingency table having 7 rows and 7 columns, the degrees of freedom is:

    a. 5
    b. 6
    c. 12
    d. 36

25. Which of the following statement is wrong? A nonparametric test

    a. can use ranked data
    b. can be one-tail or two-tail
    c. does not need a null hypothesis
    d. does not need data to be numerical measurements

26. A normal population distribution is needed for the following statistical test:

    a. The Chi-square test for Goodness-of-fit
    b. The Wilcoxon Signed Ranks test
    c. The Mann Whitney U test
    d. The one sample t-test

27. With regard to the chi-square test:

    a. it is used to test the difference between frequencies
    b. it is used as an alternative to the t-test to determine the difference between two means
    c. the greater the value of the chi-squared test, the less likely it is to be significant
    d. the null hypothesis is not required

28. A z score of -0.56 would mean that the test score:

    a. was above the mean
    b. was below the mean
    c. was equal to the mean
    d. could have been above or below the mean; the z score gives no indication of that

29. During the pre-flight check, Pilot Jones discovers a minor problem – a warning light indicates that the fuel gauge may be broken. If Jones decides to check the fuel level by hand, it will delay the flight by 45 minutes. If Jones decides to ignore the warning, the aircraft may run out of fuel before it gets to Gimli. In this situation, what would be (1) the appropriate null hypothesis, and (2) a type I error?

    a. H_0: assume that the warning can be ignored ; Type I error: decide to check the fuel by hand when there is in fact enough fuel.
    b. H_0: assume that the warning can be ignored ; Type I error: decide to ignore the warning when there is in fact not enough fuel.
    c. H_0: assume that the fuel should be checked by hand ; Type I error: decide to ignore the warning when there is in fact not enough fuel.
    d. H_0: assume that the fuel should be checked by hand ; Type I error: decide to check the fuel by hand when there is in fact enough fuel.

30. In a hypothesis testing problem:

    a. the null hypothesis will not be rejected unless the data are not unusual (given that the hypothesis is true).
    b. the null hypothesis will not be rejected unless the p-value indicates the data are very unusual (given that the hypothesis is true).
    c. the null hypothesis is also called the research hypothesis
    d. the null hypothesis is the hypothesis that we would like to prove

31. A research psychobiologist has carried out an experiment on a random sample of 15 experimental plots in a field. Following the collection of data, a test of significance was conducted and the P-value was determined to be approximately .03. This indicates that:

    a. this result is statistically significant at the .01 level.
    b. the probability of being wrong in this situation is only .03.
    c. there is some reason to believe that the null hypothesis is incorrect.
    d. If this experiment were repeated 3 per cent of the time we would get this same result.

32. Which of the following statements is correct?

    a. An extremely small p-value indicates that the actual data differs markedly from that expected if the null hypothesis were true.
    b. The p-value measures the probability of making a Type II error.
    c. The larger the p-value, the stronger the evidence against the null hypothesis
    d. A large p-value indicates that the data is inconsistent with the alternative hypothesis.

33. Here are the scores of a memory test in 14 undergraduate students: 102, 108, 104, 102, 106, 107, 115, 98, 103, 99, 109, 111, 101, 99.
    Typically, it is published that the average score for this test at the University is 103. You believe that this published claim is not true. Test this claim at the $\alpha = 0.01$ level of significance. Which of the following conclusions is correct?

    a. There is significant evidence to support that the average IQ of undergraduate students on an IQ test is more than 103
    b. There is not significant evidence to support that the average IQ of undergraduate students on an IQ test is not 103
    c. There is significant evidence to support that the average IQ of undergraduate students on an IQ test is not 103
    d. not enough information

Questions 34-36

Consider this table

Independent Samples Test		t-test for Equality of Means		
		T	df	Sig. (2-tailed)
Reaction time	Equal variances assumed	10.991	28	.000
	Equal variances not assumed	10.991	30.727	.000

34 How many people took part in this experiment?

    a. 27
    b. 28
    c. 29
    d. 30

35. Did the independent variable have an effect?

    a. Yes, but the effect was not statistically significant
    b. Yes, there is a statistically significant effect
    c. Not enough information
    d. No, it is not statistically indicant

36. What was the dependent variable?

    a. Variances
    b. Reaction time
    c. Equal assumptions
    d. Not given in table

Questions 37-40

Consider the following data and analyse the data using the Wilcoxon Signed Ranks test

participants	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Data before training	8	4	6	2	4	8	3	1	3	5	9
Data after training	7	9	3	6	3	10	6	7	8	6	7

37. What rank should be assigned to the difference in participant's 5 data?
    a. 1
    b. 1.33
    c. 1.5
    d. 2

38. What is the value of Wilcoxon's T?
    a. 2
    b. 4
    c. 15
    d. 64

39. What is the critical value that should be used in assessing the significance of the value of T at the 5% level for a two-tailed test?

    a. 0
    b. 8
    c. 10
    d. 13

40. What is the most appropriate conclusion that could be made from this Wilcoxon Signed Ranks test?

    a. There is a significant decrease in the performance after the training
    b. There is a significant increase in the performance after the training
    c. Performance is not the same before and after the training
    d. The Null hypothesis should be accepted

PSYC1036 – Research Skills 1 - Mock Examination

This question paper consists of 10 printed pages and 40 Multiple Choice Questions (MCQ).

Formula sheets and statistical tables are provided.

Each question has four possible answers, only one is correct

Student Number:………………………………………………

Age: ....

Gender (Please circle): Male/Female

Ethnicity: ....

Please note: In previous years in the RS1 exam we have found no differences in statistical ability across men and women.

----

1. A _________ represents the frequency counts in discrete categories for two nominal level variables.

   a. bar chart
   b. histogram
   c. box plot
   d. contingency table

2. What is the modal score in the following list of test scores: 55, 55, 59, 65, 65, 65, 70?

   a. 62.5
   b. 55
   c. 60
   d. 6

3. The mean, mode, and median:

   a. are all measures of central tendency
   b. are never equal
   c. are always equal
   d. both a and b.

4. The Wilcoxon Signed Ranks test is:

   a. a parametric version of the Mann-Whitney U-test
   b. a non-parametric version of the independent groups t-test
   c. a non-parametric test equivalent to the Binomial Sign Test
   d. both b and c

5. What is the Mann-Whitney U-test?

   a. a parametric version of the Wilcoxon Signed Ranks test
   b. non-parametric version of the independent groups t-test
   c. a non-parametric equivalent of the Binomial Sign Test
   d. a non-parametric equivalent of the Chi-square test of independence

6. Which of the following is a Repeated Measures design?

   a. all participants perform in all conditions
   b. each condition is repeated twice
   c. different participants perform in each condition
   d. none of the above

7. Choose the best definition for what is meant by the term ‘dependent variable’:

   a. A variable that is allowed to vary at random
   b. a predictor variable
   c. an outcome variable
   d. a variable that is manipulated by the experimenter

8. The effectiveness of a new pain-killer drug B was compared with that of drug A which had been in use for many years. Randomly selected patients were assigned to two treatments, with either drug A or drug B. What statistic should be used in order to compare the effectiveness of the two drugs?

   a. The Chi-square test for relatedness
   b. The Mann-Whitney U test
   c. The dependent t test
   d. The independent groups t test

9. The value of the degrees of freedom is best defined as:

   a. The critical value of a test statistic
   b. The number of scores free to vary in the calculation of a test statistic
   c. Always equal to N-1
   d. The number of dependent variables

Questions 10-19

Read the following scenario and use the information to help answer the questions that follow.

In a large national company the mean salary for all males in middle management with 3 to 5 years experience is £28,000. The salaries (expressed in thousands of pounds) for a random sample of 10 similarly experienced females in middle management are:

24, 27, 31, 21, 19, 26, 30, 22, 15, 36

10. What is the mean female salary (in thousands of pounds)?

    a. 25.00
    b. 25.10
    c. 28.00
    d. 26.55

11. What is the variance of the sample?

    a. 5.91
    b. 38.77
    c. 6.23
    d. 34.89

12. What is the standard deviation of the sample?

    a. 5.91
    b. 38.77
    c. 6.23
    d. 34.89

13. What is the value of ΣX^2?

    a. 348.9
    b. 63001
    c. 6300.1
    d. 6649

14. What is the estimated variance of the population?

    a. 5.91
    b. 38.77
    c. 6.23
    d. 34.89

15. What is the estimated standard deviation of the population?

    a. 5.91
    b. 38.77
    c. 6.23
    d. 34.89

16. What is the estimated standard error of the mean based on this sample?

    a. 6.23
    b. 3.16
    c. 1.97
    d. -2.9

17. What is the (one sample) *t*-score for this sample?

    a. 1.47
    b. -1.47
    c. 1.97
    d. -2.9

18. Given the original question, which critical value of *t* should be used to assess the significance of the obtained *t*-score at the two-tailed 5% level?

    a. 1.833
    b. 2.306
    c. 2.262
    d. 1.383

19. Based on your answers to questions 10 and 18, which of the following is the most appropriate conclusion to draw?

    a. There is evidence to suggest that the female salaries are significantly different from the male salaries.
    b. There is no evidence to suggest that the female salaries are significantly different from the male salaries.
    c. There is evidence to suggest that the female salaries are significantly lower than the male salaries.
    d. There is insufficient information to draw any conclusions

20. 14 people took part in a word-recall experiment under two conditions, in a quiet room and in a noisy room, using a counterbalanced design. What statistic should be used in order to compare the two conditions?

    a. The Chi-square test for relatedness
    b. The Wilcoxon Signed Ranks test
    c. The dependent t test
    d. The independent t test

21. A group of 8 boys is compared with a group of 6 girls in the number of errors made in a series of problem-solving tasks. What statistic should be used assuming that the samples do not come from a normally distributed population?
    a. The Chi-square test for Goodness-of-fit
    b. The Wilcoxon Signed Ranks test
    c. The Mann Whitney U test
    d. The Binomial Sign test

22. The degrees of freedom for a one-sample t-test with a sample of 9 participants is:

    a. 7
    b. 8
    c. 9
    d. None of these

23. The degrees of freedom for a Chi-square test when there is one independent variable with 7 levels is:

    a. 5
    b. 6
    c. 7
    d. None of these

24. In a χ^2 test for a contingency table having 7 rows and 7 columns, the degrees of freedom is:

    a. 5
    b. 6
    c. 12
    d. 36

25. Which of the following statement is wrong? A nonparametric test

    a. can use ranked data
    b. can be one-tail or two-tail
    c. does not need a null hypothesis
    d. does not need data to be numerical measurements

26. A normal population distribution is needed for the following statistical test:

    a. The Chi-square test for Goodness-of-fit
    b. The Wilcoxon Signed Ranks test
    c. The Mann Whitney U test
    d. The one sample t-test

27. With regard to the chi-square test:

    a. it is used to test the difference between frequencies
    b. it is used as an alternative to the t-test to determine the difference between two means
    c. the greater the value of the chi-squared test, the less likely it is to be significant
    d. the null hypothesis is not required

28. A z score of -0.56 would mean that the test score:

    a. was above the mean
    b. was below the mean
    c. was equal to the mean
    d. could have been above or below the mean; the z score gives no indication of that

29. During the pre-flight check, Pilot Jones discovers a minor problem – a warning light indicates that the fuel gauge may be broken. If Jones decides to check the fuel level by hand, it will delay the flight by 45 minutes. If Jones decides to ignore the warning, the aircraft may run out of fuel before it gets to Gimli. In this situation, what would be (1) the appropriate null hypothesis, and (2) a type I error?

    a. H_0: assume that the warning can be ignored ; Type I error: decide to check the fuel by hand when there is in fact enough fuel.
    b. H_0: assume that the warning can be ignored ; Type I error: decide to ignore the warning when there is in fact not enough fuel.
    c. H_0: assume that the fuel should be checked by hand ; Type I error: decide to ignore the warning when there is in fact not enough fuel.
    d. H_0: assume that the fuel should be checked by hand ; Type I error: decide to check the fuel by hand when there is in fact enough fuel.

30. In a hypothesis testing problem:

    a. the null hypothesis will not be rejected unless the data are not unusual (given that the hypothesis is true).
    b. the null hypothesis will not be rejected unless the p-value indicates the data are very unusual (given that the hypothesis is true).
    c. the null hypothesis is also called the research hypothesis
    d. the null hypothesis is the hypothesis that we would like to prove

31. A research psychobiologist has carried out an experiment on a random sample of 15 experimental plots in a field. Following the collection of data, a test of significance was conducted and the P-value was determined to be approximately .03. This indicates that:

    a. this result is statistically significant at the .01 level.
    b. the probability of being wrong in this situation is only .03.
    c. there is some reason to believe that the null hypothesis is incorrect.
    d. If this experiment were repeated 3 per cent of the time we would get this same result.

32. Which of the following statements is correct?

    a. An extremely small p-value indicates that the actual data differs markedly from that expected if the null hypothesis were true.
    b. The p-value measures the probability of making a Type II error.
    c. The larger the p-value, the stronger the evidence against the null hypothesis
    d. A large p-value indicates that the data is inconsistent with the alternative hypothesis.

33. Here are the scores of a memory test in 14 undergraduate students: 102, 108, 104, 102, 106, 107, 115, 98, 103, 99, 109, 111, 101, 99.
Typically, it is published that the average score for this test at the University is 103. You believe that this published claim is not true. Test this claim at the $\alpha = 0.01$ level of significance. Which of the following conclusions is correct?

    a. There is significant evidence to support that the average IQ of undergraduate students on an IQ test is more than 103
    b. There is not significant evidence to support that the average IQ of undergraduate students on an IQ test is not 103
    c. There is significant evidence to support that the average IQ of undergraduate students on an IQ test is not 103
    d. not enough information

Questions 34-36

Consider this table

Independent Samples Test		t-test for Equality of Means		
		T	df	Sig. (2-tailed)
Reaction time	Equal variances assumed	10.991	28	.000
	Equal variances not assumed	10.991	30.727	.000

34 How many people took part in this experiment?

    a. 27
    b. 28
    c. 29
    d. 30

35. Did the independent variable have an effect?

    a. Yes, but the effect was not statistically significant
    b. Yes, there is a statistically significant effect
    c. Not enough information
    d. No, it is not statistically indicant

36. What was the dependent variable?

    a. Variances
    b. Reaction time
    c. Equal assumptions
    d. Not given in table

Questions 37-40

Consider the following data and analyse the data using the Wilcoxon Signed Ranks test

participants	1	2	3	4	5	6	7	8	9	10	11
Data before training	8	4	6	2	4	8	3	1	3	5	9
Data after training	7	9	3	6	3	10	6	7	8	6	7

37. What rank should be assigned to the difference in participant's 5 data?
    a. 1
    b. 1.33
    c. 1.5
    d. 2

38. What is the value of Wilcoxon's T?
    a. 2
    b. 4
    c. 15
    d. 64

39. What is the critical value that should be used in assessing the significance of the value of T at the 5% level for a two-tailed test?

    a. 0
    b. 8
    c. 10
    d. 13

40. What is the most appropriate conclusion that could be made from this Wilcoxon Signed Ranks test?

    a. There is a significant decrease in the performance after the training
    b. There is a significant increase in the performance after the training
    c. Performance is not the same before and after the training
    d. The Null hypothesis should be accepted

PSYC1036 – Research Skills 1 - Mock Examination

Student Number

Please complete the following the questions using the scales provided:

1. To what extent are there gender differences in performance on this maths test?

No gender differences 1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7 -- 8 -- 9 -- 10 -- 11 *Gender differences*

2. Who do you believe performs better on this maths test?

Men perform better 1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7 -- 8 -- 9 -- 10 -- 11 *Women perform better*

Appendix E

Study 3:

Maths question types

Maths tests (high diagnosticity vs. low diagnosticity)

Manipulation check

Table E.1.

Question type (solve vs. comparison) for each maths question in Study 3.

Question Number	Question type
1	Comparison
2	Solve
3	Solve
4	Comparison
5	Solve
6	Solve
7	Solve
8	Comparison
9	Comparison
10	Comparison

MATHEMATICS (SPECIFICATION A)
Higher Tier
Paper 1 Non-calculator

UNIVERSITY OF LEEDS

Age: ....

Gender (Please circle): Male/Female

Ethnicity: ....

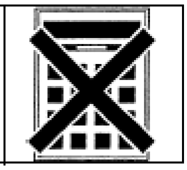
Nationality: ....

Note:

- **Previous research has shown gender differences on this test.**

For this paper you must have:
- mathematical instruments.

You must **not** use a calculator.



Time allowed: 35 minutes

Instructions
- Use black ink or black ball-point pen. Draw diagrams in pencil.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- You must answer the questions in the spaces provided. Answers written in margins or on blank pages will not be marked.
- Do all rough work in this book.

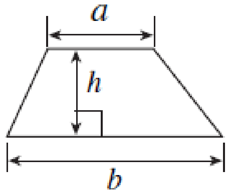
Information
- The maximum mark for this paper is **30**
- The marks for questions are shown in brackets.
- You may ask for more answer paper, graph paper and tracing paper. This must be tagged securely to this answer booklet.

Advice
- In all calculations, show clearly how you work out your answer.

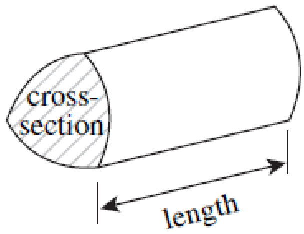
For Examiner's Use	
Pages	Mark
3	
4–5	
6–7	
8–9	
10–11	
12–13	
14–15	
16–17	
18–19	
20–21	
TOTAL	
Examiner's Initials	

Formulae Sheet: Higher Tier

Area of trapezium = $\frac{1}{2}(a+b)h$

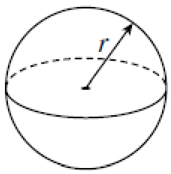


Volume of prism = area of cross-section × length



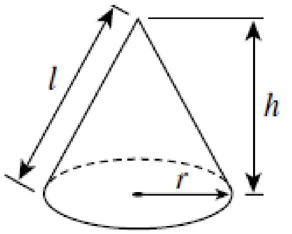
Volume of sphere = $\frac{4}{3}\pi r^3$

Surface area of sphere = $4\pi r^2$



Volume of cone = $\frac{1}{3}\pi r^2 h$

Curved surface area of cone = $\pi r l$

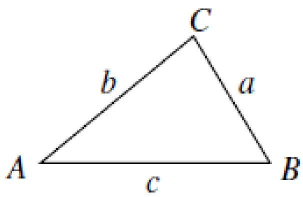


In any triangle ABC

Area of triangle = $\frac{1}{2}ab\sin C$

Sine rule $\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$

Cosine rule $a^2 = b^2 + c^2 - 2bc\cos A$



The Quadratic Equation

The solutions of $ax^2 + bx + c = 0$, where $a \neq 0$, are given by

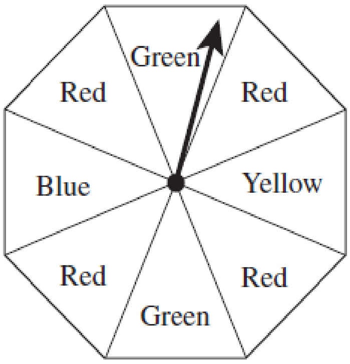
$$x = \frac{-b \pm \sqrt{(b^2 - 4ac)}}{2a}$$

Areas outside the box will not be scanned for marking

Answer **all** questions in the spaces provided.

1 (a) Oscar has a spinner with eight sections.
Four of the sections are Red, two are Green, one is Blue and one is Yellow.
He spins the spinner 200 times.
His results are shown in the table.

Colour	Red	Green	Blue	Yellow
Frequency	105	48	22	25



1 (a) (i) Explain why the relative frequency of Green is 0.24

....

....

(1 mark)

1 (a) (ii) Do the results suggest that the spinner is fair?
Explain your answer.

....

....

....

(2 marks)

Areas outside the box will not be scanned for marking

2 Work out $1\frac{7}{8} \times 2\frac{2}{5}$

Give your answer in its simplest form.

....

....

....

....

Answer .... *(3 marks)*

3 Show **clearly** that $(n + 1)^2 + (n - 1)^2 \equiv 2n^2 + 2$

....

....

....

....

(3 marks)

Areas outside the box will not be scanned for marking

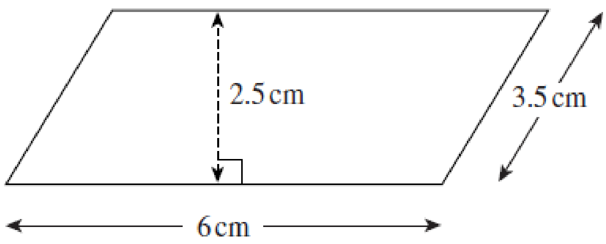
4 x, y and z represent lengths.

For each expression, put a tick in a box to show whether it represents a length, an area, a volume or none of these.

Expression	Length	Area	Volume	None of these
$3x + y + 2z$				
$x + z^2$				
$x^3 + 5y^2z$				

(3 marks)

5 The diagram shows a parallelogram.



Not drawn accurately

Calculate the area of the parallelogram.
State the units of your answer.

....

....

....

Answer .... *(3 marks)*

6 Leah, Chloe and Maya share £400 between them.
Leah receives the smallest amount of £90
The ratio of Leah's share to Chloe's share is 2 : 3

Work out how much Maya receives.

....

....

....

....

....

....

Answer £ .... *(3 marks)*

7 Work out $3\frac{3}{4} \div 1\frac{2}{3}$

....

....

....

....

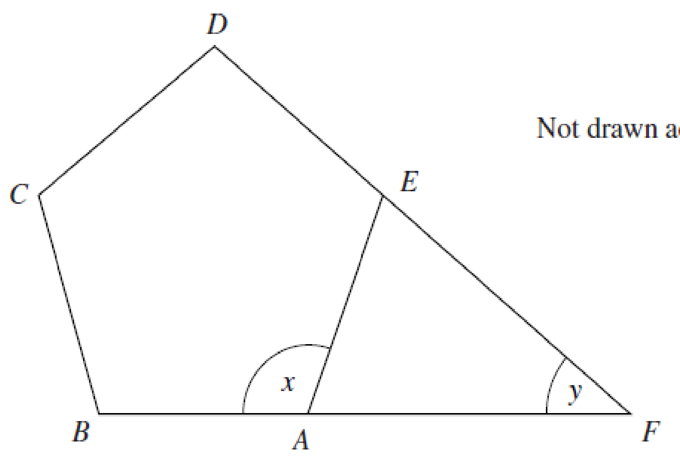
....

....

Answer .... *(3 marks)*

Areas outside the box will not be scanned for marking

8 *ABCDE* is a regular pentagon.
DEF and *BAF* are straight lines.



8 (a) Which **one** of these statements is true?

1 The exterior angle of a regular pentagon is equal to $360° \div 5 = 72°$

2 The interior angle of a regular pentagon is equal to $360° \div 5 = 72°$

3 The exterior angle of a regular pentagon is equal to $360° - 72° = 288°$

4 The interior angle of a regular pentagon is equal to $360° - 72° = 288°$

Answer .... *(1 mark)*

8 Work out the size of the angle marked y on the diagram.

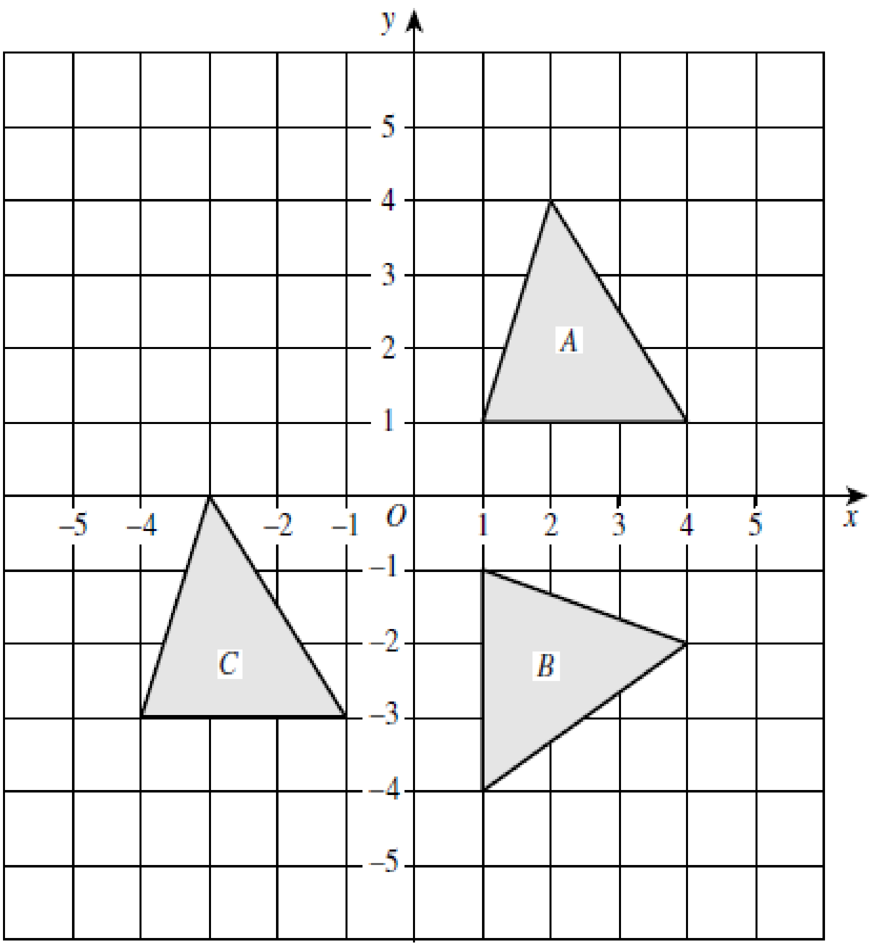
....

....

Answer .... degrees *(2 marks)*

Areas outside the box will not be scanned for marking

9 Triangles *A*, *B* and *C* are shown on the grid.



Describe fully the **single** transformation that maps triangle *A* onto triangle *B*.

....

....

....

(3 marks)

Areas outside the box will not be scanned for marking

10 A Golf Club has 600 members.
A stratified sample of members is taken, by age group.

The table shows the age grouping of the members.
Some information is given in the table.

Age group	Junior	18 – 39	40 – 59	Senior
Number of members	100			120
Number in sample	20		35	

Complete the table.

....

....

....

....

....

....

....

....

(3 marks)

END OF QUESTIONS

Centre Number						Candidate Number				

For Examiner's Use

MATHEMATICS (SPECIFICATION A)
Higher Tier
Paper 1 Non-calculator

UNIVERSITY OF LEEDS

Age: ....

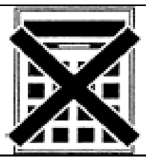
Gender (Please circle): Male/Female

Ethnicity: ....

Nationality: ....

Note:

- **Previous research has shown no gender differences on this test.**

For this paper you must have: • mathematical instruments. You must **not** use a calculator.	

Time allowed: 35 minutes

Instructions

- Use black ink or black ball-point pen. Draw diagrams in pencil.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- You must answer the questions in the spaces provided. Answers written in margins or on blank pages will not be marked.
- Do all rough work in this book.

Information

- The maximum mark for this paper is 1 30
- The marks for questions are shown in brackets.
- You may ask for more answer paper, graph paper and tracing paper This must be tagged securely to this answer booklet.

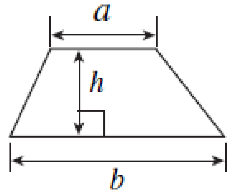
Advice

- In all calculations, show clearly how you work out your answer.

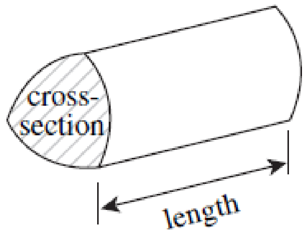
For Examiner's Use	
Pages	Mark
3	
4–5	
6–7	
8–9	
10–11	
12–13	
14–15	
16–17	
18–19	
20–21	
TOTAL	
Examiner's Initials	

Formulae Sheet: Higher Tier

Area of trapezium = $\frac{1}{2}(a+b)h$

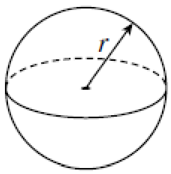


Volume of prism = area of cross-section × length



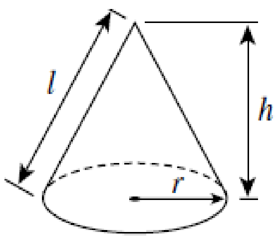
Volume of sphere = $\frac{4}{3}\pi r^3$

Surface area of sphere = $4\pi r^2$



Volume of cone = $\frac{1}{3}\pi r^2 h$

Curved surface area of cone = $\pi r l$

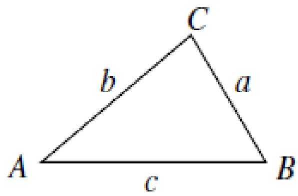


In any triangle ABC

Area of triangle = $\frac{1}{2}ab\sin C$

Sine rule $\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$

Cosine rule $a^2 = b^2 + c^2 - 2bc\cos A$



The Quadratic Equation

The solutions of $ax^2 + bx + c = 0$, where $a \neq 0$, are given by

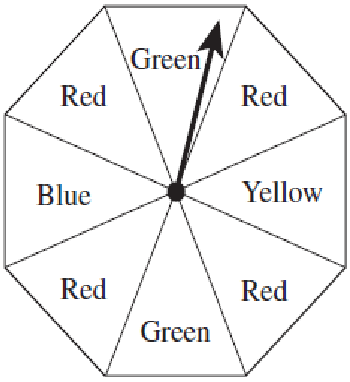
$$x = \frac{-b \pm \sqrt{(b^2 - 4ac)}}{2a}$$

Areas outside the box will not be scanned for marking

Answer **all** questions in the spaces provided.

1 (a) Oscar has a spinner with eight sections.
Four of the sections are Red, two are Green, one is Blue and one is Yellow.
He spins the spinner 200 times.
His results are shown in the table.

Colour	Red	Green	Blue	Yellow
Frequency	105	48	22	25



1 (i) Explain why the relative frequency of Green is 0.24

....

....

(1 mark)

1 (ii) Do the results suggest that the spinner is fair?
Explain your answer.

....

....

....

(2 marks)

Areas outside the box will not be scanned for marking

2 Work out $1\frac{7}{8} \times 2\frac{2}{5}$

Give your answer in its simplest form.

....

....

....

....

Answer .... *(3 marks)*

1 **3** Show **clearly** that $(n+1)^2 + (n-1)^2 \equiv 2n^2 + 2$

....

....

....

....

(3 marks)

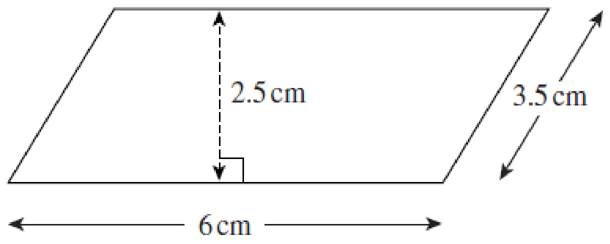
4 x, y and z represent lengths.

For each expression, put a tick in a box to show whether it represents a length, an area, a volume or none of these.

Expression	**Length**	**Area**	**Volume**	**None of these**
$3x + y + 2z$				
$x + z^2$				
$x^3 + 5y^2z$				

(3 marks)

5 The diagram shows a parallelogram.



Not drawn accurately

Calculate the area of the parallelogram.
State the units of your answer.

....

....

....

Answer .... *(3 marks)*

Areas outside the box will not be scanned for marking

6 Leah, Chloe and Maya share £400 between them.
Leah receives the smallest amount of £90
The ratio of Leah's share to Chloe's share is 2 : 3

Work out how much Maya receives.

....

....

....

....

....

....

Answer £ .... *(3 marks)*

7 Work out $3\frac{3}{4} \div 1\frac{2}{3}$

....

....

....

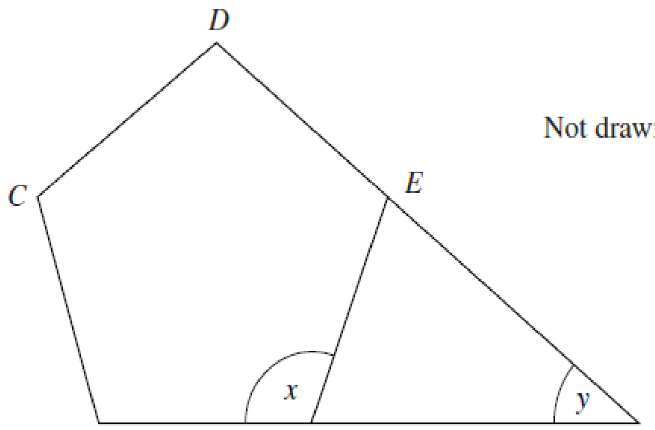
....

....

....

Answer .... *(3 marks)*

8 *ABCDE* is a regular pentagon.
DEF and *BAF* are straight lines.



Not drawn accurately

8 (a) Which **one** of these statements is true?

1 The exterior angle of a regular pentagon is equal to $360° \div 5 = 72°$

2 The interior angle of a regular pentagon is equal to $360° \div 5 = 72°$

3 The exterior angle of a regular pentagon is equal to $360° - 72° = 288°$

4 The interior angle of a regular pentagon is equal to $360° - 72° = 288°$

Answer .... *(1 mark)*

8 (ii) Work out the size of the angle marked y on the diagram.

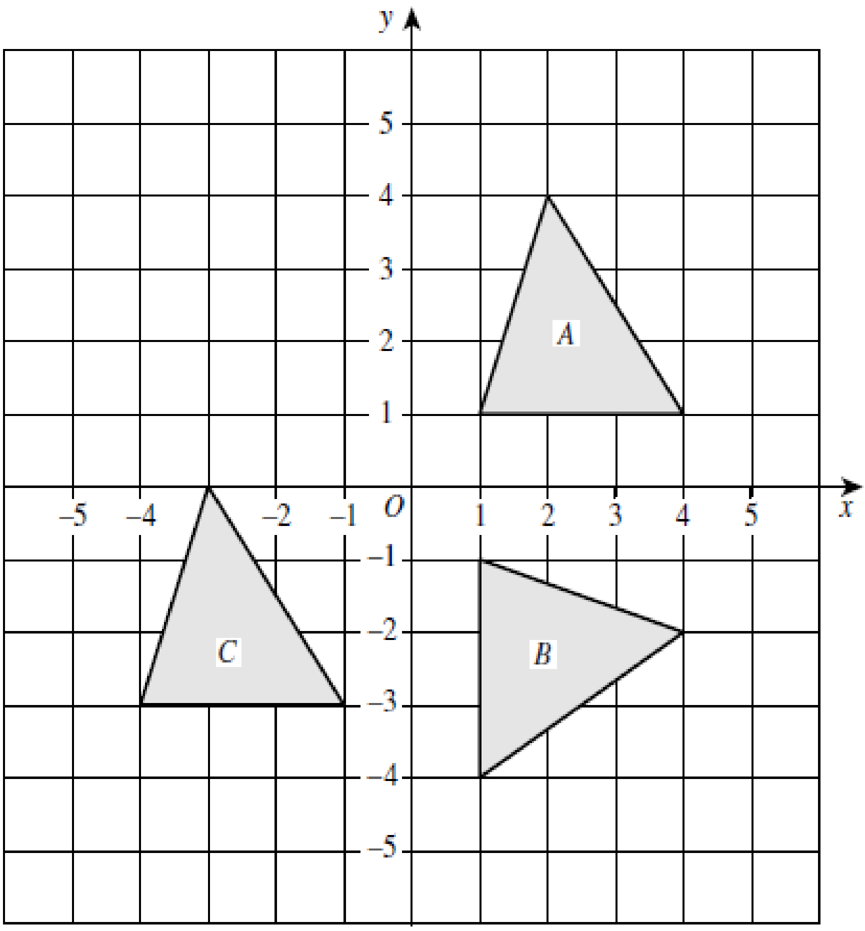
....

....

Answer .... degrees *(2 marks)*

Areas outside the box will not be scanned for marking

9 Triangles *A*, *B* and *C* are shown on the grid.



Describe fully the **single** transformation that maps triangle *A* onto triangle *B*.

....

....

....

(3 marks)

Areas outside the box will not be scanned for marking

10 A Golf Club has 600 members.
A stratified sample of members is taken, by age group.

The table shows the age grouping of the members.
Some information is given in the table.

Age group	Junior	18 – 39	40 – 59	Senior
Number of members	100			120
Number in sample	20		35	

Complete the table.

....

....

....

....

....

....

....

....

(3 marks)

END OF QUESTIONS

GCSE Mock Examination

2012/2013

Student Number

Please complete the following the questions using the scales provided:

1. To what extent are there gender differences in performance on this maths test?

No gender differences 1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7 -- 8 -- 9 -- 10 -- 11 *Gender differences*

2. Who do you believe performs better on this maths test?

Men perform better 1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7 -- 8 -- 9 -- 10 -- 11 *Women perform better*

End of Exam

Appendix F

Pilot study 5a:

Maths question selection types

Maths test

Study 5:

Maths question selection types

Maths question selection task

Table F.1.

Question type (solve vs. comparison) for each maths question in Pilot study 5a.

Question Number	Question type
1	Solve
2	Comparison
3	Solve
4	Comparison
5	Solve
6	Comparison
7	Comparison
8	Solve
9	Solve
10	Comparison
11	Comparison
12	Solve
13	Comparison
14	Solve
15	Comparison
16	Solve
17	Comparison
18	Solve

Centre Number						Candidate Number				

For Examiner's Use

MATHEMATICS (SPECIFICATION A)
Higher Tier
Paper 1 Non-calculator

UNIVERSITY OF LEEDS

For this paper you must have:
• mathematical instruments.
You must **not** use a calculator.

Time allowed: 1 hour

Instructions
- Use black ink or black ball-point pen. Draw diagrams in pencil.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- You must answer the questions in the spaces provided. Answers written in margins or on blank pages will not be marked.
- Do all rough work in this book.

Information
- The maximum mark for this paper is 54
- The marks for questions are shown in brackets.
- You may ask for more answer paper, graph paper and tracing paper. This must be tagged securely to this answer booklet.

Advice
- In all calculations, show clearly how you work out your answer.

For Examiner's Use	
Pages	Mark
3	
4–5	
6–7	
8–9	
10–11	
12–13	
14–15	
16–17	
18–19	
20–21	
TOTAL	
Examiner's Initials	

Answer **all** questions in the spaces provided.

1 $k = 9$ and $m = -4$

Work out the value of $\dfrac{5(2k-6)}{m}$

....

....

....

....

Answer .... *(3 marks)*

2 a and b are different prime numbers less than 12.

Work out **three** pairs of numbers a and b such that $\sqrt{(2a + b)}$ is a whole number.

....

....

....

....

....

....

Answer a = ... and b = ...

a = ... and b = ...

a = ... and b = ... *(3 marks)*

Areas outside the box will not be scanned for marking

3 Simplify $w^{12} \div w^4$

....

Answer .... *(1 mark)*

Rearrange $y = 3x + 2$ to make x the subject.

....

....

....

Answer .... *(2 marks)*

4 Ali, Beth and Clare take a test.

The ratio of Ali's score to Beth's score is 5 : 3
Ali scored 10 more marks than Beth.

Clare scored 7 more marks than Ali.

Work out each of their scores.

....

....

....

....

....

....

Answer Ali .... marks

Beth .... marks

Clare .... marks

(3 marks)

Areas outside the box will not be scanned for marking

5 Given that $d = 6$ and $f = -12$

work out the value of $\frac{9(d-10)}{f}$

....

....

....

....

Answer .... *(3 marks)*

6 Hence, or otherwise, write 176 as the product of its prime factors.
Give your answer in index form.

....

....

....

....

....

....

Answer .... *(3 marks)*

Areas outside the box will not be scanned for marking

7 A bag contains only red, blue and yellow counters.
There are three times as many blue counters as yellow counters.
There are 43 counters in the bag.

Some red counters are added to the bag.
There are now 50 counters in the bag.
The number of red counters has doubled.

How many yellow counters are in the bag?

....

....

....

....

....

....

Answer .... *(3 marks)*

8 $2(x + 16) + 4(x - 5)$ simplifies to $a(x + b)$

Work out the values of a and b.

....

....

....

....

Answer a = ... , b = ... *(3 marks)*

Areas outside the box will not be scanned for marking

9 You are given that $23.5 \times 64 = 1504$

Work out $\dfrac{1504}{640}$

....

Answer .... *(1 mark)*

Work out 23.5×65

....

Answer .... *(2 marks)*

10 Liz has a £20 voucher for an online music shop.
She buys ten songs costing 80p each.

What percentage of her voucher has Liz spent?

....

....

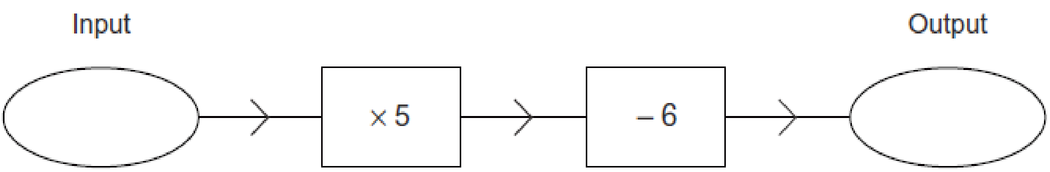
....

....

Answer ....% *(3 marks)*

Areas outside the box will not be scanned for marking

11 Here is a different number machine.



The output is equal to the input.

Work out the input.

....

....

....

Answer .... *(3 marks)*

12 Divide £600 in the ratio 9 : 6 : 5

....

....

....

....

Answer £ .... : £ .... : £ .... *(3 marks)*

13 Dan has lost weight.
He now weighs 108 kg.
He has lost 10% of his weight since March.

How much did he weigh in March?

....

....

....

....

....

....

Answer .... kg *(3 marks)*

14 Simplify $\frac{3x^2 - 19x + 20}{x^2 - 25}$

....

....

....

....

....

....

Answer .... *(3 marks)*

15 Here is a list of ingredients.

Serves 4 people	
Bacon	50 g
Minced beef	450 g
Chopped tomatoes	400 g
Button mushrooms	100 g
Beef stock	125 ml

Marco is making a meal for 14 people using these ingredients.

Work out the number of grams of minced beef he needs.

....

....

....

....

....

Answer .... g *(3 marks)*

16 Solve $4(x + 3) = 17$

....

....

Answer x = ... *(3 marks)*

Areas outside the box will not be scanned for marking

17 In a sale, a TV is reduced in price by 20%.
The sale price is £280.
After the sale, the price goes back to the original price.
Matt has £340 to spend.

Can he afford the TV when it goes back to its original price?

....

....

....

....

(3 marks)

18 Make h the subject of $2(h - y) = 5y + 3$

....

....

....

....

....

Answer .... *(3 marks)*

END OF QUESTIONS

Table F.2.

Question type (solve vs. comparison) for each maths selection question in Study 5.

Question Number	Question type	Label
1	Comparison	Type B
2	Solve	Type A
3	Solve	Type A
4	Comparison	Type B
5	Solve	Type A
6	Solve	Type A
7	Solve	Type A
8	Comparison	Type B
9	Comparison	Type B
10	Comparison	Type B

Note. Maths question selection test order 1 is shown. For each question selection test order (1-5) the questions are moved by 2 places. For example in order 2, questions 1 and 2 become 3 and 4, questions 3 and 4 become 5 and 6 etc.

Centre Number						Candidate Number				

For Examiner's Use

MATHEMATICS (SPECIFICATION A)
Higher Tier

UNIVERSITY OF LEEDS

Question Selection Task

- For this task, you will be asked to select **FIVE** maths questions to answer from the following choice of ten.
- Psychologists have identified two different types of maths questions labelled here as either **type A** or **type B**.
- When selecting the questions you wish to answer please put a cross in the box provided next to the question number.
- Once you have made your selection please contact the experimenter.

type B ☐

a and b are different prime numbers less than 12.

Work out **three** pairs of numbers a and b such that $\sqrt{(2a + b)}$ is a whole number.

....

....

....

....

....

....

Answer a = ... and b = ...

a = ... and b = ...

a = ... and b = ... *(3 marks)*

type A ☐

Simplify $w^{12} \div w^4$

....

Answer .... *(1 mark)*

Rearrange $y = 3x + 2$ to make x the subject.

....

....

....

Answer .... *(2 marks)*

type B ☐

Ali, Beth and Clare take a test.

The ratio of Ali's score to Beth's score is 5 : 3
Ali scored 10 more marks than Beth.

Clare scored 7 more marks than Ali.

Work out each of their scores.

....

....

....

....

....

....

Answer Ali .... marks

Beth .... marks

Clare .... marks

(3 marks)

type B ☐

Hence, or otherwise, write 176 as the product of its prime factors.
Give your answer in index form.

....

....

....

....

....

....

Answer .... *(3 marks)*

type A

$2(x + 16) + 4(x - 5)$ simplifies to $a(x + b)$

Work out the values of a and b.

....

....

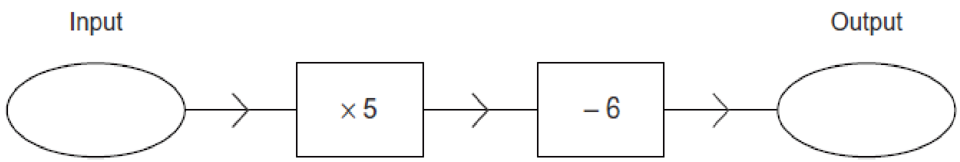
....

....

Answer a = ... , b = ... *(3 marks)*

type B

Here is a different number machine.



The output is equal to the input.

Work out the input.

....

....

....

Answer .... *(3 marks)*

type A ☐

Simplify $\dfrac{3x^2 - 19x + 20}{x^2 - 25}$

....

....

....

....

....

....

Answer .... *(3 marks)*

type A ☐

Solve $4(x + 3) = 17$

....

....

Answer x = ... *(3 marks)*

type B ☐

In a sale, a TV is reduced in price by 20%.
The sale price is £280.
After the sale, the price goes back to the original price.
Matt has £340 to spend.

Can he afford the TV when it goes back to its original price?

....

....

....

....

(3 marks)

type A ☐

Make h the subject of $2(h - y) = 5y + 3$

....

....

....

....

....

Answer .... *(3 marks)*

END OF QUESTIONS