How can we help children to learn?

Yue Li, M.Sc.

A thesis submitted in fulfilment of requirements for the degree of
Doctor of Philosophy
to
The University of Edinburgh
August 2020

Acknowledgements

I would like to express my deep and sincere gratitude to my supervisors Professor Timothy C. Bates and Dr René Mõttus, for their inspiration and support during the period in which the work presented in this thesis was undertaken: I thank especially my principal supervisor Tim, who provided invaluable guidance throughout this thesis and advice for carrying out scientific research. He taught me how to think critically, to be curious, to verify curiosity scientifically, and to be open-minded about everything. It was a great honour for me to work with him.

I am extremely grateful to my parents, for their constant love and unconditional trust and respect. I am very much indebted to my parents for their support for all my decisions. I would not have been able to finish this thesis without their support and encouragement.

谢谢我的父母给了我生命，并把我养育成人。你们无私的爱，无条件的信任、尊重和支持给了我追逐梦想的勇气和动力。同时也感谢你们在我遇到困难和挫折时一直站在我身后给我力量，没有你们的鼓励和支持，我不可能有毅力完成我的博士论文。你们是世界上最好的父母！我爱你们！

I also wish to thank all of my friends for their friendship. They encouraged me when I was frustrated, listened to me when I was upset, and accompanied me when I felt lonely. They made my life in Edinburgh colourful!

Abstract

This thesis began with a question: How can we help children to learn? I examined this question by testing four distinct factors proposed to influence learning: number awareness, mindset, stereotyping and conscientiousness.

Beginning with the effects of number awareness (Chapter 1), I focused on the approach of helping young children to understand the symbolic meaning of numbers. This idea was inspired by the study of Ramani and Siegler (2008) which tested whether playing a board game (e.g. snakes and ladders) in which their "player" was identified with a number would improve children's understanding of numbers and their numerical performance. I attempted to replicate this study in China but found no effect. Children who played the number board game did not show significant improvements in the post-manipulation numerical tests compared to children who played the coloured board game. I also noticed that children in my study showed a better baseline performance in counting and identification tests compared to children in the original study (in the U.S.). Therefore, I concluded that the board game manipulation was not helpful for raising number awareness among Chinese children. I then searched the literature for other manipulations that might help children on learning. The most prominent manipulation was based on mindset theory (Dweck, 2006), which measured whether children believe their intelligence is malleable or fixed (Chapter 2).

Therefore, the second major theme, occupying most of the thesis, consisted of eight studies on mindset. In this vein, I firstly used both a mindset manipulation and a self-reported mindset scale to test whether having a growth mindset would improve children's cognitive performance after a challenge (Chapter 3). Children were firstly asked to solve a set of moderate difficulty cognitive problems, and then they were given a mindset manipulation: the experimenter either praised children for being smart (fixed mindset manipulation) or for working hard (growth mindset manipulation) on these problems (Mueller & Dweck, 1998). Following the mindset manipulation, the children were asked to work on a set of even more difficult problems (failure), and they were given negative feedback on their performance. This was followed by a final set of problems with moderate difficulty. I failed to find a consistent association between the growth mindset manipulation and improvement in children' post-challenge performance. The only nominally significant effect of growth mindset manipulation was found in my first study, but disappeared in the other two studies. I next tested whether children's own mindsets would be associated with their post-failure performance or school grades. Again, I failed to find a significant association between children's own mindsets and their post-failure performance or school grades. I then tested whether growth mindset would be associated with better grades only across a challenging transition or growth mindset would be beneficial only for children who encountered the greatest challenge when entering university (Chapter 4). Similar to my previous studies, I failed to find any significant effect of growth mindset on grades, either across a challenging transition (from high school to university), in any subsequent year in university, nor among children who encountered the greatest challenge when entering university. Finally, I explored whether children obtained their own mindsets from their parents' growth mindsets or parents' failure mindsets (Chapter 5). I found that children's own mindsets were significantly associated with their perceptions of parents' failure mindsets. However, these perceptions were biased by their own mind. Therefore, I concluded that mindset theory was not influential in the learning process, children might obtain their mindsets from their parents' failure mindsets, and children's perceptions of their parents' failure mindsets were biased. I also suggested that mindset manipulation has its limitations and suggested that rather than addressing mindset, addressing attitudes to hard work may be a viable direction.

Next, I conducted a series of five studies to test a recent claim regarding the early origins of negative gender stereotypes about brilliance (Bian, Leslie, & Cimpian, 2017), that is proposed to reduce girls' interests in science-related subjects (Chapter 6). Consequently, the negative stereotypes would impair girls' academic performance and increase the gender gap in science-related subjects. I tested children' gender stereotypes about brilliance, kindness, and dullness, by presenting vignettes of people who were very high in these characteristics in both China and the U.K. (Chapter 7). I asked children to identify which person in two male and two female images presented to them was the person described in the vignette. The gender of the person chosen was used as an indicator of stereotyping. I failed to find the existence of a gender stereotype that brilliance was a male trait among Chinese children, but did find it among British children. For niceness, I did find that both Chinese and British children have a gender stereotype towards women. For dullness, both Chinese and British children hold a gender stereotype towards men.

The final theme in this thesis was followed by the mindset theme, which consisted of four studies testing whether conscientiousness (i.e., attitudes towards hard work) would be associated with children' grades in both primary school and high school (Chapter 8). I found that conscientiousness was a significant and consistent predictor of grades in both primary school and high school. I also tested whether teachers' conscientiousness would be associated with children's grades. However, children's perceptions of their teachers' conscientiousness were not associated with children's grades. Thus, I concluded that conscientiousness was a powerful predictor of grades, but teachers' conscientiousness might not be.

In the final chapter (Chapter 9) of this thesis, I concluded the main findings in the thesis, discussed the implications for each theory I explored, and made suggestions for further studies in the studied areas.

Lay summary

This thesis began with a question: how can we help children to learn? Over the course of the thesis, I addressed this question from four distinct theoretical approaches: number awareness, mindset, stereotyping and conscientiousness.

In the initial work, I focused on novel approaches to teaching mathematical knowledge to young children (around age 5). The idea behind this work was that some children may experience roadblocks in math learning for reasons as simple as not having entered the very first step of symbolic learning. This idea was inspired by a recent influential paper (Ramani & Siegler, 2008) testing whether playing a number board game (e.g. snakes and ladders) would help children to understand the meaning of numbers, and consequently to have better performance on numerical tasks. I attempted to replicate this study in a Chinese sample, but failed to find the same results. Namely, playing the number board game was not helpful for Chinese children. I also noticed that children in my study showed a better baseline performance in counting and identification tests compared to children in the original study (in the U.S.). Therefore, I turned to searching for other possible models that may be helpful for children's learning. Reviewing the literature suggested that a temporal model of the growth mindset theory may be the correct answer. I therefore changed my focus from a teaching approach to the mental aspect: whether children believed their intelligence could be changed or not would influence their educational attainment.

Reviewing, replicating and testing mindset theory (Dweck, 2006) may help me to understand how this works for helping children to learn. I began my investigations of mindset theory by attempting to replicate one of their widely cited papers (Mueller & Dweck, 1998). In my replication studies, I gave children a short manipulation to see if their post-manipulation performance differed by group. I also tested whether children's own implicit beliefs about intelligence were influential for their cognitive performance as well as their school grades. The results indicated that the mindset manipulation has its own limitations. Since the current era of mindset theory suggested that a growth mindset would significantly enhance children's educational attainment across a challenging transition or would be beneficial for children who encountered the greatest challenge, I tested these predictions in two empirical studies. Again, I failed to find any significant effect of growth mindset on educational attainment, either across a challenging transition (from high school to university), in any subsequent year in university, nor among children who encountered the greatest challenge when entering university. My final exploration on mindset was about the development of children's mindsets. Haimovitz and Dweck (2016) suggested that children's mindsets were adopted from their parents' attitudes towards failure (failure mindsets) rather than beliefs about intelligence (mindsets). I tested this prediction in two studies and did not find consistent results. These results lead to some initial work on a possible direction for research in this area, using what I came to believe may be the active agent in learning - namely work and persistence, rather than beliefs about the nature of the mind. In other words, it may be more valuable to test attitudes to hard work in further studies, rather than addressing mindset.

Next, I explored an additional theory that was proposed in a very recent paper (Bian et al., 2017) that the origins of negative stereotypes emerged early in childhood. These negative stereotypes would reduce children's interests in science-related subjects and impair their educational attainment. Having demonstrated whether or not the mind can grow, I turned to whether abandoning or eliminating gender stereotypes may be suitable for children to attempt. I tested children's gender stereotypes about brilliance by asking them to identify one person's gender in a story and found mixed evidence in five empirical studies and a cultural difference among children in China and the U.K. Namely, I failed to find the existence of a gender stereotype that brilliance was a male-trait among Chinese children, but did find it among British children. In addition, I found that children have a gender stereotype favouring female about niceness, and children were more likely to choose a male target as dullness in both China and the UK.

My final exploration was inspired by the mindset studies, testing whether attitudes to hard work (i.e. conscientiousness) would affect children's educational attainment. I tested children's own conscientiousness, their perceptions of teachers' conscientiousness and children's behaviours related to conscientiousness in four empirical studies in primary school and high school. I found that conscientiousness was a significant and consistent predictor of children's educational attainment in both primary school and high school. However, children's perceptions of their teachers' conscientiousness were not significantly associated with children's educational attainment.

In the final chapter, I attempted to bring all these findings together and highlight the need for new directions in further studies.

Contents

List of figures

List of tables

Chapter 1. Is playing number board games beneficial for children's mathematical ability?

At the beginning of undertaking my thesis, I viewed the theme of the thesis as focussing on approaches to helping young children to understand concepts, beginning with acquisition of the symbolic meaning of numbers. This idea was inspired by a study by Ramani and Siegler (2008), that tested whether playing a board game (e.g. snakes and ladders), in which their "player" was identified with a number, would improve children's understanding of numbers and their performance in numerical tasks.

In this chapter, I introduce how mathematical ability is defined in this thesis, discuss the cultural differences in mathematical ability between children in China and Western countries, summarise the evidence for the benefits of playing number board games on children's learning, and report an empirical study testing the effects of playing a number board game on Chinese children's mathematical ability.

1.1 The definition of mathematical ability

Children's mathematical ability at the early developmental stage could be represented by different numerical capabilities (Jordan, Kaplan, Ramineni, & Locuniak, 2009; Siegler & Booth, 2004; Whyte & Bull, 2008). Siegler and Booth (2004) suggested that one of the important measurements for testing children's mathematical ability at the early developmental stage is children's capability of estimating numbers' linear locations. Jordan et al. (2009) suggested that "*number competencies*", which indicate children's understandings of the meaning of numbers and the relationships between numbers, would be one of the important capabilities. Jordan et al. (2009) also indicated that number competencies include children's capabilities of understanding small quantities' values instantly, judging the magnitudes of numbers, comprehending counting principles, and adding and separating simple number sets (Jordan et al., 2009). Whyte and Bull (2008) suggested that "*number sense*", which refers to "*the ability to represent and manipulate numbers nonverbally*" (p. 588), could represent children's early mathematical ability. Whyte and Bull (2008) also indicated that number sense covers children's capabilities of enumerating small number sets rapidly and accurately, comparing the magnitudes of numbers, estimating, and placing number sets linearly in the mind ,based on the magnitude of each number (Whyte & Bull, 2008).

There are overlapping elements in each of the definitions of mathematical ability (e.g. counting abilities, comparison of magnitudes, and linear location estimations), but some definitions have also missed important elements. For example, the capability of number identification - which could be seen as the most basic capability in children's mathematical ability at the early developmental stage - was missing in the definition of number competencies by Jordan et al. (2009). Although the capability of number identification was included in the definition of number sense, the definition proposed by Whyte and Bull (2008) omitted the capabilities of addition and subtraction for simple number sets. Thus, none of the stated definitions was comprehensive. The definition of children's mathematical ability at an early developmental stage used in this thesis synthesises the stated definitions and includes the numerical capabilities of naming numbers (by seeing symbolic Arabic digits), counting, comparing the magnitudes of numbers, estimating numbers' locations on a line, and adding and subtracting a small number set (from 1 to 10).

1.2 Cultural differences in children's mathematical ability at the early developmental stage

The superiority of Chinese children's mathematical ability has been reported widely (Aunio, Aubrey, Godfrey, Pan, & Liu, 2008; Geary, Fan, & Bow-Thomas, 1992; Ryoo et al., 2014; Siegler & Mu, 2008; Stevenson et al., 1990; Stevenson, Lee, & Stigler, 1986). Since children's mathematical ability could be represented by various kinds of numerical capabilities (e.g., counting and estimations), however, some studies have suggested that Chinese children might not be superior at all kinds of numerical capabilities (Muldoon, Simms, Towse, Menzies, & Yue, 2011; Stevenson et al., 1990; J. Wang & Lin, 2009). For example, Stevenson et al. (1990) used a set of nine numerical tasks (including word problems, number concepts, mathematical operations, measurement and scaling, graphs and tables, spatial relations, visualisation, estimation, and speed tasks) to test whether Chinese and American children's (in the 1^{st} and 5^{th} grades in primary school) mathematical ability differed. They found that in the 1^{st} grade, Chinese children performed significantly better on all of the nine tasks compared to American children. However, in the 5^{th} grade, Chinese children were consistently significantly superior to American children on seven of the nine tasks, but were not superior at the tasks involving graphs and tables, and visualisation (Stevenson et al., 1990).

In a meta-analysis, J. Wang and Lin (2009) reviewed 16 studies to compare the differences in mathematical achievement among children in China and the U.S. Children in the reviewed studies were at different educational levels including pre-kindergarten, kindergarten, primary school (1^{st}, 3^{rd}, 4^{th} 5^{th} grades), secondary school (7^{th} and 8^{th} grades) and high school. J. Wang and Lin (2009) found that in terms of their overall mathematical performance, Chinese children were significantly outperformed by their counterparts in the U.S.. Turning to the specific numerical capabilities, Chinese children were superior to their counterparts in the U.S. in most of the numerical capabilities (e.g., geometry and functions), but not in the capabilities of computation, reasoning and estimation (J. Wang & Lin, 2009).

Another study that also provided evidence to support that Chinese children could be inferior in some of the numerical capabilities was conducted by Muldoon et al. (2011). Muldoon et al. (2011) tested whether children in the U.K. would differ from children in China when using the numerical capability of number line estimation.

Muldoon et al. (2011) asked children to estimate the linear locations of three number sets (i.e., 0 to 10, 0 to 20, 0 to 100), and used the accuracy of estimation and the slope of estimates (whether the slope was close to 1 or not) as two standards for measuring children's capability of number line estimation. Muldoon et al. (2011) found that British children showed both significantly higher estimation accuracy, and were closer (to 1) slope of their estimations on two of the number sets (i.e., 0 to 10 and 0 to 20) than the Chinese children.

Although the studies of Stevenson et al. (1990), J. Wang and Lin (2009) and Muldoon et al. (2011) all supported the notion that Chinese children were not superior on all of the numerical capabilities, each of the studies had some limitations. Specifically, Stevenson et al. (1990) collected data from two metropolitan cities which might not be representative of the general population in each country. Socioeconomic status (SES) is an important influence on academic performance and children from low SES backgrounds achieve significantly lower performance than children from middle or high SES backgrounds (Hojnoski, Caskie, & Miller Young, 2017; National Mathematics Advisory Panel, 2008; Sirin, 2016). Studies containing participants only from middle or high SES backgrounds are not representative and generalisable. A similar non-representative problem occurred in the study of J. Wang and Lin (2009), whose data was mostly collected from Hong Kong, Taiwan and Macao. Since these areas have different educational systems from those of mainland China (Koo, Kam, &

Choi, 2003), children in mainland China might perform differently when performing the same tasks and thus, these studies could not represent children in mainland China. Additionally, there was only one dataset - J. Wang and Lin (2009) - that examined the mathematical ability of children in kindergartens, which was not sufficient to make a robust conclusion for children's mathematical ability in kindergartens. In the study of Muldoon et al. (2011), there was one limitation that the Chinese children were 10 months younger than the British children. Since cognitive ability is significantly associated with mathematical achievement (Geary, 1993), and age is an important factor in the development of cognitive ability (Welsh, Nix, Blair, Bierman, & Nelson, 2010), it might be that Chinese children's cognitive ability was not as developed as the British children in the study of Muldoon et al. (2011) and thus, their mathematical achievements were different. Even though Muldoon et al. (2011) analysed children whose scores in a standard test (the British Ability Scales) were matched to those British children, and found the Chinese children's capability of number line estimation was still lower than that of the British children, the powerful effects of cognitive ability should not be neglected.

1.3 The effects of playing number board games on children's mathematical ability

Before having formal mathematical classes at school, children's mathematical ability is already at different levels (Klibanoff, Levine, Huttenlocher, Vasilyeva, &

Hedges, 2006; Ramani & Siegler, 2011). These differences occur even before kindergarten: some children can count from one to ten accurately, or even do some simple sums like one plus one; other children cannot (Klibanoff et al., 2006). Children's mathematical ability at the early developmental stage has been reported as having enduring effects on children's learning in mathematics (Duncan et al., 2007; Muldoon et al., 2011; Ostergren & Traff, 2013; Stevenson et al., 1986). Researchers found that children's mathematical ability at an early developmental stage is significantly associated with children's performance in mathematical tests in primary school or even high school (Duncan et al., 2007; Muldoon et al., 2011; Ostergren & Traff, 2013). Therefore, one way to help children who are disadvantaged in mathematics, and also for vulnerable children (e.g., from low SES families), would be increasing those children's mathematical ability at the early developmental stage.

Playing number board games was reported as an efficient approach to increasing children's mathematical ability at the early developmental stage (Ramani & Siegler, 2008): specifically a number board game called "The Great Race", which is similar to "snakes and ladders". The board includes 10 horizontally arranged squares, and each of the squares was of equal size but in different colours. From the left to the right of the board, the numbers 1 to 10 were written consecutively in each of the squares. At the two ends of the board, the words "Start" (left) and "End" (right) were written. Children were firstly asked to choose from a rabbit and a bear token as the marker of their progress on the board. Then the children were asked to spin a spinner with 1 and 2 on either side to decide how far to move. To compare with the effects of playing the number board game on children's mathematical ability, Ramani and Siegler (2008) also used a colour board game in their study. The board used in the colour board game was identical to the board used in the number board game, with the exception that it lacked any numbers on the squares. While moving the token on the board, children were asked to read loudly the numbers in each square they passed. Ramani and Siegler (2008) used four numerical tasks including number identification, numerical magnitude comparison, counting and number line estimation as measurements to test whether playing the number board game would increase children's mathematical ability compared to playing the colour board game.

After playing the two board games for 2 weeks respectively (4 sessions in total and each session took 15-20 minutes), Ramani and Siegler (2008) found that children who played the number board game showed significantly better performance on all of the 4 numerical tasks than children who played the colour board game. These significant differences were consistent 9 weeks later in the follow-up session (Ramani & Siegler, 2008). After the publication of Ramani and Siegler (2008), more supportive evidence for the benefits of playing number board games on children's mathematical ability were found, both in their own studies (Laski & Siegler, 2014; Ramani & Siegler, 2011; Ramani, Siegler, & Hitti, 2012; Siegler & Ramani, 2009), and in others (S. K. Cheung & McBride, 2016; Whyte & Bull, 2008).

Ramani and Siegler (2008) explained the mechanism of playing number board games on increasing children's mathematical ability. Specifically, the effects of playing number board games on enhancing children's capabilities of number identification and counting are explicit: all of the squares on the number board had numbers in sequence. Through playing the number board game repeatedly, where children were asked to read the numbers loudly when moving their tokens, their impressions of the name of each number and the sequence of those numbers was strengthened, and their capabilities in number identification and counting were enhanced. Children's ability to compare numerical magnitudes could be increased through playing number board games because children learned implicit clues of addition when moving their token on the board. In other words, the number of the target square children aimed to move to was based on the addition of the number on the spinner and the number on their starting square. For example, if a child was starting from the square with number 3, and the child turned 2 on the spinner, then their target square would be number 5. Children would have an impression that the number in every target square was bigger than the number in the starting square. For the effectiveness of playing number board games on children's capability of number line estimation, children could receive both visual and spatial clues when moving their tokens on the board. The numbers 1 to 10 were listed linearly on the board so children could easily observe the location of each number in a line and also get a clue that the distance between numbers 1 - 2 is shorter than the distance between 1–8. If children wanted to move their token to the number 8, they would need to turn the spinner more times than to move to the number 2.

1.4 Aims of my study

Although there are studies reporting that playing number board games is beneficial for children's mathematical ability at an early stage of development, most of the studies were conducted in Western countries (e.g. the U.S.), and only one study was conducted in Hong Kong. As stated above, Chinese children are superior to their counterparts in Western countries in terms of their overall mathematical ability, however, it is not clear whether playing number board games would have the same benefits for children in China, who are reported to have more advanced mathematical abilities than children in Western countries. My study aims to replicate the study of Ramani and Siegler (2008) to test if playing number board games would be beneficial for Chinese children. Since Chinese children might be worse at some specific numerical capabilities (especially on the number line estimation), a more specific aim of my study is to test whether playing number board games could help Chinese children's ability to estimate on a line.

1.5 An empirical study testing whether playing a board game would increase children's mathematical ability in China

Based on Ramani and Siegler (2008), there are four hypotheses in my study: 1) Children who play the number board game would show a larger improvement in the post-manipulation counting task compared to those who play the colour board game; 2) Children who play the number board game would show a larger improvement in the post-manipulation numerical magnitude comparison task compared to those who play the colour board game; 3) Children who play the number board game would show a larger improvement in the post-manipulation number identification task compared to those who play the colour board game; 4) Children who play the number board game would show higher accuracy in the post-manipulation number line estimation task compared to those who play the colour board game.

1.5.1 Method

1.5.1.1 Participants

In total, 50 children were recruited from three kindergartens in a rural area in Harbin, Heilongjiang Province, China. All of the three kindergartens were public and drawn from a catchment area where the average income is 21% below the Chinese national average income (average income 48,881 Yuan: National Bureau of Statistics of the People's Republic of China, 2017), equating to USD 7,133 (~$14,000 purchasing-power equivalent). average income 48,881 Yuan:

Children were aged between 2 years 10 months and 4 years 9 months (mean = 3 years 6 months, SD = 0.45). Thirteen children did not complete the post-manipulation numerical tasks either because they withdrew from the experiment after taking the first experimental session (N = 11) or they were absent when the post-manipulation tasks were taken (N = 2). In total, 37 children completed both the pre- and post-manipulation numerical tasks (22 children in the number board game condition, 15 children in the colour board game condition; 13 boys, 24 girls). In the follow-up session, 22 children were absent. Therefore, 15 children (10 in the number board condition, 5 in the colour board condition; 6 boys, 9 girls) completed the pre-manipulation, post-manipulation and the follow-up numerical tasks. Due to inappropriate data management, data in the follow-up sessions was missing.

1.5.1.2 Materials

Board games

Two board games were used in my study: the number board game and the colour board game. The board used in the number board game includes ten horizontally arranged squares and the numbers one to ten were written in sequence on the squares (Ramani & Siegler, 2008). The board used in the colour board game was identical to the number board, with the exception that the squares did not have numbers on them (also see Figure 1.1 for the boards used in each game; Ramani & Siegler, 2008).

Numerical tasks

As in Ramani and Siegler (2008), four numerical tasks were used to measure the children's mathematical ability (identical tasks were used in the first, fourth and fifth sessions): counting, numerical magnitude comparison, number identification, and number line estimation.

Figure 1.1 The boards used in the number board game (upper) and the colour board game (bottom) in my study

Counting: The counting task aimed to test whether children could count from 1 to 10 accurately. The number before the first error occurred was recorded as the score in the counting task (e.g. if a child counted “1, 2, 3, 4, 6, 7, 8”, their counting score would be 4).

Numerical magnitude comparison task: A booklet including 20 pages of number-pairs was shown to the children. Children were asked to choose the bigger number between each number-pair. The first two pairs in the booklet were used as two examples to explain how to do the task. For these two examples children also received feedback about the accuracy of their answers. If children provided an incorrect answer in the examples, they were asked to do the example questions repeatedly until they gave the correct answer. After the examples, children were asked to work on the remaining 18 number-pairs in the booklet, and this time they did not receive any feedback. The number-pairs were counterbalanced; namely, half of the participants received each number-pair in one order (e.g. 5 vs 9) and the other half received each number-pair in the opposite order (e.g., 9 vs 5). The number of number-pairs that the children answered correctly was recorded as the score in this task.

Number identification task: This task contained 10 cards, with each card having a random number from 1 to 10. Children were asked to identify each number shown on the card. The total number of numbers that children successfully identified was recorded as the score in this task.

Number line estimation task: The children were given 18 sheets of paper in this task (one at a time). On each paper, there was a 25 cm line, with number “0” written below the left end and number “10” written below the right end of the line. One of the numbers 1 to 9 was written above the centre of the line. Children received the nine papers in a random order, and when all of the nine numbers had been shown once, children received another set of nine papers for a second time. Children were told that they were going to play a game where they needed to help each number to find its location on the line. I firstly asked children to identify the number on the paper and then said: *“If this is where 0 goes (pointing) and this is where 10 goes (pointing), where does N go?*” (Ramani & Siegler, 2008, p. 379).

1.5.1.3 Procedure

This study was approved by the Psychology Research Ethics Committee at the PPLS, University of Edinburgh. After gaining consent from the headmaster, headteachers, parents and children themselves, the children were greeted and given a brief introduction. There were five experimental sessions in my study, with each session taking around 15 to 20 minutes. The first four sessions were taken within a two-week period. Children were asked to play each of the board games with me for 20 times in the first four sessions. After the fourth session, children had a break period of 2, 4, 6 or 9 weeks depending on the recruiting phases. Then children had a fifth session which included the 4 numerical tasks only.

At the beginning of the first session, children were asked to do the 4 numerical tasks in a room near their classroom. Each child was then assigned to either a number board game condition or a colour board game condition. I told children that they were going to play a racing game which started from the left end of the board, and the child who arrived first at the right end of the board would win a reward (a sticker). I then asked children to choose a token of a rabbit or a bear as the mark of their progress on the board. To start the game, children were told that they would need to take turns with me in spinning the spinner and, based on the number they received on the spinner, the children could move their tokens that number of spaces forward on the board. When moving, the children were asked to read loudly the number on the squares they had passed. For example, if children spun to “1” in the first turn, they needed to move their token to the square with number “1” and read “one” loudly. In the next turn, if children spun to “2”, they needed to move their token to the square with number “3” and read “two” and “three” (the passed squares). Children in the colour board game condition were also told that they needed to spin a spinner in turns with me. The spinner used in this condition had only colours that were matched with the colours of the squares on the board. I told children that they need to move their token to the nearest square that had the same colour indicated on the spinner. Similarly to the number board game condition, the children were asked to read the colour of the squares they passed when moving their token on the board.

In the second and third sessions, children in each condition played the same board game as they did in the first session. In the fourth session, children were firstly asked to play the same board game as they did in the first three sessions and then were given the identical four numerical tasks as in the first session. In the fifth session, I asked children to do the identical numerical tasks as in the previous sessions and then told children that they had reached the end of the experiment and thanked them for their participation.

1.5.2 Results

I first tested hypothesis one: that children who played the number board game would show a larger improvement in the post-manipulation counting task in comparison to the children who played the colour board game. A linear regression was used to test this hypothesis, with children's counting scores in the post-manipulation test as the dependent variable and the board game condition as the independent variable, controlling for children's counting scores in the pre-manipulation test (see Table 1.1 for the descriptive statistics). Contrary to prediction, children in the number board game condition did not perform significantly better in the post-manipulation counting test compared to those in the colour board game condition ($t = -1.57, p = .125$). Interestingly, the estimate was in the reverse direction to the expectation (β = -0.48, CI_{95} [-1.11, 0.14]).

Next, I tested hypothesis two that children who played the number board game would show a larger improvement in the post-manipulation number comparison task compared to those who played the colour board game. Again, this hypothesis was tested using a linear regression, with children's comparison scores in

Table 1.1 Means and standard deviations (in brackets) of children's scores in the counting task, their accurate percentage in the numerical magnitude comparison task, their scores in the number identification task, and their percent absolute error in the number line estimation task in my study and Ramani and Siegler (2008)

	Ramani and Siegler (2008)				**The present paper**			
	Number board game condition		**Colour board game condition**		**Number board game condition**		**Colour board game condition**	
Test session	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test
Counting	8.70	9.90	8.10	8.40	7.32 (2.92)	7.77 (2.91)	9.13 (2.03)	9.67 (1.05)
Comparison	73%	85%	68%	70%	50% (1.31)	50% (2.01)	55% (1.67)	52% (1.88)
Identification	7.00	8.20	6.10	6.30	4.55 (3.23)	5.09 (3.02)	7.87 (3.00)	7.93 (2.76)
Number line estimation (PAE)	28%	21%	28%	30%	1% (0.17)	3% (0.17)	2% (0.13)	3% (0.12)

the post-manipulation test as the dependent variable, the board game condition as the independent variable, and controlling for children's comparison scores in the pre-manipulation test. Contrary to prediction, children in the number board game condition did not show significantly better performance in the post-manipulation comparison test than children in the colour board game condition (β = 0.18, CI_{95} [-0.43, 0.79], t = 0.60, p = .552).

I next tested hypothesis three, that children who played the number board game would show a larger improvement in the post-manipulation number identification task compared to those who played the colour board game. Again, the result was contrary to the prediction, in that the children in the number board game condition did not perform significantly better in the post-manipulation identification task than the children in the colour board game condition (t = -0.43, p = .671), and the estimates were in the reversed direction (β = -0.08, CI_{95} [-0.48, 0.31]).

Finally, I tested the fourth hypothesis, that children who played the number board game would show higher accuracy in the post-manipulation number line estimation task compared to those who played the colour board game. The accuracy in the number line estimation task was measured by using the percent absolute error (PAE): (Estimate value–actual value)/number line scale. For example, number 3 should be located at 7.5 cm of the line (25/3 cm), and if a child marked number 3 at 12cm of the line, the percentage of absolute error would be (7.5-12)/25 = 18%. Contrary to prediction, children's accuracy in the number line estimation task did not significantly differ between the number board game condition and the colour board game condition in the post-manipulation task (β = -0.02, CI_{95} [-0.61, 0.56], t = -0.08, p = .935; also see Figure 1.2).

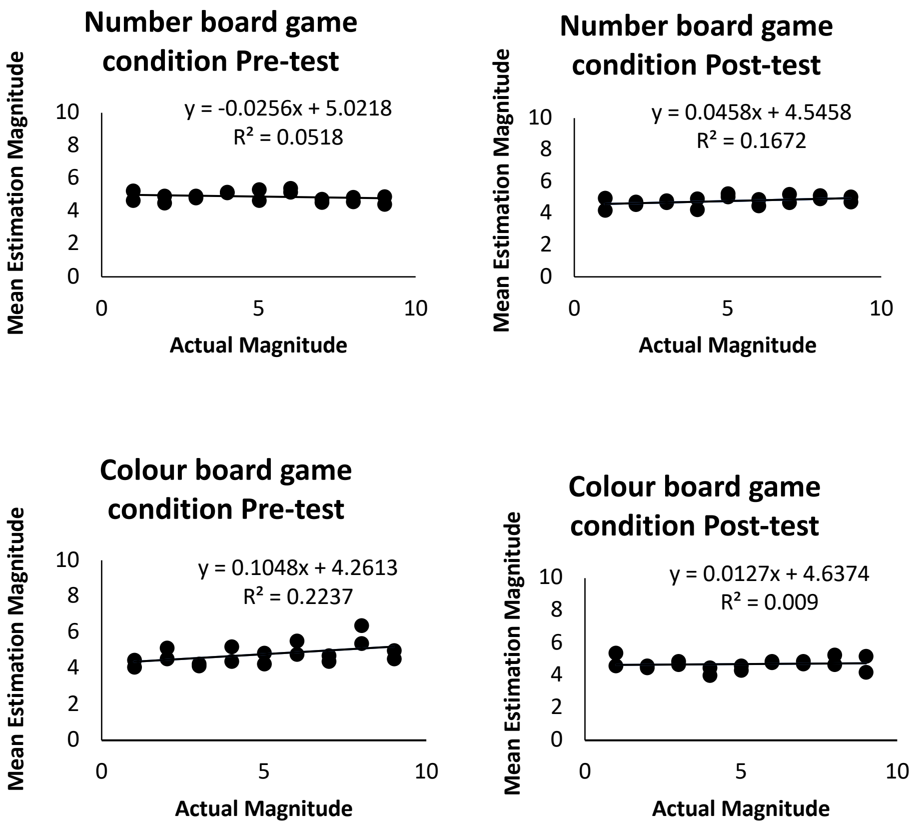

Figure 1.2 The linear models for children's number line estimates in the pre- and post-test

1.6 Discussion

Ramani and Siegler (2008) reported that playing number board games significantly increased children's numerical capabilities in counting, comparison, identification, and number line estimation. I tried to replicate their study in a Chinese sample (N = 37) of 3 years old children but did not find any significant effect. In specific, for children's capabilities in counting and identification, there was no significant improvement in the number board game condition when compared to the colour board game condition. More strikingly, children in the number board game condition performed even worse than those in the colour board game condition in these two tasks in the post-manipulation test. For the capability in comparison, the number board game did slightly improve children's capability in the post-manipulation test compared to the colour board game, however the improvement was not significant. For the capability in number line estimation, children who played the number board game did not show a significantly larger improvement than those in the colour board game.

One concern regarding my failure of replication would be that the difficulties in each of the numerical tasks were not appropriate for my participants. In specific, the difficulty in the counting task was too easy for my participants since there was a ceiling effect in the counting task (21 out of 37 children counted numbers 1 to 10 correctly). On the other hand, the magnitude comparison task was difficult for my participants. Children only guessed which one was bigger between the two numbers when they did the comparison task and thus they had a near-chance performance in the task. The number line estimation was also too difficult for my participants since children in the number board game condition got only 1% accuracy and children in the colour board game condition got 2% accuracy. It is possible that my participants (3 years old) were too young to understand the board games and the numerical tasks (consistent with their near-chance performance at post-test on two of the four tasks, with number line estimation task performance near zero).

Regarding the cultural differences in mathematical ability among my study and the original study of Ramani and Siegler (2008), my Chinese participants' capabilities were similar in counting (8.4 versus 8.2), lower in comparison (70.5% versus 52.5%), similar in identification (6.55 versus 6.21), and drastically lower in number line estimation (28% versus 1.5%). These results confirmed the previous finding that Chinese children are not better than their counterparts in Western countries at the number line estimation task (Muldoon et al., 2011; J. Wang & Lin, 2009). Since children in my study (3 years 6 months) were 15 months younger than children (4 years 9 months) in the original study of Ramani and Siegler (2008), and age is associated with the development of cognitive ability, and cognitive ability is a significant influence on mathematical performance (Deary, Strand, Smith, & Fernandes, 2007; Geary, 1993), it might be that my Chinese participants' cognitive ability was not as developed as the ability of the children in the original study, and this led to the differences in children's performance in those numerical tasks.

1.6.1 Limitations and further directions

One major limitation in my study was that participants were not randomly assigned to the two board game conditions. This led to a significant difference on children's pre-manipulation counting task and caused the negative estimate when compared the counting performance of children in the number board game condition to those in the colour board game condition in the post-manipulation task. Participants in the further studies should be randomly assigned into the number and colour board conditions. Another limitation was that my study has a small sample size (N = 37) which might explain those non-significant results. One suggestion for further studies would be recruiting more participants to increase the power to detect the effect of playing number board games on children's mathematical performance. The third limitation in my study was that individual children's SES was not measured. Although SES has been found to be an important factor in children's learning (Hojnoski et al., 2017; National Mathematics Advisory Panel, 2008; Sirin, 2016), missing individual children's SES data might not make a difference in these results since the children were recruited from a rural district of an area that was in relative poverty.

As stated, the expected effects of number board games on children's learning was not supported. The outcome of my single replication with N = 37 leaves many possibilities open: I may not have used the manipulation as designed, leading to it not working (though I have no reason to think this is the case), the result itself may not be replicable (consistent with reports from Ramani on other attempts to replicate the result in China), the theory that playing number board game would improve children's understanding of numbers and increase their numerical performance may not be correct. Understanding what has happened would require a major investment of resources in this method.

What was most impressive to me, however, was the idea that published results need to be checked, that replications need to be well powered, and that not all published finding may turn out to be valid. After communicating with the original authors, I decided not to replicate this board game experiment again, as the authors say that they have run it in China, but they could not find the same results as they did originally in the U.S., and the original authors did not expect it would work. So, I explored a different direction. I sought out another instance, more well established and widely relied upon as a tool to improve children's learning, which I could myself test using robust and well-powered methods. One theory which has such impact, which even a glancing review of the literature on psychological theory regarding children's learning will highlight, is mindset theory (Dweck, 2006). In the next chapter of this thesis, I will introduce the growing influence of mindset theory on children's learning process.

Chapter 2. Can growth mindset improve children's educational attainment? – An introduction to mindset theory

The previous chapter introduced the theoretical motivation behind the idea that playing number board games may help children learn numerical knowledge, and how I tested a replication of this idea. In this chapter, I will introduce an alternative theory that may help children's learning: mindset theory.

Mindset theory is a prominent and influential theory of children's learning which proposes that people's beliefs about the malleability of intelligence affect their learning outcomes (Dweck, 2006). In this chapter, I introduce the definition of mindset, and how mindset is proposed to affect children's learning. I will then cover three key areas: (1) describing the two different types of mindset concerning intelligence which are believed to be important for learning, (2) describing the most commonly used mindset scale and (3), describing two influential mindset manipulations. I conclude the chapter by proposing that independent replications of two key mindset findings, namely the effect of a manipulation of current activated beliefs on performance, and an association of mindset with grades in school - will be important to establish for research in this field, and I describe how I set about accomplishing these two goals.

2.1 What is mindset?

2.1.1 The historical origins of mindset

Mindset was originally presented as a social-cognitive model of motivational process to explain why people with equal ability performed differently in cognitive tasks (Dweck, 1986; Dweck & Bempechat, 1983). Through testing children's performance on various cognitive tasks (e.g. learning new materials or performing demonstrated skills) in both real-life (classroom) and laboratory settings, Dweck and

Bempechat (1983) observed that children behaved differently and had different achievements in performing the tasks, especially in the difficult tasks. Specifically, some children achieved low scores on difficult tasks since they were not able to persist on those tasks, not even on the items they had previously solved easily. By contrast, other children achieved high scores on difficult tasks because they were concentrating and kept working on those tasks. Dweck and Bempechat (1983) explained that the differences in children's behaviours and achievements on the difficult tasks were caused by different motivational patterns.

There are two motivational patterns in children's learning: The *helpless* (or *maladaptive*) pattern and the *mastery-oriented* (or *adaptive*) pattern (Dweck, 1986; Dweck & Bempechat, 1983). The helpless motivational pattern originates from the evaluations of innate ability and emphasises the *stability* of competence. With this motivational pattern, children view difficulty or failure as the indicator of their insufficient innate ability or permanent incompetence. To avoid having negative evaluations for their innate ability, children would avoid having challenging tasks and would be less likely to persist on difficult tasks (Dweck, 1986). In contrast, the mastery-oriented motivational pattern is formed based on the possession of innate ability and emphasises the *acquisition* of competence. With this motivational pattern, children would be more interested in intensifying their effort, or improving their problem-solving strategies to overcome difficulties. Moreover, children with the mastery-oriented motivational pattern would be less likely to seek excuses for their failures or to consider themselves as failures if their achievements were not as expected (Dweck & Bempechat, 1983). Consequently, children with this motivational pattern would be more likely to persist on difficult tasks to improve their strategies or to make more effort, and would be find more enjoyment in their problem-solving process compared to children with the helpless motivational pattern (Dweck, 1986). These definitions indicated that the base for forming the two motivational patterns is the conception of competence (or innate ability). Dweck (2000) defined people's conception of their innate intellectual ability as *mindset*.

2.1.2 The definition of mindset

Mindset (see Table 2.1 for the different names) has been discussed widely in the last two decades, which could largely be attributed to the efforts of Carol Dweck and her research group. This group has worked hard to explore how mindset is associated with sports performance (Atwood, 2010; Dweck, 2006; Ommundsen, 2003), leadership (Burnette, Pollack, & Hoyt, 2010; Chase, 2010), relationships (Chen, DeWall, Poon, & Chen, 2012; Knee & Petty, 2013), motivation (Bedford, 2017; Caluori, 2014; O'Neill, 2011), achievement in business (Johnston, 2017), and perhaps most centrally, learning outcomes (Blackwell et al., 2007; Claro, Paunesku, & Dweck, 2016; Dweck, 2006; Gunderson et al., 2013; Paunesku et al., 2015; Yeager et al., 2019; Yeager et al., 2016).

Dweck (2012, p. 615) defined mindset as "*people's lay beliefs about the nature of human attributes, such as intelligence or personality*". People could either believe that the nature of their attributes is fixed or malleable (Dweck, 2000). Those who believe their attributes to be fixed no matter what they do are said to have a fixed mindset; those who believe their attributes to be malleable are said to have a growth mindset (Dweck, 2006). Since these fixed and malleable beliefs are utilisable for individual attributes, people would have different mindsets about different attributes – e.g. intelligence mindset or personality mindset. As intelligence is considered an important predictor for children's learning (e.g., Deary, Strand, et al., 2007; Furnham, Chamorro-Premuzic, & McDougall, 2003; Kuncel, Hezlett, & Ones, 2004; Ritchie, Bates, & Deary, 2015) and the main aim of this thesis is finding approaches to help children to learn, mindset in this thesis refers to the intelligence mindset in specific. Additionally, by adopting the word 'mindset' as one of the key themes in this thesis, I use fixed mindset and growth mindset as the two mindset types throughout the whole thesis.

Table 2.1 The names, types and definitions of mindset, and details on usage in subsequent studies

Names	**Types**	**Definitions**	**Details on usage**
Mindset	Fixed mindset VS Growth mindset	*"People's lay beliefs about the nature of human attributes, such as intelligence or personality"* (Dweck, 2012, p. 615)	Proposed in Dweck (2012), used in Bahník and Vranka (2017); Claro et al. (2016); Dweck (2007b, 2014); Paunesku et al. (2015); Rattan, Savani, Chugh, and Dweck (2015); Zeng, Hou, and Peng (2016)
Implicit theories of intelligence	Entity theory VS incremental theory	"*Implicit conception about the nature of ability*" (Dweck & Leggett, 1988)	Proposed in Dweck and Leggett (1988), used in Abd-El-Fattah and Yates (2006); Aronson, Fried, and Good (2002); Blackwell et al. (2007); Cury, Elliot, Da Fonseca, and Moller (2006); Hong, Chiu, Dweck, Lin, and Wan (1999); Robins and Pals (2002); Romero, Master, Paunesku, Dweck, and Gross (2014); Yeager and Dweck (2012)
Theories of intelligence	Entity theory VS incremental theory	"*Different beliefs about the nature of intelligence*" (Dweck & Bempechat, 1983, p. 243)	Proposed in Dweck and Bempechat (1983), used in Zhao and Wang (2014)

2.1.3 The relationship between achievement goal and mindset

2.1.3.1 The conceptualisation of goal

Dweck and Leggett (1988) suggested that "*the goals individuals are pursuing create the framework within which they interpret and react to events*". There are two types of goals in intellectual achievement: the performance goal and the learning goal (Dweck & Leggett, 1988). People who pursue a performance goal would be more likely to be concerned with their competence, and with the favourable judgements about their competence from other people (Dweck & Leggett, 1988; Elliott & Dweck, 1988). Therefore, these people would prefer to do tasks in which they would be likely to succeed in to gain other people's positive judgements, and avoid tasks in which they would be likely to fail and therefore receive negative judgements (Elliott & Dweck, 1988). Alternatively, people who pursue a learning goal are not concerned about their competence. Instead, they are more likely to focus on the knowledge or skills they could learn or develop during the problem-solving process, the potential improvement they could make after changing their inappropriate learning strategies, and finding the most appropriate approach to solve new problems or tasks (Blackwell et al., 2007; De Castella & Byrne, 2015; Dweck & Leggett, 1988; Yeager & Dweck, 2012).

2.1.3.2 Different mindsets oriented different achievement goals

One prediction associating mindset to achievement goals is that people with a fixed mindset would pursue a performance goal and people with a growth mindset would pursue a learning goal (Dweck & Bempechat, 1983). People with a fixed mindset believe that their intelligence cannot be changed and their performance is positively associated with their intelligence level (Dweck, 2000). Therefore, the amount of effort they put into the tasks would not matter because their performance would remain at the same level as their intelligence (Dweck, 2000). In other words, superior performance means people are highly intelligent and unsatisfactory performance means people are not intelligent (Hong et al., 1999). To make themselves look intelligent, demonstrate their intelligence to other people, and gain positive judgements from other people, these people would pursue the goal of achieving superior performance rather than developing learning strategies (Hong et al., 1999). The most straightforward way to achieve superior performance is to work on easy rather than challenging tasks (Blackwell et al., 2007; De Castella & Byrne, 2015; Yeager & Dweck, 2012).

Contrarily, people with a growth mindset believe that their performance is malleable and is positively associated with the amount of effort they put into the tasks (Hong et al., 1999). If people with a growth mindset want to improve their performance, the most straightforward way is to put in more effort (Hong et al., 1999). People with a growth mindset would also not be concerned about their competence or performance, but would focus on whether the learning strategies they used are appropriate, or the amount of effort they have made for completing the tasks is sufficient (Blackwell et al., 2007). Through changing inappropriate learning strategies or spending more time on practising, people with a growth mindset could make improvements on their performance, gain more knowledge, and learn more skills (Blackwell et al., 2007; Cury et al., 2006; Hong et al., 1999).

2.1.4 The relationship between mindset, attitude to failure, attributional style and resilience to failure

People with a fixed mindset, in comparison to those with a growth mindset, would show more negative attitudes to failure, attribute their failures to innate factors such as intelligence, and have relatively low resilience to failure (Dweck, 2006). Specifically, for people with a fixed mindset, experiencing failures means that they have a low intelligence level, their potential is limited, or their future performance would be low (Haimovitz & Dweck, 2016; Yeager & Dweck, 2012). These negative beliefs lead people with a fixed mindset to worry about failure and to refuse to experience failures whenever possible (Aronson et al., 2002; Dweck, 2007a; Yeager & Dweck, 2012). People with a fixed mindset would also have a low resilience to failure, such as giving up immediately after experiencing frustration or failure, to avoid having more negative judgements about their intelligence from other people (Dweck, 2006, 2014; Hong et al., 1999). When attributing failure, people with a fixed mindset would be more likely to attribute failure to lack of ability rather than lack of effort (Hong et al., 1999).

On the other hand, people with a growth mindset are more likely to view failure in a positive way because they believe that failure could enhance their learning strategies, promote their skill acquisition, and indicate the ability-growing process that they need to revise (Aronson et al., 2002; Dweck, 2006; Haimovitz & Dweck, 2016). Therefore, failure is a part of learning, one which could not represent their potential ability nor predict their future ability (Dweck, 2006; Yeager & Dweck, 2012). When experiencing failure, these people would show a high level of resilience such as taking immediate remedial action to resolve difficult tasks or overcome setbacks (Hong et al., 1999; Yeager & Dweck, 2012). For attributions of failure, people with a growth mindset are more likely to attribute failures to lack of effort or to the use of inappropriate learning strategy rather than intelligence (Blackwell et al., 2007; Hong et al., 1999; Paunesku et al., 2015; Rattan, Good, & Dweck, 2012; Yeager & Dweck, 2012).

In conclusion, people with a fixed mindset believe that their intelligence is fixed no matter what they do, are concerned about their performance rather than learning strategy, are more likely to give up when experiencing failure or having difficulties, view failure as debilitating, and attribute their failures to lack of ability. On the other hand, people with a growth mindset believe that their intelligence is malleable, are more likely to pursue a learning goal rather than a performance goal, are more likely to persist in difficult tasks, believe failure can enhance their learning, and attribute their failures to lack of effort or to using inappropriate learning strategies.

2.1.5 A brief review of the development of mindset theory

After mindset theory had been proposed, many researchers attempted to investigate the possible effects of mindset on children's learning. Most of the mindset researchers attempted to explore the effects of mindset in two aspects: the effects of children's naturally-held mindset (e.g., Blackwell et al., 2007), and the effects of mindset manipulations (e.g., Yeager et al., 2016). To distinguish the different effects of the naturally held mindset and mindset manipulation in children's learning outcomes, I use "*children's own mindsets*" to describe children's naturally held mindset, and use "*mindset manipulation*" to describe the usage of mindset manipulation in my empirical studies.

Through empirical studies, the majority of mindset studies have reported that having a growth mindset or using a growth mindset manipulation could: improve children's motivation (Haimovitz, Wormington, & Corpus, 2011; Rattan et al., 2015), increase children's resilience to failure (Yeager & Dweck, 2012), result in higher psychological well-being (Zeng et al., 2016), enhance willpower (Dweck, 2012), reduce stereotype threat (Aronson et al., 2002; Good, Aronson, & Inzlicht, 2003), and help children to achieve better learning outcomes (Claro et al., 2016) in comparison to those who have a fixed mindset. Since the main aim of this thesis is finding appropriate approaches to help children to learn, I will focus on the possible effects of mindset on children's learning outcomes, namely, educational attainment.

2.1.6 The theories of intelligence scale

The most commonly used measurements to assess children's own mindsets would be the Theories of Intelligence scale (Dweck, 2000). This scale includes 8 items, and each item introduces an idea about whether people believe their intelligence could be changed (4 items) or is fixed (4 items). Example items include "*You have a certain amount of intelligence, and you can't really do much to change it*" and "*No matter who you are, you can significantly change your intelligence level*".

The Theories of Intelligence scale could be used in different lengths and among children at different educational levels, and each variation of the scale showed good psychometric values. For example, Blackwell et al. (2007) used the 6-item version of the Theories of Intelligence to assess junior high school students' mindsets. The internal reliability of the scale was acceptable ($\alpha = .78$) in their study (Blackwell et al., 2007). Haimovitz et al. (2011) used the 3-item version of the scale, which has good internal reliability ($\alpha = .86$), to assess 6^{th} and 8^{th} grades' mindsets. P. Li et al. (2017) also used a 3-item scale to assess 10^{th} graders' mindsets in a Chinese sample. The internal reliability of the scale in their study was high ($\alpha = .94$; P. Li et al., 2017). Bahník and Vranka (2017) used the 2-item version to test university applicants' mindsets and found that the two items were positively correlated ($r(df = 5651) = 0.52$, CI_{95} [0.50, 0.54], $p < .001$).

2.1.7 The association between children's own mindsets and educational attainment

For researchers who explored the effects of children's own mindsets in educational environments, the majority reported a positive association between growth mindset and high educational achievement (e.g., Blackwell et al., 2007; Claro et al., 2016; Cury et al., 2006; De Castella & Byrne, 2015; Romero et al., 2014). For example, Claro et al. (2016) conducted a national study in Chile, testing whether 10th graders' own mindsets predicted their achievement in standardised tests (including mathematics and language tests). They found a significant effect of growth mindset on the test achievement, in that children with a growth mindset achieved significantly higher grades in both mathematics and language tests than those with a fixed mindset. A similarly positive effect of growth mindset on their school grades was found by Romero et al. (2014). Middle schoolers with a growth mindset achieved higher grades than their counterparts with a fixed mindset. Children's own mindsets have also been demonstrated to affect their achievement in cognitive tests, in that children with a fixed mindset achieved lower scores in cognitive tests compared to those with a growth mindset (Cury, Da Fonseca, Zahn, & Elliot, 2008).

However, some researchers are dubious about the strength of the effects of children's own mindsets on educational achievement, and some have suggested that it might even be harmful to learning. For instance, Dupeyrat and Mariné (2005) tested whether mature students' own mindset scores would be associated with their school grades and found there was no significant association between children's own mindsets and their grades. This null effect of mindset was further supported by other researchers such as Bahník and Vranka (2017), Bazelais et al. (2018), Corradi, Nicolaï, and Levrau (2018) and Sisk, Burgoyne, Sun, Butler, and Macnamara (2018). Specifically, Sisk et al. (2018) meta-analysed the magnitude of the association between students' mindset scores and their achievement. They failed to find consistent effect sizes across 129 studies, but most of the included studies in their meta-analysis showed small or null effects of children's mindset on their achievement. Furthermore, the negative effect of having a growth mindset on student's achievement was found in a recent study (Bahník & Vranka, 2017). Bahník and Vranka (2017) explored the association between university applicants' scholastic aptitude test scores and their mindset scores. They found that growth mindset was negatively associated with students' scholastic aptitude test scores. Namely, students with a fixed mindset had better scores compared to those with a growth mindset, which was contrary to the mindset theory that holding a growth mindset would improve achievement.

2.1.8 The effects of mindset manipulations on children's educational attainment

Apart from the studies which explored the association between children's own mindsets and educational attainment, researchers have also explored the possible effects of mindset manipulations on children's educational attainment (e.g., Burnette, Russell, Hoyt, Orvidas, & Widman, 2017; Paunesku et al., 2015). A majority of researchers have found that inducing a growth mindset manipulation could significantly increase children's educational attainment (Blackwell et al., 2007; Bostwick & Becker-Blease, 2018). This thesis focuses on two mindset manipulations which were commonly used in mindset studies: the praise manipulation for children (e.g., Mueller & Dweck, 1998) and the mind-can-be-grown manipulation for mature students (e.g., Aronson et al., 2002; Blackwell et al., 2007).

2.1.8.1 The praise manipulation

As a specific type of feedback, praise has been seen as an influential factor in the development of children's mindsets (Macnamara & Rupani, 2017). The mechanism behind the association of praise and children's educational attainment is that a specific type of praise provides a specific kind of key information for achieving a goal (Zentall & Morris, 2010); learning goals are associated with different motivational patterns (Dweck & Bempechat, 1983; Elliott & Dweck, 1988); motivations are then associated with different levels of resilience to failure (Kamins & Dweck, 1999); and resilience levels consequently affect children's educational attainment (Chalk & Bizo, 2004).

There are two types of praise: person praise and process praise (Dweck, 2000). The person praise "*involves a global assessment based on a specific behaviour or performance, would teach children to measure themselves by their performance and would thus foster more helpless reactions to setbacks.* " (Kamins & Dweck, 1999, p. 835). Contrarily, the process praise mainly focuses on children's usage of learning strategies or effort rather than their intelligence (Dweck, 2000). In this thesis, I refer to the two types of praise as "*praise for being smart*" and "*praise for hard work*" to make their meaning explicit.

Praise was seen as an effective approach for teachers and parents to create a mindset environment for children (Dweck, 2010). Specifically, children who were praised for being smart were activated with a fixed mindset, whereas children who were praised for hard work were activated with a growth mindset (Gunderson et al., 2013). The development of a fixed mindset was due to praise for being smart, implying that performance is the reflection of intelligence, which was assessed by teachers or parents (Chalk & Bizo, 2004). Thus, children would be more likely to attribute their failures to lack of ability and would be motivated to achieve good performance to prove they have sufficient intelligence (Lam, Yim, & Ng, 2008). Meanwhile, they would give up trying after experiencing failure because failure meant they are not smart enough (Zentall & Morris, 2010). All these beliefs and behaviours might reinforce children's fixed mindsets (Dweck, 2007a).

On the other hand, the praise for hard work approach implies that the key approach to achieving success is working hard (Gunderson et al., 2013). Children who were praised for hard work would be focused on the problem-solving strategies they have used, and the possible improvements they would see after making greater effort (Lam et al., 2008). In addition, praise for hard work would lead children to attribute their successes or failures to lack of effort they had rather than to lack of ability (Gunderson et al., 2013). Because the amount of effort could be changed by children themselves, children who were praised for effort would be more likely to make greater effort after experiencing failures or having challenging tasks rather than giving up and turning to easy tasks (Gunderson et al., 2013). All of these beliefs and behaviours might reinforce children's growth mindsets (Dweck, 2007a).

In empirical studies, researchers have used the praise manipulation to test if it could affect children's development. The results are mixed. Many researchers found benefits of the praise-for-hard-work manipulation on children's development. For example, praise for hard work could increase children's interests in learning and their self-evaluation (Lam et al., 2008), motivate children to be persistent when carrying out challenging tasks (Zentall & Morris, 2010), foster children's growth mindsets (Gunderson et al., 2013), make children enjoy the problem-solving process, and most importantly, improve children's performance after they have experienced failures (Mueller & Dweck, 1998). However, some researchers argued that compared to those who were praised for hard work, children who were praised for being smart showed better performance, had higher self-efficacy, and acquired more skills (Miller, Brickman, & Bolen, 1975). The superiority of praise for being smart might be caused by the low expectations children had for their futures (Schunk, 1983).

Since most of the studies in this field were conducted in Western countries, it is not clear whether the praise manipulation would work in a different country (i.e. in China). People in different countries may hold different attitudes towards praise, and these differences could influence the efficacy of the manipulation. If Chinese people have negative attitudes toward praise and praise is rarely used in their daily life even after they achieved success, the Chinese participants would not believe what the experimenter said during the experiment due to their lack of experience for being praised. In this case, the praise manipulation would not work in China. Therefore, applying the manipulation in different countries could verify if the theory could be applied in a broader population, which would be helpful for the generalisation of the mindset theory. Before testing the possible effects of the praise manipulation on children's educational attainment, it would be reasonable to first describe Chinese children's attitudes towards praise.

2.1.8.2 Chinese people's attitudes towards praise

The usage of praise is rare in comparison to the use of punishment in China (Salili & Hau, 1994). Most Chinese parents and teachers hold traditional attitudes towards praise - that it could lead children to be proud and complacent about themselves - both of which would be harmful to their development (Salili & Hau, 1994). However, praise for children's virtues such as diligence was thought to be beneficial for the development of the traditional Chinese beliefs in children (Salili & Hau, 1994).

Contrary to the traditional attitudes towards praise, Hau and Salili (1996) reported that praising Chinese children for their intelligence, rather than praising children for effort, had positive effects on their test performance. Hau and Salili (1996) explained that for children who were praised for effort would think working hard on tasks is their duty, and praise for doing their duty would not motivate them. However, praise for intelligence would lead children to have high levels of self-evaluation about their capabilities and studying skills (Hau & Salili, 1996). Children would then have higher expectations for their future performance due to the strongly implied information about competence in the intelligence praise (Lam et al., 2008).

Although some praise-related research has been done in the last two decades in China, to my knowledge no independent study has used the praise manipulation to test whether praise for children's hard work would be more beneficial on children's educational attainment than praise for being smart. Moreover, no study has related praise to the mindset theory in Chinese culture to test whether praise for hard work and praise for being smart would lead children to have different mindset orientations. Thus, it would be insightful to conduct a study utilising the praise manipulation to test if it does have a significant effect on children's educational attainment and to test if the praise manipulation is associated with Chinese children's mindset orientations.

2.1.8.3 The different forms of mind-can-be-grown manipulation

Based on the mindset research, Dweck and her research group have designed many manipulations to encourage and promote children to have a growth mindset (Donohoe, Topping, & Hannah, 2012). These mindset manipulations had various forms, for example, Paunesku et al. (2015) used a 'reading and writing' form of the mind-can-be-grown manipulation in their study. Specifically, high school students were asked to "*read an article describing the brain's ability to grow and reorganize itself as a consequence of hard work and good strategies on challenging tasks*" (Paunesku et al., 2015, p. 786). The main focus of the growth mindset article was that everyone has the potential to be more intelligent through studying. Additionally, the article addressed the fact that although students had experienced setbacks in school, this did not mean they lacked potential, but rather that their learning opportunities would be promoted. After reading the article, students were asked to do two writing exercises to reinforce the content they had learnt from the article.

Another example of the mind-can-be-grown manipulation is the computer programme called "Brainology" that was designed by Dweck (2008a). This programme aimed to promote a growth mindset to young children, to instil the idea that people's intelligence could be grown through coping with difficulties in learning, making greater effort, and using effective learning strategies. Children would follow two animated characters in the programme to discover how our brain works, and how to make our brain smarter and stronger. There were four learning modules introducing what the bases of our brain are, how our brain works, how our brain is built, and how to boost our brain. Each module started with a brain question, and children would be asked to do their own research to solve this question. Children using the programme are also encouraged to relate the information that they have learnt in the module to their school problems. At the end of each module, children would be asked to work on brain-related puzzles, and if they succeed in completing these puzzles, they would move to the next module.

Although the mind-can-be-grown manipulation has different forms, all of the forms conveyed the same concept: that people's intelligence is malleable (Blackwell et al., 2007; Burnette et al., 2017; Paunesku et al., 2015). Therefore, all the stated forms are collectively referred to as the mind-can-be-grown manipulation in this thesis.

2.1.8.4 The association between the mind-can-be-grown manipulation and children's educational attainment

Turning to the existing studies that explored the effects of the mind-can-be-grown manipulation on children's educational attainment, most findings focused on two aspects: the effects on children's mindset orientation, and the consequential effects on children's educational attainment (Blackwell et al., 2007; Paunesku et al., 2015). For example, a positive association between the mind-can-be-grown manipulation and children's growth mindset was reported by DeBacker et al. (2018). In addition, the mind-can-be-grown manipulation consequently increased children's educational attainment (Andersen & Nielsen, 2016; Bostwick & Becker-Blease, 2018; Paunesku et al., 2015). This positive association between the mind-can-be-grown manipulation and children's educational attainment has also been verified among students who are at high risk of dropping out (Paunesku et al., 2015). Specifically, Paunesku et al. (2015) gave students a practical growth mindset manipulation (as noted above), and asked their schools to provide their official grades. They found that the mind-can-be-grown manipulation significantly increased one-third of the high-risk students' grades by 6.4 percentage points.

Similar to the praise manipulation, those studies which used the mind-can-be-grown manipulation also had a limitation concerning generalisability, in that the majority of the studies were conducted in Western countries such as the U.S. (e.g., Blackwell et al., 2007) or within Europe (e.g., Andersen & Nielsen, 2016). It is not clear whether the mind-can-be-grown manipulation would work in Eastern countries, such as China, since these countries differ with regards to culture, educational system and attitudes towards learning. In addition, a study by Burgoyne, Hambrick, and Macnamara (2020) suggested that the foundations of mindset theory may not be as firm as it proposed. Testing the theory in other countries could thus, provide further evidence regarding the accuracy and firmness of mindset theory. Thus, I reviewed literature on research that was conducted in China, to see if the mindset paradigm has been shown to work in a different country.

2.2 Does mindset work in China?

2.2.1 The cultural beliefs about learning in China and in Western countries

Children's beliefs about learning in China and Western countries are culturally different (J. Li, 2002). These cultural differences concerning learning might be due to the different emphases in the classroom in China and Western countries. Specifically, for children in Western countries, their school education is more likely to focus on developing personal characteristics such as independence or self-esteem. However, school education in China is more likely to focus on the acquisition of knowledge and skills rather than independence.

Another possible cause of the cultural differences in children's beliefs about learning in China in comparison to Western countries could be children's different attitudes towards success and failure (J. Li, 2002). Specifically, Chinese children are more likely to be hard workers since they believe that either success or failure is a part of learning, and that success is the outcome of making an effort (Heine et al., 2001; J. Li, 2002). Thus, it is more important for children to acquire skills or knowledge during the process of problem-solving rather than the results of success or failure (J. Li, 2002). However, children in Western countries are more likely to view success or failure as an outcome of learning, which would consequently decrease children's motivation to learn (J. Li, 2002).

The most important factor that leads to the cultural differences of children's beliefs about learning would be the different views towards effort and intelligence among children in China and Western countries (Tweed & Lehman, 2002). Chinese children were more likely to view effort rather than intelligence as a core component of learning (Tweed & Lehman, 2002). In Chinese beliefs, intelligence is seen as malleable and unlimited (Rattan, Savani, Naidu, & Dweck, 2012), and making greater effort means having more intelligence (Salili & Hau, 1994). Therefore, for Chinese children, to become a genius means working hard to achieve goals (J. Li, 2002). Those who are less or not intelligent could achieve the same goal as a genius does if they could make a greater effort (J. Li, 2002). However, in Western countries, intelligence is defined as a trait that is dominated by genes, and therefore, only a few people who are highly intelligent could be seen as geniuses (Rattan, Savani, et al., 2012). Additionally, effort and ability are negatively correlated in Western countries, where people with high levels of intelligence could make less effort to achieve success, compared to those with relatively low levels of intelligence (Salili & Hau, 1994). These different views about effort and intelligence lead to different attributions among children in China and Western countries. Specifically, Chinese children are more likely to attribute their failures to lack of effort, and they would change their effort rather than change the environment when experiencing failure (Heine et al., 2001). Additionally, Chinese children prefer to work harder when they notice their shortcomings (Heine et al., 2001). However, children in the West are more likely to attribute their failure to lack of ability or noncontrollable factors (e.g., teacher's teaching quality), and prefer to work harder when they focus on their strengths (Heine et al., 2001; Tweed & Lehman, 2002).

Therefore, the cultural differences of children's beliefs about learning are possibly caused by the different emphases of learning at school, different attitudes towards success and failure, and different views of effort and intelligence. Since both attitudes towards success and failure, and views about effort and intelligence are important components in the mindset theory, if we link these two factors together, it would be reasonable to deduce that Chinese children would be more likely to hold a growth mindset, whilst their Western counterparts would hold a fixed mindset.

2.2.2 The mindset theory in China

Although the mindset theory was proposed around 30 years ago, only a few studies have tested the effects of mindset on children's development in China (P. Li et al., 2017; Q. Wang & Ng, 2012; Zeng et al., 2016). For example, Zeng et al. (2016) conducted a study testing whether Chinese children's own mindsets would affect their school engagement and psychological well-being. They found that the growth mindset had significantly positive effects on children's psychological well-being and school engagement (Zeng et al., 2016). In other words, children with a growth mindset had better psychological well-being, and higher school engagement than those with a fixed mindset (Zeng et al., 2016). Another study conducted by P. Li et al. (2017) found a positive mediating effect of children's own mindsets on the association of children's prior achievement and their school engagement (including behavioural, emotional and cognitive engagement). Specifically, children who held a growth mindset more strongly associated their prior achievement with behavioural engagement in comparison to children with a fixed mindset (P. Li et al., 2017). Furthermore, a significant positive association between prior achievement and children's emotional engagement, as well as a significant positive association between prior achievement and children's cognitive engagement were only found among children with a growth mindset (P. Li et al., 2017).

In addition to the studies that tested the association between mindset and school engagement or well-being, researchers also explored whether Chinese children could distinguish mindset and school performance, and whether these two factors predicted children's problem-solving approaches to their schoolwork (Q. Wang & Ng, 2012). Chinese children were found to have distinct views of mindset and school performance. Similar to the two mindset types, Chinese children viewed school performance in two different ways: malleable or fixed, and Chinese children were more likely to believe that their school performance had more malleability than their intelligence. Additionally, Q. Wang and Ng (2012) found that the more fixed beliefs children held, the more helpless their problem-solving approaches to schoolwork would be.

Although some studies found positive effects of mindset on children's development in some psychological perspectives (e.g., psychological well-being and school engagement), no independent study has tested the association of mindset and children's educational attainment, one of the most important outcomes in children's development, among Chinese children. Furthermore, of all of the studies that investigated the effects of children's own mindsets on their development, none of them used any mindset manipulations. Thus, it is still unclear whether using a growth mindset manipulation would be beneficial for Chinese children.

2.3 The current study

There are four main drawbacks in the reviewed studies.

Firstly, previous studies used different scales to measure children's own mindsets. For example, Cury et al. (2008) used a 6-item mindset scale that was designed by Cury et al. (2006), Dupeyrat and Mariné (2005) used a 9-item mindset scale including five fixed mindset items adopted from Hong, Chiu, and Dweck (1995) and four growth mindset items whose origins were not clearly stated. For most of these scales, to my knowledge, their reliability and validity are not clarified or have not been verified in other independent studies. For those studies that used the same mindset scale - usually this was the Theories of Intelligence scale from Dweck (2000) - researchers used different items. For instance, Blackwell et al. (2007) used 6 items, but Claro et al. (2016) used only 2 items. The usage of different items leads me to doubt the validity and reliability of the results of these previous studies. For example, Claro et al. (2016) used only two (fixed mindset) items to measure participants' mindsets. They assigned participants who chose "agree" or "strongly agree" to the fixed mindset group; participants who chose "disagree" or "strongly disagree" to the growth mindset group; and participants who were uncertain about these two items to the mixed mindset group. The major problem here is that participants, and especially for those who were in the mixed mindset group, might be miscategorised. Since all items in Claro et al. (2016) were about fixed mindset, participants in the mixed group might not have clear attitudes to these two items, but if more items are provided, they may be able to provide clear attitudes to those items and to be assigned to the fixed or growth mindset group. Thus, the reliability of their results is questionable.

Secondly, the sample sizes in studies that used mindset manipulations (either praise manipulation or the mind-can-be-grown manipulation) were smaller than those tested the effects of children's own mindsets. For example, Claro et al. (2016) recruited more than 168,000 participants in their study to test whether children's own mindsets would affect their educational attainment. However, there were only 33 participants in a study testing whether using an online mindset manipulation would affect children's resilience to failure (Donohoe et al., 2012). For those studies with relatively small sample sizes, it is not convincing to generalise their findings to a larger population.

Thirdly, although mindset theory was proposed around 30 years ago, most of the studies have only been conducted in Western countries such as the U.S. or European countries (e.g., Aronson et al., 2002; Blackwell et al., 2007; Yeager et al., 2016). For those few studies that have been conducted in China, none of them used any form of the mindset manipulation, nor tested the effects of the mindset manipulation on children's educational attainment. Thus, it is unclear whether mindset only works under specific circumstances or in a specific population (e.g. the U.S.).

Finally, there are some studies (e.g., Burnette et al., 2017; Cury et al., 2006) that used both children's own mindsets and the mind-can-be-grown manipulation to investigate their effects on children's educational attainment. However, to my knowledge, there is no independent study combining children's own mindsets and the praise manipulation in one study to test their effects on children's educational attainment. Since praise is a common form of feedback that children could receive from their parents or teachers, and it starts to affect children in very early childhood (Kamins & Dweck, 1999), it would be valuable to test the effects of both the praise manipulation and children's own mindsets on children's educational attainment.

Due to the aforementioned reasons, I will conduct a series of studies that: 1) use the most commonly used mindset scale: Theories of Intelligence scale (8-item; Dweck, 2000); 2) include a relatively larger sample size (more than 600 participants) than previous studies (e.g., 33 participants); 3) are conducted in China, which is culturally different than in previous studies; and 4) use both the praise manipulation and the measurement of children's own mindsets. My main aim is to test whether the praise manipulation and children's own mindsets would affect children's cognitive performance after challenges, as well as children's educational attainment (i.e., school grades).

Chapter 3. Is growth mindset associated with cognitive performance and educational attainment?

This chapter reports four studies testing whether growth mindset would be associated with higher cognitive performance after challenges and better educational attainment among children in primary school. Since the studies reported in this chapter generated a paper that has been published by *Journal of Experimental Psychology: General* in order to maintain the consistency with the published version and to acknowledge the collaborator's effort for its publication, the personal pronoun "we" is used in this chapter.

3.1 Introduction

Mindset theory (implicit theories) predicts that children's beliefs about whether basic ability is stable (fixed mindset) or can be changed substantially (growth mindset) impact causally on their cognitive performance (Mueller & Dweck, 1998) and achievement (Dweck, 2006), including educational attainment (Blackwell et al., 2007; Dweck, 2000; Gunderson et al., 2013; Paunesku et al., 2015), with the strongest effects occurring for the most challenging material (Good, Rattan, & Dweck, 2012). These findings have been widely cited, and have been recommended for adoption into "*policy at all levels (federal, state, and local) ... to lift the nation's educational outcomes*" (Rattan et al., 2015). This call has been widely heeded in education (Yettick, Lloyd, Harwin, Riemer, & Swanson, 2016). These claims have, however, been subject to little independent replication, and there have been failures to support the theory (e.g., Bahník & Vranka, 2017; Burgoyne et al., 2020). Here, we tested the relationship of mindset to resilience to failure (Mueller & Dweck, 1998) and school grades (Blackwell et al., 2007) in three large samples.

3.1.1 Background

Mueller and Dweck (1998) is a hallmark paper on mindset manipulations. Mueller and Dweck (1998) reported six studies on children aged 9-12 years old. Four of these studies tested the effects of a mindset manipulation on subsequent task performance (studies:1, 3, 5, and 6; ns = 128, 88, 46, and 48 respectively). Mindset was manipulated by giving different forms of praise. As Dweck (2008b) explains about the Mueller and Dweck (1998) studies, "*intelligence praise instilled more of a fixed mindset, making students believe that their intelligence was a fixed trait, whereas the effort praise instilled more of a growth mindset*" (p. 57; see also, e.g., Paunesku et al. (2015); Yeager and Dweck (2012)). Mueller and Dweck (1998) manipulated mindset via these carefully crafted praise scenarios following a set of moderately difficult items (Trial 1). They praised the students for being a "hard worker," or for being "smart at these." Children in the control group received congratulations, but neither form of additional praise. All children then completed a set of more difficult items (Trial 2) and were told they performed "a lot worse" on these. This was followed by a final set of moderate difficulty items (Trial 3). The critical test was an ANOVA comparing difference-scores (Trial 3 - Trial 1) with mindset condition (growth or fixed) as a predictor. Children exposed to the growth mindset condition significantly outperformed children in the fixed mindset condition in all four experiments. Children in the growth and fixed mindset conditions differed in their scores by ~1.3 SD (~ 20 points in IQ terms). The manipulation was reported to affect all children, independent of their ability and/or ethnicity. Mueller and Dweck (1998) also reported that the fixed mindset condition impaired children's motivation for additional learning opportunities. Specifically, children who were in the fixed mindset condition had lower task enjoyment and task persistence and were more likely to attribute their failure to a lack of ability compared to those who were in the growth mindset condition (Mueller & Dweck, 1998). The method thus produced large effects, emerging reliably in each of four studies, and formed what is still acknowledged as the core "careful laboratory experiments" testing mindset theory (Paunesku et al., 2015, p. 791).

A second highly-cited report extended these findings to examine the relationship of children's mindsets to their educational learning outcomes concluding that "*Implicit theories of intelligence [mindsets] predict achievement*" (Blackwell et al., 2007, p. 246). Study 1 of this paper followed 373 children progressing into junior high school (aged around 12 years old) and observed for two years. Children were assessed at entry using a questionnaire measure of mindset (Dweck, 2000). Entry scores on mathematics were unrelated to children's mindsets, but mathematics grades at the end of the first semester of observation were correlated positively with growth mindset ($r = .12$) and math scores at the end of the second year of observation were positively associated with children's mindsets controlling for their entry scores on math ($\beta = 0.17$, $t(372) = 3.40$, $p < .05$).

Subsequent studies of the association between mindset and academic achievement, however, have yielded mixed results (Sisk et al., 2018). For instance, in a Chinese population, Zhao and Wang (2014) reported in 524 pupils aged 12-16 years, finding a significant association of mindsets with students' baseline achievement ($r = .23$). Paunesku et al. (2015) reported on 1,594 9^{th}-12^{th} grade students finding a small ($\beta = 0.06$, CI_{95} [0.03, 0.09], $t(1561) = 3.47$, $p < .001$) association of growth mindsets with pre-study GPA and finding an association of children's mindsets with final grades only when restricting analyses to the bottom $1/3^{rd}$ of participants (whereas Blackwell et al. (2007) had reported a null association with pre-study grades and a main effect in the full sample). Recently, in a large (n= 5,653) sample of university applicants, Bahník and Vranka (2017) found a small significant effect of children's mindsets on scholastic aptitude, but the direction was reversed to the prediction from mindset theory (r = - .03, CI_{95} [-0.05, -0.00], p = .040).

3.1.2 Goals of the present studies

The findings reported by Mueller and Dweck (1998) and by Blackwell et al. (2007) are clearly important if they are replicable. However, the claim that praising 9-12 year-old children for being smart versus for being a hard worker causes large (> 1 SD) impacts on their cognitive performance (Mueller & Dweck, 1998) has not, to our knowledge, been independently replicated. Likewise, while some studies have tested the prediction that growth mindsets are associated with improvement in school grades, the results in this field are mixed, as noted above.

In our study 1, we therefore began with a close replication of Mueller and Dweck (1998) study 1. We did this to establish if, with our sample, we could replicate the finding that the growth mindset manipulation is associated with better post-failure performance relative to a fixed mindset manipulation. A positive finding, even with a reduced effect size, would suggest that our population and methods are suitable to further test the theory. We therefore undertook a close replication using the manipulation, tasks and analytic approach specified in the original Mueller and Dweck (1998) study.

The similarities and differences of the present study and Mueller and Dweck (1998) study 1 are detailed in Table 3.1. Briefly, we used the same mindset manipulation (priming a fixed mindset with "*you must be smart at these*" and priming a growth mindset with "*you must be a hard worker at these*"), the items from the same cognitive measures as originally used, given for the same durations. We used the same negative feedback, and the same analyses of the data. We also used the same suite of measures of achievement goal, desire to persist, enjoyment of the problems, perceptions of the quality of performance and attributions of the causes of the performance to test how these were associated with the mindset theory.

Table 3.1 The similarities and differences between Mueller & Dweck (1998) study 1 and the present studies 1, 2 and 3

	Mueller & Dweck (1998)	**The present paper**		
	Study 1	**Study 1**	**Study 2**	**Study 3**
Participants	N= 128 (70 girls and 58 boys)	N= 190 (101 girls, 89 boys)	N= 222 (106 girls, 116 boys)	N= 211 (91 girls, 120 boys)
Age	Mean age = 10.7, SD = 0.60	Mean age = 10.48, SD = 0.51	Mean age = 11.03, SD = 0.47	Mean age = 10.70, SD = 0.54
Ethnicity	50% Caucasian, 19% African American, 31% Hispanic	100 % Chinese	100 % Chinese	100 % Chinese
Source	One public elementary school in a small midwestern town and two public elementary schools in a large north-eastern town in the U.S.	One public primary school in a north-eastern city in China.	Another public primary school in the same city as study 1.	Another public primary school in the same city as study 1.

SES	Not reported	City 21% below the Chinese national average income	City 21% below the Chinese national average income	City 21% below the Chinese national average income
Ravens tests	All trials were from Raven's Standard Progressive Matrices (SPM: Raven, 1976).	All trials were from the SPM (Raven, Raven, & Court, 2000).	All trials were from the SPM (Raven et al., 2000) and SPM Parallel trials (Styles, Raven, & Raven, 1998)	All trials were from the SPM (Raven et al., 2000) and SPM Parallel trials (Styles et al., 1998)
Tests	Three tests 1) moderate difficulty (Trial 1)* 2) more difficult (Trial 2)* 3) equal to Trial1 (Trial 3)* (* exact items were not given)	Three tests: 1) moderate difficulty (Trial 1) 2) more difficult (Trial 2) 3) equal to Trial1 (Trial 3)	Four tests: 1) moderate difficulty (Trial 1) 2) more difficult (Trial 2) 3) equal to Trial1 (Trial 3) 4) equal to Trial2 (Trial 4)	Four tests: 1) moderate difficulty (Trial 1) 2) more difficult (Trial 2) 3) equal to Trial1 (Trial 3) 4) equal to Trial2 (Trial 4)
Test lengths	Trial 1 & 3 = 10 items	Trial 1 & 3 = 10 items	Trial 1 & 3 = 12 items	Trial 1 & 3 = 12 items

	Trial 2 = 10 items	Trial 2 = 10 items	Trial 2 & 4 = 10 items	Trial 2 & 4 = 10 items
Average score on Trial 1	5.2/10 (52%) (7.9 attempted)	8.94/10 (89%) (attempts not scored)	7.64/12 (63%) (10.9 attempted)	7.61/12 (63%) (10.8 attempted)
Average score on Trial 2	1.6/10	4.4/10	5.1/10	4.5/10
Feedback rule	All participants were told that they had solved at least 80% of the problems that they answered, no matter what their actual scores were.	All participants were told that they had solved at least 80% of the problems that they answered, no matter what their actual scores were.	All participants were told that they had solved at least 80% of the problems that they answered, no matter what their actual scores were.	All participants were told that they had solved at least 80% of the problems that they answered, no matter what their actual scores were.
General praise	"Wow, you did very well on these problems. You got [numbers of problems] right. That's a really high score."	"Wow, you did very well on these problems. You got [numbers of problems] right. That's a really high score."	"Wow, you did very well on these problems. You got [numbers of problems] right. That's a really high score."	"Wow, you did very well on these problems. You got [numbers of problems] right. That's a really high score."

Fixed mindset condition	"You must be smart at these problems"	"You must be smart at these problems"	"You must be smart at these problems"	"You must be smart at these problems"
Growth mindset condition	"You must have worked hard at these problems"	"You must have worked hard at these problems"	"You must have worked hard at these problems"	"You must have worked hard at these problems"
Control condition praise	Control group received general praise only, with no additional feedback given.	No controls. To maximise effective n, all participants were allocated to either the fixed mindset or growth mindset conditions	Active control group were told "Even though we cannot change our basic ability, you work hard at hard problems and that's how we get hard things done!"	Active control group were told "Even though we cannot change our basic ability, you work hard at hard problems and that's how we get hard things done!"
Negative feedback (after Trial 2)	Participants were told they had performed "a lot worse" on the second trial of problems and had solved no more than 50%	Participants were told they had performed "a lot worse" on the second trial of problems and had solved no more than 50%	Participants were told that they had performed "a lot worse" on the second trial of problems and had solved no more than 50%	Participants were told that they had performed "a lot worse" on the second trial of problems and had solved no more than 50%

	of the problems that they answered.	of the problems that they answered.	of the problems that they answered.	of the problems that they answered.
Additional tests	None	None	Theories of Intelligence Scale (Dweck, 2000)	Theories of Intelligence Scale (Dweck, 2000)
Time allowed	4 minutes	4 minutes	4 minutes	4 minutes
Analysis	One-way ANOVA comparing change in performance (Trial 3 – Trial 1).	One-way ANOVA comparing change in performance (Trial 3 – Trial 1).	One-way ANOVA comparing change in performance (Trial 3 – Trial 1).	One-way ANOVA comparing change in performance (Trial 3 – Trial 1).

In studies 2 and 3, we extended this work by improving the methods. Specifically, we added an active control condition and expanded the post-failure measure to include a set of more difficult items. Both these additions were designed to allow us to better understand the mechanism, if any, of the mindset manipulation. By incorporating an active control condition, we were able to isolate the predicted effect of mindset from other aspects of the growth condition, such as potential experimenter-demand effects and an effort encouragement confound (Locke & Latham, 2002). We also took the opportunity to test the effect of children's own mindset on their responses to failure. According to mindset theory, one's beliefs about intelligence should have profound effects on one's achievement (Dweck, 2006, 2008a). Because children’s mindsets are deeply embodied and range from very fixed to very growth-oriented, we predicted these would have effects at least as large as those of a brief verbal manipulation.

Finally (study 4), using children’s school grades across two waves of assessment, we were able to test the claim that children’s mindsets affect their educational attainment (Blackwell et al., 2007). Specifically, growth mindset theory predicts whether children believe that basic ability can be greatly changed or is fixed and hard to change causes differences in attainment and response to failure in educational setting. We therefore tested whether growth mindsets are associated with either initial grades, improvement in grades over a semester, or improved grades in children initially scoring poorly. The similarities and differences between Blackwell et al. (2007) study 1 and our study 4 are detailed in Table 3.2. We report how we determined our sample sizes, all data exclusions (if any), all manipulations, and all measures in the studies.

Table 3.2 Summary of key hypotheses across the present studies 1, 2 and 3 with key stats for each prediction

	Study 1	**Study 2**	**Study 3**
Does a growth mindset manipulation enhance post-failure performance on the moderate difficulty items?	Yes	No	No
Statistical results	F (1,188) = 3.930, p = .049*	F(2,219) = 0.440, p = .645	F(2,208) = 2.744, p = .067 (active control did best) β = 0.32, CI_{95} [0.05, 0.60], t = 2.34, p = .020*
Does a growth mindset manipulation enhance post-failure performance on the more difficult items?	NA	No	No
Statistical results	NA	F(2,219) = 0.630, p = .534	F(2, 208) = 0.216, p = .806

Do children's mindsets predict response to failure on moderate difficulty items?	NA	No	No
Statistical results	NA	$F(1, 220) = 0.074$, $p = .786$	$F(1, 209) = 0.179$, $p = .673$
Do children's mindsets predict response to failure on more difficulty items?	NA	No (in a reversed direction)	No
Statistical results	NA	$F(1, 220) = 7.482$, $p = .007$**; $\beta = -0.13$, CI_{95} [-0.23, -0.04]	$F(1, 209) = 0.020$, $p = 0.888$
Do children's mindsets relate to grades?	NA	No	No
Statistical results	NA	Semester 1: $\beta = 0.03$, CI_{95} [-0.10, 0.16], $t = 0.42$, $p = .671$	Semester 1: $\beta = 0.04$, CI_{95} [-0.10, 0.17], $t = 0.52$, $p = .601$

		Semester 2: $\beta = 0.05$, CI_{95} [-0.11, 0.21], $t = 0.63$, $p = .530$	Semester 2: $\beta = 0.06$, CI_{95} [-0.08, 0.20], $t = 0.88$, $p = .382$
Do children's mindsets relate to changes of grades?	NA	No	No
Statistical results	NA	$\beta = 0.03$, CI_{95} [-0.06, 0.12], $t = 0.63$, $p = .532$	$\beta = 0.04$, CI_{95} [-0.04, 0.11], $t = 1.00$, $p = .319$
Do children's mindsets relate to cognitive ability?	NA	No	No
Statistical results	NA	Trial 1: $\beta = 0.12$, CI_{95} [-0.01, 0.25], $t = 1.76$, $p = .080$ Trial 2: $\beta = 0.12$, CI_{95} [-0.02, 0.25], $t = 1.72$, $p = .086$	Trial 1: $\beta = 0.12$, CI_{95} [-0.02, 0.25], $t = 1.68$, $p = .094$ Trial 2: $\beta = 0.07$, CI_{95} [-0.07, 0.20], $t = 0.99$, $p = .322$

3.2 Study 1

We first closely replicated the report that, in 9-13-year-old children, a brief mindset manipulation induces a large change in post-failure performance, as reported by Mueller and Dweck (1998). In study 1 of Mueller and Dweck (1998), children first completed a moderate difficulty trial of 10 cognitive ability items (which we refer to as Trial 1) from the Standard Progressive Matrices (SPM: Raven et al., 2000). Children were given 4 minutes for this task after which they were told that they got at least 80% correct and received one of three kinds of praise: growth ("*you must have worked hard at these problems*"), fixed ("*you must be smart at these problems*"), or control (no additional feedback). This brief laboratory manipulation of mindset using "carefully crafted scenarios" (Dweck, 2013) was followed by a second, more difficult set of SPM items (Trial 2). Children were told they did "*a lot worse*" on these, getting no more than 50% correct. Finally, children were given a further trial of 10 moderate difficulty items (Trial 3). The difference between performance on Trials 1 and 3 formed the dependent variable.

We closely followed the methods of Mueller and Dweck (1998) study 1, testing replicability of the reported effect of praise for being smart versus praise for hard work (see Table 3.1). As in Mueller and Dweck (1998) study 1, children aged 9-13 years old were tested individually. We also implemented the full set of additional measures of learning and motivation, task-persistence, task-enjoyment, self-rated performance and failure attributions as described below and as used by Mueller and Dweck (1998). Differing from Mueller and Dweck (1998), we omitted the control group and randomly assigned children to one of the two mindset conditions to maximise power. According to mindset theory, we should see the largest difference between these two groups.

We tested four classrooms of children in the same grade and school (n = 190), yielding ~85% power to detect a small effect (d = .3). We deemed this effect size the lower limit compatible with the theoretical mechanisms proposed by mindset theory, which imply a tight dependence of performance on mindset condition.

3.2.1 Method

3.2.1.1 Participants

A total of 190 children participated (100% of available children). Of these 89 were boys (mean age 10.56 years, SD = 0.51) and 101 were girls (mean age 10.41 years, SD = 0.50). All children were recruited from a large primary school in Harbin (the capital city of Heilongjiang Province, China). The school is public and draws from a catchment area 21% below the Chinese national average income (average income 48,881 Yuan: National Bureau of Statistics of the People's Republic of China, 2017), equating to USD 7,133 (~$14,000 purchasing-power equivalent). The children are thus in relative poverty (low income relative to others in their country: OECD, 2008). Low socioeconomic status has been argued to increase the influence of mindset on performance (Claro et al., 2016). Thus, we expected, if anything, a larger effect in our studies. Compensation for participation consisted of a reward of sweets at the end of the study.

3.2.1.2 Materials

Individual cognitive performance was assessed using items from sets B, E, and C of the SPM (Raven et al., 2000). Following Mueller and Dweck (1998), Trial 1 (the praise cognitive test) consisted of the first 10 items from set B (moderate difficulty items). Trial 2 (the failure test) consisted of the first 10 items from set E (more difficult items). Trial 3 (the post-failure measure) consisted of the first 10 items from set C (moderate difficulty items).

Learning and motivation were assessed using the learning and motivation questionnaire (Mueller & Dweck, 1998). Preference for learning or performance goal was assessed by an item asking children which of four options they would prefer: A: "*problems that aren't too hard, so I don't get many wrong*", B: "*problems that are pretty easy, so I'll do well*", C: "*problems that I'm pretty good at, so I can show that I'm smart*" and D: "*problems that I'll learn a lot from, even if I won't look so smart*" (Mueller & Dweck, 1998), with D scored as a learning goal, and responses A, B, or C as performance goal preference. Task-persistence, task-enjoyment, and self-rated performance were assessed via a 4-item measure described in Mueller and Dweck (1998). Items were "*How much would you like to take these problems back home to work on*?", "*How much did you like working on the first/second set of problems*?", "*How much fun were the problems*?" and "*How well did you do on the problems overall*?". Children responded on a scale from 1 (not at all) to 6 (very much).

Attributional style for performance after negative feedback was assessed as in Mueller and Dweck (1998). Children were asked to explain "why they had some trouble" with the items on Trial 2. Four slotted-disks of coloured paper were pinned together so children could rotate, exposing various amounts of each disk viewed from the front. The disks each had printed on them one of four attributions: "*I didn't work hard enough.*", "*I'm not good enough at the problems*", "*I'm not smart enough.*", or "*I didn't have enough time*.", corresponding to attributions of lack of effort, lack of ability (the average of the second and third attributions) and lack of time respectively.

Children were asked to rotate the disks to show how much each factor accounted for their failure. In addition, children were asked to weight the importance of ability and hardworking when solving the puzzles using a circle with marks from 1-36 around its circumference which they connected to divide the circle into two parts ("smart" and "hard work"), and colouring-in the smart proportion.

Whenever items were translated from English text into Chinese, the experimenter made an initial translation, which was then back translated by 5 bilingual (Chinese and English) speakers, checked for round-trip accuracy, and edited where necessary to ensure an accurate translation.

3.2.1.3 Design

This study used a between-group design. The independent variable was the mindset manipulation, with two levels: fixed mindset condition and growth mindset condition. The dependent variable was difference of scores between Trial 1 and 3.

3.2.1.4 Procedure

Study 1 was approved by the Psychology Research Ethics Committee at the School of Philosophy, Psychology and Language Sciences (PPLS), University of Edinburgh (reference number: 229-1415/3). After informed consent was gained from the headmaster, teachers, parents, and children themselves, children were asked to provide demographic information, and were then tested individually in a private room near their classroom. Testing began with a welcome, and an introduction to the testing procedures in which children were given an example item from the SPM items. Children were shown how to solve this problem and then were assigned to a mindset manipulation condition in a sequential ABAB order (95 in each condition).

After this introduction, children then completed the initial moderate difficulty trial (Trial 1), answering as many items as they could in 4 minutes. The experimenter (YL) then removed the children's answer sheets and scored their responses. All children received the same positive feedback "*Wow, you did very well on these problems. You got 7/8/9 right, That's a really high score*!". Children who correctly solved fewer than 5 items were told they got 7 items correct. Children solving 6–9 items correct were told they had got 8 items correct. Children who got all 10 items correct were told they got 9 items correct. Children randomized to the fixed mindset condition were then told "*You must be smart at these problems!*" while children in the growth mindset condition were told "*You must have worked hard at these problems!*". Children then completed the learning goals questionnaire.

The more difficult trial (Trial 2) was then administered. After 4 minutes, the test was scored, and, no matter what their performance, children were told "*Your performance was poor on that: You got less than half the items correct*". As in Mueller and Dweck (1998), children then completed the task persistence, task enjoyment, and overall self-rated performance quality questionnaires. Finally, children were asked to work on the post-failure items (Trial 3), again with a 4-minute time limit.

All children were then debriefed and were told that the more difficult trial on which they had received poor scores contained items that were appropriate for older and higher-grade children. Therefore, children in their grade who solved even a single item should be proud as they were especially hard working to have attempted and succeeded at these.

3.2.2 Results

All analyses were completed using R (R Core Team, 2019) and umx (Bates, 2018; Bates, Maes, & Neale, 2019). Standardized effect sizes are reported to aid interpretability and incorporation into subsequent meta-analyses. All data and analysis code are open-access and raw data and R analysis scripts used in all four studies are available in supplementary data at https://osf.io/u5v8f. Scores on the moderate difficulty test (Trial 1) were skewed due to ceiling effects (skew = -2.41, kurtosis = 7.66).

3.2.2.1 Does mindset manipulation affect post-failure performance?

We first tested the hypothesis that children who were in the growth mindset condition (i.e., praised for hard work) would have higher post-failure performance (Trial 3 SPM score) compared to those who were in the fixed mindset condition. We tested this hypothesis using the same one-way ANOVA approach used by Mueller and Dweck (1998), namely a difference of scores (Trial 3 - Trial 1)[1] was used as the dependent variable (DV), and mindset condition as the independent variable (IV). As in Mueller and Dweck (1998), age or gender was not controlled. This one-way ANOVA revealed a significant difference, with children in the growth mindset condition scoring higher on Trial 3 (controlling for Trial 1) compared to those in the fixed mindset condition ($F(1,188) = 3.930$, $p = .049$; $\beta = -0.28$, $CI_{95}[-0.55, 0.00]$: see Figure 3.1 and Table 3.3). Following Mueller and Dweck (1998) we also tested whether children in the two groups differed in their baseline scores (Trial 1). No

[1] Mueller and Dweck's use of difference scores as the outcome might be problematic since difference scores and the pre- and post- test scores are highly correlated (see Cook, Campbell, & Day, 1979).

significant difference was found (Trial 1: $F(1,188) = 0.129$, $p = .720$; $\beta = 0.05$, CI_{95} [-0.23, 0.34]).

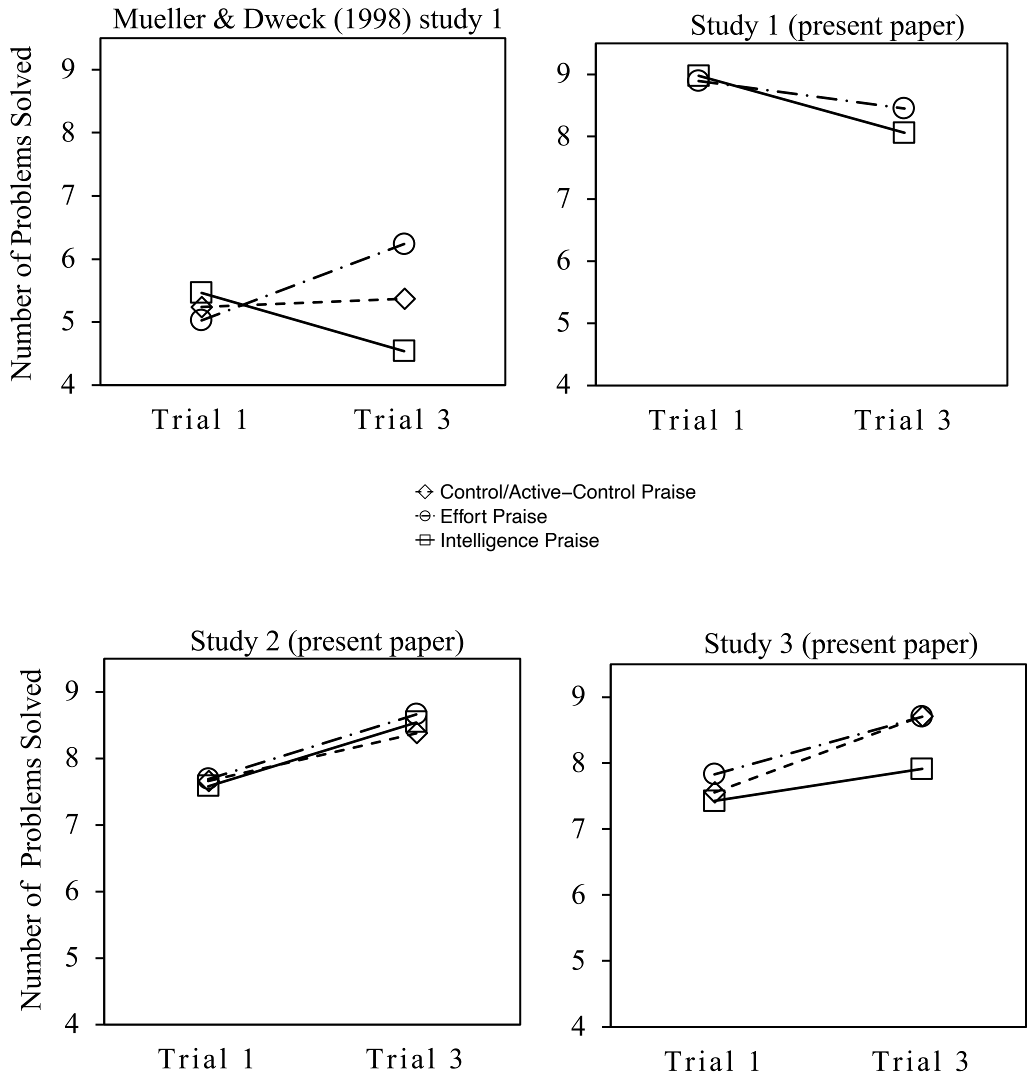

Figure 3.1: Number of problems children solved before (Trial 1) and after (Trial 3) the failure SPM test in Mueller and Dweck (1998) study 1 and the present studies 1, 2 & 3

Table 3.3 The similarities and differences between Blackwell, Trzesniewski & Dweck (2007) study 1 and the present studies 2 and 3

	Blackwell, Trzesniewski & Dweck (2007)	**The present paper**	
	Study 1	**Study 2**	**Study 3**
Participants	N= 373 (198 girls and 175 boys)	N= 222 (106 girls, 116 boys)	N= 211 (91 girls, 120 boys)
Age	7th grade	5th grade (Mean age =11.03, SD=0.47)	5th grade (Mean age=10.70, SD=0.54)
Ethnicity	55% African American, 27% South Asian, 15% Hispanic, 3% East Asian and European American.	100 % Chinese	100 % Chinese
Source	One public secondary school in New York city.	One public primary school in a north-eastern city in China.	One public primary school in a north-eastern city in China.
SES	53% children were eligible for free lunch	City 21% below the Chinese national average income	City 21% below the Chinese national average income
Academic outcomes measurement	Math grades	GPA (Math, Chinese and English grades)	GPA (Math, Chinese and English grades)

Theory of Intelligence measurement	Implicit Theories of Intelligence Scale for Children (Dweck, 2000, p.177)	Theories of Intelligence Scale (Dweck, 2000, p.178)	Theories of Intelligence Scale (Dweck, 2000, p.178)
Scale properties	6 items, each scored 1-6	6 items from the Implicit Theories of Intelligence Scale for Children plus 2 extra items, each scored 1-6	6 items from the Implicit Theories of Intelligence Scale for Children plus 2 extra items, each scored 1-6
Extra items	NA	"To be honest, you can't really change how intelligent you are"; "You can change even your basic intelligence level considerably."	"To be honest, you can't really change how intelligent you are"; "You can change even your basic intelligence level considerably."
Average mindset score	4.45 (SD = 0.97)	4.25 (SD = 0.81)	4.16 (SD = 0.89)

3.2.2.2 Does mindset manipulation affect motivation?

We also examined the hypotheses that growth mindset manipulation would: 1) lead children to pursue a learning goal rather than a performance goal, 2) increase task persistence, 3) increase children's enjoyment on solving the problems, 4) have higher self-rated performance quality, 5) attribute their failure on Trial 2 to effort rather than ability compared to those in the fixed mindset condition. As in Mueller and Dweck (1998), these hypotheses were tested using a Chi-square test (for hypothesis 1) and one-way ANOVAs (for hypotheses 2, 3 & 4, 5), with responses on these questions as the dependent variables, and mindset manipulation as the independent variable. Finally, Mueller and Dweck (1998) tested the attributions of the children for their failure to either hard work or lack of ability using a one-way ANOVA. Despite the significant effect of mindset manipulation on changes of cognitive scores (Trial 3 – Trial 1), the predicted effects on motivation were not supported by the results. Mindset manipulation was not associated with expression of a learning goal (χ^2 (1) = 0.192, p = .661), wishing to take the problems home (F(1,188) = 2.833, p = .094), finding working on the problems enjoyable (F(1,188) = 0.552, p = .459), or fun (F(1,188) = 0.229, p = .633). Neither was there any effect of mindset manipulation on perceived performance (F(1,188) = 0.733, p = .393). Participants' attributions regarding the role of ability and effort did not differ by condition (F(1,188) = .570, p = .451 and F(1,188) = .496, p = .482 respectively). The relative attribution of failure to ability versus effort also did not differ significantly (F(1,188)= .209, p = .648).

3.2.3 Study 1 discussion

The results of study 1 indicated that children in the growth mindset condition showed significantly higher post-failure performance compared to children in the fixed mindset condition. This close replication of Mueller and Dweck (1998) study 1 indicated that with the same mindset manipulation, SPM items, negative feedback, and analysis plan, we could replicate the basic finding in our population, albeit substantially reduced in magnitude. This is distinct from concluding that the effect observed is driven by the mechanism proposed by mindset theory. In planning our next studies, we were guided by a desire to incorporate methodology that would allow us to better understand the mechanism behind this effect, specifically, whether the effect was due to mindset or an effort confound. To this end, we added an active control condition.

We were also cognisant that mindset theory is designed to explain how children cope with difficult material and significant challenges (Good et al., 2012). While study 1 was a close replication of Mueller and Dweck (1998), the materials were only moderately difficult. Including more difficult material in the post-failure trials would increase the power and validity of the study.

Finally, in considering the results of study 1, it was apparent that the design ignores an important available resource: that of the children's internalised mindsets. A design that tests the effects of children's mindsets on their post-failure performance would be valuable. These considerations of the results of study 1 lead us to undertake a second replication, modified as described next.

3.3 Study 2

In constructing study 2, we wished to enhance the power of the design to better investigate the predictions of mindset theory. Increased scores in the growth condition found in study 1 provide support for mindset theory only to the extent that praise for "being a hard worker" has its effects by priming a growth mindset. However, it is also plausible that this condition primes beliefs about conscientiousness (Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007) or other non-mindset motivational effects (Locke & Latham, 2002). To test this, we introduced an active control condition. This condition was designed to isolate any effect of beliefs about intelligence from potential experimenter-demand effects, goal-setting or effort (Locke & Latham, 2002). If theories about the value of hard work (rather than the malleability of intelligence) were driving the modest effect observed in study 1 (and presumably Mueller and Dweck's large effects), then students in the growth mindset condition ought to perform no better than students in the active control condition in study 2.

The active control we wished to construct was one which could prime the fixed mindset (which should lower performance), but which would also prime hard work as something that is needed to accomplish work (but that does not and cannot "grow the mind"). If this condition were to show effects as large or larger than the classic "you must be a hard worker" prime, that would be evidence against mindset (which predicts that priming the idea that ability is fixed should impair post-failure performance), and instead support a conscientiousness or motivational model of effortful performance. To distinguish these, we created a novel active control condition derived from the mindset questionnaire item "*You can learn new things, but you can't really change your basic intelligence*". Participants in this new active control condition were told "*Even though we cannot change our basic ability, you work hard at hard problems and that's how we get hard things done*". This condition thus confirmed the fixed mindset (we cannot change basic ability – which is predicted to be harmful), while also activating the belief that hard work is required to do hard things (which is not specific to mindset theory).

Second, we took advantage of the fact that children bring very different mindsets to the experiment. Our rationale was as follows: if beliefs about the fixed or malleable nature of intelligence change response to failure, then a child's mindset should affect their post-failure performance – indeed, this is the rationale of manipulations targeting, among other things, growth mindset (Paunesku et al., 2015). If the wide range of naturally occurring variation in children's own beliefs (see further discussion near Figure 3.3 below) does not affect outcomes, this falsifies the theory. We therefore included the standard Theories of Intelligence questionnaire (Dweck, 2000), allowing us to test whether children's mindsets are associated with differences in post-failure performance. Because children's mindsets are stable and range from very fixed to very growth-oriented, we predicted these would have effects at least as large as those of the manipulation.

Third, because mindset theory predicts that a growth mindset is critically involved in responses to demanding challenges that otherwise may lead to giving up or dropping out, we wished to ensure that we tested this aspect of the theory. The classic design tests children's post-failure performance on moderate difficulty items. To extend the range of information provided by the experiment, we added an additional more difficult trial (Trial 4) containing items matched to those of Trial 2. As it is predicted that mindset most strongly affects individuals' responses to the more difficult materials (Good et al., 2012), we predicted that any effects of mindset should be most strongly reflected in responses to these more difficult items (tested as the difference in scores on Trials 2 and 4), thus maximising the opportunity to detect mindset effects on responses to failure.

Finally, to assure that the items in Trials 1 and 3 are moderately difficult for the population under test, and based on the distribution of scores in study 1, for study 2 we slightly increased the difficulty of items used in Trial 1 and increased the number of items used from 10 to 12 to assure difficulty more closely matched to that reported by Mueller and Dweck (1998).

Regarding our mindset manipulations, we hypothesised that the growth mindset manipulation (praise for hard work) would enhance children's post-failure performance on the moderate (Trial 3) and difficult (Trial 4) items relative to the active control condition. If mindset is responsible for effects, only the growth mindset condition should enhance post-failure performance; post-failure performance should be similar for the active control condition and the fixed mindset (praise for being smart) condition, and neither should positively predict post-failure performance. Regarding children's own mindsets, we hypothesised that growth mindsets would be positively correlated with post-failure performance on the moderate items (Trial 3) and more difficult items (Trial 4).

3.3.1 Method

3.3.1.1 Participants

In total, 222 pupils were recruited from a second primary school in the same city as study 1. In total, 116 boys (mean age 11.07 years, SD = .49) and 106 girls participated (mean age 11 years, SD = .45). Compensation for participation consisted of sweets at the end of the study.

3.3.1.2 Materials

Children’s mindsets were assessed using the 8-item Theories of Intelligence Scale (Dweck, 2000). When translating the items to Chinese, we used child-friendly language to ensure that children understood the items (Cain & Dweck, 1995). Example items include "*You have a certain amount of intelligence, and you can't really do much to change it*." Possible responses range from 1 (strongly agree) to 6 (strongly disagree) with high scores coded to indicate a growth mindset.

The item-sets were drawn from parallel-form versions of the SPM (Raven et al., 2000) and presented in a counterbalanced order. Trial 1 (moderate difficulty trial) included 12 (rather than 10) items from set C (rather than set B). Trial 2 (more difficult trial) consisted of the first 10 items from set E. Equivalent tests were used in the post-failure Trials 3 and 4, constructed from the parallel forms of the SPM sets C and E (Styles et al., 1998). Learning and motivation measures were given as in study 1.

3.3.1.3 Design

Study 2 used a between-group design. Two independent variables were examined: the mindset manipulation (with three levels: fixed, growth, and active control), and children's mindsets. The dependent variables were Trial 3 - Trial 1 performance, matching Mueller and Dweck (1998) dependent variable, and Trial 4 - Trial 2 performance, which should provide a larger effect of condition given the presumed association between mindset and challenge.

3.3.1.4 Procedure

Studies 2 and 3 were approved by the Psychology Research Ethics Committee at the PPLS, University of Edinburgh (reference number: 106-1516/8). The consent and welcome procedure were identical to those used in study 1. After consent, children completed the mindset measure in their classroom. Children were allocated to one of the three conditions using a sequential-ABCABC order. Testing again took place individually in a private room near their classroom. This began with children being given an example item from the SPM items and shown how to solve this problem. They then completed Trial 1 answering as many items as they could in 4 minutes and were given the feedback appropriate to their randomized condition.

As in study 1, the experimenter removed children's answer sheets, scored their responses, and gave the child positive feedback "*Wow, you did very well on these problems. You got 7/8/9 right, That's a really high score*!". Children randomized to the fixed mindset and growth mindset conditions received appropriate praise consisting of either "*You must be smart at these problems!*" or "*You must have worked hard at these problems!*". Children in the active control condition were told "*Even though we cannot change our basic ability, you work hard at hard problems and that's how we get hard things done*".

Children then completed the learning goals questionnaire. After this, Trial 2 (more difficult items) was administered. After 4 minutes, the items were scored, and, no matter what their performance, children were told "*Your performance was poor on that: You got less than half the puzzles correct*". Again, as in Mueller and Dweck (1998), children then completed the learning and motivation measures. Finally, children were asked to work on the items in Trials 3 and 4, again with 4-minute time limits for each trial.

All children were then debriefed with a procedure identical to that used in study 1.

3.3.2 Results

As in Mueller and Dweck (1998), we first tested if children's initial ability (Trial 1 scores) differed for the three mindset manipulation conditions before testing the four stated hypotheses. No difference was found ($F(2,219) = 0.057$, $p = .944$).

3.3.2.1 *Does mindset manipulation affect post-failure performance on moderately difficult items?*

As in study 1, we tested the hypothesis that the growth mindset condition would significantly improve children's post-failure performance on the moderate difficulty trial (Trial 3) compared to the fixed mindset and active control manipulation conditions. As in Mueller and Dweck (1998), we tested this hypothesis using a one-way ANOVA, with a difference of scores on the initial and final cognitive tests (Trial 3 – Trial 1) as the DV, mindset condition as the IV, and did not control for age and gender.

Contrary to prediction, there was no effect of the manipulation on the change in scores on the moderate difficulty materials ($F(2,219) = 0.440$, $p = .645$; see Figure 3.2). The classic contrast of the fixed mindset vs growth mindset conditions was also non-significant ($\beta = 0.00$, CI_{95} [-0.29, 0.30], $t = 0.03$, $p = .974$).

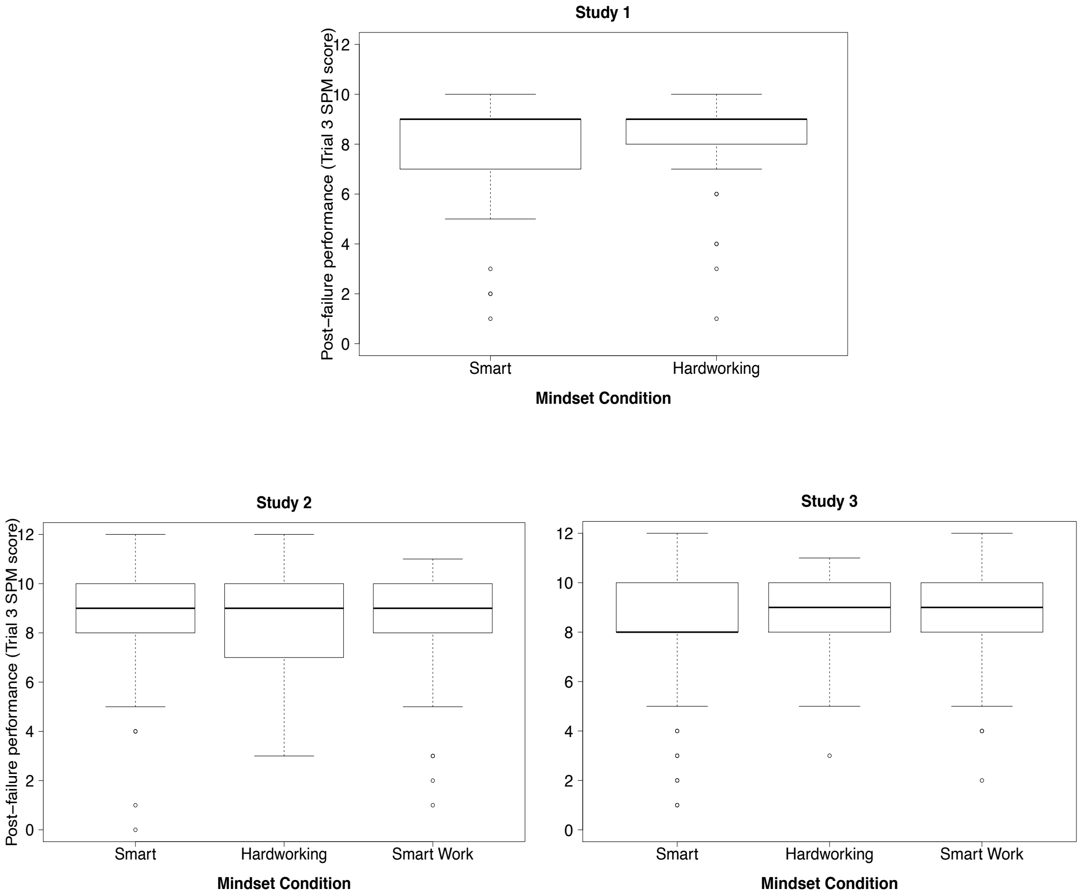

Figure 3.2: Post-failure performance (Trial 3 SPM score) for each mindset condition (shown on the x-axis), plotted separately for study 1 (top), study 2 (bottom left) & study 3 (bottom right) panes

3.3.2.2 Does mindset manipulation affect post-failure performance on more difficult items?

Next, we tested if the growth mindset manipulation would improve children's cognitive scores on the more difficult trial (Trial 4) relative to their initial scores (Trial 2). Again, this was done by using a one-way ANOVA with a difference of cognitive scores (Trial 4 – Trial 2) as the DV and mindset condition as the IV. Again, as in Mueller and Dweck (1998), age and gender were not controlled.

On the more difficult items, where mindset was predicted to most strongly reveal its effects, there was, again, no effect of the growth mindset manipulation

($F(2,219) = 0.630$, $p = .534$). The classic fixed mindset vs growth mindset conditions contrast was similarly non-significant ($\beta = 0.13$, CI_{95} [-0.10, 0.37], $t = 1.12$, $p = .264$).

3.3.2.3 Do children's mindsets correlate with post-failure performance on moderately difficult items?

We next tested whether children's mindsets affected their responses to failure on the moderate difficulty items. This was done by using a regression model with a difference of cognitive scores (Trial 3 – Trial 1) as the DV, and children's scores on the Theories of Intelligence Scale (Dweck, 2000) as the IV. Contrary to prediction, children's mindsets were unrelated to their post-failure performance on the moderate difficulty items ($F(1, 220) = 0.074$, $p = .786$, $\beta = 0.02$, CI_{95} [-0.10, 0.14]).

3.3.2.4 Do children's mindsets correlate with post-failure performance on more difficult items?

Finally, we tested whether children's mindsets impacted their responses to failure on the more difficult items. Again, this hypothesis was tested by using a regression model with a difference of cognitive scores (Trial 4 – Trial 2) as the DV, and children's scores on the Theories of Intelligence Scale (Dweck, 2000) as the IV.

Children's mindsets were significantly linked to their post-failure performance on the more difficult items ($F(1, 220) = 7.482$, $p = .007$). However, this effect was in the reverse direction to that predicted by theory ($\beta = -0.13$, CI_{95} [-0.23, -0.04]). If this result could be replicated in further studies, this would suggest that holding a growth mindset *harms* response to more difficult items.

3.3.3 Study 2 discussion

Summarising the results of study 2, we did not find support for any effects of the mindset manipulation on children's responses to either moderate difficulty (Trial

3) or more difficult (Trial 4) items. We also found no evidence for any effects of children's mindsets on their performance on the moderate items. Moreover, when it came to the more difficult material, we found support for a *harmful effect* of growth-oriented mindsets on scores. Thus, contrary to Mueller and Dweck (1998), we found no positive effects of growth mindset on response to failure.

We took these null outcomes seriously, and wished to run a third study, exactly replicating study 2, in an independent sample to gather more evidence regarding whether a growth mindset manipulation can improve children's post-failure performance (or if it might even harm it), as well to further explore the role of children's mindsets on performance in this task. This is presented next, and exactly follows the analytic approach used above in study 2.

3.4 Study 3

Study 3 was executed identically to study 2, testing the same hypotheses and under the same ethical consent.

3.4.1 Method

3.4.1.1 Participants

In total, 212 children participated. One male participant was removed from the analyses. This student had consistent exceptionally low grades scoring, for example, 9.2 SDs below the class average for Chinese. Their mindset was 3.75, close to the class average. Of the 211 remaining participants, 120 were boys (mean age 10.78 years, SD = 0.58) and 91 were girls (mean age 10.60 years, SD = 0.46).

3.4.1.2 Materials

The materials used in study 3 were identical to those used in study 2.

3.4.1.3 Design

The experiment design was identical to those in study 2.

3.4.1.4 Procedure

Numbers in the fixed mindset, growth mindset and active control condition were 70, 71, and 70 respectively. All procedures were identical to those of study 2.

3.4.2 Results

As before, we formulated the same four hypotheses listing in study 2. Before testing these hypotheses, we first tested whether children's initial cognitive ability (Trial 1) differed in three mindset manipulation conditions. Again, no significant difference was found ($F(2,208) = 0.747$, $p = .475$).

3.4.2.1 Does mindset manipulation affect post-failure performance on moderately difficult items?

We tested the prediction that the growth mindset condition would improve post-failure performance, relative to the fixed mindset and active control conditions. Again, this was done using a one-way ANOVA with the difference in scores on the initial and final ability tests (i.e., Trial 3 – Trial 1) as the DV and mindset condition as the IV. As in Mueller and Dweck (1998), age and gender were not controlled.

The overall test for differences among the levels of the mindset manipulation was not significant ($F(2,208) = 2.744$, $p = .067$). As in study 2, the contrast of fixed mindset vs growth mindset conditions was non-significant ($\beta = 0.18$, CI_{95} [-0.09, 0.46], $t = 1.32$, $p = .189$). Interestingly, contrary to our study 2 and to Mueller and Dweck (1998), performance in the active control condition was significantly improved ($\beta = 0.32$, CI_{95} [0.05, 0.60], $t = 2.34$, $p = .020$) relative to the fixed mindset condition.

3.4.2.2 Does mindset manipulation affect post-failure performance on more difficult items?

We next tested if the classic growth mindset manipulation might raise performance on more difficult items – the stated purpose of mindset manipulations. As in study 2, a one-way ANOVA was conducted with a difference of cognitive scores (Trial 4 – Trial 2) as the DV, and mindset condition as the IV. As in study 2, no significant effect of the manipulation was found ($F(2, 208) = 0.216$, $p = .806$). A contrast of the fixed mindset versus growth mindset conditions showed no effect ($\beta = 0.03$, CI_{95} [-0.20, 0.25], $t = 0.24$, $p = .810$).

3.4.2.3 Do children's mindsets correlate with post-failure performance on moderately difficult items?

We next tested whether children's mindsets might impact their post-failure performance on the moderate difficulty materials. As in study 2, this hypothesis was tested using a regression model with a difference of cognitive scores (Trial 3 – Trial 1) as the DV and children's scores on the Theories of Intelligence Scale (Dweck, 2000) as the IV. Again, as in study 2, the hypothesis was not supported ($F(1, 209) = 0.179$, $p = .673$; $\beta = -0.02$, CI_{95} [-0.14, 0.09]).

3.4.2.4 Do children's mindsets correlate with post-failure performance on more difficult items?

Finally, we tested if children's mindsets would affect their responses to the more difficult materials by using regression with a difference of initial and final score (Trial 4 – Trial 2) as the DV, and children's scores on the Theories of Intelligence Scale (Dweck, 2000) as the IV. Contrary to prediction, children's mindsets were not associated with their performance on the more difficult materials ($F(1, 209) = 0.020$, $p = .888$; $\beta = -0.01$, CI_{95} [-0.10, 0.09]).

3.4.3 Study 3 discussion

The results did not support any effect of growth mindset on children's post-failure performance, either on moderate or more difficult material. The sole significant beneficial effect in the results was a higher score for children in the active control condition (relative to the fixed mindset condition). While we would not make too much of this finding, it is in the reverse direction to that predicted by the growth mindset theory – the children were primed for a fixed mindset, and this should theoretically have reduced their performance. There was no evidence found for any effects on the more difficult material. Likewise, there was no association of children's mindsets on any outcome.

Next (study 4) we examined the association of children's mindsets with their school grades before discussing the results of all four studies.

3.5 Study 4

Children's mindsets are predicted to enhance educational attainment and a central motivation for mindset manipulations is expected improvements in educational attainment (Blackwell et al., 2007; Dweck, 2006; Paunesku et al., 2015).

As noted in the main introduction, a seminal report supporting the role of children's mindsets on educational attainment was provided by Blackwell et al. (2007). This study reported no association between children's mindset and their mathematics grades on entry, but, controlling for these initial grades, children's mindsets correlated significantly with grades two years later. Moreover, the effect was general (rather than being restricted to students with poor initial performance). As noted above, subsequent studies of this association have yielded reversed results (e.g. $r = -.03$ in a study of over, 5,600 university applicants: Bahník & Vranka, 2017), through to small positive associations with initial grades, or associations only in the bottom 1/3rd of participants (e.g. Paunesku et al., 2015).

To test the relationship of mindset to grades and grade change, we used data on the grades of all children tested in study 2 and study 3. Our expectations for this study, were as follows. First, based on Paunesku et al. (2015) and Zhao and Wang (2014), we predicted a positive association of children's growth mindset with their initial GPA. Second, longitudinally, and following Blackwell et al. (2007), we predicted a positive association of growth mindsets with improvements in grades across a semester. Third, based on Paunesku et al. (2015), we had a subsidiary or more restricted hypothesis that this improvement might be larger for the children with lower initial grades (i.e. a mindset × initial GPA interaction). Fourth, we wished to test if having a growth mindset is associated with higher intelligence test scores. As Dweck stated about the Mueller and Dweck (1998) study, "*Since this was a kind of IQ test, you might say that praising ability lowered the student's IQs. And that praising their efforts raised them*" (Dweck, 2006, p. 73). Additionally, many mindset interventions teach students that their brain is like a muscle and can grow smarter to induce a growth mindset. Thus, we were interested in testing whether in fact the belief that ability can grow if one believes it can, is reflected in the data, i.e., if children who believe they can grow their basic cognitive ability have done so.

We tested these four predicted associations using the children from Studies 2 and 3. Across our two studies, we have a comparable number (433 compared to 373) of children, or a comparable age (around 11 years old in both studies). We observed the children for one semester, rather than two years, but Blackwell et al. (2007) reported a significant effect after just one semester (r = .12). As in Blackwell et al. (2007), children's mindsets were assessed at entry using a questionnaire measure (Dweck, 2000). We recorded not only initial scores on mathematics but also English and Chinese grades.

3.5.1 Methods

3.5.1.1 Participants

Participants were all 433 pupils from studies 2 (n= 222) and 3 (n= 211) as described above.

3.5.1.2 Materials

All children in the sample are formally assessed by their school twice each semester. With permission, we obtained children's grades in their three core classes (English, Chinese, and mathematics) for the semesters preceding and following our mindset measures. This allowed us to test both the association of mindset with initial GPA and change in performance across time. Children's mindsets were assessed using scores on the 8-item Theories of Intelligence Scale (Dweck, 2000) as described in studies 2 and 3. Cognitive ability was assessed using scores on the Trial 3 (set-E SPM) items ascertained in the first phases of studies 2 and 3.

3.5.2 Analyses

To maximise power, and because children's grades in the three subjects correlated highly, we formed a GPA measure for each child for each semester, based on the factor scores on a 1-factor model of grades. For both studies, this 1-factor CFA model of grades fit well (e.g. for study 2 CFI = 1; TLI = 1; RMSEA = 0). Subject loadings on this factor were also high (e.g. 0.80, 0.79, and 0.87 and 0.70, 0.86, and

0.90 for Math, Chinese, and English in semesters 1 and 2 respectively for study 2). Similar results obtained for the children in study 3. Factor-score GPAs were used to test predicted associations of children's mindsets with grades within and across semesters.

3.5.2.1 Do our participants show typical variation and means of mindset scores?

As shown in Figure 3.3, children in studies 2 and 3 displayed the full range of mindset scores, which appeared normally distributed. Mean scores were in keeping with previous reports: compared to the children studied in Blackwell et al. (2007), our children had slightly more fixed mindset, e.g. the mean mindset score in our study 3 was 4.16 (CI_{95} [4.04, 4.29]), compared to 4.45 (*CI* not reported) reported in the 12-year-olds studied by Blackwell et al. (2007). The standard deviation in study 4 suggests little if any restriction of range to suppress associations of differences in children's mindsets with attainment or change in attainment.

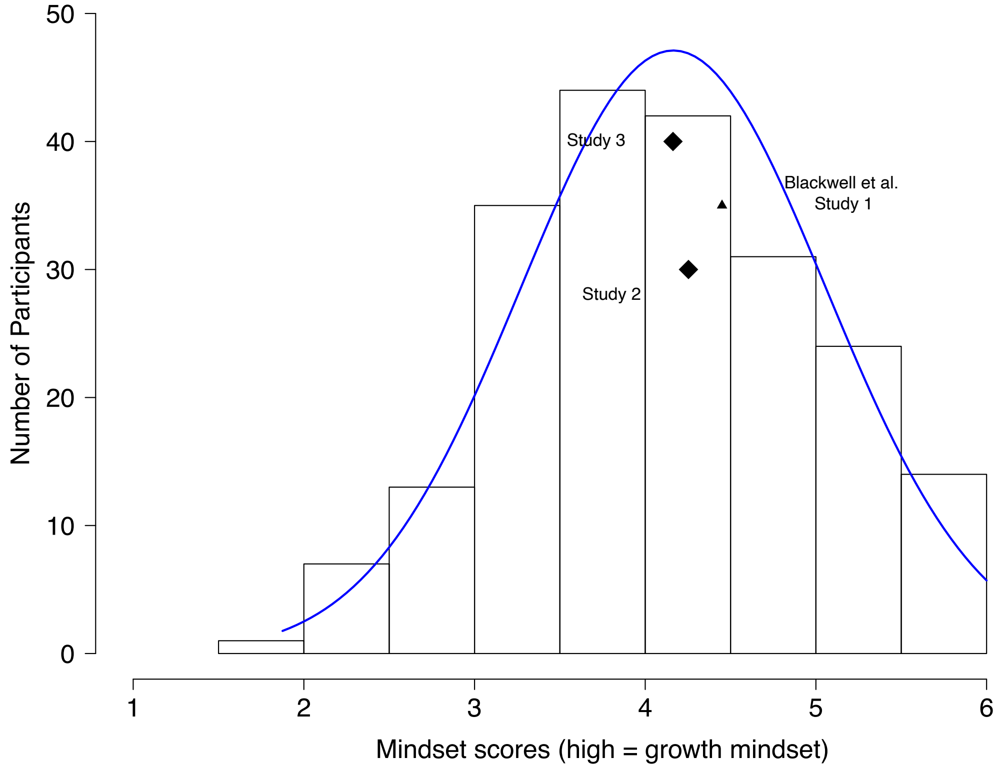

Figure 3.3: Mean mindset (Dweck, 2000) scores for children tested in study 2 and study 3 in the present report, as well as the mean score from Blackwell et al. (2007) study 1 for comparison. A histogram of all scores from the present report, along with a corresponding superimposed normal curve are also presented

3.5.2.2 Do children's mindsets predict grades?

Regression was used to test the hypothesis that school grades would be associated with children's mindsets. As in Blackwell et al. (2007), we did not control for age and gender in this analysis. For children in study 2, mindset was unrelated to initial GPA (β = 0.03, CI_{95} [-0.10, 0.16], t = 0.42, p = .671). Nor were children's mindsets associated with GPA in semester 2 (β = 0.05, CI_{95} [-0.11, 0.21], t = 0.63, p = .530). Thus, our first hypothesis was not supported. Adding cognitive ability to the model left these associations unchanged (e.g. controlling for Trial 1 scores: β = -0.01,

CI_{95} [-0.14, 0.11], t = -0.21, p = .834 and β = 0.01, CI_{95} [-0.14, 0.17], t = 0.18, p = .855 for semester 1 and 2 respectively). Cognitive ability was a highly significant predictor of GPA in both semesters (e.g. controlling for Trial 1 scores, semester 1 β = 0.35, CI_{95} [0.22, 0.47], t = 5.4, p < .001; semester 2 β = 0.25, CI_{95} [0.09, 0.41], t = 3.16, p = .002). In addition, because Blackwell et al. (2007) found a positive association between children's mindsets and math ability in specific, we therefore tested our first hypothesis in single school subject levels instead of averaged as GPA. Again, children's mindsets yielded only null effects on attainment (p-values 0.883 for English, 0.872 for Chinese, and 0.356 for mathematics). Furthermore, a cognitive ability × mindset interaction (testing the hypothesis that growth mindset would translate cognitive performance into greater GPA outcomes in children with lower ability scores), was non-significant for GPA in semester 1 and 2 (β = 0.04, CI_{95} [-0.09, 0.16], t = 0.58, p = .560 and β = 0.05, CI_{95} [-0.13, 0.24], t = 0.60, p = .553 respectively).

Similar null results obtained in the children tested in study 3: children's mindsets failed to predict initial GPA (β = 0.04, CI_{95} [-0.10, 0.17], t = 0.52, p = .601). Nor were they associated with GPA in semester 2 (β = 0.06, CI_{95} [-0.08, 0.20], t = 0.88, p = .382). Adding a control for cognitive ability level did not change these results (e.g. controlling for Trial 1 scores, β = 0.00, CI_{95} [-0.13, 0.13], t = -0.06, p = .955 and β = 0.01, CI_{95} [-0.12, 0.13], t = 0.14, p = .888 for semester 1 and 2 respectively). Once again, cognitive ability scores were strong predictors of GPA in both semesters (e.g. controlling for Trial 1 scores, β = 0.36, CI_{95} [0.23, 0.49], t = 5.38, p < .001 in semester 1). In models substituting single school subjects for GPA, children's mindsets were unrelated to attainment (p values .876, .552, and .504 for English, Chinese, and mathematics respectively). Again, as in study 2, we tested the effect of the cognitive ability × mindset interaction on GPA outcomes. No significant result was found for GPA in either semester 1 (β = - 0.06, CI_{95} [-0.21, 0.10], t = -0.74, p = .458) or semester 2 (β = -0.02, CI_{95} [-0.17, 0.13], t = -0.23, p = .816).

3.5.2.3 Do children's mindsets enhance learning across time?

We next tested the prediction that children with a growth mindset would show GPA improvement (final GPA, controlling for initial GPA), either as a main effect, or, if only children gaining lower scores in semester 1 showing any benefit of their mindsets (Paunesku et al., 2015), as an interaction with initial GPA. This prediction was tested in a regression predicting GPA in semester 2 from children's mindsets and initial GPA × children's mindsets. Age and gender were not controlled.

For children in study 2, neither hypothesis was supported: There was no significant effect of mindset on GPA change (β = 0.03, CI_{95} [-0.06, 0.12], t = 0.63, p = .532). In addition, there was no interaction of children's mindsets (β = -0.06, CI_{95} [-0.19, 0.07], t = -0.87, p = .387). Similarly, in study 3, there was no main effect of children's mindsets on GPA change (β = 0.04, CI_{95} [-0.04, 0.11], t = 1.00, p = .319) and no initial GPA × children's mindsets interaction (β = -0.05, CI_{95} [-0.13, 0.03], t = -1.18, p = .238).

3.5.2.4 Might children's mindsets have highly restricted across-time effects, specific to single school subjects?

We next examined the possibility that children's mindsets may have a highly specific effect, interacting on a course-by-course basis with low semester 1 grades such that while, in most children, their mindsets would be unrelated to grades. For the lowest-performing children in each subject, growth mindsets would trigger the predicted effort and hard work response which would improve grades in that subject by the end of the semester. For the children in study 2, this predicted interaction failed to emerge. In all cases these subject × children's mindsets interaction effects were non-significant (β = 0.03, CI_{95} [-0.09, 0.14], t = 0.47, p = .640; β = 0.01, CI_{95} [-0.08, 0.10], t = 0.25, p = .803; β = 0.06, CI_{95} [-0.06, 0.19], t = 0.99, p = .323 for English, Chinese, and mathematics, respectively). Similarly, for the children in study 3, course-by-course tests for initial grade × children's mindsets effects on final grades also were not supported for any subjects: β = 0.02, CI_{95} [-0.05, 0.10], t = 0.56, p = .578; β = 0.02, CI_{95} [-0.07, 0.12], t = 0.50, p = .620; β = 0.05, CI_{95} [-0.04, 0.14], t = 1.04, p = .297 for English, Chinese, and mathematics respectively.

3.5.2.5 Do children's mindsets predict baseline reasoning ability?

We tested the hypothesis that growth mindset would be associated with higher cognitive ability using regression models, again not controlling for age and gender. Contrary to prediction, children's mindsets were not significant associated with cognitive ability as measured by either the moderate difficulty (β = 0.12, CI_{95} [-0.01, 0.25], t = 1.76, p = .080) or more difficult (β = 0.12, CI_{95} [-0.02, 0.25], t = 1.72, p = .086) baseline tests. Similar results were obtained for the children in study 3: children's mindsets were unrelated to scores on either the moderate difficulty (β = 0.12, CI_{95} [-0.02, 0.25], t = 1.68, p = .094) or more difficult (β = 0.07, CI_{95} [-0.07, 0.20], t = 0.99, p = .322) baseline tests.

3.5.3 Discussion of study 4

We found no evidence for growth mindset promoting higher grades or higher cognitive ability scores. Children's mindsets were unrelated to their initial grades and were unrelated to their change in GPA. Likewise, the possibility that children's mindsets effects could appear, but only in children doing less well at the beginning of the semester (Paunesku et al., 2015), was not supported. We were surprised also to find no association of children's mindsets with cognitive ability scores, as these are stable (Deary, 2012) and we expected the chronic developmental influence of children's belief in the malleability of intelligence to have some association with their manifested ability. The mean ability scores of children with a growth mindset did not differ from those with mixed or fixed mindsets.

To interpret the full set of findings in an integrated fashion, we next synthesise the findings from Studies 1-4 in a brief general discussion.

3.6 General Discussion

Mindset was predicted to have a major influence on determining children's resilience to failure as well as influencing real-world outcomes in the form of school grades. Mindsets and mindset manipulation effects on both grades and ability, however, were largely non-significant, or even reversed from the theorised direction. In study 1 we found a significant effect of the growth mindset condition on post-failure performance. This was not replicated in Studies 2 or 3. In no study did we find any effects on post-failure performance on the more challenging materials (contrary to the prediction from mindset theory). The only significant effect of mindset manipulation across studies 2 and 3 was that in study 3, children in the active control condition showed improved scores on the moderate difficulty material relative to the growth mindset and fixed mindset conditions. As these participants were primed for an implicit fixed mindset, this effect contradicts the idea that beliefs about ability being fixed are harmful. At best, this supports a role for explicit exhortations to exert effort as potentially improving performance on moderate difficulty (but not more difficult) tasks. This effect, however, is predicted by both personality (Roberts et al., 2007) and motivation theory (Locke & Latham, 2002).

Turning to the effects of children's internalised mindsets, we found only one significant outcome, but this was in the reverse direction, with a growth mindset appearing to harm post-failure performance in study 2 (but not in study 3). Finally, in study 4, which examined predicted linkages between children's mindsets and grades and progress in school, we failed to find any support for growth mindsets promoting higher grades, either as a main effect, or in interaction with initial scores, or in subjects in which children were struggling.

In summary, we studied relationships of mindset manipulations as well as children's internalised mindsets on their responses to failure and their school performance. We found little or no support for the idea that growth mindsets are beneficial for children's responses to failure or school attainment. Our findings across multiple substantial studies with active controls as well as real-life outcomes across time suggests mindset has no impact on school grades, response to challenge, or goal orientation. Namely, that implicit mindsets about the nature of intelligence have near-zero effects on grades and no effect on general cognitive ability. In the specific case of responses to failure, neither children's internalised mindset nor activated beliefs about whether intelligence is or is not fixed impacted on performance. The data collected are compatible with an effect of praising hard work on increased effort, but not with any increase in performance on difficult tasks, again, in line with data on incentive effects (Gignac, 2018).

3.6.1 Limitations

One limitation suggested by reviewers regards the ethnicity of our participants, contrasting the discovery samples which were US-based, while our participants were Chinese. Relatedly, a reviewer at another journal hypothesised that our Chinese participants likely had uniformly growth-oriented mindsets due to living in a collectivist culture. They suggested that this would account for the higher PISA (Programme for International Student Assessment) scores in China. As shown in Figure 3.3, however, our participants were not clustered around a growth orientation and were normally distributed across a wide range of scores with a mean in keeping with previous US-based reports. Moreover, ethnic differences have previously been examined and reported as being unrelated to mindset effects in the original mindset studies (Mueller & Dweck, 1998) and mindset theory has also been used for two decades in Asian samples, including seminal papers such as Hong, Chiu, Dweck, and Sacks (1997), and continuing in current reports (Zeng et al., 2016). This suggests an expectation among mindset experts that the theory should work in Chinese participants, and we were unable to find any statement to the contrary. Additionally, the effects of mindset are not couched in terms of ethnicity but in terms of universal developmental processes linking mindsets to realised cognitive and educational attainment. As such they should hold in all children. Related to this question of sample composition, a reviewer hypothesised that our Chinese participants were too wealthy to show the effects. As noted in the participants description in study 1, our participants were not wealthy and, in fact, were significantly impoverished, even relative to the Chinese median income. Low socioeconomic status is predicted to increase, not nullify the influence of mindset on performance (Claro et al., 2016). Finally, a reviewer suggested that the experience of failure may have been insufficiently severe to elicit effects of mindset. As perception of failure is largely a matter of feedback, we disagree: Participants in the study were often distressed to receive such negative feedback.

Our samples, then, appear suitable for revealing mindset effects if they exist: our participants were children near-identical in age to those reported in Mueller and Dweck (1998), they lacked material resources – argued to magnify mindset effects (Claro et al., 2016) – and showed a range of mindset scores and attainment scores. Rather than being uniformly growth-oriented, the sample showed a full normal range of mindsets and was slightly more fixed-minded on average than in previous samples. This, again, should have increased our power to create group-differences in the mindset manipulation studies, and the wide variation in mindset should have revealed similarly large effects of mindset on responses to failure and in educational attainment. The failure to show significant growth versus fixed mindset condition effects in the lab or effects of mindset on grades appears to be strong evidence against mindset theory.

3.6.2 Future directions

As the purpose of mindset manipulations in school is to impact how children are taught (Paunesku et al., 2015), given these null outcomes, additional independent studies testing the theory are needed. Other outcomes attributed to mindset should also be tested for replicability, e.g. the role of mindset on willpower (Job, Walton, Bernecker, & Dweck, 2013), as well as claims about the general applicability of the theory to domains broad as personal relationships and sporting success (Dweck, 2006). Future work on mindset should remove the confound of encouraging hard work and conscientiousness - which is a known influence on attainment (Rosander &

Backstrom, 2014). Additionally, since experimenter expectations can significantly alter experimental results (e.g., Doyen, Klein, Pichon, & Cleeremans, 2012), a double-blind experimental design could be considered in further studies. Also, further studies should use “strong tests” (cf. Platt, 1964) to test for the effect of the praise manipulation used in Mueller and Dweck (1998). A strong test would enable independent researchers to easily disprove the theory if, in fact, it is false (Platt, 1964). One such test may be to apply one praise condition (e.g., praise for being smart) on a group of participants and then switch to apply the other condition (e.g., praise for hard work) on the same group of participants, we should expect to see that participants' performance would change when the conditions are switching. Finally, given widespread and costly policy and real-world educational implications, we encourage an ‘emptying of the file drawer’ to account for non-reported studies.

For the majority of teachers who report believing mindset matters, 80% of whom say they have been unable to make effective changes in their own classes (Yettick et al., 2016), the present results may provide a simple answer to this apparent disparity: learning does not require (Finn et al., 2014) or cause (Ritchie et al., 2015) changes in basic ability, but does require prosaic teaching practices such as systematic practice and feedback via appropriate testing (Lindsey, Shroyer, Pashler, & Mozer, 2014).

3.6.3 Context of the research

Across over 600 children we found no evidence to support mindset theory. The children were 9-13 years old, living in poverty, and had a normal distribution of mindsets, all of which should have increased the chances of observing impacts of mindsets if they existed. Instead, we found that the children's naturally held mindsets did not predict performance on cognitive tests, grades, or improvement in academic achievement. This lack of a relationship persisted for low-achieving children. Further, we found no evidence that a growth mindset condition improved children's performance on cognitive tests following failure. In all cases, including examinations of low-achieving sub-groups, we found that growth mindset either had no effect on performance or appeared to be explained by motivation to work hard rather than beliefs about the malleability of intelligence (i.e., mindset). We encourage further independent studies to test mindset theory and suggest controlling for confounding variables such as experimenter demands and effort encouragement.

Chapter 4. Does growth mindset promote better educational attainment among undergraduate students?

The last chapter introduced four empirical studies testing the effects of mindset on children's cognitive performance after challenges and children's educational attainment in primary school. This chapter reports two empirical studies testing the association of mindset and children's educational attainment across a challenging transition from high school to university. These studies also generated a paper which has been submitted to *Intelligence*. I gratefully acknowledge the assistance and contributions of student project workers to study 1 reported here, especially Adrian Carr, Lidia Brookmann, Martine McKenna, Kari Taylor, and Jessica Lane. To maintain the consistency with the published version and to acknowledge the collaborator's effort for its publication, the personal pronoun "we" is used in this chapter.

4.1 Introduction

Developing interventions to raise students' educational attainment is of high interest for educational psychologists. Mindset theory, an influential model in this field, suggests that students who believe that basic intelligence is fixed have greatly reduced educational attainment (Blackwell et al., 2007). The theory has influenced teaching (Yettick et al., 2016), business (Bock, 2015), philanthropy (e.g., Gates, 2015), and the public mind (Dweck, 2006). Despite this wide influence, the central predictions of the theory regarding the effect of beliefs about intelligence on educational attainment have been subject to little independent replication. Here we report two studies testing if growth and fixed beliefs about intelligence are, in fact, associated with educational attainment, focussing on prediction of university grades across a challenging transition.

A growth mindset refers to the belief that one can substantially change one's basic ability. This contrasts with a fixed mindset, which refers to the belief that, while one can learn, basic ability is fixed (Dweck, 2006). The proposed mechanism linking these beliefs to educational attainment involves how individuals interpret and react to challenges. Believing basic intelligence is fixed is predicted to cause people to avoid attempting difficult tasks as, the theory suggests, these can only make one look bad (Yeager & Dweck, 2012). Instead, fixed mindset individuals are predicted to "document" their success by engaging in tasks with low likelihood of failure (Yeager & Dweck, 2012). These beliefs about the nature of intelligence are predicted to have both immediate and long-term effects. In the short term, growth mindset is predicted to raise task performance, e.g., on an IQ test, after a challenge (Mueller & Dweck, 1998), but see Y. Li and Bates (2019). In the long term, mindset is predicted to improve educational attainment, at least across a challenging transition (Blackwell et al., 2007). Here we focused on whether beliefs about basic ability show this predicted link to educational attainment.

Mindset is readily measured, with common scales using as few as 2-items such as "*You can change even your basic intelligence level considerably*" (Blackwell et al., 2007; Dweck, 2000; Haimovitz & Dweck, 2016). These scales have been widely used in subsequent studies to test the association between mindset and educational attainment, but the results have been inconsistent. (e.g., Bahník & Vranka, 2017; Bazelais et al., 2018; Blackwell et al., 2007; Claro et al., 2016; Y. Li & Bates, 2019). For example, Blackwell et al. (2007), study 1, reported that growth mindset predicted improvement in mathematics grades across what was described as a "challenging transition" to junior high school (β = 0.17, t(372) = 3.40, p < .05). By contrast, Y. Li and Bates (2019) found non-significant associations of mindset and improvement in grades among primary school students (β = 0.03, CI_{95} [-0.06, 0.12], $t = 0.63$, $p = .532$ for their study 2 and $\beta = 0.04$, CI_{95} [-0.04, 0.11], $t = 1.00$, $p = .319$ for their study 3). In addition, Bazelais et al. (2018) reported that mindset was not associated with average grades in college after controlling for high school grades ($F(2, 293) = .265$, $p = .767$, partial $\eta^2 = .002$). More strikingly, Bahník and Vranka (2017) found a negative association between growth mindset and scholastic aptitude among university applicants. These findings raised one question in our mind: Does the association between mindset and grades exist, either on average or across a challenging transition? We conducted two substantial studies to explore these questions. The data presented constitute all data we have collected on mindset and grades. These results, we hope are of direct value, and will be of use in future meta-analyses.

Study 1 tested the simple association between mindset and self-reported university grades. Study 2 tested the association of mindset with official grades, beginning with entrance qualifications, and continuing across the challenging transition from high school to first year university grades, and on through each year to graduation. In study 2 we were also able to test for a predicted interaction of mindset with entry grades, such that the effects of mindset are restricted to students who are most likely to struggle when entering university (Paunesku et al., 2015). Both studies were approved by the Psychology Research Ethics Committee at the School of Philosophy, Psychology and Language Sciences, University of Edinburgh.

4.2 Study 1: Testing association of mindset and educational attainment in 246 undergraduates.

In study 1, we examined the association between undergraduates' mindset and their grades in university. Many students find university to be challenging, especially in the first year. The big transition from high school to university brings difficulties to students both in learning and living, which further causes a high dropout rate in the following year. For example, 6.3% of students who enrolled in UK universities in the academic year 2016/2017 later dropped out altogether (HESA, 2019). The subsequent years in university are also challenging to students seeking grades required to enter their preferred honours course, and involve a steep learning curve year on year, as new and more complex tasks such as scientific writing, statistics, dissertation projects, and the self-management required to complete course work.

In study 1, we tested the association of mindset with grades, hypothesising that students with a growth mindset would have better grades than those with a fixed mindset. Regarding the effect size, since university provides what is widely reported as a highly challenging transition (Briggs, Clark, & Hall, 2012), often associated with significant learning-related stress and risk of dropout, we hypothesised that we would find an association of at least $r = .2$ effect, comparable to that reported by Blackwell et al. (2007). The study had power of 88% to detect an effect of this size (two sided). Ideally, one would control for university entry grades but in this initial study we simply examined the raw association of mindset with grades among students in the midst of the challenging transitions presented by university (note: control for entry grades and record-based grades for each year were added in the much larger study 2 reported below).

Another factor that shares some similar constructs with mindset in changing student's motivation and behaviour is locus of control (Nallapothula et al., 2020). Locus of control refers to whether individuals tend to attribute their success and failure to internal (own) or external (environmental) factors (Rotter, 1966). For those who tend to attribute their success and failure to internal factors, they believe that they can achieve some improvements by exerting effort and have positive attitudes towards failure, which resemble the growth mindset beliefs (Burgoyne, Hambrick, Moser, & Burt, 2018; Nallapothula et al., 2020). By contrast, for those who tend to attribute their success and failure to external factors, they believe that exerting effort would not change their achievement, which resemble the fixed mindset beliefs (Burgoyne et al., 2018). The theoretical correlation between mindset and locus of control has been verified in an experimental study: growth mindset is significantly positively correlated ($r = 0.24$, $p < 0.01$) with internal locus of control (Burgoyne et al., 2018). Since locus of control shares some features with mindset theory, when exploring the effects of mindset on student's educational attainment, it is necessary to distinguish the effects of locus of control from mindset effects. Therefore, in addition to mindset, we included a locus of control scale (Findley & Cooper, 1983) in this study.

We also included two additional traits that have been suggested to be associated with educational attainment: Self-esteem (Leary & Baumeister, 2000) and grit (Duckworth, Peterson, Matthews, & Kelly, 2007). We hoped that controlling for some of these non-mindset traits might enhance power to detect a true association between mindset and grades, since this approach would help us to distinguish the possible effects of these traits from mindset effects. In addition, we wished to test the simple associations between these non-mindset traits and grades in the present data. Our second hypothesis was thus that including self-esteem, locus of control and grit as covariates would allow the predicted association between mindset and educational attainment to emerge more strongly.

4.2.1 Methods

4.2.1.1 Participants

In total, 308 students were invited to take part in the study. All students were in their second or subsequent year of the undergraduate participation program at a single UK University. Of 246 students who consented, 68 were men and 178 were women (mean age = 21.43, SD = 4.27).

4.2.1.2 Materials

Mindset was assessed using the 8-item Theories of Intelligence scale (Dweck, 2000). Example items include "*You have a certain amount of intelligence, and you can't really do much to change it.*" Responses were recorded on a Likert scale (from 1: strongly agree to 6: strongly disagree). Responses were reversed where appropriate and summed to form a mindset score for each participant with high scores indicating a growth mindset. Self-esteem was measured using the 10-item self-esteem scale (Rosenberg, 1965). Locus of control was measured using the 29-item locus of control scale (Rotter, 1966) and grit was measured using the 8-item short grit scale (Duckworth & Quinn, 2009). Students entered their most recent year's final letter grades (received 4 months prior to testing). These were recoded into numerical scores corresponding to the university grade bands used in study 2 (ranging from 0-100).

4.2.1.3 Procedure

Students completed consent, and then proceeded to complete the online survey. This included demographic information comprising age and gender. This was followed by the Theories of Intelligence scale, Rosenberg self-esteem scale, Rotter's locus of control scale, and the grit scale followed by the letter grade received for each course, giving their most recent year's grade in each case.

4.2.2 Results

We first tested the hypothesis that growth mindset would be positively associated with higher grades. This was tested using linear regression, with average grade as the dependent variable and mindset as the independent variable. Contrary to prediction, growth mindset was associated with worse, not better grades (β = -0.02, CI_{95} [-0.16, 0.12]) and the effect was not significant (t = -0.26, p = .792; also see Figure 4.1 and Table 4.1). Adding age and gender as covariates did not change the null association of mindset and grades (β = -0.02, CI_{95} [-0.16, 0.12], t = -0.28, p = .778).

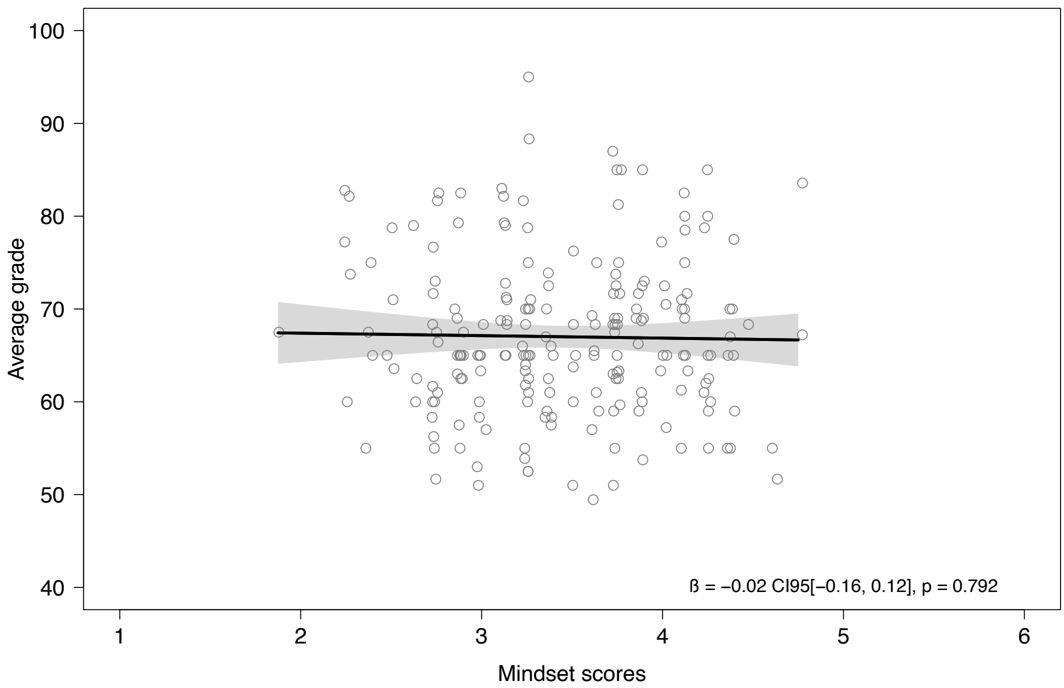

Figure 4.1 The association between students' mindset scores and average grade in university in study 1. Confidence bands indicate 95% confidence intervals

We next tested association of the other scales measured in this study, testing if they were associated with grades, and whether including them might reveal an association of mindset with grades. Grit and locus of control were not associated with grades (β = 0.01, CI_{95} [-0.13, 0.15], t = 0.17, p = .864 and β = -0.01, CI_{95} [-0.16, 0.13], t = -0.15, p = .878 respectively). By contrast, self-esteem was significantly associated with grades (β = 0.26, CI_{95} [0.11, 0.40], t = 3.54, p < .001), supporting theories linking this trait to academic achievement, perhaps as an effect rather than a cause (Leary & Baumeister, 2000). Importantly, adding these covariates did not change the association of mindset with grades (β = 0.00, CI_{95} [-0.14, 0.13], t = -0.03, p = .973).

Table 4.1 The similarities and differences between Blackwell, Trzesniewski & Dweck (2007) study 1 and the present studies 1 and 2

	Blackwell, Trzesniewski & Dweck (2007)	**The present paper**	
	Study 1	**Study 1**	**Study 2**
Participants	N= 373 (198 girls and 175 boys)	N= 246 (178 women, 68 men)	N= 582 (446 women, 136 men)
Age	7th grade students	University students (Mean age = 21.43, SD = 4.27)	University students (Mean age =21.88, SD = 3.18)
Source	One public junior high school in New York city.	One research-intensive university in the UK.	One research-intensive university in the UK.
Educational attainment measurement	6th grade math scores 7th grade math scores 8th grade math scores	One-year grade (received approximately 4 months prior to testing)	University entry grades and grades in each year of university
The key challenging transition	6th grade to 7th grade	High school to university	High school to university

Other challenging transitions	7^{th} grade to 8^{th} grade	1^{st} – 2^{nd} year transition 2^{nd} – 3^{rd} year transition 3^{rd} – 4^{th} year transition	1^{st} – 2^{nd} year transition 2^{nd} – 3^{rd} year transition 3^{rd} – 4^{th} year transition
Mindset scale	Implicit Theories of Intelligence Scale for Children (Dweck, 2000, p.177)	Theories of Intelligence Scale (Dweck, 2000, p.178)	Theories of Intelligence Scale (Dweck, 2000, p.178)
Number of items in the mindset scale	6 items	8 items	8 items
Average mindset score	4.45 (SD = 0.97)	3.46 (SD = 0.60)	3.92 (SD = 0.92)
Is mindset associated with grades across a challenging transition?	Yes	No	No
Statistical results	$\beta = 0.17$, $t(372) = 3.40$, $p < .05$	$\beta = -0.02$, $CI_{95}[-0.16, 0.12]$, $t = -0.26$, $p = .792$	$\beta = -0.05$, $CI_{95}[-0.14, 0.05]$, $t = -0.95$, $p = .345$

4.2.3 Study 1 discussion

The major finding of study 1 was a lack of association of mindset with grade. Indeed, the (non-significant) effect was in the reverse direction to that predicted. This null result was robust to covariates. A similar lack of association of mindset with grades was reported by Blackwell et al. (2007), but these authors reported a significant association of mindset once students were progressing across the “challenging transition” into 7th grade. This lack of an association is also in-line with a growing set of studies reporting no support for a relationship between growth mindset and better educational attainment (e.g., Bahník & Vranka, 2017; Bazelais et al., 2018). Moreover, the slight negative association has been found in other samples (Y. Li & Bates, 2019). We can be increasingly confident, then, that mindset does not lead to any average increase in grades, with similar results in the present study, and other studies of the cross-sectional association of mindset and attainment, including even Blackwell et al. (2007). As the study covers what many students find to be a highly challenging transition, the null association is, however, even more surprising, and contrary to mindset theory.

Selection effects can potentially influence effect sizes. If it were the case, however, that students with a growth mindset are more likely to apply for university, mindset scores in our participants should tend to be high. But they are not: participants’ scores in our study 1 were normally distributed with a mean of 3.46 (SD = 0.60), i.e., more fixed than was reported for Blackwell et al. (2007), see Figure 4.2. It is unlikely therefore that the null effect in our study 1 is due to selection. A second possibility is that our results would have been significant if, like Blackwell et al. (2007), we were able to control for entry grades. Also, in this study we relied on self-reported grades rather than official transcripts to measure educational attainment. Self-reported grades have somewhat lower construct validity than transcripts (Kuncel, Credé, & Thomas, 2016). If the challenging transition of university activates effects of growth mindset promoting educational attainment, then this should have been apparent as a main effect of mindset on attainment, independent of initial grades.

To address these limitations, we conducted a second major study, in which we controlled for transcript-based entry grades, and recorded mindset in year 1 along with transcript-based grades across each of the next four years for three near-complete cohorts of students entering a psychology program. These changes allow us to clarify whether the null result of study 1 reflect the use of self-reported grades or absence of control for entry grades, or if there is in fact null or very slight negative association of mindset with grades across a challenging transition.

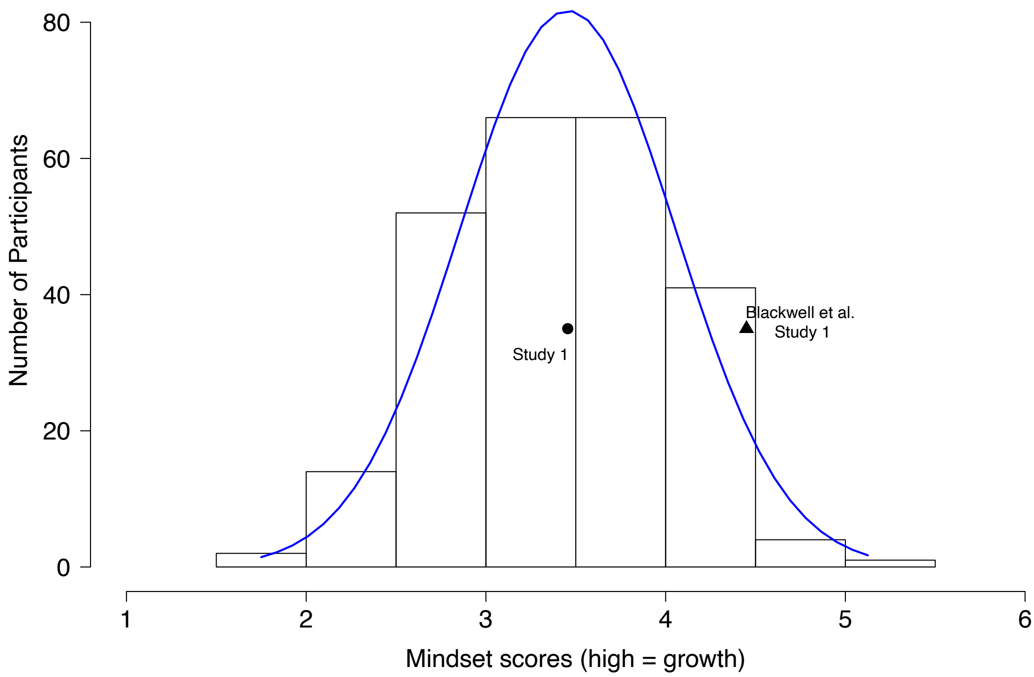

Figure 4.2 Distribution of mindset scores observed in study 1, with mean score plotted, along with the mean mindset score for Blackwell et al. (2007), study 1

4.3 Study 2: Does mindset associate with undergraduate students' performance across a challenging transition (high school to university) and beyond?

In study 1, we tested the association of mindset with university grades, finding a null result. In study 2, as noted above, we set out to test this association using three near-complete cohorts of students enrolled in an introductory psychology course, with available transcript records of their entry grades and their average grade in each year of university. This offered not only greater power (n = 586; 99% power to detect an effect of r = .2) and improved measurement precision, but an opportunity to test whether mindset relates to specifically to first year grades (putatively the greatest challenge is the transition from high school to university), but also across the challenge of entrance into an honours program. Finally, with entry grades available, we could also test the hypothesis that mindset effects are magnified in students entering with low grades (Paunesku et al., 2015).

Each of the university years is challenging for students. The first year in university is the perhaps particularly challenging. Compared to high school, university carries a much heavier work load, and students have to work not only in the class but also outside of class and with far less supervision. The second year is a threshold year for students, as grades in second year determine whether they are able to progress to the honours program, or have to take out a non-honours bachelor degree. The third year, again, presents unique challenges. The material studied increases sharply in complexity such as learning foundational knowledge to in-depth subject-specified knowledge and students undertake new and challenging assessments including a literature review and running a project. In fourth year, students are again challenged, now needing to complete a major dissertation project. Thus, each of the four years represents a distinct challenge, with the first year being perhaps the closest match to the challenging transition for our students as identified by Blackwell et al. (2007).

Our hypotheses were as follows. First, it is not clear whether there is an association between mindset and entry grades, but we were interested in testing this. Therefore, the first hypothesis tested was that mindset would be associated with students' entry grades. Second, following Blackwell et al. (2007), we predicted that growth mindset would be associated with higher grades at the end of year 1 of university, controlling for student's entry grades (hypothesis two). Extending this hypothesis, because, as noted above, each year of university presents an escalating series of challenges to students, we predicted positive associations of year 2, 3, and 4 grades with growth mindset, controlling for entry grades (hypotheses three, four, and five). Finally, and following the "current era" model of growth mindset (Dweck & Yeager, 2019; Paunesku et al., 2015), we predicted that mindset would most strongly predict grades in those students encountering the greatest challenge (those joining the university with the lowest entry grades), i.e., a negative interaction between mindset and entry grades predicting students' first year grades (hypothesis six).

4.3.1 Methods

4.3.1.1 Participants

Our participants consisted of 586 undergraduate students entering and completing a bachelor's degree at a research-intensive university. Four students were recorded as having first-year numerical grades more than 5 SDs below the mean of the sample (due to personal circumstances). These participants were removed from the analyses. In total, 446 female and 136 male students were studied across their 4-year degree (mean age 21.88, SD = 3.18).

4.3.1.2 Materials

Mindset measure: Mindset was measured using the 8-item Theories of Intelligence scale (Dweck, 2000). Grades: When applying for university, students' high school exit qualifications were made available. These consist of grades from a range of national tests (e.g. A-level). The letter grades that students achieved were converted to uniform numerical scores based on the tariff table provided by the Universities and Colleges Admissions Service (UCAS, 2019). Grade was calculated for each student for each year using course records (range from 0-100).

4.3.1.3 Procedure

Students provided consent and completed the mindset scale in the first semester of their degree. Further consent was gained from the Psychology Research Ethics Committee at University of Edinburgh to access required records. Thus, a data frame consisting of students' mindset scores, entry grades, average grade in each year of university was assembled.

4.3.2 Results

As in Blackwell et al. (2007), we first tested if students' pre-challenge grades (i.e., entry grades) were associated with their growth mindset. This was done using a linear regression, with entry grades as the dependent variable, mindset as the predictor, and controlling for age and gender. Growth mindset was not significantly associated with students' entry grades (β = -0.01, CI_{95} [-0.11, 0.09], t = -0.24, p = .808).

Next, we tested hypothesis two, that growth mindset would be associated with higher grades at the end of year 1 of university, controlling for entry grades (i.e., with change in grades across this challenging transition). This was again tested using a linear regression with first year grade as the dependent variable, mindset scores as the independent variable, and entry grades as covariate. Growth mindset was slightly negatively linked to change in grades (β = -0.05, CI_{95} [-0.14, 0.05]), in the reverse direction to expectation, and non-significant (t = -0.95, p = .345). Controlling for student's age and gender did not change the null association of mindset and year 1 average grade (β = -0.04, CI_{95} [-0.14, 0.06], t = -0.81, p = .417).

We next tested whether growth mindset would be associated with grades in years 2, 3, and 4, controlling for entry grades (hypotheses three, four, and five). This was tested using three linear regressions, with average grade in each of years 2 3 and 4 as the dependent variable respectively, mindset scores as the independent variable, and, again, controlling for entry grades. For second year, the effect of mindset was non-significant (β = 0.02, CI_{95} [-0.10, 0.13], t = 0.30, p = .763). Likewise, at third year no significant effect was found (β = -0.04, CI_{95} [-0.19, 0.11], t = -0.54, p = .591), and this was repeated at fourth year, where the effect of mindset was again estimated as negative (β = -0.04, CI_{95} [-0.22, 0.14]) and non-significant (t = -0.48, p = .634). In summary, no support for any significant association of growth mindset and grade was found either across the most challenging transition year from high school to university, nor at any subsequent year in university. Indeed, growth mindset was negatively associated with grades in year 1, year 3 and year 4 (results were not statistically significant; also see Figure 4.3).

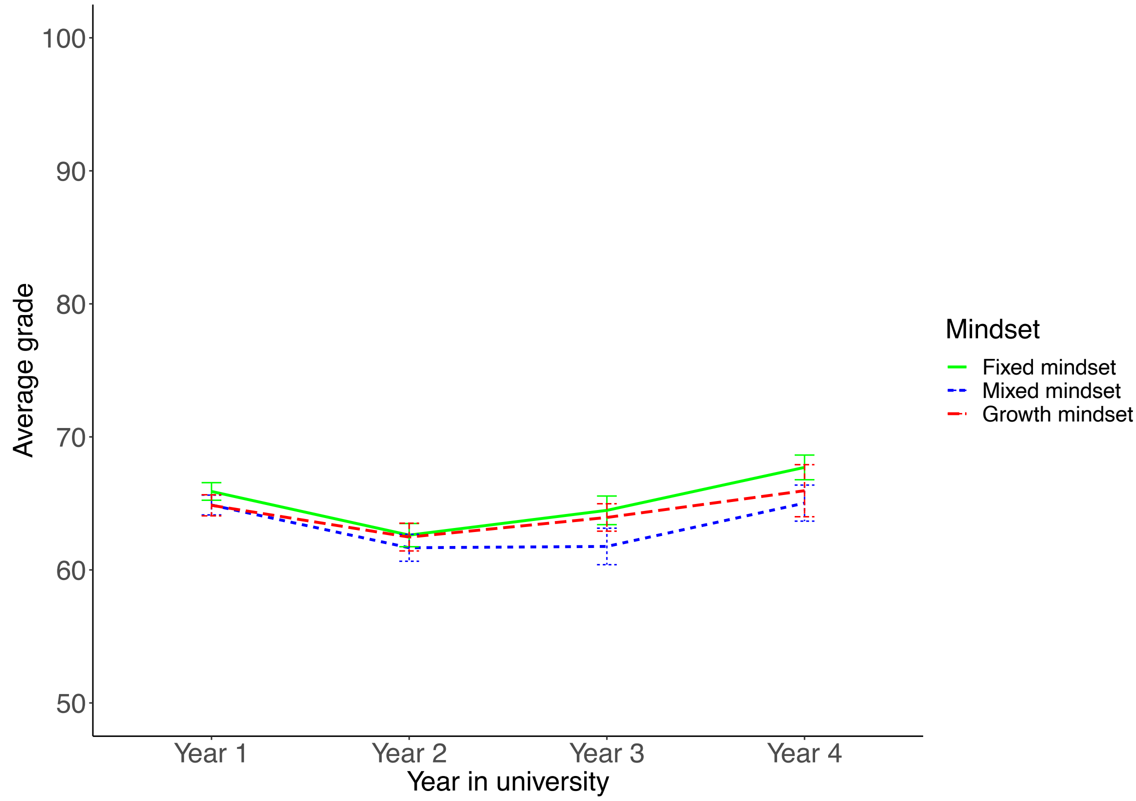

Figure 4.3 Grade trajectories across year of university in study 2 separately for students with fixed, mixed, or growth mindset. Note: Mindset was binned into three quantile groups: fixed (mindset <= 30); mixed (30> mindset <36); and growth (mindset >= 36). The sample sizes for fixed, mixed and growth mindset groups were n= 223, 190, and 169 respectively. Error-bars show the standard error of measurement at each time.

Finally, we tested hypothesis six, that mindset would predict grades for those students encountering the greatest challenge (those joining the university with the lowest entry grades). This was tested using a linear regression with average grade at the end of first year in university as the dependent variable, mindset and the interaction of mindset and entry grades as the independent variables, and entry grade as covariate.

Contrary to prediction, the interaction of mindset and entry grades was not significant (β = -0.03, CI_{95} [-0.12, 0.07], t = -0.54, p = .592).

4.3.3 Study 2 discussion

Study 2 yielded four main findings of interest. First, as did Blackwell et al. (2007), we failed to find support for an association of mindset with initial grades (in our case entry grades) taken prior to a major transition. Second, unlike Blackwell et al. (2007), we failed to find the predicted association of growth mindset with improved grades across the challenging transition from high school to university. Instead, the estimated association was non-significant and in the reverse direction to that predicted by mindset theory. Third, we found no support for association of mindset with change in grades at any subsequent challenging transition. Fourth, and contrary to the prediction that mindset would be especially effective in participants with low initial grades (Paunesku et al., 2015), we found no interaction of mindset × low entry grades on improvement in grades across the challenging transition to the first year of university. Our results thus did not support any association of growth mindset with educational attainment. We discuss these four findings briefly before concluding with a joint discussion of the impact of both studies for mindset theory.

Similar to participants in study 1 in Blackwell et al. (2007), our participants in study 2 were tested across a challenging transition. Their entry grades were controlled, and their grades across a university degree were known. Our participants were thus a suitable sample to detect a significant positive effect of growth mindset on grades if present, but no such effect emerged. Instead, we found a non-significant negative association of mindset with grades across the most challenging transition from high school to university, which is consistent with our null finding in study 1, and our previous work (Y. Li & Bates, 2019). The subsequent years in university are also challenging, but the associations of mindset with grades across those transitions repeated the null result across the most challenging transition. These findings are incompatible with mindset theory and could not support the prediction that growth mindset activates behaviours that causes better grades even across a challenging transition. The lack of interaction (mindset x low entry grades) effect on year 1 grades is consistent with our null main effect. The strongest benefit of growth mindset should emerge in this interaction effect, but it did not. We next discuss the overall findings.

4.4 Joint discussion

Mindset has often been presented as playing a critical role and having a powerful impact on educational attainment (see Sisk et al., 2018). Across two studies, we tested the association between mindset and educational attainment (total 832 undergraduate students). Study 1 used the standard mindset scale to test whether having a growth mindset was associated with better grades. No support was found for this prediction. Study 2 tested whether growth mindset predicted higher university grades across a series of challenging transitions (high school to university, and transitions within university), also examined any effects were apparent in those who achieved relatively low score at entry. Mindset, however, was not significantly associated with grades at any point. In addition, all effect sizes (except the second year) we obtained were in the reversed direction to the predicted associations in mindset theory. Likewise, growth mindset did not significantly predict higher grades even among students who achieved relatively low score at entry.

In study 1, we had a smaller sample, with self-reported grades, no entry grades available, tested across only one year, and not always for the initial transition year into university (which is predicted to activate mindset most strongly). In study 2 we had a large sample, with transcript documentation across the challenging transition (from high school to university), and a series of transitions subsequent to this. Thus, our participants in study 2 are well suited for a strong test of the proposed association between mindset and educational attainment. Though an effect of mindset on grades similar, or even larger than the mindset effect reported by Blackwell et al. (2007), was predicted, we failed to find evidence of any effect. Despite one of the most important implications of mindset theory for real-life outcomes being that growth mindset promotes educational attainment, we found no evidence for substantial (or significant) effects of growth mindset on better educational attainment. The lack of relationship between mindset and educational attainment is in keeping with Sisk et al. (2018), our own previous studies (Y. Li & Bates, 2019) and those of others (Bahník & Vranka, 2017; Bazelais et al., 2018; Sriram, 2014).

In trying to understand the differences between our results and those of Blackwell et al. (2007), two explanations seem relevant. First, growth mindset may activate learning across a challenging transition, but such effects may wash out very rapidly. Therefore, if researchers test for mindset effects after the transition, the effects would be absent. However, this explanation is inconsistent with previous findings that mindset has enduring (at least one academic year) effects on educational attainment (Blackwell et al., 2007; Gunderson et al., 2018; Park, Gunderson, Tsukayama, Levine, & Beilock, 2016). An alternative explanation is that mindset is not reliably associated with grades even during a challenging transition.

One point of interest is that both study 1 and 2 confirmed that mindset is not associated with grades prior to (or indeed during) a period of transition or challenge.

This finding is itself one which should be a reason for caution (Munafo, Zammit, & Flint, 2014). Mindset theory predicts beneficial effects, activated most strongly under challenge (Blackwell et al., 2007; Dweck, 2006; Yeager & Dweck, 2012). Test scores confirm that, at any given time, many individuals struggle with learning, and are encounter significant challenges. Thus mindset ought to be generating a significant main (average) effect on grades: Whenever some individuals are experiencing a learning challenge, mindset should be improving the outcomes of those in the challenged group who have a growth mindset, leading to a main effect (see also Munafo et al., 2014 for a similar argument in psychiatry). The lack of a main effect at baseline in studies like ours or those of Blackwell et al. (2007) is, thus, counter to the statistical expectation of main effects in the presence of an unmeasured interaction. The lack of any support for an interaction (indeed the interaction was estimated in the wrong direction in our study 2) is still further evidence against mindset working to raise grades, either in general or for a more limited time.

In summary, we find that mindset does not appear to influence educational attainment. Having a growth mindset did not promote educational attainment across a challenging transition from high school to university, nor in a series of transitions in university. In addition, for students who encountered the greatest challenge when entering university, growth mindset did not increase their educational attainment. More importantly, our results suggested that having a growth mindset might harm student’s educational attainment. Thus, beliefs about the nature of intelligence appear to be simply that: knowledge about a scientific topic and unrelated to learning.

Chapter 5. Two empirical studies testing whether parents' failure mindsets are associated with children's intelligence mindsets

The last two chapters reported two published papers testing the effects of mindset on children's cognitive performance and their educational attainment in primary school and university. Since no significant mindset effect was found on either children's cognitive performance or on children's educational attainment, I concluded that the mindset theory might be incorrect. This chapter aims to examine the development of children's mindset to test if the mindset theory is correct or not. If the development of the mindset theory is incorrect, the mindset theory would consequently have some problems. Haimovitz and Dweck (2016) suggested that the development of a child's mindset was related to their parent's attitudes to failure, rather than what their parent's believed about intelligence. I wish to test the prediction of Haimovitz and Dweck (2016) in this chapter. To distinguish the concepts of attitudes to failure and beliefs about intelligence, I refer them as failure mindset and intelligence mindset respectively in this chapter. I gratefully acknowledge the assistance and contributions of Eilidh Kirkpatrick to study 2 reported here.

5.1 Introduction

In the last two decades, mindset researchers (e.g., Blackwell et al., 2007; Claro et al., 2016; Dweck, 2006; Hong et al., 1999; Mueller & Dweck, 1998; Yeager & Dweck, 2012) have generated a lot of interest in intelligence mindsets (whether one's intelligence is fixed, or can be greatly changed). These researchers have investigated the potential effects of intelligence mindsets on learning outcomes such as school grades (Claro et al., 2016; Paunesku et al., 2015), and explored the possible development of intelligence mindsets in children's development (Haimovitz & Dweck, 2016; Park et al., 2016). It has also been reported that both teachers and parents play an important and influential role in the development of children's intelligence mindsets (Dweck, 2014; Haimovitz & Dweck, 2016; Park et al., 2016). For example, teachers' instructional practices were found to have significant effects on children's intelligence mindsets (Park et al., 2016). For teachers who were more likely to focus on children's performance or intelligence (the performance orientation), their students were more likely to have a fixed intelligence mindset; whereas the students who had teachers who focused on learning strategies or efforts (the learning orientation) rather than performance were more likely to have a growth intelligence mindset (Park et al., 2016). The relationship between learning versus performance orientation and intelligence mindset has also been tested in other studies (e.g., Blackwell et al., 2007; De Castella & Byrne, 2015; Elliott & Dweck, 1988; Yeager & Dweck, 2012).

Turning to parental influence on children's development, a recent study by Haimovitz and Dweck (2016) reported that children's intelligence mindsets were not associated with parents' intelligence mindsets, but significantly associated with parents' failure mindsets (whether parents believe failure is debilitating or enhancing). In study 1 of Haimovitz and Dweck (2016), 73 parent-child dyads were recruited from two schools in San Francisco. Parents were asked to report their intelligence mindsets, their failure mindsets, and their perceptions of children's competence in school via an online survey. Children (66% were in the 5^{th} grade) were asked to report their intelligence mindsets and their perceptions of parents' learning versus performance orientation. Haimovitz and Dweck (2016) found that children's intelligence mindsets were not significantly associated with their parents' intelligence mindsets ($\beta = 0.17$, $p = .162$), but were significantly associated with their parents' failure mindsets ($\beta = 0.24$,

$p = .38$). They also tested the association between parents' failure mindsets, parents' intelligence mindsets and children's perceptions of parents' learning versus performance orientation. For parents who believed failure is debilitating, they were more likely to focus on children's performance rather than children's learning strategies ($\beta = 0.37$, $p = .002$). However, parents' intelligence mindsets were not associated with children's perceptions of their parents' learning versus performance orientation ($\beta = 0.01$, $p > .250$).

To further understand the mechanism behind the association of parents' failure mindsets and children's intelligence mindsets, Haimovitz and Dweck (2016) tested the possible mediating effects of children's perceptions of their parents' learning versus performance. They found that parents' failure mindsets indirectly affected children's intelligence mindsets before controlling parents' learning versus performance orientation in the mediation model ($\beta = 0.41$, $p = .038$, $CI_{95} = [0.092, 0.501]$). After controlling for parent's learning versus performance orientation, the mediation effect was not significant ($\beta = 0.16$, $p > .250$). This result indicated that parents' learning versus performance orientation directly mediated the association of parents' failure mindsets with children's intelligence mindsets. Thus, Haimovitz and Dweck (2016) concluded that children's intelligence mindsets were significantly associated with their parents' failure mindsets, and parents' learning versus performance orientation could mediate the association directly.

Haimovitz and Dweck (2016) also tested whether children's perceptions of their parents' intelligence and failure mindsets were accurate in their study 3a. They found that children's perceptions of their parents' failure mindsets were significantly correlated with parents' self-reported failure mindsets ($\beta = 0.30$, $p = .002$). By contrast, children's perceptions of their parents' intelligence mindsets were not significantly associated with parents' self-reported intelligence mindsets ($\beta = 0.11, p > .250$). These findings suggested that children had better accuracy when rating for their parents' failure mindsets compared to their ratings for parents' intelligence mindsets.

Although a significant association of parents' failure mindsets and children's intelligence mindsets was found by Haimovitz and Dweck (2016), to our knowledge, this study has not been replicated in any independent studies, nor in different cultures (the study was originally conducted in the U.S.). Therefore, it is not clear whether the association between parents' failure mindsets and children's intelligence mindsets is consistent in other studies or other cultures. I conducted two empirical studies to test whether parents' failure mindsets are associated with children's intelligence mindsets in both Chinese and British populations.

5.2 Study 1

Study 1 was adapted from study 1 in Haimovitz and Dweck (2016), which aimed to test whether children's intelligence mindsets would be associated with parents' intelligence mindsets, and parents' learning versus performance orientation. To ensure the experimental process could be more easily manipulated, and to recruit as many participants as possible, I made six methodological adjustments in the study. First, instead of using self-reported surveys from both parents and children, I asked children to report their own intelligence mindsets, their own failure mindsets, their perceptions of parents' failure mindsets, and their perceptions of parents' learning versus performance orientation in my study. Second, I used the 8-item Theories of Intelligence Scale (Dweck, 2000) rather than the 4-item shortened scale ($\alpha = 0.77$; reported in Haimovitz & Dweck, 2016) to measure children's intelligence mindsets to achieve a higher reliability ($\alpha = 0.82$ to 0.97; reported in De Castella & Byrne, 2015). Third, children's failure mindsets were not tested in study 1 of Haimovitz and Dweck (2016), therefore, it is not clear if children's failure mindsets would mediate the association between parents' failure mindset and children's intelligence mindsets. Thus, in my study, children's failure mindsets would be assessed by using the failure mindset scale that was originally used in the parent survey in study 1 of Haimovitz and Dweck (2016). Fourth, I included the children's perception of parents' failure mindset scale which was originally used in study 3a in Haimovitz and Dweck (2016) to measure how children perceived their parents' failure mindsets. Fifth, as teachers play an important role in children's development, to measure the effects of their teachers on children's intelligence mindsets I adapted a children's perception of teachers' failure mindset scale from the parents' version. Finally, to maximise the effects of failure mindsets, I did not test the children's ratings for their learning versus performance orientation in this study.

There are five hypotheses in this study: 1) children's growth intelligence mindsets would be positively associated with their perceptions of parents' failure-is-enhancing mindsets; 2) children's growth intelligence mindsets would be associated with their perceptions of parents' learning orientation; 3) children's failure-is-enhancing mindsets would be positively associated with their growth intelligence mindsets; 4) children's perceptions of parents' failure mindsets would be positively associated with children's own failure mindsets; 5) children's perceptions of teachers' failure mindsets would be positively associated with children's intelligence mindsets.

5.2.1 Method

5.2.1.1 Participants

In total, 227 pupils (118 boys, 109 girls) were recruited from a public primary school in Harbin, China. Participants were aged from 7 years 8 months to 11 years 4 months old (M = 10.09, SD = 0.52).

5.2.1.2 Materials

Intelligence mindset scale: This 8-item Theories of Intelligence scale (Dweck, 2000) measured whether children have a fixed (believing their intelligence is fixed) or growth (believing their intelligence can be changed) intelligence mindset. Example items included "*You have a certain amount of intelligence, and you can't really do much to change it*" and "*You can always substantially change how intelligent you are*". Responses were made on a 6-point Likert Scale, ranging from 1 (strongly agree) to 6 (strongly disagree). Items that assessed a fixed intelligence mindset were reverse scored, so that higher scores reflected a growth intelligence mindset.

Failure mindset scale: This 6-item scale measured whether children themselves believed failure is enhancing or debilitating (Haimovitz & Dweck, 2016). Example items included "*Experiencing failure enhances my performance and productivity*", and "*The effects of failure are negative and should be avoided*". Responses were made on a 6-point Likert Scale, ranging from 1 (strongly agree) to 6 (strongly disagree). Items that assessed a failure-is-debilitating mindset were reverse scored, so that higher scores reflected a failure-is-enhancing mindset.

Children's perceptions of their parents' failure mindset scale: This 4-item scale measured how children perceived their parents' attitudes about whether failure is enhancing or debilitating (Haimovitz & Dweck, 2016). Example items include "*My* *parents think failure hurts my learning*", and "*My parents think failure can help me grow*". Responses were made on a 6-point Likert Scale, ranging from 1 (strongly agree) to 6 (strongly disagree). Items that assessed a failure-is-debilitating mindset were reverse scored, so that higher scores reflected a perception of failure-is-enhancing mindset.

Children's perceptions of their parents' learning versus performance orientation scale: This 8-item scale was used to test whether children perceived their parents as having a learning (concerned about children's learning strategy and effort) or performance (concerned about children's performance and intelligence) orientation in learning process (Haimovitz & Dweck, 2016). Example items include "*My parents would be pleased if I could show that school is easy for me*" and "*My parents would like me to do hard work, even if I make mistakes*". Responses were made on a 6-point Likert Scale, ranging from 1 (strongly agree) to 6 (strongly disagree). Items that assessed a performance orientation were reverse scored, and higher scores reflected a perception of learning orientation.

Children's perceptions of their teachers' failure mindset scale: This 4-item scale measured whether children perceived their teachers as having a failure-is-enhancing or failure-is-debilitating mindset (Haimovitz & Dweck, 2016). Example items include "*My main course teacher thinks failure can help me to learn"* and "*My main course teacher thinks failure hurts my learning*". Responses were made on a 6-point Likert Scale, ranging from 1 (strongly agree) to 6 (strongly disagree). Items that assessed a failure-is-debilitating mindset were reverse scored, and higher scores reflected a perception of failure-is-enhancing mindset.

Whenever items were translated from English into Chinese, I made an initial translation, which was then back translated by 5 bilingual (Chinese and English) speakers, checked for round-trip accuracy, and edited where necessary to ensure an accurate translation.

5.2.1.3 Procedure

Children were tested in their classroom during normal school hours. After gaining consent, I introduced the aim and procedure of the experiment to children. Then children were asked to answer the survey in the order of intelligence scale, failure mindset scale, children's perceptions of their parents' failure mindset scale, children's perceptions of their parents' learning versus performance orientation scale, and children's perceptions of their teachers' failure mindset scale. Children had been told that there is no right or wrong answer in this survey, and their answers would not be marked as formal grades.

5.2.2 Results

Before testing the five hypotheses, I first tested whether my participants show typical variation and means of mindsets scores. As shown in Figure 5.1, children who participated in my study showed a full range of mindset scores, which appeared to be normally distributed (mean = 4.14, range from 1.63 to 5.75; also see Table 5.1 for descriptive statistics of each scale used in this study).

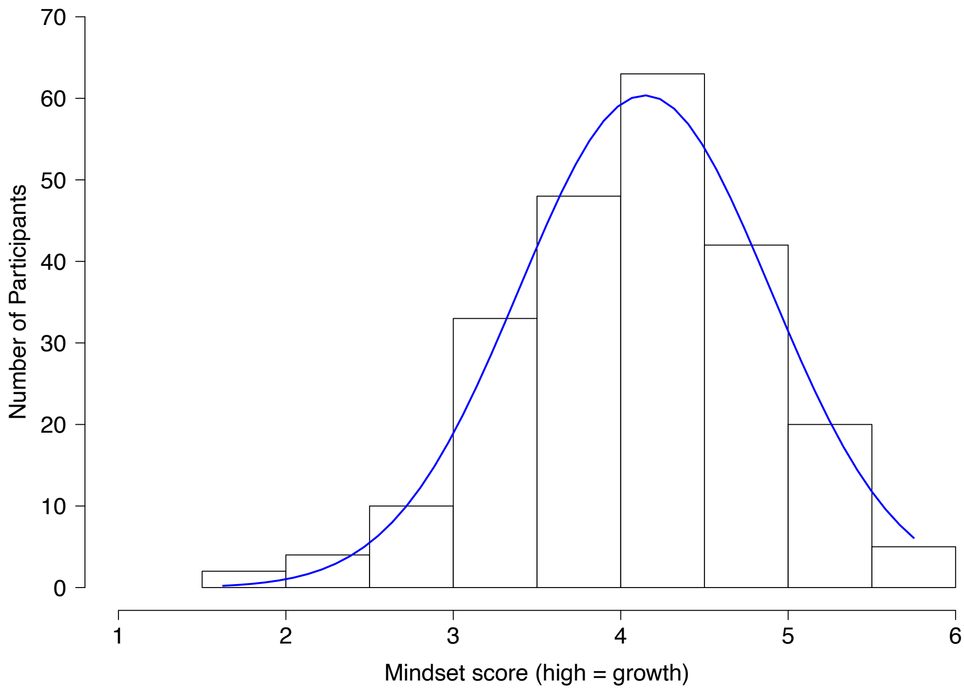

Figure 5.1 The distribution of children's mindset scores in study 1

Table 5.1 Means and standard deviations of children's intelligence mindsets, children's failure mindsets, children's perceptions of their parents' failure mindsets, children's perceptions of their parents' learning versus performance orientation, and children's perceptions of their teachers' failure mindsets in study 1

	Mean	**Standard deviation**
Intelligence mindset	4.14	0.75
Failure mindset	4.45	0.93
Children's perceptions of their parents' failure mindsets	4.30	1.21
Children's perceptions of their parents' learning versus performance orientation	3.76	0.63
Children's perceptions of their teachers' failure mindsets	4.51	1.15

As predicted, children's growth intelligence mindsets were significantly positively associated with their perceptions of parents' failure-is-enhancing mindsets (β = 0.15, CI_{95} [0.02, 0.28], t = 2.27, p = .024; see Figure 5.2). However, children's intelligence mindsets were not significantly associated with their perceptions of parents' learning versus performance orientation (β = 0.08, CI_{95} [-0.05, 0.21], t = 1.18, p = .240).

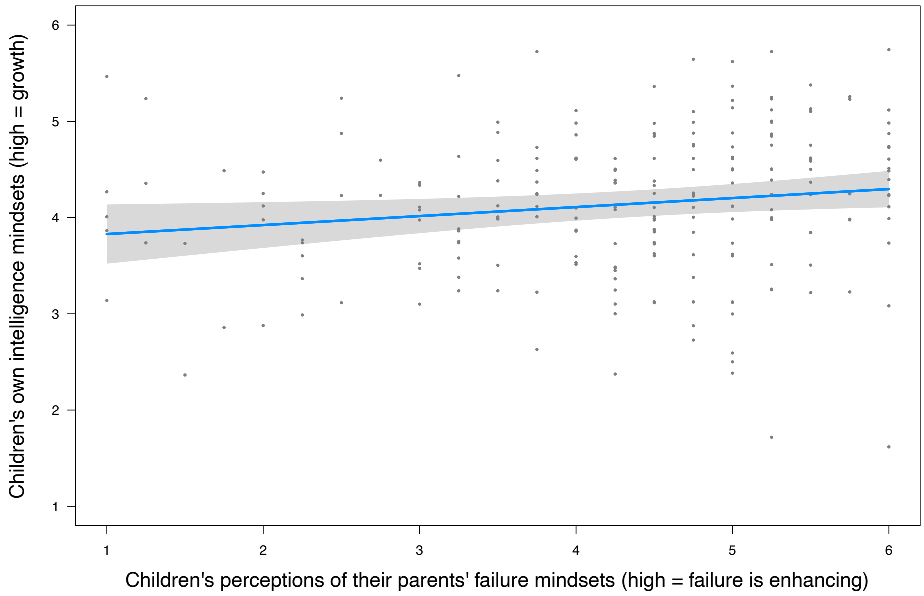

Figure 5.2 The association of children's perceptions of their parents' failure mindsets (x axis) and children's intelligence mindsets (y axis) in study 1

As predicted, children's failure-is-enhancing mindsets significantly predicted their growth intelligence mindsets (β = 0.24, CI_{95} [0.11, 0.37], t = 3.67, p < .001). In addition, children's perceptions of parents' failure mindsets were significantly positively associated with children's own failure mindsets (β = 0.56, CI_{95} [0.45, 0.67], t = 10, p < .001; also see Figure 5.3). For the effects of teachers' failure mindsets on children's intelligence mindsets, a significant association was found (β = 0.17, CI_{95} [0.04, 0.30], t = 2.52, p = .012).

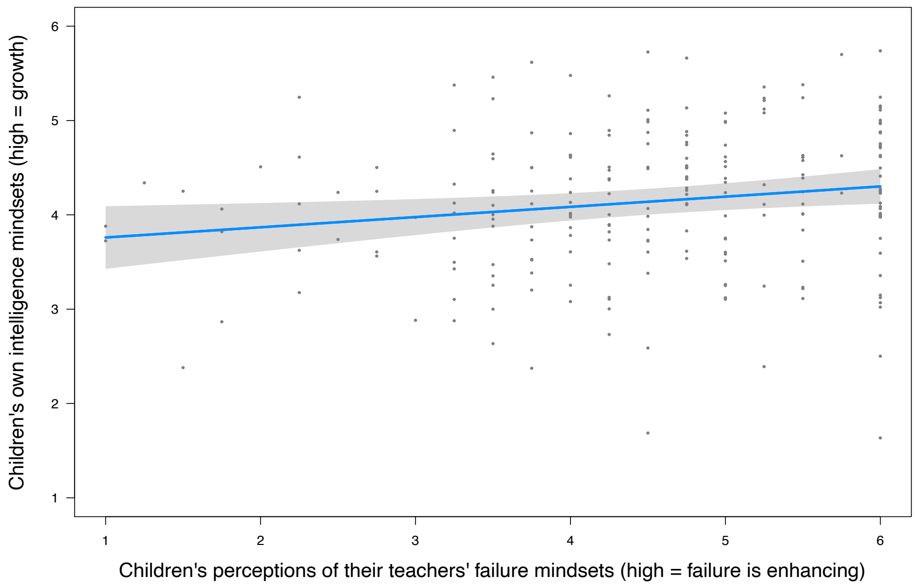

Figure 5.3 The association of children's perceptions of their teachers' failure mindsets (x axis) to children's own intelligence mindsets (y axis) in study 1

5.2.3 Study 1 discussion

By asking children to self-report their own intelligence and failure mindsets, to rate their parents' and teachers' failure mindsets and to rate their parents' learning versus performance orientations, I report the following five findings. First, in the children's ratings, parents' failure-is-enhancing mindsets significantly promoted their children's growth intelligence mindsets. Second, neither parents focusing on learning or performance significantly influence children's intelligence mindsets. Third, children who had a failure-is-enhancing mindset were significantly more likely to have a growth intelligence mindset. Fourth, parents' failure-is-enhancing mindsets were significantly associated with children's failure-is-enhancing mindsets. Finally, teachers' failure-is-enhancing mindsets were significantly associated with children's growth intelligence mindsets.

Therefore, my findings supported the notion that if children perceive their parents as having a failure-is-enhancing mindset, children would be more likely to have a growth intelligence mindset, as predicted in Haimovitz and Dweck (2016). On the other hand, if children perceived their parents as having a failure-is-debilitating mindset, children would be more likely to have a fixed intelligence mindset. Contrary to Haimovitz and Dweck (2016), the predicted mechanism that parents' learning orientation mediates the effects of failure-is-enhancing mindsets on children's growth intelligence mindsets was not supported.

One limitation in this study was that children were recruited from four classrooms and thus there were only four teachers that were rated by children. Children still perceived their teachers as having a full range of failure mindsets - these findings raised the limitation of response bias in this study. In other words, children's responses to their parents and teachers might be biased by children's own attitudes rather than their parents' or teachers' actual attitudes. I then conducted a second study to test if children's perceptions accurately represent their parents' actual attitudes.

5.3 Study 2

In this study, I wished to test the association of children's perceptions of their parents' failure mindsets and parents' self-reported failure mindsets. I closely replicated study 3a of Haimovitz and Dweck (2016), by using the same experimental materials and recruiting the same age (i.e., 10 years old) child participants. As in the Haimovitz and Dweck (2016) study 3a, I made the same hypotheses: 1) children's perceptions of their parents' failure mindsets would be significantly associated with their parents' self-reported failure mindsets; 2) children's perceptions of their parents' intelligence mindsets would not be significantly associated with their parents' self-reported intelligence mindsets; 3) parents' self-report failure mindsets would not be associated with children's perceptions of parents' intelligence mindsets and; 4) parents' self-reported failure mindsets would not be associated with their self-reported intelligence mindsets.

5.3.1 Methods

5.3.1.1 Participants

In total, 97 parent-child dyads were recruited in this study. Participants were recruited from a supermarket (*n* = 4) and through the research assistant's friends and family (*n* = 93) in the United Kingdom. The child participants (42 boys and 46 girls, 9 unknown) were aged between 6 and 14 years old (*M* = 10.36, *SD* = 2.20). For the adult participants (mean age = 31.00, *SD* = 15.03), 41 identified as girl, 19 as boy (37 participants did not have their gender recorded).

5.3.1.2 Materials

The present study used the same materials as in study 3a in Haimovitz and Dweck (2016). All responses were recorded using a 6-point Likert Scale from 1 (strongly disagree) to 6 (strongly agree).

This study used two surveys: the parents' self-reported survey and children's survey. The parents' self-reported survey had two subscales including parents' failure mindset scale and parents' intelligence mindset scale. Children's survey included child perceptions of parents' intelligence mindset scale and child perceptions of parents' failure mindset scale.

Parents' failure mindset scale: This scale included 6 items measuring whether parents believed failure is enhancing or debilitating. Example items included "*Experiencing failure enhances my performance and productivity*", and "*The effects of failure are negative and should be avoided*". Items that assessed a failure-is-enhancing mindset were reverse scored, and higher scores indicated a failure-is-debilitating mindset.

Parents' intelligence mindset scale: This 2-item scale was a shortened version of the 8-item intelligence mindset scale used in study 1. Both items "*You have a certain amount of intelligence, and you really can't do much to change it*" and "*You can learn new things but you can't really change how intelligent you are*" indicated a fixed mindset. The two item's scores were averaged, and higher scores indicated a fixed intelligence mindset.

Children's perceptions of their parents' intelligence mindset scale: This scale included 3 items measuring whether children perceived their parents have a fixed (2 items) or growth (1 item) mindset. An example item is "*My parents think you can learn new things but you can't change how smart you really are*". Items related to a growth mindset were reverse scored, and a higher score indicated children perceived their parents as having a fixed mindset.

Children's perceptions of their parents' failure mindset scale: This scale was identical to that used in study 1. However, due to different dataset recording issue (the data in study 2 was recorded by the research assistant), I reverse scored items that assessed a perception of failure-is-enhancing mindset in study 2. Higher scores indicated that children perceived their parents as having a failure-is-debilitating mindset.

5.3.1.3 Procedure

All parents were given a written information sheet and consent form for themselves and their children. Participants were told that there was no time limit to complete the survey. Parents were asked to answer the items measuring their own attitudes to failure, followed by those measuring their beliefs about intelligence. Children were asked to answer the questions on their own but were also told that they could ask for their parents' assistance if needed. Children firstly were asked to rate their parents' failure mindsets, followed by items about their parents' intelligence mindsets.

5.3.2 Results

For the association between children's perceptions of their parents' failure mindsets and parents' self-reported failure mindsets, contrary to prediction, no significant association was found ($\beta = 0.17$, CI_{95} [-0.04, 0.37], $t = 1.63$, $p = .105$). Adding control variables of children's age and gender did not change the null result ($\beta = 0.15$, CI_{95} [-0.06, 0.37], $t = 1.44$, $p = .154$). However, children's perceptions of their parents' intelligence mindsets were significantly associated with parents' self-reported intelligence mindsets ($\beta = 0.48$, CI_{95} [0.30, 0.66], $t = 5.29$, $p < .001$; see Figure 5.4). Controlling for children's age and gender changed the stated association to be non-significant ($\beta = -0.22$, CI_{95} [-0.6, 0.16], $t = -1.13$, $p = .261$).

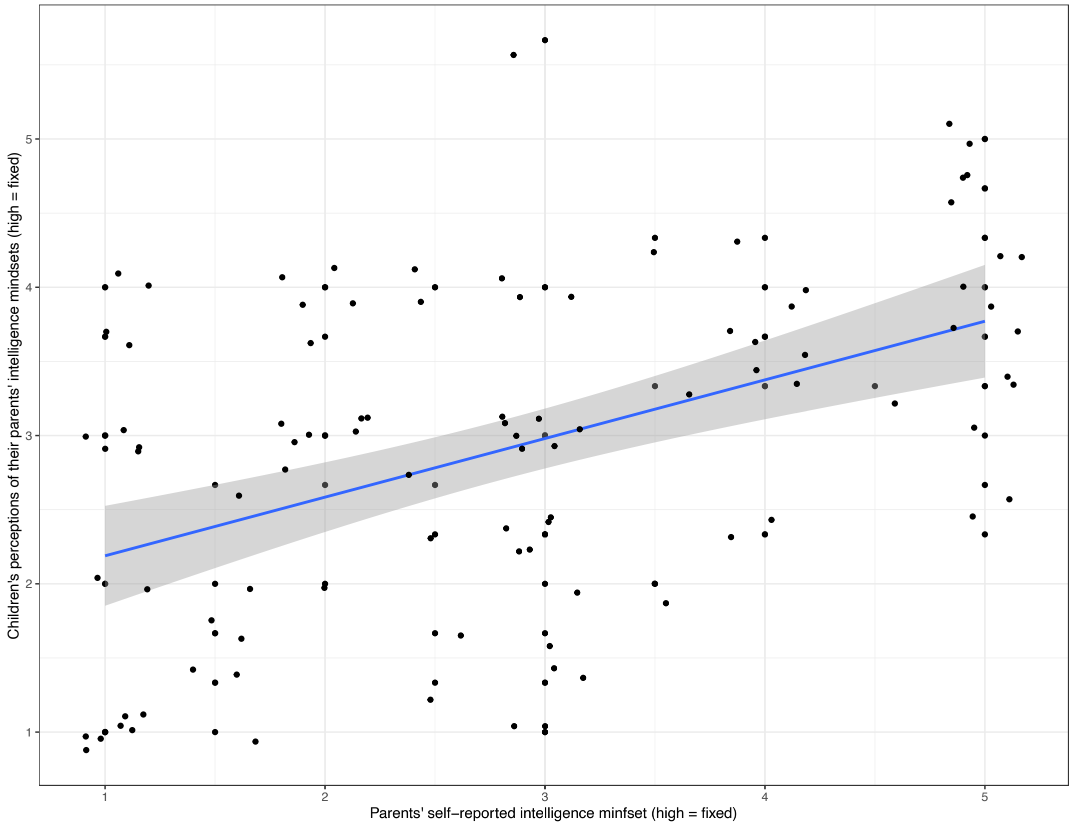

Figure 5.4 The association of parents' self-reported intelligence mindsets (x axis) and children's perceptions of their parents' intelligence mindsets (y axis) in study 2

There was no significant association between parents' failure mindsets and children's perceptions of their parents' intelligence mindsets, as predicted, (β = 0.06, CI_{95} [-0.14, 0.27], t = 0.62, p = .539). Controlling for children's age and gender did not change the null result (β = 0.05, CI_{95} [-0.18, 0.29], t = 0.45, p = .656). Contrary to prediction, there was a significant association between the association of parents' intelligence mindsets and their own failure mindsets (β = 0.26, CI_{95} [0.06, 0.45], t = 2.58, p = .011).

5.3.3 Study 2 discussion

By using the same materials and same-aged child participants as used in the original study of Haimovitz and Dweck (2016), I expected to find similar results as in the original study, but I did not. Children's perceptions of their parents' failure mindsets were predicted to be significantly associated with parents' self-reported failure mindsets, and children's perceptions of their parents' intelligence were predicted to not be associated with parents' self-reported intelligence mindsets. However, my findings were contrary to those of Haimovitz and Dweck (2016), in that children's perceptions of parents' failure mindsets were not significantly associated with their parents' self-reported failure mindsets, and children's perceptions of parents' intelligence mindsets were significantly associated with their parents' self-reported intelligence mindsets. I also found that parents' self-reported failure mindsets were significantly associated with their self-reported intelligence mindsets. These results jointly support that the development of children's intelligence mindsets was from parents' intelligence mindsets, but not from parents' failure mindsets.

5.4 General discussion

Haimovitz and Dweck (2016) reported that children's intelligence mindsets were significantly affected by their parents' failure mindsets, but not by parents' intelligence mindsets. I tested this prediction in two empirical studies and found contrary findings. In study 1, I asked children to report their own intelligence mindsets, their own failure mindsets and to rate their parents' failure mindsets, their parents' learning versus performance orientation and their teacher's failure mindsets. I found that children's perceptions of their parents' failure mindsets were significantly associated with children's own intelligence mindsets and children's own failure mindsets. A similar significant effect was found between children's perceptions of their teachers' failure mindsets and children's own intelligence mindsets. However, children's intelligence mindsets were not significantly associated with their perceptions of parents' learning versus performance orientation. In study 2, I used identical materials and recruited similarly-aged of participants as in Haimovitz and Dweck (2016) to test whether children's perceptions were associated with their parents' self-reported beliefs and whether children's perceptions of their parents' failure mindsets were associated with their own intelligence mindsets. I found that children's perceptions of their parents' failure mindsets were not significantly associated with their parents' self-reported failure mindsets. Namely, children's perceptions of their parents' failure mindsets were not accurate. More strikingly, I found a significant association between parents' self-reported intelligence mindsets and children's perceptions of their parents' intelligence mindsets, which was contrary to the findings in Haimovitz and Dweck (2016).

Haimovitz and Dweck (2016) explained that parents' failure mindsets (rather than parents' intelligence mindsets) significantly affected children's intelligence mindsets because parents' reactions to their children's failures were associated with parents' learning versus performance orientation. When reacting to their children's failures, parents who have a failure-is-debilitating mindset would be more likely to focus on children's performance or children's innate ability rather than the learning strategies children have used or the efforts children have made. These reactions transferred parents' implicit beliefs to explicit behaviours and concerns. And through observing parents' performance-focused behaviours or concerns, children would also pursue a performance orientation rather than a learning orientation in learning, which would further lead children to hold a fixed intelligence mindset. If these relationships are robust, a significant association between parents' learning versus performance orientation and children's intelligence mindsets should have emerged in my studies, but it did not. Instead, my findings suggested that compared to parent's failure mindsets, parents' intelligence mindsets were more influential on children's intelligence mindsets.

5.4.1 Limitations and further directions

One limitation in my studies might be that children's intelligence mindsets and parents' intelligence mindsets were not measured in the same study. Namely, I measured children's intelligence mindsets, but did not ask parents to report their own intelligence mindsets in study 1. Although parents' intelligence mindsets were measured in study 2, children's own intelligence mindsets were not measured. Therefore, I could not test whether parents' intelligence mindsets were associated with children's intelligence mindset as predicted in study 1 of Haimovitz and Dweck (2016). Also, since I only asked children to rate their parents' and teachers' failure mindsets in study 1, there might be a bias, in that children's perceptions of their parents' and teachers' failure mindsets might be biased by children's own attitudes. This limitation might be the cause of the inconsistent results between my study 1 and the original study of Haimovitz and Dweck (2016). The third limitation might be that my two studies were conducted in different countries, which might have led to a potential cultural bias when rating parents and teachers' attitudes and beliefs.

Since these two studies were the first two (to my knowledge) close replications testing the development of children's intelligence mindsets, further independent studies testing the same association are called for. Due to the stated limitations in the two studies, further studies should assess both parents' and children's intelligence mindsets, failure mindsets and learning versus performance orientations in one study to get more objective results. Also, further studies should eliminate the potential cultural difference by using participants from the same culture.

Chapter 6. The association of gender stereotypes and children's academic performance – A brief literature review

The previous theme introduced the prominent theory of mindset and its effects on children's learning. Since I failed to find any significant effects of mindset on children's learning, I sought out alternative proposals or potential remediations which are involved in addressing current problems in children's learning. To this end, I turned my research focus to the current concerns about women's participation in Science, Technology, Engineering, and Mathematics (STEM) fields. Specifically, the idea that gender stereotypes about brilliance may account for at least part of the gender gap in STEM participation, and that these differences may have their origins in children as young as 6 years old (Bian et al., 2017).

In this chapter, I introduce the definition and types of stereotype, how gender stereotypes are formed in China, the existing evidence of gender differences in children's educational attainment, and how gender stereotypes affect children's educational attainment.

6.1 What are stereotypes?

6.1.1 Definition of stereotypes

When talking about people from different countries (e.g. China, Italy and Germany), which words or phrases would you use to describe them? If you think, for example, Chinese people are shy and conservative, Italians are romantic and passionate, Germans are efficient and demanding, you may hold stereotypical views.

An early definition of stereotypes was proposed by Lippmann (1922), who suggested that a stereotype is people constructing a picture - about a world that people could not reach - inside of their head, and the picture construction process is partially influenced by their culture. Lippmann (1922) also proposed that there are three features of stereotypes: stereotypes are incorrect, their origins are illogical, and they are too rigid to accept new information. McCauley, Stitt, and Segal (1980) tested whether the features of stereotypes were accurate in an empirical study which explored racial stereotypes among college students. A conclusion was drawn that racial or national stereotypes emerge only when people hold fallacies about skin colour or birth place consciously or unconsciously (McCauley et al., 1980). This finding verified the first feature of stereotypes, that stereotypes are incorrect (Katz & Braly, 1933).

On the other hand, researchers argued that the features of stereotypes defined by Lippmann (1922) were not precise, as not all of these features have been tested in empirical studies (e.g., Brigham, 1971; Fagot, Leinbach, & O'Boyle, 1992; Jussim, Nelson, Manis, & Soffin, 1995). Therefore, an alternative definition that stereotypes should be generalisations or beliefs about a group of people that distinguish them from other groups was proposed (Jussim et al., 1995; McCauley et al., 1980). This definition seemed simplified because it did not need to fulfil all of the three features stated in the previous definition (McCauley et al., 1980). Additionally, this definition provided a piece of important information, in that people from the same group would have similar characteristics or behaviours (Jussim et al., 1995). Based on the group similarities, researchers could make predictions of people in a specific group (McCauley et al., 1980).

6.1.2 The types of stereotypes

People may hold stereotypes about different aspects, such as gender, ethnicity, age and religion (e.g., Fagot et al., 1992; Katz & Braly, 1933; Lindholm, 2017). Compared to other kinds of stereotypes, gender stereotypes have been investigated for around 80 years in psychology (Katz & Braly, 1933), and are claimed to be associated with many psychological traits (e.g., Cvencek, Meltzoff, & Greenwald, 2011; Lippa, 2010). For example, researchers have reported that men and women are stereotypically thought to have different personality traits. Specifically, women are stereotyped to be agreeable, neurotic, anxious, sensitive, and emotional; but men are thought to be conscientious, open, extraverted, and emotionally stable (Del Giudice, Booth, & Irwing, 2012; Löckenhoff et al., 2014; Weisberg, Deyoung, & Hirsh, 2011; Williams, Satterwhite, & Best, 1999). Additionally, gender stereotypes exist in social roles, as women are stereotyped to be more needed in family-related roles, whereas men are stereotyped to be important in work-related roles (e.g., Lauzen, Dozier, & Horan, 2008; Okimoto & Heilman, 2012). Furthermore, men and women are stereotyped as having different levels of academic performance or educational attainment (e.g., Cvencek, Meltzoff, et al., 2011; Ma, 2008; Spencer, Steele, & Quinn, 1999). This will be discussed in detail in section 6.3.

6.2 Gender stereotypes in China

6.2.1 The origins of gender stereotypes in China

Gender stereotypes in China reflect traditional gender roles which are rooted in Confucian philosophy (Bowen, Wu, Hwang, & Scherer, 2007; S. S. Liu, Comer, & Dubinsky, 2001). Confucius was born almost 2600 years ago (around 551 B.C.), but his philosophy is influential in both ancient and modern China. Confucianism has explicit gender preference towards men, shown in many of their famous claims (Bowen et al., 2007). For example, one stereotyped claim is that a lack of talent is a virtue for women (Bowen et al., 2007). More specifically, Confucianism posited that women who are talented or knowledgeable, but also considerate and virtuous, were rare. For those rare women who are talented or knowledgeable, in the Confucian philosophy, most of them would do bad things to shame their family rather than doing good things to honour their family. Therefore, Confucianism did not encourage women to develop their intelligence by engaging in learning activities, such as going to school.

Another claim relating to gender in Confucianism is that men are responsible for external (outside the household) work while women are responsible for internal (within the household) work (C. Li, 2000). In other words, men are expected to find a job in the community or society to earn money and fulfil the living demands of the whole family, whilst women are expected to stay at home to do daily housework (such as cooking or cleaning). This gender-stereotyped lifestyle is still quite common in some rural areas in China (Cao & Chai, 2018; Tatli, Ozturk, & Woo, 2017; Yun, 2012). By staying at home and doing housework, women could not partake in any social position, other than playing the role of daughter, wife or mother in their own family (Yun, 2012). As their family roles were not as high-ranking as those of men, women were seen as appendages of or subservient to men (F. M. Cheung, 1996; Yun, 2012). Moreover, the Confucianism gender norm stated that women needed to obey men in accordance with the "Three Obediences". These were as follows: until married with a man, a woman needed to obey her father; after marrying with a man, a woman should obey her husband; if her husband passes away, a woman should obey her sons (Bowen et al., 2007; Yun, 2012). All of the above claims lead to a gender stereotype that men played a more important and dominant roles than women, both in society and family life in ancient China (F. M. Cheung, 1996; C. Li, 2000; Raphals, 1998; Tatli et al., 2017).

The famous ‘yin-yang’ binary theory in Confucianism is also related to a gender stereotype. In Chinese, ‘yin’ originally represents the shady part that sunshine cannot reach, and ‘yang’ represents the sunny side (Yun, 2012). Confucians use the ‘yin-yang’ binary theory to describe the complementary, contrary, and correlated relationship of paired things (e.g. dark and light, night and day) in nature (Yun, 2012). In the binary pair, ‘yin’ usually represents the negative, evil, and weak part, whilst ‘yang’ represents the positive, kind, and strong part (Raphals, 1998). Since ‘yin’ and ‘yang’ have a contradictory but linked relationship, Confucians also used this binary pair to describe the two genders (Raphals, 1998; Yun, 2012). Specifically, ‘yin’ is used to represent women, and ‘yang’ represents men (C. Li, 2000; Yun, 2012). Additionally, because ‘yin’ also represents weakness, and ‘yang’ represents strength, in the binary theory, a gender stereotype emerged that women were weak and inferior, whilst men were strong and superior (Raphals, 1998; Yun, 2012).

Turning to the possible explanations of how such Confucianist claims or theories formed, the traditional masculine orientation in the patrilineal society would play an important role (Tatli et al., 2017). Specifically, men could dominate the society. They kept their family names, family honour, and fortune (Bowen et al., 2007; F. M. Cheung, 1996). Their power and rights would not be reduced or eliminated after getting married, and they had the duty to look after their elder parents in the patrilineal society (Hannum, Kong, & Zhang, 2009). Rather than dominating, women were dominated in society. They were not expected to work outside the home, they needed to take their husbands’ surname, and would no longer belong to their original family any more after marriage (Lee, 1995). Living and studying in such a masculine culture would influence Confucians' negative thoughts towards women, which further led their claims to be more favourable towards men.

6.2.2 The current situation of gender stereotypes in China

Although feudal traditions have somewhat slackened over time, the gender stereotype that women are inferior to men still exists. One present-day expression of the Confucianism perspectives about gender in modern China is in relation to the sex ratio at birth (SRB). According to the World Bank (2017), the normal range of sex ratio at birth (SRB) is around 1.03:1 to 1.07:1 (male births versus female births). However, the SRB in China exceeded the normal range from the 1980s to the 2010s (World Bank, 2017). The SRB increased from 1.07:1 (in 1982) to 1.17:1 (in 2007), and then decreased slightly to 1.15:1 (in 2017). This tendency indicates that Chinese parents have had more male offspring than female offspring in the last three decades (World Bank, 2017). One possible explanation for the increased SRB after the 1980s would be the establishment of the one-child policy in China, introduced in 1979 (Ebenstein, 2010). Parents were allowed to have more than one child per family before 1979, and therefore, they might be less likely to abort a female offspring if this was their first child as they had opportunities to have a male offspring. After the government announced the one-child policy, parents could have only one child per family. If the first child was female, parents were more likely to abort her in order to have the opportunity to have a male. If this was the case, then this could be the reason for the SRB being unequal after 1979.

Another current reflection of Confucianism on gender relates to school enrolment rates in modern China. According to the National Bureau of Statistics of the People's Republic of China (1998, 2015), the average school enrolment of women in all kinds of schools (i.e. primary school, high school, and university) has increased over the last three decades (from around 43% in 1980 to around 48% in 2014), but is still less than men's average school enrolment (around 57% in 1980 to 52% in 2014). In the history of Chinese education, men were allowed to be enrolled in the mainstream education system (in the form it took at that time) around three thousand years ago, but women were only first allowed to accept formal education at school less than 100 years ago (J. Liu & Carpenter, 2005). After the new Chinese government was established in 1949, leaders made efforts to promote education and provide opportunities for all children to be educated (Rong & Shi, 2001). For example, a government policy stipulated that the first nine years of education was to be compulsory and free (of tuition fees) for all children (Rong & Shi, 2001; L. Song, Appleton, & Knight, 2006). Additionally, the one-child policy also increased the opportunity for women to be educated (F. Liu, 2006). However, education was seen as a form of investment (F. Liu, 2006). Therefore, when a family had limited financial resources and more than one offspring, parents preferred to invest in their male offspring, rather than female, to ensure that their investments were sound and valuable (L. Song et al., 2006). Thus, the school enrolment rates of women were lower than men.

Additional evidence from interviews also suggested the gender stereotypes favouring men still exist in Chinese culture. In interviews, most teachers believed that men were superior to women in mathematics and science, and the teachers attributed this to men's naturally-held advantages in spatial and visual ability (Tsui, 2007). As in the West (Flore & Wicherts, 2015), the evidence is mixed on the effect of the gender stereotype threat on mathematical performance, with reports of significant effects (J.

Song, Zuo, & Yan, 2016), and failures to replicate (Tsui, Xu, Venator, & Wang, 2016). As the traditional gender stereotypes seem to remain rooted in current Chinese culture, I expect to find as strong or stronger effects in the Chinese sample.

6.3 The association of gender stereotypes and children's educational attainment

6.3.1 Gender differences in educational attainment

Numerous researchers have explored whether boys and girls differed in their educational attainment. The majority of studies indicated that the trend of gender differences in educational attainment favours girls rather than boys (e.g., Batool, 2012; Duckworth & Seligman, 2006; Lai, 2010; Ma, 2008; Mau & Lynn, 2001; McNabb, Pal, & Sloane, 2002). For instance, Lai (2010) compared 7235 middle schoolers' scores in five subjects (Mathematics, English, Chinese, Chemistry and Physics) over three years (including six semesters), and their scores in the Middle School Graduation Exam (MSGE) as well as the High School Entrance Exam (HSEE). Girls were superior to boys in all semester tests, the MSGE and the HSEE, and the only exception where boys were superior to girls was found in Physics in the HSEE (Lai, 2010). This gender difference where girls outperform boys in educational attainment was also found in a longitudinal study (Duckworth & Seligman, 2006). By observing 140 eighth-grade children over two years, Duckworth and Seligman (2006) found that girls had better school-reported GPAs (~ 0.5 SD) than boys. Girls outperforming boys in their educational attainment has been reported among children in different kinds (i.e. mainstream and religious) of school (Batool, 2012).

Female advantages in educational attainment were not consistent across time. For example, Mau and Lynn (2001) analysed over 10,080 American college graduates'

Scholastic Aptitude Test (SAT, scores before entering college) or the American College Test (ACT, scores before entering college) and their Grade Point Average (GPA, scores in college). They found that, before entering college, boys had higher aptitude scores than girls. However, when they were in college, girls performed better than boys. This female superiority in educational attainment has been confirmed among university students in other studies (e.g., Khwaileh & Zaza, 2011). The study of Mau and Lynn (2001) not only confirmed the gender differences existing in educational attainment and standardised tests (e.g. SAT), but also suggested the trajectory of the gender differences. Namely, the trajectory of gender differences in educational attainment means that the advantage changes hands from men to women, and vice versa, across time.

Contrary to the above findings, J. S. Matthews, Ponitz, and Morrison (2009) have argued that the gender differences in children's achievement has been eliminated in recent years. In a 5-year longitudinal study, J. S. Matthews et al. (2009) examined whether 5 year old boys and girls differed in general knowledge, mathematics ability, literacy, sound awareness, and vocabulary aspects. They did not find any significant results of gender differences in these five aspects. However, the achievement tasks used in J. S. Matthews et al. (2009) were cognitive ability tasks rather than grades from official school examinations, therefore, they could conclude that there were no gender differences in cognitive ability among children in kindergartens, but could not make any conclusion about academic achievement. Although cognitive ability is a strong predictor of academic achievement (Deary, Strand, et al., 2007), the influences of other factors (e.g., personality traits) on academic achievement could not be neglected. Therefore, the results in J. S. Matthews et al. (2009) are not convincing evidence of null gender differences in academic achievement.

In summary, although there are arguments about whether gender differences exist in educational attainment or not, the majority of researchers in this field have found the existence of gender differences in educational attainment. Among those who did find a gender difference in educational attainment, some researchers not only explored whether boys and girls differed in overall educational attainment, but also tested whether there were gender differences in specific school subjects such as mathematics, science, or reading.

6.3.2 Gender differences in individual school subjects

For the studies that tested whether gender differences existed in individual school subjects (e.g. mathematics or reading), researchers found inconsistent results (e.g. Duckworth & Seligman, 2006; Lynn & Mikk, 2009; Ma, 2008; Reilly, Neumann, & Andrews, 2015). On the one hand, researchers found that boys performed better than girls in science-related subjects (such as mathematics or physics), and girls performed better than boys in non-science related subjects (such as reading; e.g., Bedard & Cho, 2010; Lynn & Mikk, 2009; Ma, 2008; Machin & McNally, 2005; Tsui, 2007; Wilberg & Lynn, 1999). These findings have been supported both longitudinally and internationally.

For example, by comparing 15-year-old children's mathematics, reading and science scores in PISA (Program for International Student Assessment) across 41 countries, Ma (2008) found that girls had higher reading scores than boys in 40 countries, but boys achieved higher scores in mathematics (in 36 countries) and science (in 25 countries) than girls. Lynn and Mikk (2009) confirmed that girls outperforming boys in reading was consistent among different cohorts across nations. Specifically, Lynn and Mikk (2009) obtained children's reading scores in PISA in three cohorts (2000, 2003 and 2006) across different nations (27 countries in the 2000 cohort, 40 countries in the 2003 cohort, and 56 countries in the 2006 cohort). Girls had superior reading skills consistently in each of the three cohorts, across almost all nations. The gender difference in mathematics, where boys outperform girls, has been confirmed in a meta-analysis (Reilly et al., 2015). Children's mathematics and science scores (at grade 4, 8 and 12) were taken from the U.S. National Assessment of Educational Progress data. In total, over 2 million children were recruited from 1990 to 2011 and significant gender differences favouring boys were found both in mathematics and science in this national representative study (Reilly et al., 2015).

Similar to the overall performance, Machin and McNally (2005) found a trajectory for gender differences in individual subjects. Specifically, Machin and McNally (2005) analysed four national datasets (four cohorts: the earliest cohort was from the National Child Development Study, the second earliest cohort was from the British Cohort Study, and there were two more cohorts from the National Pupil Database) at age 11 (the end of primary school) and 16 (the end of secondary school)). They found that at age 11, girls and boys had different performances in mathematics and reading, where girls performed better in reading, and boys were superior in math. However, the direction of gender differences in mathematics shifted at age 16, where boys (in two cohorts) performed worse (9.7% points) than girls (Machin & McNally, 2005). These findings indicated that the male advantage in mathematics could change across time.

Furthermore, the gender differences in individual subjects were reported to exist not only in mathematics, science, and reading, but also in other school subjects such as history (Wilberg & Lynn, 1999). For example, in a cross-national (26 countries) study, children (around 15 years old) were asked to report their grades in history and took four tests (including knowledge about economics, general history, costume, and ships) intended to measure their historical knowledge. Wilberg and Lynn (1999) found that boys had more historical knowledge than girls, but their self-reported grades in history were not as good as girls in 22 countries. Therefore, it seems appropriate to conclude that boys are superior in science-related subjects (e.g. mathematics), and girls are superior in subjects that unrelated to science (e.g. reading).

The contradictory evidence for gender differences in individual subjects could be divided into two broad arguments: girls are superior to boys in science and non-science related subjects and the gender differences do not continuously exist. A considerable number of researchers have suggested that girls outperform boys in not only in non-science related subjects (e.g. reading), but also in science-related subjects (e.g. Ding, Song, & Richardson, 2006; Duckworth & Seligman, 2006; Fergusson & Horwood, 1997; Kenney-Benson, Pomerantz, Ryan, & Patrick, 2006; Pomerantz, Altermatt, & Saxon, 2002). For instance, researchers reported that girls had better grades in mathematics, science, language, and social studies than boys in elementary school (Pomerantz et al., 2002). The higher performance of girls in mathematics has been supported in a longitudinal study with slightly older children (Duckworth & Seligman, 2006). Girls (in the 8th grade) performed better than boys not only in overall educational attainment, but also in all individual subjects, including basic and advanced mathematics (Duckworth & Seligman, 2006). For children in high school, Ding et al. (2006) found girls continued to perform better than boys in mathematics.

There are also researchers who have argued that gender differences in individual school subjects do not exist continuously (e.g., Hyde, Lindberg, Linn, Ellis, & Williams, 2008; Lachance & Mazzocco, 2006; Lindberg, Hyde, Petersen, & Linn, 2010; Paulsen & Johnson, 1983; Scheiber, Reynolds, Hajovsky, & Kaufman, 2015). For instance, Paulsen and Johnson (1983) examined over 300 children's mathematical ability (in the 4^{th}, 8^{th}, and 11^{th} grade). At each grade, gender differences in mathematical ability did not emerge. This finding was supported in subsequent studies that involved younger or similar aged children. For example, in a three-year longitudinal study, Lachance and Mazzocco (2006) reported that the gender differences in mathematical skills were negligible among kindergarten pupils (over 200). It did not appear to make a difference whether children were tested by the standardised psychometric tests (e.g. the Test of Early Mathematical Ability – Second Edition (TEMA-2)) or on the basic mathematics problems (e.g. counting, addition and subtraction). In a national representative study (involving over 7 million participants), Hyde et al. (2008) found similar results, in that there were no gender differences in mathematical skills among pupils from grade 2 to 11. The non-existence of gender differences in mathematical ability has been reported in countries other than the U.S. (Alkhateeb, 2001).

In sum, the above studies suggest that there continues to be a debate concerning the extent, if any, of gender differences in educational attainment (or in individual subjects). The majority of studies where no gender difference in educational attainment was reported were conducted in Western cultures. This leads to the question: could gender differences in educational attainment be influenced by culture? To explore this question, conducting a comparison study with a sample from an Eastern country (i.e., China) would be valuable. For those researchers who did find a gender difference in educational attainment (and/or in mathematics), studying how these gender differences are formed, or how to decrease or even eliminate these differences would be important. One factor that could contribute to these gender differences in educational attainment is gender stereotypes (e.g. Tsui, 2007; M. T. Wang & Degol, 2017).

6.4 The influence of gender stereotypes in educational attainment

6.4.1 Possible causes of gender differences in educational attainment

The gender differences in educational attainment (especially the differences in science-related school subjects) may lead to a gender gap in later careers (Bian et al., 2017; M. T. Wang & Degol, 2017). Although the gender gap in career choices has decreased in recent years, researchers continue to report that women are less likely to choose jobs related to Science, Technology, Mathematics, and Engineering (STEM) in comparison to men (Meyer, Cimpian, & Leslie, 2015).

When researchers tried to explain the possible causes of gender differences in educational attainment, they suggested various factors. One of the possible factors was gender differences in cognitive ability (Benbow, 2010). Cognitive ability was found to be a strong predictor of educational attainment (Deary, Strand, et al., 2007; Kuncel et al., 2004). If girls and boys showed different strengths in different aspects of cognitive ability (e.g., spatial or reasoning ability), they might perform differently in individual subjects (Spelke, 2005). This explanation has been confirmed by Benbow (2010): gender differences in cognitive ability were shown to influence children's mathematical achievement. However, Benbow's study has a problem with generalisation, since their participants were talented children.

On the other hand, some researchers argued that cognitive ability might not explain the gender differences in educational attainment since the existence of gender differences in cognitive ability is still equivocal (M. T. Wang & Degol, 2017). Other factors such as self-discipline (Duckworth & Seligman, 2006), interests (Su & Rounds, 2015), competitiveness (Niederle & Vesterlund, 2010), self-control (Wu, Kung, Chen, & Kim, 2016), the country's economic status (Stoet & Geary, 2018), or social interactions with other students or lecturers (Clifton, Perry, Roberts, & Peter, 2008) would be possible explanations for gender differences in educational attainment.

6.4.2 The association of gender stereotypes and gender differences in educational attainment

Researchers have found that, to some extent, gender stereotypes could explain the gender differences in educational attainment, especially in mathematics performance (e.g. Else-Quest, Hyde, & Linn, 2010; J. Song et al., 2016; Tiedemann, 2000; Tsui, 2007; M. T. Wang & Degol, 2017). Spencer et al. (1999) proposed that women's mathematics performance is threatened by a negative gender stereotype (i.e., mathematics-gender stereotype), namely that women have less ability than men in mathematics. Following this study, researchers expanded the idea to a broader context. Specifically, researchers suggested that doing science-related subjects, such as STEM, requires innate talent (Leslie, Cimpian, Meyer, & Freeland, 2015; Meyer et al., 2015) and women are stereotyped to have limited innate talent and to be inferior in science-related subjects in comparison to men in those subjects (e.g., Ma, 2008; Meyer et al., 2015). Therefore, a gender stereotype that science-related subjects are men's domains had been formed (Lindberg et al., 2010). This stereotype is widely held among young children, as well as by their parents and teachers (Furnham, Reeves, & Budhani, 2002; Hyde et al., 2008; Lindberg et al., 2010; Tsui, 2007).

The mathematics-gender stereotype would not only decrease women's interest in science-related subjects, but could also impair their performance in those subjects (Doyle & Voyer, 2016; Spencer et al., 1999). The association of the gender stereotype with educational attainment has been tested in various studies across cultures (J. Song et al., 2016; Tsui, Xu, & Venator, 2011), and across time (e.g., Bian et al., 2017; Cvencek, Meltzoff, et al., 2011; Leslie et al., 2015; Meyer et al., 2015; Spencer et al., 1999; M. T. Wang & Degol, 2017). For instance, a cross-national study found an association of the mathematics-gender stereotype with educational attainment (Nosek et al., 2009). Specifically, 8th grade children (across 34 countries) were more likely to associate science with men as opposed to women, and this male-science association predicted the gender differences in both science and mathematics performance (Nosek et al., 2009). The association did not change when researchers used adults rather than young children as participants (Nosek & Smyth, 2011).

The influence of the mathematics-gender stereotype has also been tested in Eastern countries, but the findings are inconsistent. For example, J. Song et al. (2016) examined how the mathematics-gender stereotype affected Chinese high school students' mathematics performance. A significant influence of the mathematics-gender stereotype was found only on female students' mathematics performance and not on that of their male counterparts (J. Song et al., 2016). However, Tsui et al. (2011) found no effect of the mathematics-gender stereotype on women's mathematics performance among university students. This again is a potential consequence of the one-child policy, which has provided a gender-neutral environment for recent generations.

6.5 The present studies

As stated above, gender stereotypes may affect children's educational attainment, and these effects could be influential and might be continuous throughout the lifespan of children. Thus, it would be valuable to detect when children start to acquire these stereotypes. If an exact time or period that children acquire the negative stereotypes could be found, then parents would be able to pay more attention at this time or during this period, and use appropriate approaches to decrease the negative effects of gender stereotypes. Consequently, this would decrease the negative effects of gender stereotypes, and in turn would help children to learn.

A recent influential paper published in Science (Bian et al., 2017) suggested that children adopted gender stereotypes about brilliance as early as age 6. Based on this study, I conducted a series of five empirical studies to test the earliest point when gender stereotypes about brilliance emerge, and their effects on children's educational attainment. I will introduce my hypotheses, experimental materials, experimental procedure, and main findings in detail in the next chapter.

Chapter 7. Five empirical studies testing the emergence of a gender stereotype about brilliance among children in China and the U.K.

This chapter introduces five empirical studies testing the emergence time of a gender stereotype about brilliance. Additionally, to my knowledge, the coinciding and complementary possibility – that whether children have gender stereotypes about low intellectual ability - has not been previously explored. If this is the case, would men or women have been stereotyped as having low intellectual ability? I generated data from both Chinese and British populations to explore these possibilities. All five studies reported in this Chapter will be turned into a peer-reviewed paper. Therefore, to maintain consistency with the published version and to acknowledge the collaborator's effort for its publication, the personal pronoun "we", instead of "I", is used in this chapter.

7.1 Introduction

Recently, it was reported that a stereotype of very-high intellectual ability as a male trait emerges as young as age 6, with boys and girls choosing images of men as more likely than women to match a person described as being "*really, really smart*" (Bian et al., 2017). Information about such stereotypes informs policy targeting gender gaps in Science, Technology, Engineering, and Mathematics (STEM) and related occupations (Leslie et al., 2015). It is important that findings are widely replicated, especially when they are being translated into policy (Open Science Collaboration, 2015). For this reason, we set out to test replicability of study 1 task (i) of Bian et al. (2017).

In task (i) of study 1, Bian et al. (2017) tested three groups of children aged 5, 6 and 7 years. A total of 96 children were tested, with equal numbers (n = 16) of boys and girls at each age. The method, described by Bian et al. (2017, p. S3), is as follows. "*One story was about a "really, really smart" person, and the other was about a "really, really nice" person. After telling the story, the experimenter laid out 4 pictures in a line (2 females and 2 males, randomly interspersed) and asked the child to guess which one of the 4 people might be the person in the story. If children chose a person of the same gender as themselves (e.g., if a girl picked a woman), they were assigned a score of 1 for that trial; otherwise, they received a 0.*" (Bian et al., 2017, p. S3).

Bian et al. (2017) found that 5-year-old children showed overwhelming same-gender positivity i.e., associating both brilliance and niceness with their own-gender. This same-gender preference has been demonstrated repeatedly in young participants (Cvencek, Greenwald, & Meltzoff, 2011, 2016; Dunham, Baron, & Banaji, 2016). In older groups, however, girls were much less likely than boys to associate brilliance with female targets e.g. at age 6-years, 48% of girls and 65% of boys chose a same-gender target as most likely to be very brilliant (Wald $\chi^2 = 8.10$, $p = .004$; Bian et al., 2017). Interestingly, a mirror-image change emerged for niceness, with older girls (67% at age 6 and 62 % at age 7) being much more likely than older boys (40% at age 6 and 43% at age 7) to attribute niceness to a target of their own-gender. These large (20%) gender differences were subsequently replicated in their study 2 of Bian et al. (2017), with a slightly larger sample (n = 144). The conclusion drawn was that environmental effects were leading boys and girls to adopt both of these stereotypes of women as being nicer and of men being more intellectually capable.

Bian et al. (2017) tested and ruled out potential confounds, such as the age of the participant depicted in the vignette (tests with adult and child stimuli yielded similar results), as well as race and ethnicity effects. They concluded that "*Contrary to the idea that the development of the stereotypes investigated here varies by racial/ethnic group, we found that race/ethnicity did not significantly moderate the key three-way interaction among trait, gender, and age*". The same non-significant moderating effect was reported in relation to socioeconomic status (SES), in that "*SES did not significantly moderate the key three-way interaction among trait, gender, and age*" (Bian et al., 2017, p. S6). Race, ethnicity, and SES, then, are unlikely to influence the results of a replication, and do not show in the theory of socialisation underlying gender stereotype formation.

One further potential moderating variable is culture. Bian et al. (2017) reported on what they interpreted as a culturally driven effect in American children. In our studies, we tested 10-year-old children in China (study 1) and as a comparison, we tested children in the same age range in the U.K. (studies 2, 3 and 4). It is possible that gender stereotyping both differs between cultures, and within a culture over time (Madon et al., 2001). For instance, Tsui (2007) found that most Chinese teachers continue to believe that men tend to be superior in mathematics and science, attributing this to better male spatial and visual ability. As in the West (Flore & Wicherts, 2015), evidence is mixed on the effect of gender stereotype threat on mathematics performance, with reports of significant effects (J. Song et al., 2016), and failures to replicate (Tsui et al., 2016). Given, if anything, stronger traditional gender stereotypes in Eastern cultures, we expected to find as strong (or stronger) effects in the Chinese sample. Regarding the older age of our sample (10 years old), we also expected similar-sized effects to those reported for 7 year-olds by Bian et al. (2017). We based this expectation on the suggestion by Bian et al. (2017) that these stereotypes increase over time, and by using an older age group, they have more time to take effect and develop.

We also set out to replicate the stereotype effect for female niceness. This provides a useful internal control for the replication. Unlike brilliance, or at least unlike cognitive ability, for which data show negligible mean differences in ability (Deary, Irwing, Der, & Bates, 2007), studies of gender differences in personality traits linked to niceness, such as agreeableness or sensitivity (e.g. Hyde, 2005; Lippa, 2010; Weisberg et al., 2011). Instead, it has been consistently reported that women score as more sensitive, warm and agreeable than men (Del Giudice et al., 2012; Weisberg et al., 2011). These effects also replicate across nations (Lippa, 2010). For niceness then, as opposed to brilliance, there is some evidence for genuine mean differences. We hypothesised both based on Bian et al. (2017) and wider literature on gender differences in personality, that boys and girls would pick a female target as most likely to be very nice.

Finally, to deepen our understanding of the apparent brilliance effect, we wished to examine stereotypes regarding the pole opposite to brilliance: namely "very low" intellectual ability. If stereotyping for brilliance works by shifting the likelihood of brilliance in the "male" direction, then, logically, it should also be the case that a vignette about a person with very low ability will be stereotyped by children as more likely to be female. We termed this the "trait shift model" of male brilliance stereotyping. A distinct possibility, in the opposite direction, would be that men are stereotyped as being both more likely to be bright and more likely to be very slow-minded. This competing hypothesis can be derived from data showing higher variance in IQ among men compared to women (Deary, Irwing, et al., 2007). If the stereotype of brilliance as a male trait is replicated, adding an assessment of low-ability gender stereotyping, then, would allow us to cast additional light on the processes involved in generating such an effect.

7.2 Study 1

In our study 1, we wished to test all of the three gender stereotypes about smartness, niceness and slow-mindedness in a Chinese sample involving 10-year-old children. The experimental materials and procedure used in this study closely followed task (i) study 1 in Bian et al. (2017), with three major differences noted below (see also Table 7.1, tabulating points of similarity and difference between Bian et al. (2017) and studies 1, 2, 3 and 4 in the present paper). First, in Bian et al. study 1 task (i), different sets of four images were used in each of the two stories (Bian, personal communication). When conducting the study, we were not aware of this, and in our study 1, we had wrongly interpreted the method described in the supplement (*"After telling the story, the experimenter laid out 4 pictures in a line (2 females and 2 males, randomly interspersed"*(Bian et al., 2017, p. S3)) as implying the same images were recycled across the stories. We therefore used only one set of images. Second, Bian et al. (2017) also presented the smart and nice conditions in a random order. To reduce possible effects of previous conditions on the critical brilliance stereotype effect, we always presented the "*really, really smart*" condition first in all cases. For this reason, the images were always novel to children at the point of answering the brilliance stereotype questions, and thus the differences in methods cannot affect the data for the key task underlying the core hypothesis about brilliance. It does however, introduce a sequential-choice effect for the nice and slow-minded conditions, which we would discuss below. Finally, children were asked to answer a paper-based survey in a classroom setting, rather than individually in a laboratory or classroom. Thus, the stories were printed on a sheet of paper, with the four target images printed in a row at the bottom of the page.

Our three hypotheses in study 1 were as follows. First, we hypothesised that girls would be significantly less likely than boys to select a same-gender target as a match for the narrative of a really smart person, replicating the finding of Bian et al. (2017). Second, we predicted that boys and girls would select a female target as a match to the narrative of a really nice person significantly more often than they would choose a male target, replicating the niceness stereotype favouring women (Bian et al., 2017). Third, based on the trait-shift model outlined above, we hypothesised that women would be chosen as most likely to be slow-minded. Our alternative experimental hypothesis is that both boys and girls would select a male target as most likely to be slow-minded, as predicted from higher male-variance theory.

Table 7.1 Similarities and differences between Bian et al. (2017) study 1 task (i) and the present studies 1, 2, 3 and 4

	Bian et al. (2017)	**The present paper**			
	Study 1 task (i)	**Study 1**	**Study 2**	**Study 3**	**Study 4**
Participants	N = 32 (Equal numbers of boys and girls) Two additional groups of n = 32 tested at ages 5 and 6.	N = 227 (118 boys, 109 girls)	N = 100 (52 boys, 48 girls)	N = 200 (111 boys, 89 girls)	N = 210 (119 boys, 91 girls)
Age	Mean age = 5.55, 6.50, and 7.44 for each age group respectively.	Mean age = 10.09.	Mean age = 7.96.	Mean age = 10.04.	Mean age = 8.03.
Ethnicity	78% European American, 7% Asian American, 5% African American, 3% Latino or Hispanic, and 7% multi-racial.	100% Chinese	100% were from England	85% were from England, 4.5% were from Wales, 2% were from Northern Ireland, 8% were from Scotland and 0.5%	100% were from England

				did not provide their nationality.	
Test situation	Individually tested in a quiet room in the lab or at their school.	Children were asked to answer the survey individually in their classroom.	Children were asked to do an online survey.	Children were asked to do an online survey.	Children were asked to do an online survey.
Story stimuli	Two stories: One was about a "really, really smart" person, one was about a "really, really nice" person.	Two matched stories: One was about a "really, really smart" person, one was about a "really, really nice" person.	Two matched stories: One was about a "really, really smart" person, one was about a "really, really nice" person.	Two matched stories: One was about a "really, really smart" person, one was about a "really, really nice" person.	Two matched stories: One was about a "really, really smart" person, one was about a "really, really nice" person.
Slow-minded story	Not used	A third story about a "really, really slow-minded" person was used.	A third story about a "really, really slow-	A third story about a "really, really slow-minded" person was used.	A third story about a "really, really slow-minded" person was used.

		Back-translated to assure fidelity.	minded" person was used.		
Photo stimuli	Photos of adult men (n=2) and women (n=2) **Different photo-sets for nice and smart stories**	Photos of adult men (n=2) and women (n=2) targets drawn from a search engine and matched for attractiveness and professional status. **Same photos reused between the three stories**	Photos of adult men (n=2) and women (n=2) **Different photo-sets for nice and smart stories**	Photos of adult men (n=2) and women (n=2) **Different photo-sets for nice and smart stories**	Photos of adult men (n=2) and women (n=2) **Different photo-sets for nice and smart stories**
DV	A choice of a target image of the same gender as the participant was scored 1, otherwise, 0.	A choice of a target image of the same gender as the participant was	A choice of a target image of the same gender as the participant	A choice of a target image of the same gender as the participant was	A choice of a target image of the same gender as the participant was

		scored 1, otherwise, 0.	was scored 1, otherwise, 0.	scored 1, otherwise, 0.	scored 1, otherwise, 0.
Exclusions	19 children were excluded for matching fewer than 4/6 pre-screens regarding meaning of bright and nice. Three children for refusal to finish and one for stereotype > 2.5 SDs away from the mean.	No exclusions.	No exclusions.	No exclusions.	No exclusions.
Order of testing	Smart and nice stories were **presented in randomised order**. Half of the children heard the smart	Smart story **presented first**, "nice" story given second, followed by slow-minded story.	Smart, nice and slow-minded stories **presented in randomised order**.	Smart, nice and slow-minded stories **presented in randomised order**.	Smart, nice and slow-minded stories **presented in randomised order**.

	story first and the other half heard the nice story first.		32% children saw the smart story first, 32% children saw the nice story first and 36% children saw the slow-minded story first.	26% children saw the smart story first, 34.5% children saw the nice story first and 39.5% children saw the slow-minded story first.	30% children saw the smart story first, 31.9% children saw the nice story first and 38.1% children saw the slow-minded story first.

7.2.1 Method

7.2.1.1 Participants

A total of 227 children (118 boys, 109 girls) were recruited from a public primary school in China. Children were aged from 7 years 8 months to 11 years 4 months old (mean age = 10.09, SD = 0.52). Socioeconomically, the region in which we recruited was 21% below the Chinese national average income in 2017 (average income 48,881 Yuan: National Bureau of Statistics of the People's Republic of China, 2017).

7.2.1.2 Materials

The two stories in task (i) study 1 in Bian et al. (2017) describing a "*really, really smart*" person and a "*really, really nice*" person were used in this study. The stories were translated into Chinese by the experimenter (YL), and a back-translation into English was then made by a bilingual PhD student, and this was iterated until the Chinese story was back-translated without appreciable loss of fidelity from the original. The novel third story, describing a "*really, really slow-minded*" person was derived by modifying the smart story. All three stories are presented in Appendix A.

In choosing the response-options, i.e., male and female target images, we aimed to match the characteristics of the stimuli reported in task (i) study 1 of Bian et al. (2017). These were not described in detail, but a closely related study (task (ii) in study 1) described the images used as being of "*white men and women, normed for attractiveness… and professional dress in a sample of 29 adults recruited via Amazon's Mechanical Turk*" (p. S3). We therefore followed the criteria described in task (ii) study 1 of Bian et al. (2017) as the source of searching criteria for the images used in the present study. Also, because Bian et al. (2017) did not specify the specific image searching engine they used in their studies, we used an internet search (i.e. "Google images") as the search engine for images. Therefore, to match our sample, we sourced images of Chinese men and women, normed for attractiveness and in professional dress. We used the keywords "Chinese", "professional dress ID photos", "blue background colour" and "face front" to decrease the possible image background confounds. The experimenter selected ten female images and ten male images ranked highly in this search. These were then shown to four female Chinese PhD students at University of Edinburgh, who were asked to each select four images (two of each gender) which they thought matched on attractive and professional appearance from the original 20 images. The four images selected as matches most often were chosen as the final stimuli (2 male and 2 female images).

7.2.2 Procedure

Children were tested in their classrooms. The stories were presented in the same session as a brief survey for a separate study. All of the studies in the present paper were approved by the Research Ethics Committee of University of Edinburgh. After gaining consent, the experimenter asked the children to fill out a demographic survey including their age, gender, and their most recent year's school grades. Then children were asked to read the first story, about a "*really, really smart*" person. They then had to select one image from the four they had in front of them that they felt was the most likely to be the person described in the story. They were then asked to read the second story (a "*really, really nice*" person), and asked to pick the most likely target image (the children were told that each image could be chosen more than once). Finally, they read the third story (a "*really, really slow-minded*" person), and identified the target image most likely to match the story. At the end of the session, oral praise and thanks were given to participants.

7.2.3 Results

We first tested hypothesis one, that girls would be significantly less likely to select a same-gender target as a match for the narrative of a really smart person than boys would be. This was examined using a binomial general linear model. Children's gender was used as a predictor of the sex of the target image selected, with age as a covariate. A male target image was selected as "*really, really smart*" by 79 boys and 39 girls. By contrast, 39 boys and 70 girls chose a female target. Thus, the hypothesised same-gender effect was not supported, with a non-significant estimate for participant gender as a predictor of the sex of the target image (β = -0.07, *OR* = 0.93, CI_{95} [-0.63, 0.50], z = -0.23, p = .817; also see the upper left graph in Figure 7.1). Instead, both boys and girls were significantly biased toward favouring a target of their same-gender as a match for the "*really, really smart*" vignette (β = 1.34, *OR* = 3.82, CI_{95} [0.78, 1.92], z = 4.61, p < .001).

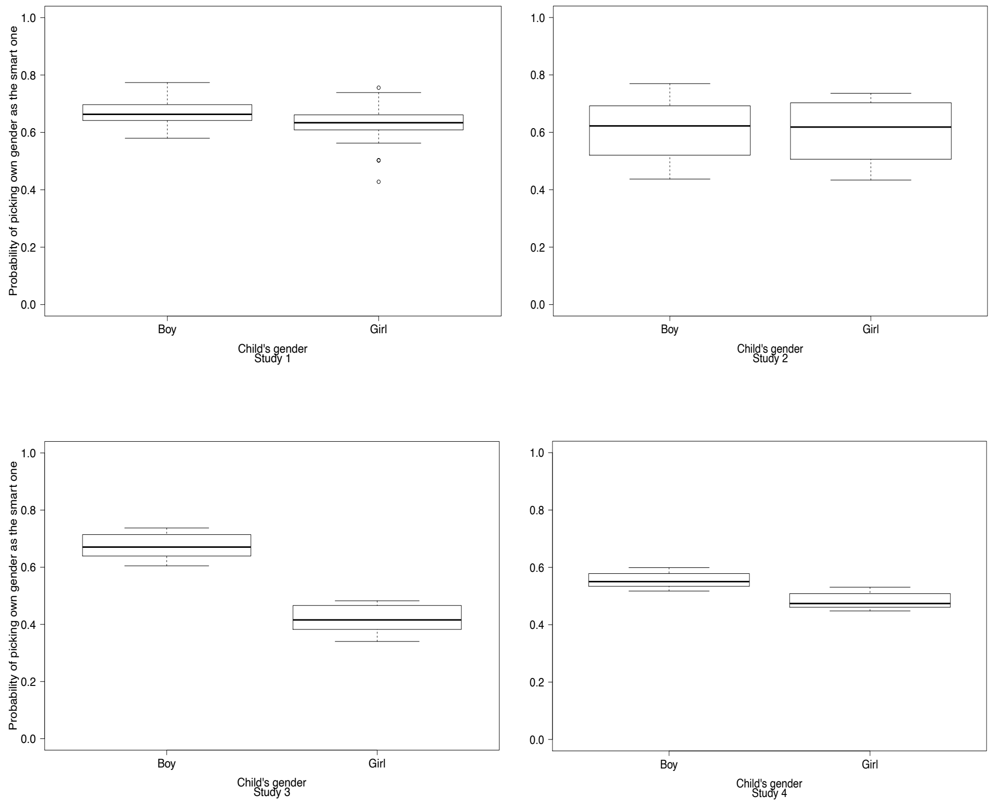

Figure 7.1 Gender stereotypes regarding beliefs about smartness in studies 1, 2, 3 and 4 (participant gender was used as a predictor of the sex of the target image selected, with age as a covariate)

We next tested hypothesis two, that both boys and girls would select a female target as a match for the narrative of a "*really, really nice*" person significantly more often than they chose a male target. Again, a binomial general linear model was used, the target image selected to be nice as the dependent variable, children's gender as the independent variable, and age as a covariate. As predicted, the results indicated that both boys (82 out of 118) and girls (86 out of 109) were biased toward selecting a female target as "*really, really nice*" ($\beta = 2.10$, $OR = 8.17$, $CI_{95}[1.49, 2.73]$, $z = 6.65$,

$p < .001$, also see the upper left graph in Figure 7.2). There was no evidence for a gender difference in this bias, with children of either gender showing a strong bias toward a female target as niceness ($\beta = 0.44$, $OR = 1.56$, $CI_{95}[-0.18, 1.08]$, $z = 1.39$, $p = .164$).

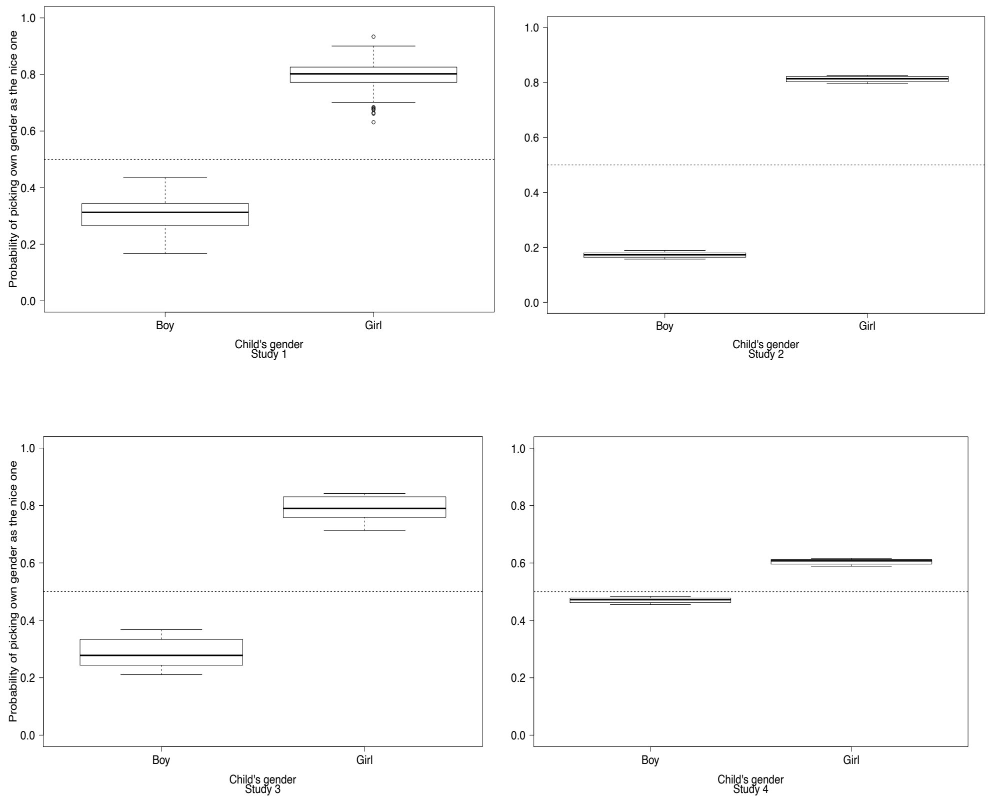

Figure 7.2 Gender stereotypes regarding beliefs about niceness in studies 1, 2, 3 and 4 (participant gender was used as a predictor of the sex of the target image selected, with age as a covariate)

Finally, we tested hypothesis three, that both boys and girls would both be more likely to select a female target as most likely to be "*really, really slow-minded*". Given the failure to replicate the predicted male-brilliance stereotype, we were curious to see what might emerge for this low intellectual ability stereotype. A majority of both boys (76 out of 118) and girls (92 out of 109) chose a male target as the "*really, really slow-minded*" person. A binomial general linear model with the gender of the chosen target as the dependent variable, children's gender as the independent variable and age as a covariate indicated that this bias against men was large and significant (β = -2.22, *OR* = 0.11, CI_{95}[-2.89, -1.59], z = -6.71, p < .001). The difference in bias against men (64% of boys and 84% of girls choosing a man) itself was significant, suggesting that the prejudiced stereotype against men was stronger in girls (β = -1.10, *OR* = 0.33, CI_{95} [-1.78, -0.46], z = -3.29, p < .001).

7.2.4 Study 1 discussion

There were three interesting findings in our study 1. First, we found that, in a sample of over 200 Chinese children, 10-year-old boys and girls did not generally associate brilliance with the male gender. Rather, they associated this positive trait with their own-gender. Second, both genders were biased to view niceness as a female trait. Finally, both genders stereotyped extreme low intellectual ability as a male trait. Despite replicating the bias toward women for niceness, we were unable to replicate the bias toward men for brilliance, as reported by Bian et al. (2017). We instead accumulated some evidence for a bias against men when it comes to stereotypes about extremely low intellectual ability.

Turning to the core hypothesis which we failed to replicate among Chinese children, rather than women showing a bias to stereotype brilliance as a male trait, we found that both genders stereotyped brilliance as an own-gender trait. Clearly, more studies are required to test this effect in children of different ages, and across different cultures, to establish reliable findings. One possibility raised on reflection is that the age of our sample may explain the difference. By age 10, children are sufficiently reflective to modify their interpretation of the task, despite having, in the case of girls, internalised a negative same-gender stereotype, such that they can consciously attribute the positive attribute to their own-gender, while unconsciously retaining a bias against their gender. We see problems with this argument, however. It seems unable to account for the successful replication of an opposite-gender bias for niceness in boys. Traits included in the "*really, really nice*" vignette, such as being friendly and helpful, are rated positively and are valued in others (D. M. Buss & Barnes, 1986). On the self-awareness hypotheses, then, 10-year-old boys should once again attribute this positive trait to themselves. Similarly, for very low intellectual ability, both genders should attribute this negative trait to the opposite gender, which they did not. Thus, age does not seem to be able to straight-forwardly account for the pattern of results.

For the striking finding regarding low intellectual ability, procedural effects seem unlikely to have driven the outcome. While the slow-minded question was asked last in our study 1, and we used the same images for each story, it might be thought that earlier choices might have subjectively reduced the available stimuli, making it more likely that the choice was made from the remaining two unselected stimuli. While this highlights an important potential confound in sequential testing designs, where participant's choices can be driven by components in the design, it does not seem to account for the present finding. Whilst girls were more likely to choose a female target as most likely to be brilliant, and also as most likely to be very nice (leaving both male targets unselected), boys were more likely to choose a male as most likely to be brilliant, and a female target as most likely to be nice, leaving them with one male and one female target as yet unchosen. If sequential testing effects were active, these could, then, account for an anti-male bias in girls, but not in boys. It may be, then, that children have a genuine stereotype of men as more likely to be very low in intellectual ability, despite, in this sample, expressing a strong same-gender bias for extremely high intellectual ability.

When planning our following studies, we decided to test the replicability of the stated bias effects in younger children in Western cultures, and to test 10-year-old Western children to establish the pattern at that age in those cultures. We also realised that we had wrongly interpreted the method in Bian et al. (2017) as implying the same images were recycled across the stories. Thus, we decided to use different sets of images for each story in our following studies. These considerations lead us to undertake a second replication, modified as introduced below.

7.3 Study 2

As outlined above, in constructing our study 2, we made five changes in the methodology to enhance the power of the design. First, we recruited participants from a Western culture (i.e. the U.K.) that is similar to the culture in the original study of Bian et al. (2017). Second, as in Bian et al. (2017), we provided different sets of images for each of the three stories in our study 2. More specifically, since study 2 would be conducted in the U.K., to be more methodologically close to the original study, we adopted the identical two sets of images in task (i) of Bian et al. (2017) for the smart and nice stories. For the slow-minded condition, we adopted one set of images used in task (iii) of Bian et al. (2017). Third, instead of presenting the stories in a "smart, nice and slow-minded" order, we presented the three stories randomly in our study 2 to avoid a sequential-choice effect. Fourth, to maximise our opportunity to detect a gender stereotype about brilliance, in study 2, we matched our participants' age to the oldest age group (i.e., age 7) in Bian et al. (2017). Finally, we wished to expand the scope of our study, to test whether children at young childhood would have gender differences in occupations and whether a gender stereotype about brilliance would be associated with children's career interests. A 6-item career interests scale was added in our study 2 to test this.

There were five hypotheses in our study 2. First, girls would be significantly less likely than boys to select a same-gender target as a match for the narrative of a really smart person. Second, boys and girls would select a female target as a match to the narrative of a really nice person significantly more often than they would choose a male target. Third, both boys and girls would select a female target as most likely to be slow-minded. Fourth, there would be gender differences in career interests among children at 7 years old. Finally, based on Leslie et al. (2015), we hypothesise that a gender stereotype about brilliance would be significantly associated with children's career interests.

7.3.1 Method

7.3.1.1 Participants

In total, 100 parent-child dyads in England were recruited from Prolific (an online participant pool). Our child participants (52 were boys and 48 were girls) were aged from 6 years 3 months to 9 years 2 months old (mean age = 7.96, SD = 0.86). Among adult participants, 53% were mothers. Participants' family socioeconomic status (SES) was measured by asking the highest educational qualification attained by a parent in the family and the highest-earning parent's employment type. 54% of families had a parent who had achieved a bachelor or a higher degree and 64% of families had a parent working as a professional, administrator or official.

7.3.1.2 Materials

The smart, nice and slow-minded stories were identical to those used in our study 1. All of the images used for each of the three stories were adopted from the original study of Bian et al. (2017).

The 6-item career interests scale was modified from the items list in the "Universal sex differences involving stratification and work and occupations" table (Ellis, 2011, p. 556). The first five items assessed what kind of jobs children wish to take when they grow up (i.e., "I would like to work as an executive manager", "I would like to work in law enforcement", "I would like to work as a nurse", "I would like to work as a scientist or engineer" and "I would like to work as a university professor"). Children were asked to indicate the extent of their interests for each item from a 5-point Likert scale: 1: "Not at all interested", 2: "Not very interested", 3: "Somewhat interested", 4: "Quite interested" and 5: "Very interested". The sixth item in the career interests scale assessed children's expected work-life balance when growing up (i.e., whether children would like to spend more time at work or with parenting and childcare). Responses were recorded on a 7-point Likert slider with 1 equal to working long hours and 7 equals to spending more time with parenting and childcare.

7.3.1.3 Procedure

After giving consent, a welcome message and instructions for the study were shown to parents. Parents were informed that the online survey included two sections and that they needed to answer the first section and their children needed to answer the second section. Then parents were asked to answer eight questions including parents' gender and age, parent's educational status, parent's employment status, children's year of birth and month of birth, children's gender and children's key stage 1 national tests scores in 2018.

Once parents finished their section, they saw a reminding message showing that their children needed to come to the screen. Once children were sitting down in front of the screen, parents were asked to click the continue button to start the second section. Children were first instructed that they would see three brief stories about a person, and for each story, they needed to select an image that they thought to be the person described. Following the instruction, the first story about a "*really, really smart*" person with a set of four images on the same page was shown to children. Once children chose an image and clicked the arrow to the next page, they saw the second story about a "*really, really nice*" person with a second set of four images. Again, once children chose an image and clicked the arrow to the next page, they saw the third story about a "*really, really slow-minded"* person and a third set of four images. After reading the three stories, a new page with the 6-item career interests scale was shown to children. They were asked to indicate the extent of their interests for the first five items and their preference on a slider for the sixth item. Finally, children saw a debriefing message showing that they had finished all the items and the adult participants would receive £1.50 as compensation once the experimenter verified their answers.

7.3.2 Results

As in study 1, we first tested hypothesis one that girls would be significantly less likely than boys to select a same-gender target as a match for the narrative of a really smart person. Again, this hypothesis was tested using a binomial general linear model, with children's gender used as a predictor of the sex of the target image selected, and age as a covariate. Thirty-two boys and nineteen girls selected a male target image as most likely to be "*really, really smart*". By contrast, 20 boys and 29 girls chose a female target. Similar to our study 1, the first hypothesis that boys would show higher same-gender effect than girls was not supported (β = -0.18, *OR* = 0.84, CI_{95} [-1.02, 0.65], z = -0.43, p = .670; also see the upper right graph in Figure 7.1). Instead, both boys and girls were significantly more likely to select a target of their same-gender as a match for the really smart vignette (β = 0.88, *OR* = 2.41, CI_{95} [0.07, 1.70], z = 2.12, p = .034).

Secondly, we tested hypothesis two that both boys and girls would select a female target as a match to the narrative of a really nice person significantly more often than they would choose a male target. As predicted, a majority of both boys (43 out of 52) and girls (39 out of 48) selected a female target as a match of the "*really, really nice*" person (β = 3.01, *OR* = 20.3, CI_{95} [2.04, 4.10], z = 5.75, p < .001; also see the upper right graph in Figure 7.2). No gender difference was found for this bias, indicating that both boys and girls showed a strong bias towards a female target for niceness (β = -0.05, *OR* = 0.95, CI_{95} [-1.10, 1.00], z = -0.10, p = .922).

Thirdly, we tested hypothesis three that both boys and girls would select a female target as most likely to be slow-minded. Contrary to prediction, a majority of boys and girls (36/52 and 38/48 respectively) chose a male target as a match for the "*really, really slow-minded*" person (β = -2.17, *OR* = 0.11, CI_{95} [-3.15, -1.28], z = -4.59, p < .001). In addition, there was no gender difference in this bias (β = -0.58, *OR* = 0.56, CI_{95} [-1.53, 0.34], z = -1.22, p = .222).

Next, we tested hypothesis four that there would be gender differences in career interests among children at 7 years old. This hypothesis was tested using six linear regressions, with children's scores in each of the career interest items as the dependent variables respectively, children's gender as the independent variable, and age as a covariate. There were significant gender differences in the career interest of working in law enforcement (β = -0.48, CI_{95} [-0.86, -0.09], t = -2.47, p = .015) and working as a nurse (β = 0.96, CI_{95} [0.61, 1.32], t = 5.38, p < .001). However, no significant gender difference was found in any other career interests or the work-life balance question (working as an executive manager: p = .764, working as a scientist or engineer: p = .300, working as a university professor: p = .539, and the work-life balance: p = .484).

Finally, we tested hypothesis five that a gender stereotype about brilliance would be significantly associated with children's career interests. This hypothesis was tested using six linear regressions, with children's scores in each of the career interest items as the dependent variables respectively, the interaction of children's gender and their selections on the smart story as the independent variable, and age as a covariate. Contrary to prediction, the interaction of children's gender and their selections on the smart story was not significantly associated with any of the six career interests (p = .603 for working as an executive manager; p = .667 for working in law enforcement; p = .282 for working as a nurse; p = .700 for working as a scientist or engineer, p = 1.000 for working as a university professor, and p = .412 for spending more time at work versus parenting activities).

7.3.3 Study 2 discussion

There were five findings in our study 2. First, within a population that had a similar age range (7-year old) and a similar culture (U.K.) to that of the original study of Bian et al. (2017), we failed to find support for a gender stereotype towards men for brilliance. Instead, we replicated our results in study 1, that both boys and girls were more likely to associate brilliance with their own-gender. Second, we again verified that both boys and girls viewed niceness as a female-trait, as reported in our study 1 as well as in Bian et al. (2017). Third, similar to our study 1, both boys and girls were more likely to select a male target rather than a female target as a match for the slow-minded person. Fourth, we did find evidence to support the existence of gender differences in career interests, but only for two kinds of career (i.e., working in law enforcement and working as a nurse). Finally, a gender stereotype about brilliance was not significantly associated with children's career interests.

Our findings in study 2 might reduce the possibility that culture would affect the replicability of a gender stereotype about brilliance. Since the sample size in our study 2 was relatively small and children were younger (7 versus 10-year-old) than our study 1, it was not clear whether British children at an older age would show a similar tendency for gender stereotype about brilliance. Therefore, we conducted a third study, with children from Western culture, but with a similar age as in our study 1 (i.e., 10 years old).

7.4 Study 3

Study 3 was a replication study of study 2, with two minor changes. First, to increase the diversity of participants in our study, we recruited children from across the whole of the U.K. rather than only in England. Second, as noted above, we aimed to recruit children from 7 to 12 years old to test if older Western children would show a similar bias about brilliance as in studies 1 and 2. We also made five hypotheses in study 3, which were identical to those in study 2.

7.4.1 Method

7.4.1.1 Participants

In total, 200 parent-child dyads were recruited from Prolific. Child participants (85% from England, 111 were boys and 89 were girls) were aged from 6 years 3 months to 13 years 2 months old (mean age = 10.04, SD = 2.20). For the family SES, 51% of families had a parent who had achieved a bachelor or a higher degree, and 65% of families had a parent working as a professional, administrator or official.

7.4.1.2 Materials

The materials used in study 3 were identical to those used in study 2, except that the two questions asking for parents' age and gender were deleted from the survey.

7.4.1.3 Procedure

All of the experimental procedures were identical to those of study 2. Participants were compensated with 50p once their answers were verified by the experimenter.

7.4.2 Results

As in studies 1 and 2, we first tested hypothesis one that girls would be significantly less likely than boys to select a same-gender target as a match for the narrative of a really smart person. Interestingly, 75 boys and 52 girls selected a male target image as most likely to be "*really, really smart*", which was contrary to our previous findings. By contrast, 36 boys and 37 girls chose a female target for smartness. Thus, the hypothesis that girls would be significantly less likely than boys to select a same-gender target as a match for the narrative of a really smart person was supported (β = -1.10, *OR* = 0.33, CI_{95} [-1.70, -0.52], z = -3.70, p < .001; also see the lower left graph in Figure 7.1). In addition, there was a gender difference in this bias (β = 0.40,

OR = 1.50, CI_{95} [-0.18, 0.98], z = 1.35, p = .177). Since we used identical materials as in study 2 but found contrary findings, this difference might be caused by the broader nationality and older age among children in study 3. Thus, we did a further analysis with restricted nationality and age. Specifically, we restricted the analysis to those children who lived in England only and were younger than 9 years 2 months old. In total, 39 boys and 23 girls were included in the analyses. A consistent result that a majority of both boys (24 out of 39) and girls (14 out of 23) selected a male target as a match for a "*really, really smart*" person was found, but this bias was not significant (β = -0.92, OR = 0.40, CI_{95} [-2.01, 0.12], z = -1.70, p = .088).

Secondly, we tested hypothesis two that boys and girls would select a female target as a match to the narrative of a really nice person significantly more often than they would choose a male target. As predicted, 79 (out of 111) boys and 70 girls (out of 89) selected a female target as a match for the "*really, really nice*" person (β = 2.21, OR = 9.12, CI_{95} [1.57, 2.89], z = 6.60, p < .001; also see the lower left graph in Figure 7.2). There was no gender difference for this bias (β = 0.46, OR = 1.58, CI_{95} [-0.20, 1.14], z = 1.35, p = .176).

Thirdly, we tested hypothesis three that both boys and girls would select a female target as most likely to be slow-minded. Similar to studies 1 and 2, this hypothesis was not supported, with 70 boys and 65 girls choosing a male target as the "*really, really slow-minded*" person (β = -1.53, OR = 0.022, CI_{95} [-2.15, -0.93], z = -4.91, p < .001). Additionally, no significant gender difference was found for this bias (β = -0.46, OR = 0.63, CI_{95} [-1.08, 0.14], z = -1.48, p = .138).

Next, we tested hypothesis four that there would be gender differences in children's career interests. Gender differences were significantly found in three kinds of career interests including working as an executive manager (β = -0.30, CI_{95} [-0.58, -0.03], t = -2.16, p = .032), working as a nurse (β = 0.68, CI_{95} [0.42, 0.94], t = 5.18, p < .001), and working as a scientist or engineer (β = -0.57, CI_{95} [-0.84, -0.29], t = -4.11, p < .001). There was no gender difference in career interests of working in law enforcement, working as a university professor, or work-life balance (p value was .265, .712, and .194 respectively).

Finally, we tested hypothesis five that a gender stereotype about brilliance would be significantly associated with children's career interests. Similar to study 2, no significant result was found for the interaction effect on any of the career interests (p values were range from .056 to .781).

7.4.3 Study 3 discussion

Similar to studies 1 and 2, we found that both boys and girls were biased towards female gender for niceness, and male gender for slow-mindedness. More strikingly, contrary to studies 1 and 2, we found that both boys and girls were biased towards men for smartness, which was compatible with the original study of Bian et al. (2017). This gender stereotype towards men for smartness was consistent even after restricting children's age and nationality to match with those in our study 2. Due to the completely contrary findings in the two British samples, we decided to run the fourth study, in an independent sample, to gather more evidence regarding whether children were biased towards male or female gender for smartness.

7.5 Study 4

In study 4, we made five improvements to make the methodology more rigorous than our previous studies. First, we swapped the two sets of images for the smart and nice stories used in our studies 2 and 3 to reduce the possible impacts of the use of images on children’s selections. Additionally, instead of presenting the four images randomly for each of the three stories, we presented each set of images in two specific orders (i.e., “man, woman, man, woman” or “woman, man, woman, man”) in our study 4. This improvement aimed to reduce a possibility that children might select one gender superior the other gender in the orders of “man, man, woman, woman” or “woman, woman, man, man” when presenting the images randomly. Third, instead of presenting the story and response-options in the same page, we separated them into two individual pages to make our design be closer to the original study of Bian et al. (2017). This improvement could also reduce the possibility that children would not read the story but just select an image based on their first-impressions on the images. Fourth, before answering the career interests scale, children were asked to use three words to indicate the three kinds of person described in each story to check their understandings of the stories. Finally, since participants in studies 2 and 3 slightly differed in nationality and age, to test whether findings in study 2 or 3 were more reliable it was necessary to follow one of the two studies’ standards to recruit participants. We decided to follow study 2 to recruit participants, and thus child participants would be those who were living in England and aged around 7 years old.

As in studies 2 and 3, we made five hypotheses for study 4.

7.5.1 Method

7.5.1.1 Participants

In total, 210 parent-child dyads were recruited from Prolific. Child participants (119 were boys and 91 were girls) were aged from 6 years 4 months to 9 years 3 months old (mean age = 8.03, SD = 0.88). For the family SES, 56% of families had a parent who had achieved a bachelor or a higher degree, and 65% of families had a parent working as a professional, administrator or official.

7.5.1.2 Materials

The materials used in study 4 were identical to those used in studies 2 and 3, except that the sets of images used for the smart and nice stories were swapped.

7.5.1.3 Procedure

The experimental procedure was identical to those used in studies 2 and 3. Participants were compensated with 50p once their answers were verified by the experimenter.

7.5.2 Results

As in studies 2 and 3, we first tested hypothesis one that girls would be significantly less likely than boys to select a same-gender target as a match for the narrative of a really smart person. Similar to study 3, a majority of both boys (66 out of 119) and girls (47 out of 91) selected a male target as most likely to be "*really, really smart*". However, this same-gender effect was not significant (β = -0.28, *OR* = 0.76, CI_{95} [-0.83, 0.27], z = -0.99, p = .320; see the lower right graph in Figure 7.1). Additionally, there was no gender difference in this bias (β = 0.18, *OR* = 1.20, CI_{95} [-0.37, 0.74], z = 0.65, p = .517).

Secondly, we tested hypothesis two that boys and girls would select a female target as a match to the narrative of a really nice person significantly more often than they would choose a male target. As predicted, 63 boys and 55 girls chose a female target for niceness (β = 0.54, *OR* = 1.72, CI_{95} [-0.01, 1.10], z = 1.91, p = .056; also see the lower right graph in Figure 7.2). There was no gender difference for this bias (β = 0.31, *OR* = 1.36, CI_{95} [-0.25, 0.86], z = 1.08, p = .279).

Thirdly, we tested hypothesis three that both boys and girls would select a female target as most likely to be slow-minded. Similar to studies 2 and 3, this hypothesis was not supported, with 89 boys and 69 girls choosing a male target as a match for the "*really, really slow-minded*" person (β = -2.24, *OR* = 0.11, CI_{95} [-2.90, -1.62], z = -6.87, p < .001). Again, no gender difference was found for this bias (β = -0.07, *OR* = 0.93, CI_{95} [-0.72, 0.56], z = -0.22, p = .827).

Finally, we tested hypothesis four that there would be gender differences in career interests among children at 7 years old and whether a gender stereotype about brilliance would be significantly associated with children's career interests. Two significant gender differences were found in children's interests in working as a nurse (β = 0.73, CI_{95} [0.48, 0.98], t = 5.68, p < .001) and working as a scientist or engineer (β = -0.61, CI_{95} [-0.88, -0.35], t = -4.60, p < .001). There was no significant gender difference in other career interests including working as an executive manager (p = .796), working in law enforcement (p = .135), working as a university professor (p = .726), and work-life balance (p = .157). Similar to studies 2 and 3, gender stereotype about brilliance was not significantly associated with any of children's career interests (p was from .523 to .999).

7.5.3 Study 4 discussion

As in studies 1, 2 and 3, both boys and girls had a gender stereotype towards female gender for niceness, and male gender for slow-mindedness (see Table 7.2 for the comparison of results in study 1 of Bian et al. (2017) and in present studies 1, 2, 3, 4). For the core hypothesis of gender stereotype about smartness, as in study 3, a majority of both boys and girls chose a male target as a match for the "*really, really* *smart*" person. This finding together with findings in study 3 suggested that there was a gender stereotype about brilliance favouring men among British children.

Interestingly, our studies consistently found that both boys and girls had a gender stereotype about slow-mindedness favouring men, not only in China but also in the U.K. These findings could confirm the higher male-variance theory among British children. Moreover, based on these results, it could logically be predicted that girls would be more likely to be viewed as having average ability than boys. This prediction was then tested in the following study.

Table 7.2 The genders children selected for a match of the smart, nice and slow-minded person in study 1 of Bian et al. (2017) and the present studies 1, 2, 3 and 4

	Bian et al. (2017)	**The present paper**			
	Study 1	**Study 1**	**Study 2**	**Study 3**	**Study 4**
Smartness	5yrs: Own-gender 6yrs: Male gender 7yrs: Male gender	Own-gender	Own-gender	Male gender	Male gender
Niceness	5yrs: Own-gender 6yrs: Female gender 7yrs: Female gender	Female gender	Female gender	Female gender	Female gender
Slow-mindedness	NA	Male gender	Male gender	Male gender	Male gender

7.6 Study 5

This follow-up study used a similar method to that in task (i) study 1 of Bian et al. (2017). Specifically, children would read a story about a person who has average intellectual ability and then they would be asked to select an image from a set of four images as a match of the described person.

Based on the results in study 4, our first hypothesis was that a female target would be more likely to be selected as a match of the person who has average ability than a male target. Since participants in studies 2 and 4 had the same nationality, but showed different gender stereotypes about brilliance, we wished to test if those participants would also have different gender stereotypes about average ability. Based on the results in studies 2 and 4, we made a second hypothesis that children who participated in study 4 would be more likely to select a female target as having average ability than those who participated in study 2.

7.6.1 Method

7.6.1.1 Participants

In total, 268 (223 boys and 45 girls, mean age = 8.03, SD = 0.87) children participated in the follow-up study. There were 84 children who had participated in study 2 and 184 who had children participated in study 4.

7.6.1.2 Materials

The story used in this follow up study was also derived by modifying the smart story in Bian et al. (2017): "*There are lots of people at the place where I work. But there is one person who is really special. This person is really, really average in how smart they are. They are right in the middle. This person figures out how to do things about as quickly and comes up with answers about as fast and about as good as most other people. With regards to being smart, this person is right in the middle*". The response options for this story were adopted from task (iii) study 1 of Bian et al. (2017). There was a validation question in this study asking children to indicate what kind of person the story was about from four options: "*very bright*", "*very slow-minded*", "*averagely bright*", "*very unhappy*".

7.6.1.3 Procedure

After giving consent, a welcome message and the study instruction were shown to children. Then children were first asked to read a story describing a person who has average ability and then were asked to select one image from four options that they thought to be the person described in the vignette in the following page. This was followed by the validation question. Participants were compensated with 10p once their answers were verified by the experimenter.

7.6.2 Results

Firstly, we tested hypothesis one that a female target would be more likely to be selected as a match of the person who has average ability than a male target. This hypothesis was tested using a binomial general linear model, with children's gender used as a predictor of the sex of the target image selected, and age as a covariate. In total, 113 boys and 24 girls selected a male target as a match of the person who has average ability. Contrary, 110 boys and 21 girls selected a female target as having average ability. The difference between choosing a male or female target was not significant (β = -0.17, *OR* = 0.84, CI_{95} [-0.82, 0.48], z = -0.51, p = .609) and there was no gender difference in children's choices (β = -0.10, *OR* = 0.90, CI_{95} [-0.75, 0.54], z = -0.32, p = .750).

Next, we tested hypothesis two that children who participated in study 4 would be more likely to select a female target as having average ability than those who participated in study 2. Again, this hypothesis was tested using a binomial general linear model, with the sex of the target image children selected as the dependent variable, the study children had previously participated in as the independent variable, children's gender and age as covariates. Children who participated in study 4 were not significantly more likely to select a female target as having average ability than those who participated in study 2 ($\beta = 0.06$, $OR = 1.06$, CI_{95} [-0.22, 0.35], $z = 0.44$, $p = .661$).

7.6.3 Study 5 discussion

This follow-up study tested whether children would be more likely to select a female target as having the average ability than a male target. Contrary to prediction, a majority of both boys and girls selected a male target as a match of the person who has average ability. However, this effect was not statistically significant. Our findings in studies 2 and 4 showed that both boys and girls viewed men as having extremely high or low-ability, and findings in study 5 indicated that men are also viewed as having average ability. It seemed that regardless of the magnitude of cognitive ability, men were always the gender that children selected relating to intellectual ability. We next interpreted all of our findings from study 1 to 5 in a brief discussion.

7.7 General discussion

Bian et al. (2017) suggested that children started to have a gender stereotype about brilliance favouring men at age 6, as well as a gender stereotype about niceness favouring women. We tested these predictions across four studies, over 700 children aged from 6 to 13 years old, in both China and the U.K.. We confirmed that both boys and girls were biased towards women regarding niceness. However, for the gender stereotype about brilliance, our findings were mixed. We tested this prediction in children with different ages (i.e., 7 and 10 years old) and across different cultures (i.e., China and the U.K.) to establish reliable findings. We did verify the gender stereotype in two of our British samples (studies 3 and 4) that men were more likely to be chosen as the smartest person by both genders. However, we failed to replicate this effect in the Chinese sample (study 1) and one British sample (study 2): rather than women showing a bias to stereotype brilliance as a male trait, both genders stereotyped brilliance as an own-gender trait. More strikingly, for our exploration about gender stereotypes regarding the pole opposite to brilliance, we found consistent results that both boys and girls chose men as more likely to be slow-minded across all of the four studies.

Turning to the existence of gender differences in children's career interests and the possible effects of gender stereotypes about brilliance on children's career interests, we tested these two predictions across three studies (studies 2, 3 and 4) with over 500 British children. A significant gender difference in the career interest in working as a nurse (favouring women) was consistently found across all of the three studies. There were gender differences in other kinds of career interests such as working as an executive manager or working as a scientist or engineer, but these results were not consistent. In addition, our findings did not support that children's gender stereotypes about brilliance were significantly associated with their career interests.

To begin with the successful replication that both genders were stereotyped against men as being more likely to be really nice. Given considerable data supporting a genuine difference favouring women in traits such as agreeableness, sensitivity etc. (Costa, Terracciano, & McCrae, 2001; Del Giudice et al., 2012), it is possible that this effect, which both Bian et al. (2017) and the present studies were able to demonstrate, reflects a true mean difference, internalised as an accurate stereotype (Jussim et al., 2016; Löckenhoff et al., 2014) shared by both genders of children.

We next turn to the core hypothesis that women would show a bias to stereotype brilliance as a male trait. One account which may explain our null findings is based on reported associations of national economic development indices and their association with interests. In a recent global survey of STEM interest, Stoet and Geary (2018) reported that differences in economic development correlated strongly with gender differences in STEM interest. Paradoxically, women's science self-efficacy, engagement in STEM subjects and uptake of STEM employment was lowest in countries with the highest levels of gender-equality (e.g. Iceland) compared to countries with low gender-equality (e.g. Jordan). Stoet and Geary (2018) resolved this apparent paradox by proposing that high levels of economic development allow individuals to choose school subjects and occupations that reflect their own interests and strengths, rather than being constrained by economic demands. Speculatively, it is possible, then, that the present result reflects relatively low self-determination in China relative to the USA and Britain. This would also, however, alter the interpretation of the US results. It would shift from the internalisation of biased stereotypes to one revealing interests and internal preferences over development.

Finally, turning to our finding regarding bias concerning extreme low intellectual ability, namely that both genders most often picked a male as most likely to be "*really, really slow-minded*", this stereotype of men as more likely to be low in intellectual ability is, like the niceness stereotype, consistent with a genuine gender difference. Men are more likely to attain very low scores on measures of intellectual ability (Deary, Irwing, et al., 2007). This outcome might not be driven by an intuitive awareness, since we did not find a bias against women at the "*really, really smart"* end of the distribution, at least not in our Chinese sample and one of the British samples. Given this, we are hesitant to conclude that the Chinese data reported here support a sensitivity in children to a relatively small difference in variance in intellectual ability, but this might be true in the British samples.

7.7.1 Limitations and further directions

One limitation in the present studies might be that there was a potential selection bias when choosing the four images as the response-options for the Chinese participants. The experimenter who initially chose 10 images is female, and the other four raters voting for the four most professional and attractive images are also female. Thus, the images were chosen based on female preference rather than a more balanced male-female preference. The other limitation in the present studies would be that our study 1 used the same set of four images for all of the three stories, which might lead a potential sequential effect of children's choices on the nice and slow-minded stories. The study suggests one main direction for future work: we were not able to test 5 and 6-year-old children, and future work should address this in both Chinese and British samples.

7.7.2 Conclusion

Our results suggest that, while stereotyping of women as more likely to be "*really, really nice*" may exist, a stereotype of men as more likely to be "*really, really smart"* does not appear to exist in 10-year old children in China, but does exist in children in the U.K.. As such, additional replication attempts with younger Chinese children are warranted, preferably in a pre-registered and collaborative effort. Additional studies of gender-differences in interests and tests of competing explanations for differences in life choices (Bian et al., 2017; Deary, Irwing, et al., 2007; Del Giudice et al., 2012; Stoet & Geary, 2018) are also warranted.

Chapter 8. Does conscientiousness predict children's educational attainment?

The last two chapters explored the association between the gender stereotype about brilliance and children's educational attainment in both the UK and China. I did not find a significant effect of the gender stereotype about brilliance on children's educational attainment in any of my studies, but instead found a cultural difference about holding the gender stereotype about brilliance among the British and Chinese children. These failures of replication discouraged me from continuing in this field. I then looked back to the mindset literature and tried to find if anything in the methods could be improved or if anything was neglected. After revisiting those mindset studies, I realised that there was a component that had been neglected in previous studies: the attitude towards hard work. Individuals' attitudes towards hard work were used as the contrary component of the attitudes towards intelligence (i.e., natural ability) in the mindset theory (Dweck, 2006), but this component on its own may be an active agent in learning - namely work and persistence rather than beliefs about the nature of the mind. In other words, attitudes to hard work, rather than mindset, may have more important impacts in people's life. Therefore, the last theme in this thesis is focusing on the effects of hard work (i.e., conscientiousness) on children's educational attainment.

In this chapter, I define personality and personality traits, three widely-known personality models, three personality measurements; briefly summarise the existing evidence of the association between conscientiousness and children's educational attainment; and report four empirical studies testing the effects of children's own conscientiousness, their perceptions of teacher's conscientiousness, and classroom chaos on children's educational attainment. I gratefully acknowledge the assistance and contribution of student project worker Ke Liu to study 4 reported here.

8.1 What is personality?

8.1.1 The definition of personality

Lay people and psychologists would define personality differently. For lay people, personality may simply mean "*the whole integrated pattern of behaviour which distinguishes one man from another*" (Lazarus & Opton, 1967). For psychologists, the definition of personality is more complicated. Psychologists define personality mostly depending on their research aims or the specific perspective they are interested in, thus, there is no agreement on the definition of personality. Here I introduce a few definitions that were made by the most famous researchers in the personality field.

Allport (1937) viewed personality as integration of psychological processes and defined personality as "*the dynamic organisation within the individual of those psychological systems that determine his unique adjustments to the environment*" (p. 48). Cattell (1950) explored personality from the trait perspective and he defined personality as that "*which permits a prediction of what a person will do in a given situation*" (p. 2). Taking a broad view, Pervin and John (1996, p. 4) defined personality as it "*represent[ing] those characteristics of the person that account for consistent patterns of feeling, thinking, and behaving*", which provided three specific aspects that personality researchers could focus on. Brody and Ehrlichman (1998) also were focused on the inside psychological process of individuals and defined personality as "*those thoughts, feelings, desires, intentions, and action tendencies that contribute to important aspects of individuality*" (p. 3). Based on the integration idea of Allport, McAdams and Pals (2006) proposed a modern and developed definition that personality is "*an individual's unique variation on the general evolutionary design for human nature, expressed as a developing pattern of dispositional traits, characteristic adaptations, and integrative life stories complexly and differentially situated in culture*" (p. 212).

8.1.2 What is a personality trait?

Nowadays, exploration of personality is mainly based on the trait perspective (Mottus, 2017). *Trait* could be simply and vaguely defined as the unique and defining features of personality (A. H. Buss, 1989), or it could involve some specificities that "*[define] what a person will do when faced with a defined situation*" (Cattell, 1979, p. 14). A more specific definition of trait would be that a trait is the tendency of a person behaving consistently in different situations (Brody & Ehrlichman, 1998). In the more complicated and detailed context, a trait could be "*an inferred relatively enduring organismic (psychological, psychobiological) structure underlying an extended family of behavioural dispositions*" (Tellegen, 1991) or "*a mental structure that may be inferred from observable behaviour to account for regularity and consistency in behaviour*" (Hogan, Johnson, & Briggs, 1997). The last two definitions indicate an intercorrelated association between trait and behaviour, in that traits could predict some behaviours and behaviours could be the manifestation of traits.

8.2 Personality models

8.2.1 Eysenck's Big Three model

The structure of personality is hierarchical (Costa & McCrae, 1995; Eysenck, 1990). Eysenck (1953, 1990) proposed that a typical personality structure includes four levels of hierarchies: the specific response level (the bottom level), which includes singly occurring behaviours or cognitions in people's daily life; the habitual response level (the second-bottom level), which includes habitual behaviours or cognitions; the trait level (the third level), which includes a group of traits that are intercorrelated with habitual behaviours; and the factor level (the top level), which includes a few high-order factors or dimensions, with each factor or dimension including a group of intercorrelated traits.

Based on the above hierarchical structure, Eysenck (1990) established the "Big Three" personality model which includes the three dimensions of Extraversion – Introversion (E), Neuroticism (N), and Psychoticism (P). Each dimension is constructed by a group of intercorrelated traits. In specific terms, the dimension of Extraversion – Introversion was interpreted as the observed correlations between dominance, surgency, sociability, activeness, assertiveness, sensation-seeking, liveliness, and venturesomeness and carefree behaviours or cognitions. The dimension of Neuroticism was defined by the intercorrelations between anxiety, depression, feelings of guilt, low self-esteem, tension, irrationality, shyness, moodiness, and emotional behaviour or cognition. The dimension of Psychoticism referred to the correlations between antisocial, unempathic, creative, tough-minded, aggressive, cold, egocentric, impersonal and impulsive behaviours or cognitions. Eysenck (1992) was confident in his "Big Three" model and suggested that for those researchers who suggested more dimensions other than the E, N, and P, those other dimensions are either components of his three dimensions or combinations of two of the three dimensions.

8.2.2 The "Big Five"

Although some psychologists agreed with the "Big Three" model of Eysenck (1992), many psychologists argued that there should be more than three dimensions, and therefore a few more personality models were proposed (Cattell, 1948; Goldberg, 1990; McCrae & Costa, 2008). Among these alternative models, one of the most widely-known models would be the "Big Five" (Goldberg, 1990), which suggests that there are five dimensions in personality. The foundation of the "Big Five" was built on work by several hallmark researchers (Goldberg, 1993) and most of those works used a lexical approach (i.e., clustering trait-descriptive terms in natural language; John, Angleitner, & Ostendorf, 1988).

The exploration of the "Big Five" was started by Allport and Odbert (1936), who classified more than 18,000 trait-descriptive terms into four lists and suggested that terms (around 4,500) in their first list were the most stable traits. Based on the lists of Allport and Odbert (1936), Cattell (1945, 1948) constructed a set of 35 bipolar traits that he thought could represent the whole personality sphere and developed rating scales to contrast those traits. Following Cattell (1945, 1948), Fiske (1949) reanalysed the 35 traits, applied them in his own studies and discovered five replicable dimensions among the self-reported, peer-rated and observational studies. This finding could be seen as the initial formation of the "Big Five" and the five dimensions were labelled as "*Confident self- expression*", "*Social adaptability*", "*Conformity*", "*Emotional control*" and "*Inquiring intellect*" (Fiske, 1949).

After the initial exploration, psychologists found more evidence to support the five-dimension structure. For example, based on Cattell and Fiske's studies, Tupes and Christal (1961) conducted a series of eight studies to test the intercorrelations of

Cattell's 35 traits. Across eight samples (participants had different ages, multiple educational levels, various kinds of working experience, and different lengths of acquaintanceship), Tupes and Christal (1961) found five dimensions that were replicable in all of the eight studies. Tupes and Christal (1961) labelled the five replicable dimensions as "*Surgency*", "*Agreeableness*", "*Dependability*", "*Emotional stability*" and "*Culture*".

Following the findings of Tupes and Christal (1961), psychologists provided further evidence to support the five-dimension structure (e.g., Digman & Inouye, 1986; Norman, 1963). The most compelling evidence was provided by Goldberg (1990). Goldberg (1990) analysed trait terms from a large and representative pool of English terms, tested the stability of the clustered traits in both self-reported and peer-rated studies, analysed the results using ten different factor-analytic approaches, and found a stable personality structure which consisted of five dimensions. Goldberg (1990) referred to his five-dimension structure as the "Big Five". By adapting Tupes and Christal (1961), Goldberg (1990) labelled his dimensions as "*Surgency (Factor I)"*, "*Agreeableness (Factor II)"*, "*Conscientiousness (Factor III)"*, "*Emotional stability (Factor IV)"* and *"Intellect (Factor V)"*. Afterwards, the initial five labels were changed to "*Extraversion", "Agreeableness", "Conscientiousness", "Emotional Stability versus Neuroticism",* and *"Intellect or Imagination"*, which are more similar to the dimensions labelled in the Five-Factor model (Saucier & Goldberg, 1998).

Goldberg (1993) proposed that each of the five dimensions in the "Big Five" is a bipolar dimension, which includes multiple groups of traits and each trait is constructed by discrete behaviours. *Surgency* (*Extraversion*) refers to an individual's preferences regarding being alone and staying with others (Brody & Ehrlichman, 1998). This dimension includes positive traits such as spirit, playfulness, expressiveness, self-esteem and courage, and negative traits such as aloofness, silence, shyness, inhibition and pessimism (Goldberg, 1993). *Agreeableness* is about how individuals react to other people's characteristics (Brody & Ehrlichman, 1998). This dimension includes positive traits such as cooperation, empathy, courtesy, flexibility and morality, and negative traits such as overcriticalness, rudeness, irritability, distrust and prejudice (Goldberg, 1993). *Conscientiousness* is about how individuals perform tasks (Brody & Ehrlichman, 1998). This dimension includes positive traits such as organisation, efficiency, dependability, precision and persistence, and negative traits such as negligence, inconsistency, forgetfulness, recklessness and aimlessness (Goldberg, 1993). *Emotional stability* (*Neuroticism*) is related to an individual's emotional status, which includes positive traits of placidity and independence, and negative traits such as insecurity, fear, instability, emotionality and envy (Goldberg, 1993). *Intellect* (*Openness to Experience*) is about an individual's appreciation for having creative ideas (McCrae & Costa, 1990). It consists of positive traits such as intellectuality, insight, creativity, curiosity and sophistication and negative traits including shallowness, unimaginativeness, imperceptiveness and stupidity (Goldberg, 1990). From the above descriptions, it can be said that the five dimensions in the "Big Five" are related to general features of *Power*, *Love*, *Work*, *Affect* and *Intellect* (Peabody & Goldberg, 1989).

Although the "Big Five" model was founded by using the lexical approach in the English language, subsequent studies have found supportive evidence in multiple languages (e.g., Church, Reyes, Katigbak, & Grimm, 1997; Cui & Wang, 2004; Szirmák & De Raad, 1994). However, the lexical approach has its own limitations.

For example, some specific traits including openness to fantasy, aesthetics, feelings and actions can be well-represented by phrases, sentences or passages, but cannot be represented well by single words in natural language (McCrae, 1990). In addition, the traits that are identified by the lexical approach belong to multiple levels of the personality hierarchy (John, Hampson, & Goldberg, 1991). Namely, those traits could be found in extremely narrow levels or extremely broad levels (Costa & McCrae, 1995). The terms in the broad levels could covary with terms in the narrow levels, but terms in the narrow levels might not covary well with each other (Costa & McCrae, 1995). Therefore, when psychologists carried out factor analysis for those clusters of terms, only terms in the broader levels would account for the major share of covariance and thus the five broad dimensions emerged (Goldberg, 1990). Psychologists tried to develop an alternative approach to avoid the stated limitations of the lexical approach when exploring the personality structure. One of the successful approaches is the traditional approach, which was based on confirmatory factor analyses of standard personality questionnaires (McCrae & Costa, 1985).

8.2.3 The Five-Factor model

Similar to the "Big Five", the Five-Factor model (FFM) also proposes that there are five dimensions in personality structure (McCrae & Costa, 1985; McCrae & John, 1992). The most important difference between the "Big Five" and the FFM is that these two models were based on two different research approaches: the five dimensions in the "Big Five" were identified by using the lexical approach (i.e., analysing hundreds and thousands of trait terms to identify the most representative ones) and the five dimensions in the FFM were identified based on analyses of the standard personality questionnaires (McCrae & Costa, 1985).

Costa and McCrae (1976) analysed the contents and structures of personality questionnaires published between the 1960s to the 1970s. They found two prominent and consistent dimensions in those questionnaires: *Neuroticism* and *Extraversion* (Costa & McCrae, 1976). A few years later, Costa and McCrae (1978) recognised that there was another important dimension in personality structure and referred to this dimension as *Openness to Experience*. Costa and McCrae (1978) suggested that Openness to Experience corresponds to the dimension of *Intellect* in the "Big Five", but it is slightly broader than *Intellect* since *Openness to Experience* includes additional traits of unconventionality and behavioural flexibility. Therefore, a three-dimension model including *Neuroticism*, *Extraversion* and *Openness to Experience* (the NEO model) emerged.

When applying the NEO model into empirical studies, McCrae and Costa (1985) found that there were some traits that could not be encompassed by the three dimensions. For example, locus of control and persistence, as two personality traits, do not fit in any of the three stated dimensions (Costa & McCrae, 1980). Costa and McCrae (1980) firstly used the term *control* to represent the uncovered traits (e.g., persistence). Then they found that the trait *conscientiousness*, which was reported in Norman's five-factor structure (Norman, 1963), corresponded to control. McCrae and Costa (1985) adopted this term and defined their new dimension as *Conscientiousness*. In addition, McCrae and Costa (1985) realised that although Neuroticism relates to some emotional adjectives such as irritable and jealous, agreeableness - which was reported by Norman (1963) - was rarely represented in personality questionnaires. Therefore, a fifth dimension of *Agreeableness* was added into the NEO model, and that was completed the formation of the Five-Factor model (McCrae & Costa, 1985).

Unlike to the models that used *factors, dimensions* and *traits* to distinguish the different levels in the hierarchical structure of personality, Costa and McCrae (1995) used *domains* to describe the five broad dimensions and *facets* to describe the underlying traits in each domain. More specifically, the Openness to Experience domain contains facets of fantasy, aesthetics, feelings, actions, ideas and values; the Conscientiousness domain contains facets of competence, order, dutifulness, striving for achievement, self-discipline and deliberation; the Extraversion domain contains facets of warmth, gregariousness, assertiveness, activity, excitement seeking and positive emotions; the Agreeableness domain contains facets of trust, straightforwardness, altruism, compliance, modesty and tender-mindedness; and the Neuroticism domain contains facets of anxiety, angry hostility, depression, self-conscientiousness, impulsiveness and vulnerability (Costa & McCrae, 1995).

Since the FFM was built on empirical evidence and contains most of the aspects that psychologists wish to explore, the FFM became one of the prime theories in personality research (Mottus, 2017). This model has been examined in more than 50 countries and has obtained substantial support (F. M. Cheung & Leung, 2016; McCrae, Terracciano, & 79 Members of the Personality Profiles of Cultures Project, 2005). Due to the popularity and high validity of the FFM, all studies reported in this chapter are based on the FFM.

Recently, some researchers argued that psychologists should not lean on the FFM to describe or predict individuals' behaviours since the domains in the FFM did not predict important life outcomes as well as the facets did (Hogan & Sherman, 2020). This argument did provide the suggestion that when using the FFM, researchers need to consider both the effects of the domains and the effects of facets. But since facets and domains are intercorrelated (i.e., facets formed domains and domains formed the FFM), neither facets nor domains are meaningless.

8.3 Personality measurements

8.3.1 The Big Five Inventory

The Big Five Inventory (BFI) is designed for the task of measuring the prototypical features of each dimension in the "Big Five" efficiently and flexibly (John & Srivastava, 1999). The BFI is a short measurement that contains 44 items in the English version (Benet-Martínez & John, 1998). Each item in the BFI contains a short phrase and the core term in the phrase is based on the "prototypical makers" of the "Big Five" (John, 1989). For example, the term of "*aloofness*" in *Surgency* served as a core of the item "*can be cold and aloof*" and the term of "*emotionality*" in *Emotional stability* became the core part of the item "*is emotionally stable, not easily upset*" (John, Donahue, & Kentle, 1991).

Although the BFI has very short phrases and there are eight or ten items for each dimension of the "Big Five", it covers most of the facets that were defined in the FFM (John & Srivastava, 1999). For example, the *Agreeableness* subscale in the BFI has only eight items but these items cover at least five (out of six) facets in the *Agreeableness* domain of the FFM (John & Srivastava, 1999). The BFI also has good psychometric properties (Benet-Martínez & John, 1998). For example, the alpha reliabilities of the BFI subscales ranged from .75 to .90 and the test-retest reliabilities ranged from .80 to .90 (after three months from the first test) in American and Canadian samples (Benet-Martínez & John, 1998). After its publication, the BFI has been translated into different languages including Chinese, Dutch, German, Italian, Lithuanian, Portuguese, Spanish and Swedish language (Berkeley Personality Lab, 2019), and has been validated in multiple countries (Schmitt, Allik, McCrae, & Benet-Martínez, 2016).

8.3.2 The NEO-PI-R

The Revised NEO Personality Inventory was an outgrowth of a self-report questionnaire - the NEO Inventory - that measured six personality facets underlying three personality domains: *Neuroticism*, *Extraversion* and *Openness to Experience* (McCrae & Costa, 1983). After noticing the underrepresentation of *Agreeableness* and *Conscientiousness* in the NEO Inventory, and based on the analyses of the common factors in previous questionnaires (e.g., the 16 Personality Factor questionnaire by Cattell, Eber, and Tatsuoka (1970)), Costa and McCrae (1985) developed the NEO Personality Inventory (NEO-PI) to measure the facets underlying all five domains of Neuroticism, Extraversion, Openness, Conscientiousness and Agreeableness in the FFM. Although the NEO-PI was a comprehensive measurement for the FFM, it had a major limitation: There were no facet scales for Agreeableness or Conscientiousness (Costa & McCrae, 2008). Specifically, there were six facet scales for each domain of Neuroticism, Extraversion and Openness to Experience, but only one facet scale for each of the Agreeableness and Conscientiousness domains. Therefore, a revised version of the NEO-PI, the NEO-PI-R, was developed (Costa & McCrae, 2008). The NEO-PI-R is a much more comprehensive personality measurement since it added more facets for Agreeableness and Conscientiousness and measures different variables including individuals' characteristics, emotions, experiences, attitudes, and motivations (Costa & McCrae, 2008).

The NEO-PI-R contains 240 items: Each of the five domains consists of 48 items and each facet is assessed by 8 items (G. Matthews, Deary, & Whiteman, 2009).

Similar to the BFI, the NEO-PI-R also has good psychometric properties. For example, the internal consistency of the NEO-PI-R ranged from .50 to .87 in a college sample and the test-retest reliability ranged from .67 to .86 in a longitudinal data (McCrae, Kurtz, Yamagata, & Terracciano, 2011). The NEO-PI-R has also been tested in multiple countries (e.g., McCrae, 2002; McCrae, Terracciano, & Personality Profiles of Cultures, 2005) and across different ages (e.g., De Fruyt, Mervielde, Hoekstra, & Rolland, 2000; Terracciano, McCrae, Brant, & Costa, 2005). The psychometric properties of the NEO-PI-R were robust across almost all studies (McCrae et al., 2011).

In order to measure the five domains quickly and efficiently, Costa and McCrae (1992) developed a short version of the NEO-PI-R: The NEO-Five Factor Inventory (NEO-FFI). The NEO-FFI contains only 60 items (12 items for each domain) that are selected from the original list of 240 items in the NEO-PI-R (Costa & McCrae, 1992). Each of the subscales in the NEO-FFI is also as reliable (e.g., the alpha reliability for each of the five subscales ranges from .72 to .87) as those in the NEO-PI-R (Egan, Deary, & Austin, 2000).

8.3.3 The Behaviour Indicators of Conscientiousness

Unlike to the BFI and the NEO-PI-R, the Behaviour Indicators of Conscientiousness (BIC) scale only measures the domain of conscientiousness rather than all five domains and facets (Jackson et al., 2010). The BIC identified 185 behaviours that are associated with 11 underlying facets including "*avoiding work, organisation, impulsivity, antisocial behaviours, cleanliness, laziness, industrious, punctuality, formalness, responsibility, and appearance*" (Mike et al., 2015, p. 660). Jackson et al. (2010) reported that the reliability of the BIC was good (ranges from .65

to .90) and the correlations between each facet was mostly acceptable (ranges from 0 to .53).

There are two advantages of the BIC. First, the BIC places importance on behaviours in personality research: behaviour could be the mediation of the relationship between psychological processes and life outcomes or it could be the manifestations of thoughts or feelings related to conscientiousness (Furr, 2009). In addition, the BIC scale suggested that there might be five more facets including industriousness, orderliness, impulsive control, reliability, and conventionality underlying conscientiousness (Jackson et al., 2010). This may provide an opportunity for psychologists to reconsider the structure of the underlying facets of conscientiousness.

8.4 Existing evidence of the association between personality domains and educational attainment

Personality is influential on many life outcomes including health (Roberts, Walton, & Bogg, 2005; Smith, 2016), marital success (Bouchard, Lussier, & Sabourin, 1999), job performance (Hurtz & Donovan, 2000; Salgado, 1997) and most importantly for the present thesis, educational attainment (Dumfart & Neubauer, 2016; Poropat, 2009). Based on the FFM, each of the five domains has a different influence on educational attainment: some domains such as Conscientiousness and Openness to Experience are positively associated with educational attainment (e.g., Dumfart & Neubauer, 2016; Poropat, 2009; Zhang & Ziegler, 2016; Zhou, 2015), and some domains such as Neuroticism are negatively associated with educational attainment (Komarraju, Karau, & Schmeck, 2009).

Among all domains in the FFM, conscientiousness has been found to be the strongest predictor of educational attainment (Bratko, Chamorro-Premuzic, & Saks, 2006; Butcher, Ainsworth, & Nesbitt, 1963; Chamorro-Premuzic & Furnham, 2003; Dumfart & Neubauer, 2016; Ivcevic & Brackett, 2014; O'Connor & Paunonen, 2007; Poropat, 2009). The association between conscientiousness and educational attainment is robust across studies using different personality measurements (e.g., Busato, Prins, Elshout, & Hamaker, 2000; De Fruyt & Mervielde, 1996; Duff, Boyle, Dunleavy, & Ferguson, 2004; O'Connor & Paunonen, 2007). For example, in a meta-analysis, O'Connor and Paunonen (2007) reviewed studies that tested the association between personality domains and educational attainment among college and university students using multiple personality measurements including the NEO-PI-R (Costa & McCrae, 1992), the 16 Personality Factor questionnaire (Cattell et al., 1970), the Personal Style Inventory (Lounsbury & Gibson, 1998) and the NEO-FFI (Costa & McCrae, 1992). Conscientiousness showed the highest correlation with educational attainment (the mean correlation was .24) compared to Agreeableness, Neuroticism, Extraversion, and Openness to Experience (the highest correlation was .06 among the other four dimensions; O'Connor & Paunonen, 2007). In another meta-analysis study, Poropat (2009) reviewed studies using personality measures only based on the FFM. Namely, studies that used measures relating to other personality models such as the 16 Personality Factor questionnaire were excluded in this meta-analysis (Poropat, 2009). Compared to the other four domains, conscientiousness still showed the highest correlation with educational attainment (sample-weighted correlation was .19, effect size was .46), which was slightly lower than the correlation between intelligence and educational attainment (sample-weighted correlation was .23, effect size was .52).

The positive association between conscientiousness and educational attainment has been found consistently across different educational levels (Barbaranelli, Caprara, Rabasca, & Pastorelli, 2003; Chamorro-Premuzic & Furnham, 2003; Furnham et al., 2003; Komarraju et al., 2009; Laidra, Pullmann, & Allik, 2007; Wagerman & Funder, 2007). For example, Laidra et al. (2007) tested the relationship between the five domains in the FFM and students' educational attainment in primary and secondary school (from grade 2 to grade 12). They found that conscientiousness had a significantly positive correlation with educational attainment through all grades (Laidra et al., 2007). Chamorro-Premuzic and Furnham (2003) tested the association between the five personality domains and educational attainment among university students throughout their three-year degree. The correlation between conscientiousness and students' educational attainment in each year of university was the highest (range from .33 to .34) compared to the correlations between each of the other four domains in the FFM (e.g., the correlation of Neuroticism and students' educational attainment ranged from -.28 to -.32; Chamorro-Premuzic & Furnham, 2003). Therefore, among all domains in the FFM, conscientiousness is the most powerful predictor of educational attainment across all educational levels (Dumfart & Neubauer, 2016).

8.5 Is the teacher's personality associated with students' educational attainment?

Apart from parents, teachers are the most important people in children's learning processes and development (Goldhaber, 2002). Studies have examined the teacher's influence on students' educational attainment from different aspects including the teacher's interpersonal behaviours (Den Brok, Brekelmans, & Wubbels, 2010), teacher's attitudes (Ulug, Ozden, & Eryilmaz, 2011), teacher's feedback (Núñez et al., 2014; Siewert, 2011), and teacher's expectations (Speybroeck et al., 2012). However, only a few studies examined the association between the teacher's personality and students' educational attainment (Hashim, Alam, & Yusoff, 2014; Kim, Jorg, & Klassen, 2019; Tahir & Shah, 2012).

Although the teacher's personality is an individualised and interpersonal process, it could be expressed externally by teacher behaviour, values, beliefs, and attitudes (Stronge, Tucker, & Hindman, 2004). The expressed behaviours, values, beliefs and attitudes could influence both the teacher's job performance (Salgado, 1997) and the students' learning environment (Tahir & Shah, 2012). For example, if a teacher is highly conscientious and extroverted, this teacher would be effective, tend to have high self-efficacy, be less likely to be biased, and be more likely to have social interactions with students (Hashim et al., 2014; Kim et al., 2019). With these positive features, teacher's teaching quality would be enhanced and students' motivation and engagement in learning would be encouraged (Hashim et al., 2014; Mojavezi & Tamiz, 2012; Rimm-Kaufman & Hamre, 2010; Stronge et al., 2004). Thus, the teacher's personality is important in students' learning (Tahir & Shah, 2012).

Among the studies that tested the effects of the teacher's personality on students' educational attainment, the results are mixed (Hashim et al., 2014; Kim, Dar-Nimrod, & MacCann, 2018; Tahir & Shah, 2012). For example, Hashim et al. (2014) tested the association between teacher's personality and students' English proficiency among English as a foreign language (EFL) students. They found that teacher's personality was significantly positively associated with EFL students' proficiency in English communication (Hashim et al., 2014). However, the questionnaire Hashim et al. (2014) used to assess teacher's personality (i.e., *"The teacher understands my English learning problems and abilities", "The teacher puts an effort to know me personally", "The teacher is always concerned for me", and "The teacher seems very sincere in teaching"*) was not personality questionnaire. The questionnaire they used did not contain any domain-related items and it was not possible to clarify the most effective dimension(s) that was (were) highly associated with students' educational attainment. Instead, their questionnaire was more likely to be a teaching quality questionnaire. Therefore, the results of the study by Hashim et al. (2014) might not be convincing. Tahir and Shah (2012) conducted a more convincing study to test the association between teacher's personality and students' educational attainment in college. Tahir and Shah (2012) asked students to rate their teacher's personality in the BFI and found that Extraversion, Agreeableness, Conscientiousness and Openness to Experience (correlations ranged from .09 to .52) were positively associated with students' educational attainment, and Neuroticism (correlation was -.43) was negatively associated with students' educational attainment (Tahir & Shah, 2012). Although Tahir and Shah (2012) used a valid personality questionnaire, they only asked students to rate their teacher's personality and did not ask teachers themselves to report their personality. Students' ratings of teachers' personalities might be biased by students' thinking, and thus might not be a good representation of teachers' personalities, but of the students' personalities. Following these studies, a much more convincing study would be the one conducted by Kim et al. (2018). Kim et al. (2018) assessed teacher's personality by using the BFI, and asked both teachers themselves (self-report) and students (observers) to rate teacher's personality. Contrary to previous studies, Kim et al. (2018) found that there was no association between teacher's personality and students' educational attainment in secondary school (year 7 to 9).

8.6 The current studies

Although substantial studies have found that conscientiousness is positively associated with educational attainment, most of the studies were conducted in Western countries (e.g., in the U.S.) and only a few studies were conducted in China (Zhang & Ziegler, 2016; Zhou, 2015). Thus, there is not sufficient data to test if culture would influence the association between conscientiousness and educational attainment. To provide informative evidence for researchers to examine whether culture could influence the association between conscientiousness and educational attainment, most of the studies reported in this chapter were conducted in China.

Second, a considerable number of studies showed that conscientiousness, compared to the other domains in the FFM, is the most influential domain on educational attainment. However, most of those studies used personality questionnaires that contained all the five domains based on the FFM and less study explored if any behaviours related to conscientiousness are associated with educational attainment in specific. The studies reported in this chapter aimed to fill this research gap. Therefore, two personality questionnaires that contain all five domains based on the FFM and one questionnaire designed to measure behaviours related to conscientiousness were used.

Finally, although some studies have tested the association between teacher's personality and students' educational attainment, the results were mixed. The studies reported in this chapter aimed to provide supplemental evidence for the research in this field.

8.7 Study 1

8.7.1 Introduction

This study aimed to test the simple relationship between personality domains and educational attainment among Chinese students. This study is the basis for conducting any further studies that test whether specific behaviours related to conscientiousness are associated with educational attainment. In other words, if personality is not associated with educational attainment among Chinese students, there is no need to conduct any further studies to test those more specific hypotheses.

In Chinese schools, there are two formal assessments in each semester: the mid-term and the end-of-term assessments. Both assessments are important because teachers and students need to take time to prepare and work hard for these two assessments (Zhang & Ziegler, 2016). The mid-term assessment is usually held by individual schools and the end-of-term assessment is sometimes held by several linked schools as a joint examination. Thus, the exam questions in the mid-term assessment are normally given by teachers in individual schools and the exam questions in the end-of-term assessment are sometimes given by teachers from the linked schools. Therefore, the difficulty and uncertainty of the end-of-term assessment might be higher than in the mid-term assessment.

Based on the existing evidence such as Dumfart and Neubauer (2016); Komarraju et al. (2009); Poropat (2009); Zhang and Ziegler (2016); Zhou (2015), I made five hypotheses that correspond to the five domains in the FFM: 1) Openness to Experience would be positively associated with grades; 2) Conscientiousness would be positively associated with grades; 3) Extraversion would be positively associated with grades; 4) Agreeableness would be positively associated with grades; and 5) Neuroticism would be negatively associated with grades.

8.7.2 Method

8.7.2.1 Participants

In total, 222 participants (116 boys, 106 girls; mean age = 11.03, SD = 0.47) were recruited from a large primary school in Harbin (the capital city of Heilongjiang Province, China). The school is public and draws from a catchment area 21% below the Chinese national average income and the children are in relative poverty. All data in this study were collected in the spring term in 2016.

8.7.2.2 Materials

Educational attainment: After gaining permission, students' end-of-term grades for the semesters preceding and following the personality measurement in three core classes (i.e., English, Chinese, and mathematics) were obtained from headteachers.

Personality measurement: The Big Five Inventory Chinese version (Benet-Martínez & John, 1998; John, Donahue, et al., 1991; John, Naumann, & Soto, 2008) was used to measure students' personality. There are 48 items in the Chinese BFI: the first 44 items are identical to those items in the English version (e.g. "*is talkative*") and the last 4 items are "*having high self-esteem*", "*is traditional and conventional*", "*is playful and loves to amuse others*", and "*is ambitious and likes to conform others to my own intentions*". Responses were recorded on a Likert scale (from 1: strongly disagree to 5: strongly agree).

8.7.2.3 Procedure

All studies reported in this chapter were approved by the Psychology Research Ethics Committee at the PPLS, University of Edinburgh. After gaining consent from headteacher, teachers, parents and students themselves, students were asked to complete the questionnaire in their classroom. Students were firstly asked to complete the demographic questionnaire including their age, gender, the end-of-term grades in the most recent semester, and then were asked to complete the BFI. After taking their end-of-term exams in the semester in which their personality was measured, students' end-of-term grades were reported by head teachers.

8.7.3 Results

I first tested hypothesis one that Openness to Experience would be positively associated with grades. This was tested using two linear regressions, with students' GPA in the semester preceding (GPA1) and following (GPA2) the personality measure as the dependent variable respectively, and their scores in the BFI-Openness subscale as the independent variable, controlling for their age and gender. Contrary to the prediction, Openness to Experience was not significantly associated with GPA in the semester preceding the personality measure ($\beta = 0.10$, CI_{95} [-0.03, 0.23], $t = 1.48$, $p = .140$), nor with GPA in the semester following the personality measure ($\beta = 0.13$, CI_{95} [-0.02, 0.28], $t = 1.76$, $p = .081$).

Next, I tested hypothesis two that conscientiousness would be positively associated with grades. Again, this was tested using two linear regressions, with students' GPA in the semester preceding (GPA1) and following (GPA2) the personality measure as the dependent variable respectively, and their scores in the BFI-Conscientiousness subscale as the independent variable, controlling for their age and gender. As predicted, conscientiousness was significantly positively associated with GPA in both the semester preceding the personality measure (β = 0.16, CI_{95} [0.02, 0.29], t = 2.35, p = .020) and the semester following the personality measure (β = 0.18, CI_{95} [0.03, 0.33], t = 2.42, p = .017).

Thirdly, I tested hypothesis three that Extraversion would be positively associated with grades. Contrary to prediction, Extraversion was not significantly associated with GPA in either the semester preceding the personality measure (β = 0.05, CI_{95} [-0.08, 0.18], t = 0.77, p = .442) or the semester following the personality measure (β = 0.02, CI_{95} [-0.14, 0.17], t = 0.22, p = .828).

I next tested hypothesis four that Agreeableness would be positively associated with grades. As expected, Agreeableness was significantly associated with GPA in both the semester preceding the personality measure (β = 0.19, CI_{95} [0.06, 0.33], t = 2.95, p = .004) and the semester following the personality measure (β = 0.20, CI_{95} [0.05, 0.35], t = 2.59, p = .010).

Finally, I tested hypothesis five that Neuroticism would be negatively associated with grades. Contrary to prediction, Neuroticism was not significantly associated with GPA in either semester (β = -0.06, CI_{95} [-0.19, 0.07], t = -0.86, p = .389 for the semester preceding the personality measure and β = -0.08, CI_{95} [-0.23, 0.07], t = -1.02, p = .307 for the semester following the personality measure).

8.7.4 Study 1 discussion

There were two findings in this study: 1) Conscientiousness and Agreeableness were significantly positively associated with students' educational attainment, and 2) Openness to Experience, Neuroticism and Extraversion were not significantly associated with student's educational attainment. These findings were compatible with the previous finding that personality domains, especially Conscientiousness and Agreeableness, plays an important role in educational attainment (Dumfart & Neubauer, 2016; Poropat, 2009). Since this study was conducted in China, it could provide evidence to support the claim that the effects of personality on educational attainment is consistent across different countries. However, since the difficulty levels of the mid-term assessment and the end-of-term assessment might be different, it was not clear if using mid-term grades as the outcome would change the predicted association between personality and educational attainment. Therefore, a second study was conducted.

8.8 Study 2

8.8.1 Introduction

As noted above, the second study aimed to provide supplementary evidence for the association between personality and educational attainment among Chinese students. Therefore, this study was a replication of study 1, with an independent sample. Since the second study also aimed to test whether using the mid-term grades would influence the association between personality and educational attainment, the dependent variable in the semester following the personality measure was changed to the mid-term grades.

There were five hypotheses in study 2, which were identical to those used in study 1.

8.8.2 Method

8.8.2.1 Participants

In total, 212 participants were recruited from the same primary school as in study 1. One male participant was removed from the data analyses due to his consistent low grades (e.g., 9.2 SDs below the class average for Chinese). There were 120 boys and 91 girls (mean age = 10.70, SD = 0.54) among the remaining participants. All data in this study were collected in the autumn term in 2016.

8.8.2.2 Materials

The materials used in study 2 were identical to those used in study 1, with one exception that students' grades in the semester following the personality measure were from the mid-term assessment.

8.8.2.3 Procedure

The procedure in study 2 was identical to that in study 1.

8.8.3 Results

As in study 1, I first tested hypothesis one that Openness to Experience would be positively associated with grades. Again, this was tested using two linear regressions, with students' end-of-term GPA in the semester preceding the personality measure (GPA1) and students' mid-term GPA in the semester following the personality measure (GPA2) as the dependent variable respectively, and their scores in the BFI-Openness subscale as the independent variable, controlling for students' age and gender. Similar to study 1, Openness to Experience was not significantly associated with students' end-of-term GPA in the semester preceding the personality measure ($\beta = 0.09$, CI_{95} [-0.04, 0.23], $t = 1.37$, $p = .172$), nor with students' mid-term GPA in the semester following the personality measure ($\beta = 0.11$, CI_{95} [-0.02, 0.25], $t = 1.68$, $p = .094$).

Secondly, I tested hypothesis two that conscientiousness would be positively associated with grades. As predicted, conscientiousness was significantly positively associated with students' end-of-term GPA in the semester preceding the personality measure (β = 0.24, CI_{95} [0.10, 0.37], t = 3.54, p < .001), as well as with students' mid-term GPA in the semester following the personality measure (β = 0.19, CI_{95} [0.06, 0.33], t = 2.92, p = .004).

Thirdly, I tested hypothesis three that Extraversion would be positively associated with grades. Contrary to the prediction, Extraversion was not significantly associated with students' end-of-term GPA (β = 0.08, CI_{95} [-0.05, 0.21], t = 1.17, p = .243), nor significantly associated with students' mid-term GPA (β = 0.05, CI_{95} [-0.08, 0.18], t = 0.72, p = .474).

Next, I tested hypothesis four that Agreeableness would be positively associated with grades. Contrary to the prediction, Agreeableness was not significantly associated with either students' end-of-term GPA (β = 0.08, CI_{95} [-0.05, 0.22], t = 1.19, p = .235) or students' mid-term GPA (β = -0.01, CI_{95} [-0.15, 0.12], t = -0.21, p = .833).

Finally, I tested hypothesis five that Neuroticism would be negatively associated with grades. Contrary to the prediction, Neuroticism was not significantly associated with either students' end-of-term GPA (β = -0.10, CI_{95} [-0.23, 0.04], t = -1.43, p = .155) or their mid-term GPA (β = -0.08, CI_{95} [-0.21, 0.06], t = -1.12, p = .264).

I further conducted a random effect meta-analysis to combine the effects of all the five personality domains on students' educational attainment in studies 1 and 2. The model consisted of 20 effect sizes ($Q(df = 19)$ = 39.05, p = .004; also see Figure 8.1). In total, 6 of the 20 effect sizes (33%) were significantly different from zero with positive directions, indicating that these personality domains were related to better educational attainment significantly. The remaining 14 effect sizes (67%) were not significantly different from zero, indicating that these personality domains were not significantly associated with educational attainment. The meta-analytic average coefficient between the five personality domains in the FFM and educational attainment was $\beta = 0.08$, CI_{95} [0.03, 0.12]. Thus, the findings in studies 1 and 2 jointly supported that personality (based on the FFM) is significantly associated with students' educational attainment, which are in line with the previous findings (e.g., Dumfart & Neubauer, 2016; Komarraju et al., 2009; Poropat, 2009; Zhang & Ziegler, 2016; Zhou, 2015).

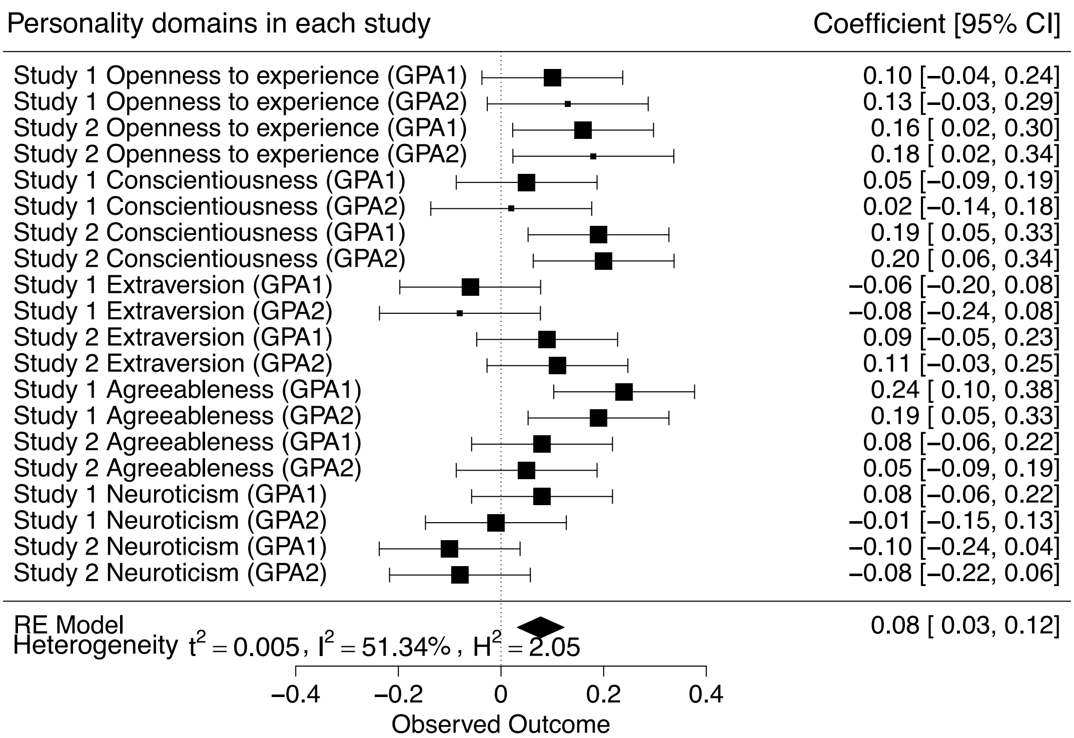

Figure 8.1 The meta-analysis results of the effects of the five personality domains (based on the FFM) on students' GPA in the semester preceding (GPA1) and following (GPA2) the personality measure in studies 1 and 2

8.8.4 Study 2 discussion

Similar to study 1, conscientiousness was significantly positively associated with both the end-of-term and the mid-term grades, but Openness to Experience, Neuroticism, Extraversion were not significantly associated with grades. Contrary to study 1, Agreeableness was not significantly positively associated with either the end-of-term or the mid-term grades in study 2. Therefore, these findings confirmed that conscientiousness is the most stable and the strongest predictor of students' educational attainment compared to the other four domains in the FFM (Bratko et al., 2006; Butcher et al., 1963; Chamorro-Premuzic & Furnham, 2003; Dumfart & Neubauer, 2016; Ivcevic & Brackett, 2014; O'Connor & Paunonen, 2007; Poropat, 2009). In addition, different types of assessment (i.e., end-of-term or mid-term test) did not affect the strong association between conscientiousness and educational attainment.

Although a consistently and significantly positive association between conscientiousness and educational attainment was found in Studies 1 and 2, these studies only measured the student's own personality. Some researchers suggested that as the most important person in the students' learning process, the teacher's personality also had influential effects on students' educational attainment (Hashim et al., 2014; Kim et al., 2018). Thus, a third study that aimed to test the association between teacher's personality and students' educational attainment was conducted.

8.9 Study 3

8.9.1 Introduction

This study aimed to provide informative evidence for the association between teacher's personality and students' educational attainment. Since studies 1 and 2 found that Conscientiousness has the strongest effect on educational attainment compared to Openness to Experience, Extraversion, Agreeableness and Neuroticism in the FFM, this study would be focused on the effects of teacher's conscientiousness on students' educational attainment. In addition, previous studies suggested that the classroom environment is positively associated with students' engagement and school achievement (Asbury, Almeida, Hibel, Harlaar, & Plomin, 2008; Dotterer & Lowe, 2011; Downer, Rimm-Kaufman, & Pianta, 2007; Pianta, la Paro, Payne, Cox, & Bradley, 2002; Walker & Plomin, 2006). This study also aimed to test whether the classroom environment (in this case classroom chaos) would be associated with students' educational attainment.

There were two hypotheses in this study: 1) teacher's conscientiousness would be positively associated with students' educational attainment and 2) classroom chaos would be negatively associated with students' educational attainment.

8.9.2 Method

8.9.2.1 Participants

In total, 100 parent-children (52 boys, 48 girls) dyads were recruited from Prolific (an online participant pool). In order to collect consensual educational attainment, adult participants were restricted to those lived in England and had a child who took the key stage 1 (end of Year 2) National Curriculum assessments in 2018. Child participants were aged from 6 years 3 months to 9 years 3 months old (mean = 7.96, SD = 0.86).

8.9.2.2 Materials

Family socioeconomic status (SES): Family SES was assessed by asking the participating parent to provide the highest educational level between the two parents in the family and the employment type of the highest earning parent in the family.

Teacher's conscientiousness: The Teacher's conscientiousness scale was adapted from the conscientiousness subscale of the NEO-FFI (Costa & McCrae, 1992). The conscientiousness subscale of the NEO-FFI includes 12 items and each of the items describes a behaviour, attitude or feeling related to conscientiousness (Costa & McCrae, 1992). An example item is "*I keep my belongings neat and clean*". In this study, all 12 items were adapted from the first-person's view to a third person's (students) view (see Appendix B and C for mathematics and English teachers' conscientiousness scale respectively). Example items are "*My mathematics teacher kept his/her belongings neat and clean*" and "*My English teacher often came into situations without being fully prepared*". Responses were recorded on a five-point Likert scale (from 1: strongly disagree to 5: strongly agree).

Classroom chaos: The classroom chaos scale was adapted from the 15-item Confusion, Hubbub, and Order Scale (CHAOS; Matheny, Wachs, Ludwig, & Phillips, 1995). Items in the CHAOS scale that could be used in the classroom environment were selected and adapted by me. In total, there were 6 items in this scale (also see Appendix D and E for mathematics and English classroom CHAOS respectively). An example item is "*In my mathematics class, teacher could talk to us without being interrupted*". Responses were recorded on a three-point Likert scale (from 1: "not true", 2: "quite true" to 3: "very true").

Educational attainment: Students' key stage 1 (end of year 2) grades in English reading, English grammar, and mathematics were reported by the participating parents.

8.9.2.3 Procedure

This study was an online study. After gaining consent, parents were asked to fill out a demographic questionnaire including the participating parent's age, gender, the highest educational level that a parent of the family had achieved, the highest earning parent's employment type, the participating child's age and gender, and the participating child's key stage 1 grades in English reading, English grammar, and mathematics. Following this, parents were instructed to ask their children to come to the screen to answer the following questions. Children firstly needed to rate their mathematics and English teachers' conscientiousness. This was followed by the classroom chaos scale. Once children finished all the ratings, they saw an end-of-survey page including a message of thanks. The adult participants received compensations of £1.50.

8.9.3 Results

Before testing the two hypotheses, I firstly did four parallel analyses to test the number of factors included in each scale. The Parallel analysis suggested there was one factor in the mathematics teacher's conscientiousness scale (adjusted eigenvalue was 5.66), the English teacher's conscientiousness scale (adjusted eigenvalue was 5.96), and the mathematics classroom chaos scale (adjusted eigenvalue was 2.73). However, for the English classroom chaos scale, the Parallel analysis suggested that there were two factors: classroom noise (the primary factor; adjusted eigenvalue was 2.20) and classroom order (the second factor; adjusted eigenvalue was 1.13).

Next, I tested hypothesis one that teacher's conscientiousness would be positively associated with students' educational attainment. This was tested using three linear regressions, with children's mathematics, English reading and grammar grades as the dependent variable respectively, children's perceptions of their mathematics and English teacher's conscientiousness as the independent variable respectively, and children's age and gender as covariates. Contrary to the prediction, children's perceptions of their mathematics teacher's conscientiousness were not significantly associated with their mathematics grades ($\beta = 0.02$, CI_{95} [-0.19, 0.22], $t = 0.16$, $p = .870$). In addition, children's perceptions of their English teacher's conscientiousness were not significantly associated with either English reading grades ($\beta = 0.02$, CI_{95} [-0.18, 0.23], $t = 0.24$, $p = .811$), or grammar grades ($\beta = -0.07$, CI_{95} [-0.28, 0.13], $t = -0.74$, $p = .463$; also see Figure 8.2).

Finally, I tested hypothesis two that classroom chaos would be negatively associated with students' educational attainment. This was done using three linear regressions, with children's mathematics, English reading and grammar grades as the dependent variable respectively, children's ratings of their mathematics and English classroom chaos (including classroom noise and classroom order factors) as the independent variable respectively, children's age and gender as the covariates. Contrary to the prediction, children's ratings of their mathematics classroom chaos were not associated with their mathematics grades ($\beta = -0.03$, CI_{95} [-0.23, 0.17], $t = -0.27$, $p = .788$). Additionally, neither children's ratings of their English classroom noise ($\beta = -0.01$, CI_{95} [-0.22, 0.19], $t = -0.12$, $p = .905$) nor classroom order ($\beta = -0.09$, CI_{95} [-0.30, 0.11], $t = -0.93$, $p = .355$) were associated with their English reading grades. Children's ratings of their English classroom noise were not significantly associated with their English grammar grades (β = 0.10, CI_{95} [-0.10, 0.30], t = 0.96, p = .339). However, as predicted, children's ratings of their English classroom order were significantly negatively associated with English grammar grades (β = -0.20, CI_{95} [-0.40, 0.00], t = -2.01, p = .047).

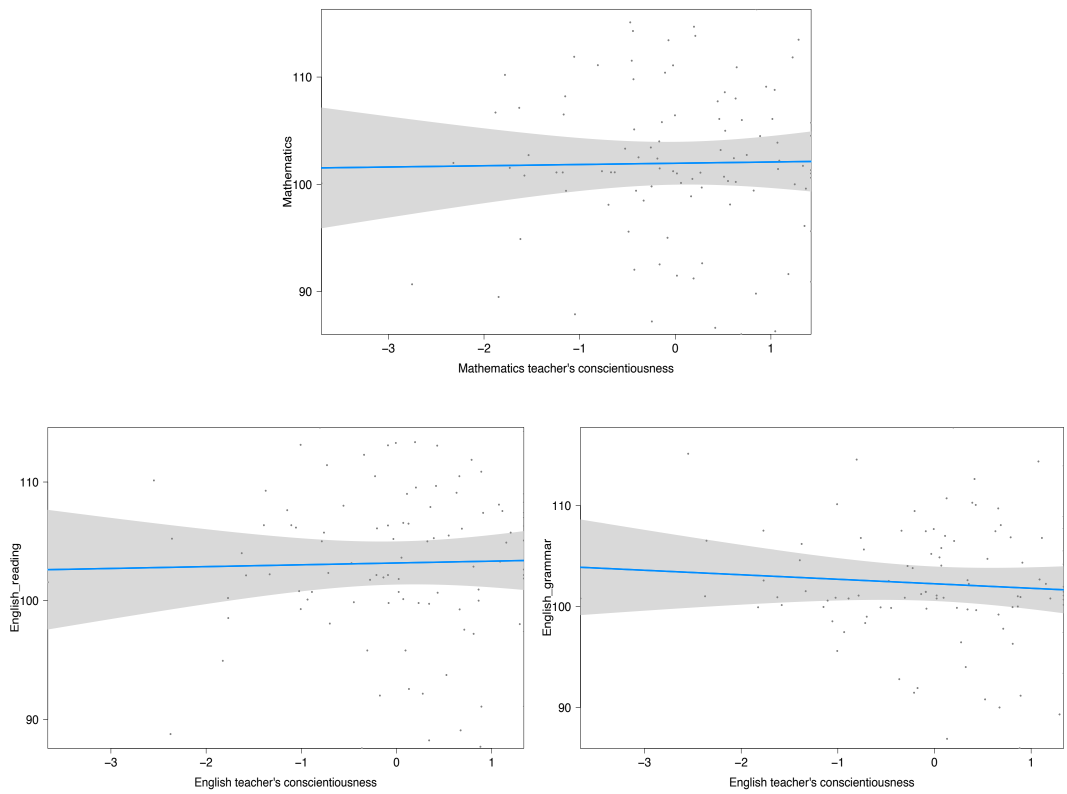

Figure 8.2 The associations between students' perceptions of their mathematics teacher's conscientiousness and students' mathematics grades (upper), students' perceptions of their English teacher's conscientiousness and students' English reading grades (bottom left), and students' perceptions of their English teacher's conscientiousness and students' English grammar grades (bottom right)

8.9.4 Study 3 discussion

There were two findings in this study. First, students' perceptions of their teacher's conscientiousness were not significantly associated with their educational attainment. This finding was contrary to that of Hashim et al. (2014) and Tahir and Shah (2012), but in line with Kim et al. (2018). Second, students' ratings of classroom chaos were not significantly associated with their educational attainment, with one exception that student's ratings of their English classroom order were significantly negatively associated with grades in grammar.

For the two studies that found a positive association between the teacher's personality and students' educational attainment, Hashim et al. (2014) asked students to rate their teacher's personality but as stated in Section 8.5, the questionnaire they used to measure teacher's personality was more likely a teaching-quality questionnaire rather than a personality questionnaire. Therefore, their positive finding could only show that teacher's teaching quality had positive effects on students' educational attainment, but it could not explain the effects of teachers' personality on students' educational attainment. Whereas Tahir and Shah (2012) also asked for students' perceptions of their teacher's personality and used an appropriate questionnaire based on the FFM (the BFI). I used the same approach as Hashim et al. (2014) and Tahir and Shah (2012) (i.e., asking students to rate their teacher's personality) and measured personality with an appropriate measurement (i.e., the NEO-FFI), but all results were null. One reason to explain the differences between previous findings and my findings might be that participants in my study were much younger than participants in the study by Hashim et al. (2014) and of Tahir and Shah (2012). Specifically, both Hashim et al. (2014) and Tahir and Shah (2012) asked mature college students to rate their teacher's personality, but I asked primary school students to rate their teacher's personality. In another study that was conducted by Kim et al. (2018), they also asked young students (secondary school) to rate their teacher's personality and found a null result. Therefore, it might be that mature students are more biased by their own opinion compared to young students when rating their teacher's personality. Since this study did not measure the student's own personality but only measured students' perceptions of their teacher's personality, it was not possible to test whether students' perceptions were biased by their own opinion. Therefore, I conducted a fourth study that aimed to assess both students' own personality and their perceptions of teachers' personality.

8.10 Study 4

8.10.1 Introduction

To increase the power of this study, I made three changes in the method. First, although Studies 1 and 2 found that conscientiousness was positively associated with student's educational attainment, it was not clear which specific behaviour related to conscientiousness was associated with educational attainment. If specific behaviours related to conscientiousness could be identified, psychologists would be able to develop interventions to enhance these behaviours and therefore to improve students' educational attainment. This study aimed to use a behaviour-specific questionnaire to identify the specific behaviours related to conscientiousness that are associated with educational attainment. In addition, to provide more comprehensive data of students' own conscientiousness, this study would use the conscientiousness subscale in the NEO-PI-R (48 items) instead of using the conscientiousness subscale in the NEO-FFI (12 items). In addition, since Studies 1 and 2 only measured students' own personality and Study 3 only measured student's perceptions of their teacher's personality, it was not clear whether students' perceptions were biased by their own opinion. This study aimed to solve this problem by measuring both students' own conscientiousness and their perceptions of teachers' conscientiousness. Finally, to clarify whether the age of participants would be a moderator of the association between students' perceptions of their teachers' conscientiousness and educational attainment, this study would recruit participants from a higher educational level (i.e., high school) than study 3.

There were three hypotheses in this study: 1) students' own conscientiousness would be positively associated with their educational attainment, 2) students' perceptions of teacher's conscientiousness would be positively associated with student's educational attainment, and 3) student's perceptions of their teacher's conscientiousness would not be correlated with students' own conscientiousness.

8.10.2 Method

8.10.2.1 Participants

In total, 227 (91 girls and 123 boys, mean age = 16.71, SD = 1.73) participants were recruited from a public high school in Wuhan, China. Seven participants did not provide any demographic information of gender, age, GPA, grades in mathematics, English and Chinese. Four participants reported their age were more than 5 SDs lower than the average age. One participant did not report their gender.

8.10.2.2 Materials

Teacher's conscientiousness was assessed using the identical measurement as in study 3.

Students' own conscientiousness: There were two measurements for assessing students' own conscientiousness: the conscientiousness subscale of the NEO-PI-R (Costa & McCrae, 1992) and the BIC scale (Jackson et al., 2010). The conscientiousness subscale of the NEO-PI-R includes 48 items (e.g., "*I have a clear set of goals and work toward them in an orderly fashion*"). Responses were recorded on a five-point Likert scale (from 1: strongly disagree to 5: strongly agree). The BIC includes 185 items that are associated with 11 underlying facets of conscientiousness (Jackson et al., 2010). An example item is "*Organise books by height, author, or genre*". Respondents were asked to indicate the frequency of taking part in each of the behaviours stated in the scale on a five-point scale (from 1: never performed the behaviour to 5: performing the behaviour quite often).

Educational attainment: Students' most recent year's GPA was reported by themselves.

8.10.2.3 Procedure

After gaining consent, students were given a brief introduction to this study. Then they were asked to answer a series of questionnaires including a demographic questionnaire (age, gender, class, the most recent year's GPA, and grades in English, mathematics and Chinese courses), the BIC, the conscientiousness subscale of the NEO-PI-R, and their perceptions of their headteacher's conscientiousness.

8.10.3 Results

Similar to study 3, I firstly did two Parallel analyses for the BIC and the teacher's conscientiousness scales used in this study. The Parallel analysis suggested that the BIC scale had three factors including organisation (the primary factor; adjusted eigenvalue was 6.22), responsibility (the second factor; adjusted eigenvalue was 2.67) and anti-social behaviour (the third factor; adjusted eigenvalue was 1.67; also see Table 8.1). For the teacher's conscientiousness scale, the Parallel analysis suggested a two-factor model which included a positive conscientiousness factor (the primary factor; adjusted eigenvalue was 5.91) and a negative conscientiousness factor (the second factor; adjusted eigenvalue was 1.53).

Table 8.1 Factor loadings of the BIC scale

	Factor 1: Organisation	Factor 2: Responsibility	Factor 3: Anti-social behaviour
Appearance	0.678		
Appearance 2	0.602		
Cleanliness	0.736		
Cleanliness 2	0.686		
Formality 2	0.590		
General	0.637	0.452	0.320
Industriousness	0.556	0.368	
Industriousness 2	0.594	0.366	
Organisation	0.786	0.307	
Organisation 2	0.721		
Punctual	0.307	0.565	
Punctual 2		0.583	
Responsibility		0.784	
Responsibility 2		0.833	
Anti-social			0.674
Impulsivity 2			0.638
Laziness			0.633
Laziness 2			0.769
Anti-social 2		-0.343	-0.370
Formality	0.390	0.445	
Impulsivity			0.306
Avoid-work			0.479
Avoid-work 2			0.442

Next, I tested hypothesis one that students' own conscientiousness would be positively associated with their educational attainment. This was done using two linear regressions, with student's GPA as the dependent variable, their scores on the conscientiousness subscale of the NEO-PI-R and scores on the BIC as the independent variable respectively, and age and gender as covariates. As predicted, student's own conscientiousness was significantly associated with their GPA (β = 0.20, CI_{95}[0.05, 0.35], t = 2.58, p = .011). In addition, responsibility, as the second factor of the BIC scale, was significantly associated with students' GPA (β = 0.21, CI_{95} [0.05, 0.36], t = 2.61, p = .010), but the other two factors (organisation and anti-social behaviour) in the BIC were not significantly associated with students' GPA (organisation: β = 0.04, CI_{95} [-0.12, 0.21], t = 0.51, p = .614 and anti-social behaviour: β = 0.01, CI_{95} [-0.16, 0.18], t = 0.11, p = .916; also see Figure 8.3).

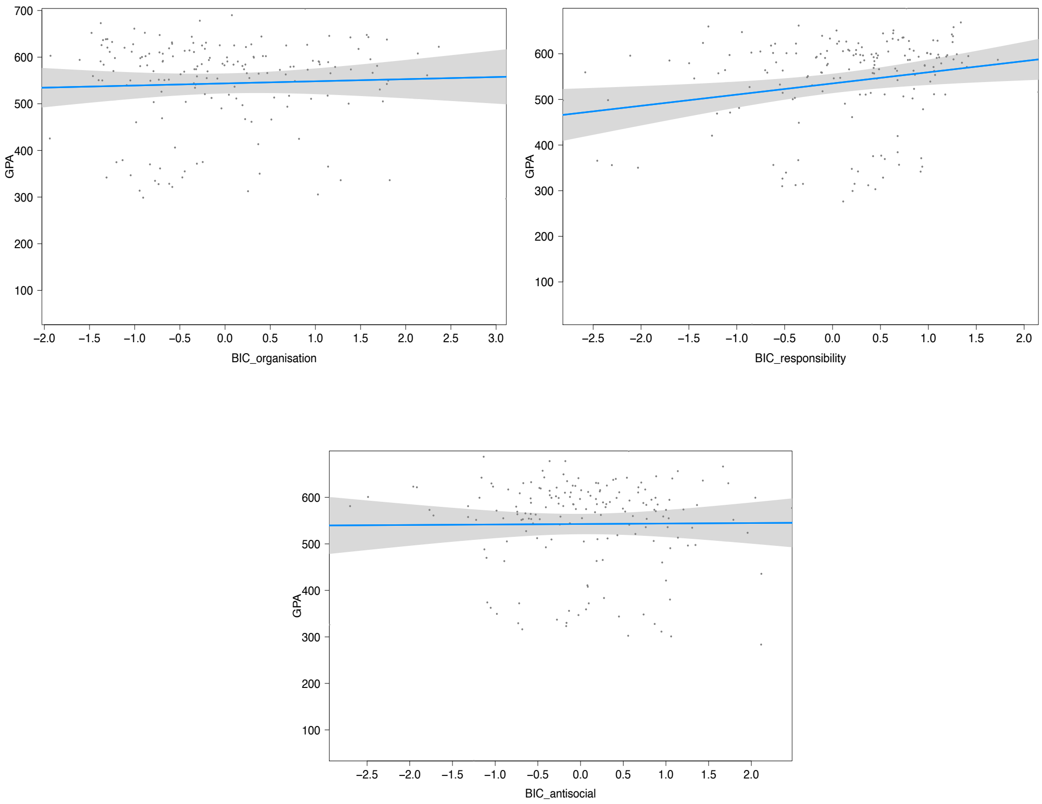

Figure 8.3 The association between the organisation factor (upper left), the responsibility factor (upper right), the anti-social behaviour factor (bottom) and students' GPA

I then tested hypothesis two that students' perceptions of teacher's conscientiousness would be positively associated with student's educational attainment. Again, this was tested using a linear regression, with student's GPA as the dependent variable, student's perceptions of their teacher's conscientiousness as the dependent variable, and age and gender as covariates. Contrary to prediction, students' perceptions of their teachers' conscientiousness were not significantly associated with students' GPA (for the positive factor; $\beta = 0.07$, CI_{95} [-0.08, 0.22], $t = 0.87$, $p = .383$ and for the negative factor: $\beta = 0.05$, CI_{95} [-0.11, 0.20], $t = 0.59$, $p = .558$).

Since participants were recruited from seven classes, only seven teachers were rated. To explore whether controlling for the number of teachers would change the results, I conducted a mixed effect model by using the lmer function of the lme4 package in R. Students' GPA was used as the dependent variable, the positive and negative factors of teacher's conscientiousness were fixed effects, and the term (1|class) was a random effect (i.e., lmer(GPA ~ TeacherCpos + TeacherCneg + (1|Class))). The linear mixed model results suggested that the differences between classes explain near-zero variance after the variance explained by the fixed effects. In addition, neither the positive nor the negative factor of teacher's conscientiousness was significantly associated with student's GPA (for the positive factor: $\beta = 6.67$, CI_{95} [-9.06, 22.39], $t = 0.83$ and for the negative factor: $\beta = 3.89$, CI_{95} [-10.88, 18.66], $t = 0.52$; also see Table 8.2 for the descriptive statistics of the mixed effect model results).

Table 8.2 Descriptive statistics for the mixed effect model (SEs are shown in the parentheses)

	Dependent variable (GPA)
Teacher's conscientiousness (positive)	6.669 (8.048)
Teacher's conscientiousness (negative)	3.890 (7.561)
Constant	542.048*** (8.004)
Observations	175
Log Likelihood	- 1053.718
Akaike Inf. Crit.	2117.436
Bayesian Inf. Crit.	2133.260

Note: *p<0.05; **p<0.01; ***p<0.001

Finally, I tested hypothesis three that students' perceptions of their teacher's conscientiousness would not be correlated with students' own conscientiousness. Contrary to prediction, students' perceptions of their teacher's conscientiousness showed a marginally significant correlation with the students' own conscientiousness scores (r = 0.23, p = .060 for the positive factor of teachers' conscientiousness and r = - 0.03, p = .070).

8.10.4 Study 4 discussion

Similar to Studies 1 and 2, student's own conscientiousness was significantly positively associated with their educational attainment. In addition, a significantly positive association between responsible behaviours and higher educational attainment was found. For the association between students' perceptions of their teacher's conscientiousness, the results were null. Further, students' perceptions of their teacher's conscientiousness were positively correlated with their own conscientiousness, but the correlations were only marginally significant. To provide an integrated interpretation of all findings in the four studies, I will next present a brief general discussion.

8.11 General discussion

Personality, especially conscientiousness, has been found as the strongest and the most stable non-cognitive predictor of educational attainment (e.g., Chamorro-Premuzic & Furnham, 2003; De Fruyt & Mervielde, 1996; Dumfart & Neubauer, 2016; Poropat, 2009; Zhang & Ziegler, 2016; Zhou, 2015). Across four studies, I tested the effects of students' personality (especially conscientiousness) and students' perceptions of their teacher's conscientiousness on students' educational attainment. Studies 1 and 2 used a self-reported personality questionnaire (the BFI) to test whether the five domains in the FFM (i.e., Openness to Experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism) were associated with educational attainment among primary school students. Results showed that conscientiousness was the most significant and stable predictor of students' educational attainment compared to the other four domains in the FFM. Study 3 tested the relationship between students' perceptions of their teacher's conscientiousness and students' educational attainment. No significant association was found. Study 4 tested both the effects of students' own conscientiousness and the effects of students' perceptions of their teacher's conscientiousness on students' educational attainment. Students' own conscientiousness was significantly positively associated with their educational attainment, but their perceptions of the teacher's conscientiousness were not. In addition, when identifying the specific behaviours that are associated with educational attainment, a significant association between behaviours related to responsibility and students' educational attainment was found. Study 4 also tested whether students' perceptions of their teacher's conscientiousness were objective reflections or were biased by students' thinking. A positive correlation between students' perceptions of their teacher's conscientiousness and their own conscientiousness was found (but the result was not significant: with the minimum *p*-value was .060), indicating that student's ratings were more likely to be biased by their own thinking.

Across all studies reported in this chapter, we can confidently conclude that conscientiousness is consistently and positively associated with educational attainment in both primary school and high school in China. These findings are in line with previous conducted in Western countries (e.g., Bratko et al., 2006; Butcher et al., 1963; Chamorro-Premuzic & Furnham, 2003; Dumfart & Neubauer, 2016; Ivcevic &

Brackett, 2014; Laidra et al., 2007; O'Connor & Paunonen, 2007; Poropat, 2009). When trying to identify the specific behaviours related to conscientiousness that are associated with educational attainment, behaviours underlying the responsibility factor were found to be significantly positively associated with educational attainment. This finding suggested that students who were responsible, punctual, organised and industrious were more likely to achieve high educational attainment at school, which is in line with other studies (Martel, McKelvie, & Standing, 1987; Zhou, 2015) .

Turning to the influence of teacher's conscientiousness on students' educational attainment, as in Tahir and Shah (2012), I used observer's ratings and an appropriate personality questionnaire (i.e., the NEO-FFI) to assess teacher's conscientiousness, but found a contrary result that there was no effect of teacher's conscientiousness on students' educational attainment either in the UK or in China. Therefore, it might be that teacher's personality is correlated with their job performance (Salgado, 1997), but teacher's personality does not have any direct impact on students' educational attainment.

8.11.1 Limitations and further directions

There were three limitations in my studies. First, socioeconomic status (SES) has been found to be an important influence in educational attainment (e.g., Sirin, 2016), but it was not well controlled in studies 1, 2 and 4. Although students in studies 1 and 2 were recruited from a city in relative poverty, individual students' SES might be different and might cause a gap in educational attainment. Therefore, further studies are needed to collect the SES of individual students.

Second, teachers' conscientiousness was only rated by their students, but their self-reported conscientiousness was not measured. McCrae et al. (2004) suggested that although the tendencies of broad personality domains were consistent between self-reported and observers' ratings, there was a certain amount of discrepancy in the estimates between self-reported and observer's ratings. Thus, further studies could include both self-reported measurement and observer's ratings when assessing teacher's conscientiousness in order to obtain more comprehensive and accurate data.

Chapter 9. General discussion

How to help children to learn is recognised as a major issue in education. In this thesis, I addressed this question by exploring a possible approach to increase children's mathematical knowledge, by testing the effects of mindset on children's cognitive ability and educational attainment, by investigating the earliest emergence of gender stereotypes about brilliance among children and the influence of these stereotypes on children's expressed career interests, and by examining the association between conscientiousness and children's educational attainment. In this final chapter, I first briefly summarise the main findings from the thesis, then discuss the implications for each theory, and finally, highlight the need for new directions in reality.

9.1 Summary of findings

In Chapter 1, I tested whether playing number board games could enhance children's mathematical knowledge at the early developmental stage. At this point, I intended that the theme of my thesis would be approaches that aid children in understanding concepts. My initial study, however – a replication of the work of Ramani and Siegler (2008) on the impact of playing number board games on acquisition of the symbolic meaning of numbers indicated that, at least in my sample of Chinese children, this activity was not significantly associated with increased performance in either counting, comparison, identification or a number line estimation task. In addition, children in my sample had higher counting and identification capabilities than their same-aged counterparts in the original paper my work was attempting to replicate (which was based on data from poor American children). By contrast, in my sample, number line estimation capability lagged behind. This highlighted the need to ensure appropriate age groups, test difficulty, and sample size.

My exploration then turned to investigate a prominent theory that has been proposed to be influential in children's learning: the mindset theory (Dweck, 2006). Across three series of studies (a total of 1,780 participants), I tested the possible effects of mindset on children's cognitive performance after challenges and educational attainment (Chapter 3), the association of mindset and educational attainment across a challenging transition or among disadvantaged children (Chapter 4), and the origins of mindset (Chapter 5). In Chapter 3, I used both praise manipulation and a self-reported mindset scale to test whether having a growth mindset would improve children's cognitive performance after challenges or increase children's educational attainment. Results showed that praising children for being smart was not harmful to their cognitive performance after challenges compared to praising children for hard work. For the association of children's own mindsets and their educational attainment, no significant result was found. These results suggested that implicit mindsets about the nature of intelligence have near-zero effects on grades and no effect on general cognitive ability. In Chapter 4, I tested whether mindset was only activated when children were experiencing challenging transitions or among disadvantaged children when entering university. Again, there was no significant association between mindset and children's educational attainment either across a challenging transition from high school to university, in any subsequent transitions in university, nor among children who achieved relatively low scores when entering university. Thus, beliefs about the nature of intelligence appear to be simply as knowledge about a scientific topic but not related to learning. In Chapter 5, I explored whether children gained their intelligence mindsets from parents' intelligence mindsets or parents' failure mindsets. Haimovitz and Dweck (2016) reported that children's mindsets were gained from parents' failure mindsets rather than from parents' intelligence mindsets. However, my findings were contrary, indicating that genetic factors might play an important role in the formation of children's intelligence mindsets.

The third theme in my thesis focused on the period in development when gender stereotypes about brilliance among young children in China and the U.K. emerge (Chapter 7) and whether gender stereotypes about brilliance were associated with children's expressed career interests. My results suggested that at age 10, Chinese children did not have a gender stereotype about brilliance in favour of men, but their counterparts in the U.K. already viewed brilliance as a male trait. In addition, I consistently found that both Chinese and British children had gender stereotypes towards niceness and dullness. In other words, both girls and boys believed that girls were nicer than boys, and boys were duller than girls. For the association of gender stereotypes about brilliance and children's career interests, no significant association was found. As such, additional replication attempts with younger Chinese children and studies of gender-differences in interests and tests of competing explanations for differences in life choices are warranted.

The final theme in this thesis examined the association between personality traits and children's educational attainment (Chapter 8). Conscientiousness was consistently and significantly associated with children's educational attainment in all of my samples recruiting from both primary school and high school. I also tested whether teachers' conscientiousness would be associated with children's educational attainment, but no significant results were found.

9.2 Implications and further directions for each theory

9.2.1 The effects of playing number board games on children's learning in mathematics

Although playing number board games has been reported to be an effective and low-cost approach to improving children's mathematical knowledge in counting, identification, comparison and number line estimation (Ramani & Siegler, 2008), as shown in Chapter 1, playing number board games was not helpful for Chinese children's learning in mathematics. As argued in Chapter 1, the contradictory results might have resulted from either my use of participants who were too young (3 year 6 months on average compared to 4 years 9 months in Ramani and Siegler (2008)) to understand the board games and the numerical tasks (consistent with their near-chance performance at post-test on two of the four tasks, with number line estimation task performance near zero). This raises the issue of why, on two of the mathematical tasks involving counting and number recognition, my sample was scoring at levels near those reported by Ramani and Siegler (2008). Yet still I found no effect on these tasks. One possibility is that the Chinese children's number awareness at the early developmental stage was based on memorising rather than understanding. This would account for the large discrepancy between their performance counting from 1 to 10 compared to their near-chance performance in comparing magnitudes. Chinese parents apparently are able to teach children to memorise the numbers but this is not accompanied by an understanding of the meanings of numbers until children had formal mathematics classes at school, leading to the discrepancy observed in my (small) sample of 3 year olds.

Although I did not find any significant effects of playing number board games on children's mathematical knowledge at the early developmental stage, there are three suggestions for further studies. First, researchers could increase the difficulty of the counting task (e.g., from counting from 1-10 to 1-20) to avoid the possible ceiling effect among Chinese children. Second, researchers could recruit children at 4 years old as those in Ramani and Siegler (2008) to avoid the inequality of cognitive development among children, and allow children to have more ability to understand the board games and the numerical tasks (especially the number line estimation task). Third, further studies could enlarge the sample size to increase the power to detect an effect of playing number board games on children's mathematical knowledge.

9.2.2 The effects of growth mindset on children's learning

Growth mindset is positively associated with cognitive performance after challenges (Mueller & Dweck, 1998) and educational attainment (Blackwell et al., 2007). However, this thesis did not find significant and consistent evidence to support the association of growth mindset with children's cognitive performance after challenges, nor the association of growth mindset with children's educational attainment, even across challenging transitions or among disadvantaged children. These null results are incompatible with the mindset theory, but are in line with an increasing volume of research which has found that growth mindset was not (even negatively) associated with educational attainment in recent years (Bahník & Vranka, 2017; Bazelais et al., 2018; Sisk et al., 2018).

These contradictory results have raised three questions in my mind. First, it is not clear how mindset is associated with children's educational attainment. Mindset refers to people's beliefs about the malleability of their intelligence (Dweck, 2006), however, it seems that beliefs about one's intelligence were independent from the development of intelligence as reported in Chapter 3. Intelligence as the most powerful predictor of educational attainment is a product of genetic effects in interaction with resources linked to socioeconomic status (Bates, Hansell, Martin, & Wright, 2016; Deary, Spinath, & Bates, 2006; Tucker-Drob & Bates, 2016). Thus, if an association of mindset with educational attainment exists, this association should not be related to intelligence. This is in line with a recent claim made by (Dweck & Yeager, 2019) that mindset is a *meaning system* which integrates variables including goals, attributions, helplessness, and effort beliefs that could guide people's motivations and behaviours. However, this thesis failed to support the claim since no significant association of mindset and children's learning versus performance goals, attributions, task persistence and task enjoyableness was found in my study. For the variable of effort beliefs in the meaning system, it is more likely a personality trait of Conscientiousness, which has been verified as a strong non-cognitive predictor of educational attainment in this thesis as well as in a considerable number of studies (e.g., Dumfart & Neubauer, 2016; Noftle & Robins, 2007; Poropat, 2009), rather than a random variable to be included in mindset.

Second, it is not clear under what circumstances people's mindsets would be activated. Researchers suggested that difficult tasks (Mueller & Dweck, 1998), challenging transitions (Blackwell et al., 2007) or deficits in achievement (Paunesku et al., 2015) could activate people's mindsets. This thesis tested each of these three possibilities but did not find that any of the possibilities did activate children's mindsets (Chapters 3 and 4). When Chapter 3 was published as a paper, there was a reviewer who argued that the difficult tasks used in those studies were not hard enough (see section 9.3.2 for details), meaning that participants had not been challenged sufficiently and their mindsets were not fully activated. I had three responses to this. First, the tasks used in Chapter 3 were not trivially easy: in fact, they shifted participants from getting most items correct (in phase 1) to failing 50% or more of items. Second, and most critically to us, the negative feedback used in the Mueller and Dweck (1998) method focussed on the experience of failure: participants were told "*Your performance was poor on that: You got less than half the puzzles correct*". My participants experienced this as harsh negative feedback and were upset to hear their performance was so poor. If these responses could not prove that my participants were being challenged in the experiments, then the mindset theory should specify the specific standards or indicate a particular circumstance that mindset could be fully activated. In addition, the theory should clarify whether the negative feedback or the puzzle itself is more important to create a challenging environment for participants.

A final concern regarding mindset theory is its status as scientifically falsifiable. As documented by Burgoyne et al. (2020); Macnamara and Rupani (2017); and Sisk et al. (2018), proponents of mindset have for decades made bold claims for the effects of mindset, and this is to be commended: Dweck in particular made specific claims capable of falsifiability, clearly linking effects on learning and grades to implicit theories of intelligence. Given that the relationship of mindset to academic achievement is central to the theory, particularly when students undergo challenges, my results showing both cross-sectional and longitudinal null and reverse-effects, and previous studies failing to find support for the basic premises of mindset theory (e.g., Burgoyne et al., 2020) provide a test case: can mindset theory be falsifiable? For mindset to be a scientific theory (Popper, 1963) and for the scientific community to function programmatically, mindset theory must be able to be falsifiable, and the community to reject, rather than protect, the theory when this has happened. Post-hoc revisions discounting effects of participant's own mindsets on core outcomes in favour of variable outcomes, relaxation of the nature of interventions to include wide ranging complex traits and items chosen because they appear to have worked, caveating potential replications with the proviso that effects of any given intervention are not warranted to work outside the exact environment in which they were initially observed, severely reducing claimed effect sizes, dropping claims of effect durability, and a non-delimited set of auxiliary moderators such as culture, cohort, age, school-district, SES, classroom learning climate, etc. do not serve to allow researchers to refine where the theory works. Rather, they render the theory unfalsifiable.

For further research in the mindset field, it might be valuable to choose a set of tasks as exactly difficult (i.e., children could solve 5.2/10 items correctly in the moderately difficult tasks and 1.6/10 items correctly in the most difficult tasks) as those used in Mueller and Dweck (1998). Future research might also focus on how educational messaging can be most effectively delivered to young children. For instance, in line with the reported protocol, children received praise for being smart or for hard working only once in the experimental process of my mindset studies, and the praise was given by an unfamiliar experimenter. Independent of the validity of mindset theory, research on the required repetitions for effective communication would be of value. Studies examining the frequency of praise, even as a reinforcer would be of value. The differences in the source of the communication might also be valuable to study: Might praise be efficient and trusted by young children if no one else had ever given similar praise to them? Or if the source is a known teacher, or a respected "white coat" (Rienzo, Rolfe, & Wilkinson, 2015)? Further studies could consider inclusion of a condition that children would receive praise for being smart or for hard work repeatedly from their teachers and parents for one week and test the differences in children's cognitive performance afterwards.

9.2.3 Are gender stereotypes about brilliance associated with children's career interests?

Bian et al. (2017) reported that children start to view brilliance as a male trait as early as age 6 years. This thesis provided evidence suggesting that there might be cultural differences in gender stereotyping about brilliance (Chapter 7). An interesting question which raised itself in my mind is why the U.K. had higher gender equality than China (rank 15 to rank 103 out of 149 countries; World Economic Forum, 2018), but children in the U.K. reported more gender-stereotypes in their mind compared to their counterparts who I studied in China? As discussed in Chapter 7, one possibility is that countries with high levels of economic development and high level of gender equality allow children to choose school subjects and occupations based on their own interests and strengths, rather than being constrained by economic demands (Stoet & Geary, 2018). Thus, my results might reflect relatively low self-determination in China relative to the U.K. and due to this low self-determination, in reality, there might be more women working in the STEM related occupations in China compared to the U.K.. Based on reality, my Chinese participants reported fewer gender-stereotypes compared to their counterparts who I studied in the U.K. (Jussim, Harber, Crawford, Cain, & Cohen, 2005).

Although I tried to replicate the task (i) in study 1 of Bian et al. (2017) closely, a few changes were made in my studies. One of the most important changes was that the Chinese participants in my study (10 years old) were older than participants (5, 6 and 7 years old) in the original study of Bian et al. (2017). The older participants might realise the study aim was about gender, so their responses might be biased. Therefore, one suggestion for further studies is to recruit Chinese participants at the same age as those in the original study of Bian et al. (2017). In addition, the images used in my studies and the studies of Bian et al. (2017) had confounding variables such as the person's dressing style which might lead participants to have biased choices. Therefore, another suggestion for further studies is using other experimental stimuli or approach rather than using images as the response-options. For instance, one possible approach could be used in further studies is asking participants to write their own stories about a smart, a nice and a slow-minded person, and testing if participants are more likely to use a name of woman or man in each of their stories.

9.2.4 The power of Conscientiousness

When comparing the dimensions of Openness to Experience, Extraversion, Agreeableness and Neuroticism in the FFM, Conscientiousness has been found to be the most powerful and stable personality predictor of children's educational attainment (Dumfart & Neubauer, 2016; Poropat, 2009). This thesis provided further evidence of this association of Conscientiousness with children's educational attainment. Moreover, this thesis narrowed the effects of Conscientiousness from a dimension level to a facet level. Namely, this thesis identified that behaviours related to the facet of responsibility were significantly associated with children's educational attainment, which was in line with previous studies (Martel et al., 1987; Zhou, 2015).

Since my study was, at least to my knowledge, the first study using the BIC scale to test the association of Conscientiousness and participants' educational attainment, further replications using the same scale, in different countries, and with participants in different ages are needed. In addition, compared to those unrealistic, blank-slate-philosophy-based mindset manipulations and those effective, but extremely expensive and long-term manipulations such as the Perry Preschool Programme (Schweinhart & Weikart, 2016), it seems that developing manipulations that could change children's personalities to higher levels of conscientiousness and agreeableness would be better approaches for researchers seeking to improve the educational outcomes of children.

9.3 Conclusion

To sum up, in a small study, I could not replicate the effect of playing number board games on children's mathematical knowledge at age 3. In three much larger studies, I failed to replicate that having a growth mindset manipulation could enhance children's cognitive performance after challenges or their educational attainment. In two substantial studies, I could not replicate that holding a growth mindset could increase children's educational attainment either across a challenging transition or among disadvantaged children when entering university. In a series of five studies, I found that at least in my samples, Chinese participants did not have a gender stereotype about brilliance towards men at age 10, but British participants did. Finally, across four substantial studies, I found that Conscientiousness strongly predicted my Chinese participants' educational attainment in both primary school and high school. The studies presented in the current thesis extend previous works to a broad global context, indicate that some approaches educators are using might be incorrect, and provide a suggestion that educators need to emphasise on the importance of personality traits to help children to learn.

Appendices

Appendix A – Gender stereotypes about brilliance, niceness and slow-mindedness stories

Story 1: Brilliance

There are lots of people at the place where I work. But there is one person who is really special. This person is really, really smart. This person figures out how to do things quickly and comes up with answers much faster and better than anyone else. This person is really, really smart.

Story 2: Niceness

There are lots of people at the place where I work. But there is one person who is really special. This person is really, really nice. This person likes to help others with their problems and is friendly to everyone at the office. This person is really, really nice.

Story 3: Slow-mindedness

There are lots of people at the place where I work. But there is one person who is really special. This person is really, really slow-minded. This person figures out how to do things slowly and comes up with answers much slower and worse than anyone else. This person is really, really slow-minded.

Appendix B - Mathematics teacher's conscientiousness scale

1. My mathematics teacher kept his/her belongings neat and clean.
2. My mathematics teacher was pretty good about pacing themselves so as to get things done on time.
3. My mathematics teacher often came into situations without being fully prepared.
4. My mathematics teacher performed all the tasks assigned to them conscientiously.
5. My mathematics teacher had a clear set of goals and worked toward them in an orderly fashion.
6. My mathematics teacher wasted a lot of time before settling down to work.
7. My mathematics teacher worked hard to accomplish his/her goals.
8. When my mathematics teacher made a commitment, I could count on them to follow through.
9. Sometimes my mathematics teacher was not as dependable or reliable as they should be.
10. My mathematics teacher was a productive person who always got the job done.
11. My mathematics teacher never seemed to be able to get organised.
12. My mathematics teacher strived for excellence in everything they did.

Appendix C - English teacher's conscientiousness scale

1. My English teacher kept his/her belongings neat and clean.
2. My English teacher was pretty good about pacing themselves so as to get things done on time.
3. My English teacher often came into situations without being fully prepared.
4. My English teacher performed all the tasks assigned to them conscientiously.
5. My English teacher had a clear set of goals and works toward them in an orderly fashion.
6. My English teacher wasted a lot of time before settling down to work.
7. My English teacher worked hard to accomplish his/her goals.
8. When my English teacher made a commitment, I could count on them to follow through.
9. Sometimes my English teacher was not as dependable or reliable as they should be.
10. My English teacher was a productive person who always got the job done.
11. My English teacher never seemed to be able to get organised.
12. My English teacher strived for excellence in everything they did.

Appendix D - Mathematics classroom CHAOS

1. We had a regular routine in our mathematics class.
2. In my mathematics class, teacher could talk to us without being interrupted.
3. It was a real zoo in our mathematics class.
4. There was often a fuss going on in our mathematics class.
5. You couldn't hear yourself think in our mathematics class.
6. The atmosphere in our mathematics class was calm.

Appendix E - English classroom CHAOS

1. We had a regular routine in our English class.
2. In my English class, teacher could talk to us without being interrupted.
3. It was a real zoo in our English class.
4. There was often a fuss going on in our English class.
5. You couldn't hear yourself think in our English class.
6. The atmosphere in our English class was calm.

References

Abd-El-Fattah, S. M., & Yates, G. C. R. (2006). *Implicit theory of intelligence scale: Testing for factorial invariance and mean structure*. Paper presented at the Australian Association for Research in Education Conference, Adelaide, South Australia.

Alkhateeb, H. M. (2001). Gender differences in mathematics achievement among high school students in the United Arab Emirates, 1991-2000. *School Science and Mathematics, 101*(1), 5-9. doi:10.1111/j.1949-8594.2001.tb18184.x

Allport, G. W. (1937). *Personality: A psychological interpretation*. London: Constable.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs, 47*(1).

Andersen, S. C., & Nielsen, H. S. (2016). Reading intervention with a growth mindset approach improves children's skills. *Proceedings of the Natlonal Academy of Sciences of the United States of America, 113*(43), 12111-12113. doi:10.1073/pnas.1607946113

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*(2), 113-125. doi:10.1006/jesp.2001.1491

Asbury, K., Almeida, D., Hibel, J., Harlaar, N., & Plomin, R. (2008). Clones in the classroom: A daily diary study of the nonshared environmental relationship between monozygotic twin differences in school experience and achievement. *Twin Research and Human Genetics, 11*(6), 586-595. doi:10.1375/twin.11.6.586

Atwood, J. R. (2010). *Mindset, motivation and metaphor in school and sport: Bifurcated beliefs and behavior in two different achievement domains*. Paper presented at the The Annual Meeting of the American Educational Research Association, Denver.

Aunio, P., Aubrey, C., Godfrey, R., Pan, Y., & Liu, Y. (2008). Children’s early numeracy in England, Finland and People’s Republic of China. *International Journal of Early Years Education, 16*(3), 203-221. doi:10.1080/09669760802343881

Bahník, Š., & Vranka, M. A. (2017). Growth mindset is not associated with scholastic aptitude in a large sample of university applicants. *Personality and Individual Differences, 117*, 139-143. doi:10.1016/j.paid.2017.05.046

Barbaranelli, C., Caprara, G. V., Rabasca, A., & Pastorelli, C. (2003). A questionnaire for measuring the Big Five in late childhood. *Personality and Individual Differences, 34*(4), 645-664. doi:10.1016/s0191-8869(02)00051-x

Bates, T. C. (2018). umx: A package for SEM in R (Version 1.9.0) [R package]. Edinburgh, UK: github.com.

Bates, T. C., Hansell, N. K., Martin, N. G., & Wright, M. J. (2016). When does socioeconomic status (SES) moderate the heritability of IQ? No evidence for g × SES interaction for IQ in a representative sample of 1176 Australian adolescent twin pairs. *Intelligence, 56*, 10-15. doi:10.1016/j.intell.2016.02.003

Bates, T. C., Maes, H., & Neale, M. C. (2019). umx: Twin and path-based structural equation modeling in R. *Twin Research and Human Genetics, 22*(1), 27-41. doi:10.1017/thg.2019.2

Batool, S. (2012). Gender differences in the academic achievement of mainstream and religious school students. *Journal of Educational Sciences and Psychology, 2*(1), 76-80.

Bazelais, P., Lemay, D. J., Doleck, T., Hu, X., Vu, A., & Yao, J. (2018). Grit, mindset, and academic performance: A study of pre-university science students. *EURASIA Journal of Mathematics, Science and Technology Education, 14*(12). doi:10.29333/ejmste/94570

Bedard, K., & Cho, I. (2010). Early gender test score gaps across OECD countries. *Economics of Education Review, 29*(3), 348-363. doi:10.1016/j.econedurev.2009.10.015

Bedford, S. (2017). Growth mindset and motivation: A study into secondary school science learning. *Research Papers in Education, 32*(4), 424-443. doi:https://doi-org.ezproxy.is.ed.ac.uk/10.1080/02671522.2017.1318809

Benbow, C. P. (2010). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences, 11*(02). doi:10.1017/s0140525x00049244

Benet-Martínez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75*(3), 729-750. doi:10.1037/0022-3514.75.3.729

Berkeley Personality Lab. (2019). Retrieved from https://www.ocf.berkeley.edu/~johnlab/bfi.htm

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science, 355*(6323), 389-391. doi:10.1126/science.aah6524

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*(1), 246-263.

Bock, L. (2015). *Work rules!: Insights from inside Google that will transform how you live and lead*. New York: Twelve.

Bostwick, K. C., & Becker-Blease, K. A. (2018). Quick, easy mindset intervention can boost academic achievement in large introductory psychology classes. *Psychology Learning & Teaching, 17*(2), 177-193. doi:https://doi.org/10.1177/1475725718766426

Bouchard, G., Lussier, Y., & Sabourin, S. (1999). Personality and marital adjustment: Utility of the Five-Factor Model of personality. *Journal of Marriage and the Family, 61*(3). doi:10.2307/353567

Bowen, C., Wu, Y., Hwang, C., & Scherer, R. F. (2007). Holding up half of the sky? Attitudes toward women as managers in the People's Republic of China. *The International Journal of Human Resource Management, 18*(2), 268-283. doi:10.1080/09585190601102455

Bratko, D., Chamorro-Premuzic, T., & Saks, Z. (2006). Personality and school performance: Incremental validity of self- and peer-ratings over intelligence.

Personality and Individual Differences, 41(1), 131-142. doi:10.1016/j.paid.2005.12.015

Briggs, A. R. J., Clark, J., & Hall, I. (2012). Building bridges: Understanding student transition to university. *Quality in Higher Education, 18*(1), 3-21. doi:10.1080/13538322.2011.614468

Brigham, J. C. (1971). Ethnic stereotypes. *Psychological Bulletin, 76*(1), 15-38. doi:10.1037/h0031446

Brody, N., & Ehrlichman, H. (1998). *Personality psychology: The science of individuality*. London: Prentice Hall International.

Burgoyne, A. P., Hambrick, D. Z., & Macnamara, B. N. (2020). How firm are the foundations of mind-set theory? The claims appear stronger than the evidence. *Psychological Science, 31*(3), 258-267. doi:10.1177/0956797619897588

Burgoyne, A. P., Hambrick, D. Z., Moser, J. S., & Burt, S. A. (2018). Analysis of a mindset intervention. *Journal of Research in Personality, 77*, 21-30. doi:10.1016/j.jrp.2018.09.004

Burnette, J. L., Pollack, J. M., & Hoyt, C. L. (2010). Individual differences in implicit theories of leadership ability and self-efficacy: Predicting responses to stereotype threat. *Journal of Leadership Studies, 3*(4), 46-56. doi:10.1002/jls.20138

Burnette, J. L., Russell, M. V., Hoyt, C. L., Orvidas, K., & Widman, L. (2017). An online growth mindset intervention in a sample of rural adolescent girls. *British Journal of Educational Psychology, 88*(3), 428-445. doi:10.1111/bjep.12192

Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences, 29*(6), 1057-1068. doi:10.1016/s0191-8869(99)00253-6

Buss, A. H. (1989). Personality as traits. *American Psychologist, 44*(11), 1378-1388.

Buss, D. M., & Barnes, M. (1986). Preferences in human mate selection. *Journal of Personality and Social Psychology, 50* 559-570.

Butcher, H. J., Ainsworth, M., & Nesbitt, J. E. (1963). Personality factors and school achievement. *British Journal of Educational Psychology, 33*(3), 276-285. doi:10.1111/j.2044-8279.1963.tb00590.x

Cain, K. M., & Dweck, C. S. (1995). The relation between motivational patterns and achievement cognitions through the elementary school years. *Merrill-Palmer Quarterly, 41*, 25-52.

Caluori, N. (2014). *Mindset and motivation.* (Master). United States Military Academy, New York.

Cao, X., & Chai, Y. (2018). Gender role–based differences in time allocation. *Transportation Research Record: Journal of the Transportation Research Board, 2014*(1), 58-66. doi:10.3141/2014-08

Cattell, R. B. (1945). The description of personality: Principles and findings in a factor analysis. *The American Journal of Psychology, 58*(1), 69-90.

Cattell, R. B. (1948). The primary personality factors in women compared with those in men. *British Journal of Psychology, 1*, 114-130

Cattell, R. B. (1950). *Personality: A systematic theoretical and factual study*. New York: McGraw-Hill.

Cattell, R. B. (1979). *Personality and learning theory. Vol. 1, The structure of personality in its environment*. New York: Springer.
Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16 PF)*. Institute for Personality and Ability Testing.
Chalk, K., & Bizo, L. A. (2004). Specific praise improves on-task behaviour and numeracy enjoyment: A study of year four pupils engaged in the numeracy hour. *Educational Psychology in Practice, 20*(4), 335-351. doi:https://doi.org/10.1080/0266736042000314277
Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality, 37*(4), 319-338. doi:10.1016/s0092-6566(02)00578-0
Chase, M. A. (2010). Should coaches believe in innate ability? The importance of leadership mindset. *Quest, 62*(3), 296-307. doi:https://doi.org/10.1080/00336297.2010.10483650
Chen, Z., DeWall, C. N., Poon, K. T., & Chen, E. W. (2012). When destiny hurts: Implicit theories of relationships moderate aggressive responses to ostracism. *Journal of Experimental Social Psychology, 48*(5), 1029-1036. doi:10.1016/j.jesp.2012.04.002
Cheung, F. M. (1996). Gender role development. In S. Lau (Ed.), *Growing up the Chinese way: Chinese child and adolescent development*. China: Chinese University Press.
Cheung, F. M., & Leung, K. (2016). Indigenous personality measures. *Journal of Cross-Cultural Psychology, 29*(1), 233-248. doi:10.1177/0022022198291012
Cheung, S. K., & McBride, C. (2016). Effectiveness of parent–child number board game playing in promoting Chinese kindergarteners' numeracy skills and mathematics interest. *Early Education and Development, 28*(5), 572-589. doi:10.1080/10409289.2016.1258932
Church, A. T., Reyes, J. A., Katigbak, M. S., & Grimm, S. D. (1997). Filipino personality structure and the big five model: A lexical approach. *Journal of Personality, 65*(3), 477-528. doi:10.1111/j.1467-6494.1997.tb00325.x
Claro, S., Paunesku, D., & Dweck, C. S. (2016). Growth mindset tempers the effects of poverty on academic achievement. *Proceedings of National Academy of Sciences of the United States of America, 113*(31), 8664-8668. doi:10.1073/pnas.1608207113
Clifton, R. A., Perry, R. P., Roberts, L. W., & Peter, T. (2008). Gender, psychosocial dispositions, and the academic achievement of college students. *Research in Higher Education, 49*(8), 684-703. doi:10.1007/s11162-008-9104-9
Cook, T. D., Campbell, D. T., & Day, A. (1979). *Quasi-experimentation: Design & analysis issues for field settings* (Vol. 351). Boston: Houghton Mifflin.
Corradi, D., Nicolaï, J., & Levrau, F. (2018). Growth mindset and its predictive validity—do migration background and academic validation matter? *Higher Education*, 1-14. doi:10.1007/s10734-018-0286-6
Costa, P. T., Jr., & McCrae, R. R. (1976). Age differences in personality structure: A cluster analytic approach. *Journal of Gerontology, 31*(5), 564-570. doi:10.1093/geronj/31.5.564

Costa, P. T., Jr., & McCrae, R. R. (1978). Objective personality assessment. In M. Storandt, I. C. Siegler, & M. F. Elias (Eds.), *The clinical psychology of aging*. Boston: Springer.

Costa, P. T., Jr., & McCrae, R. R. (1980). Still stable after all these years: Personality as a key to some issues in adulthood and old age. In P. B. Baltes & O. G. Brim (Eds.), *Life span development and behavior* (Vol. 3, pp. 65-102).

Costa, P. T., Jr., & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Odessa: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment, 64*(1), 21-50. doi:10.1207/s15327752jpa6401_2

Costa, P. T., Jr., & McCrae, R. R. (2008). The revised NEO Personality Inventory (NEO-PI-R). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment: Volume 2 — Personality measurement and testing*. London: SAGE Publications Ltd.

Costa, P. T., Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81*(2), 322-331. doi:10.1037/0022-3514.81.2.322

Cui, H., & Wang, D. (2004). The development and applicability of the" Big Five" personality model. *Psychological Science-Shanghai, 27*(3), 545-548

Cury, F., Da Fonseca, D., Zahn, I., & Elliot, A. (2008). Implicit theories and IQ test performance: A sequential mediational analysis. *Journal of Experimental Social Psychology, 44*(3), 783-791. doi:10.1016/j.jesp.2007.07.003

Cury, F., Elliot, A. J., Da Fonseca, D., & Moller, A. C. (2006). The social-cognitive model of achievement motivation and the 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology, 90*(4), 666-679. doi:10.1037/0022-3514.90.4.666

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2011). Measuring implicit attitudes of 4-year-olds: The preschool implicit association test. *Journal of Experimental Child Psychology, 109*(2), 187-200. doi:10.1016/j.jecp.2010.11.002

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2016). Implicit measures for preschool children confirm self-esteem's role in maintaining a balanced identity. *Journal of Experimental Social Psychology, 62*, 50-57. doi:10.1016/j.jesp.2015.09.015

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-gender stereotypes in elementary school children. *Child Development, 82*(3), 766-779. doi:10.1111/j.1467-8624.2010.01529.x

De Castella, K., & Byrne, D. (2015). My intelligence may be more malleable than yours: The revised implicit theories of intelligence (self-theory) scale is a better predictor of achievement, motivation, and student disengagement. *European Journal of Psychology of Education, 30*(3), 245-267. doi:10.1007/s10212-015-0244-y

De Fruyt, F., & Mervielde, I. (1996). Personality and interests as predictors of educational streaming and achievement. *European Journal of Personality, 10*(5), 405-425. doi:10.1002/(sici)1099-0984(199612)10:5<405::Aid-per255>3.0.Co;2-m

De Fruyt, F., Mervielde, I., Hoekstra, H. A., & Rolland, J. P. (2000). Assessing adolescents' personality with the NEO PI-R. *Assessment, 7*(4), 329-345. doi:10.1177/107319110000700403

Deary, I. J. (2012). Intelligence. *Annual Review of Psychology, 63*, 453-482. doi:10.1146/annurev-psych-120710-100353

Deary, I. J., Irwing, P., Der, G., & Bates, T. C. (2007). Brother-sister differences in the g factor in intelligence: Analysis of full, opposite-sex siblings from the NLSY 1979. *Intelligence, 35*(5), 451-456. Retrieved from <Go to ISI>://000249214800004

Deary, I. J., Spinath, F. M., & Bates, T. C. (2006). Genetics of intelligence. *European Journal of Human Genetics, 14*(6), 690-700. doi:10.1038/sj.ejhg.5201588

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13-21. doi:10.1016/j.intell.2006.02.001

DeBacker, T. K., Heddy, B. C., Kershen, J. L., Crowson, H. M., Looney, K., & Goldman, J. A. (2018). Effects of a one-shot growth mindset intervention on beliefs about intelligence and achievement goals. *Educational Psychology, 38*(6), 711-733. doi:10.1080/01443410.2018.1426833

Del Giudice, M., Booth, T., & Irwing, P. (2012). The distance between Mars and Venus: Measuring global sex differences in personality. *PLoS One, 7*(1), e29265. doi:10.1371/journal.pone.0029265

Den Brok, P., Brekelmans, M., & Wubbels, T. (2010). Interpersonal teacher behaviour and student outcomes. *School Effectiveness and School Improvement, 15*(3-4), 407-442. doi:10.1080/09243450512331383262

Digman, J. M., & Inouye, J. (1986). Further specification of the five robust factors of personality. *Journal of Personality and Social Psychology, 50*(1), 116-123

Ding, C. S., Song, K. I. M., & Richardson, L. I. (2006). Do mathematical gender differences continue? A longitudinal study of gender difference and excellence in mathematics performance in the U.S. *Educational Studies, 40*(3), 279-295. doi:10.1080/00131940701301952

Donohoe, C., Topping, K., & Hannah, E. (2012). The impact of an online intervention (Brainology) on the mindset and resiliency of secondary school pupils: A preliminary mixed methods study. *Educational Psychology, 32*(5), 641-655. doi:https://doi.org/10.1080/01443410.2012.675646

Dotterer, A. M., & Lowe, K. (2011). Classroom context, school engagement, and academic achievement in early adolescence. *Journal of Youth and Adolescence, 40*(12), 1649-1660. doi:10.1007/s10964-011-9647-5

Downer, J. T., Rimm-Kaufman, S. E., & Pianta, R. C. (2007). Engagement in learning. *School Psychology Review, 36*(3), 413-432

Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS One, 7*(1), e29081. doi:10.1371/journal.pone.0029081

Doyle, R. A., & Voyer, D. (2016). Stereotype manipulation effects on math and spatial test performance: A meta-analysis. *Learning and Individual Differences, 47*, 103-116. doi:10.1016/j.lindif.2015.12.018

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*(6), 1087-1101. doi:10.1037/0022-3514.92.6.1087

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short grit scale (grit-s). *Journal of Personality Assessment, 91*(2), 166-174. doi:10.1080/00223890802634290

Duckworth, A. L., & Seligman, M. E. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology, 98*(1), 198-208.

Duff, A., Boyle, E., Dunleavy, K., & Ferguson, J. (2004). The relationship between personality, approach to learning and academic performance. *Personality and Individual Differences, 36*(8), 1907-1920. doi:10.1016/j.paid.2003.08.020

Dumfart, B., & Neubauer, A. C. (2016). Conscientiousness is the most powerful noncognitive predictor of school achievement in adolescents. *Journal of Individual Differences, 37*(1), 8-15. doi:10.1027/1614-0001/a000182

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428-1446.

Dunham, Y., Baron, A. S., & Banaji, M. R. (2016). The development of implicit gender attitudes. *Developmental Science, 19*(5), 781-789. doi:10.1111/desc.12321

Dupeyrat, C., & Mariné, C. (2005). Implicit theories of intelligence, goal orientation, cognitive engagement, and achievement: A test of Dweck's model with returning to school adults. *Contemporary Educational Psychology, 30*(1), 43-59. doi:https://doi.org/10.1016/j.cedpsych.2004.01.007

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040-1048.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*: Psychology Press.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York: Ballantine Books.

Dweck, C. S. (2007a). The perils and promises of praise. *Educational Leadership, 65*(2), 34-39.

Dweck, C. S. (2007b). The secret to raising smart kids. *Scientific American Mind, 18*(6), 36-43.

Dweck, C. S. (2008a). Brainology. Retrieved from https://www.nais.org/magazine/independent-school/winter-2008/brainology/

Dweck, C. S. (2008b). Mindsets: How praise is harming youth and what can be done about it. *School Library Media Activities Monthly, XXIV*(5), 55-58.

Dweck, C. S. (2010). Even geniuses work hard. *Educational Leadership, 68*(1), 16-20.

Dweck, C. S. (2012). Mindsets and human nature: Promoting change in the Middle East, the schoolyard, the racial divide, and willpower. *American Psychologist, 67*(8), 614-622. doi:10.1037/a0029783

Dweck, C. S. (2013). QnAs with Carol S. Dweck. Interview by Sujata Gupta. *Proceedings of the National Academy of Science USA, 110*(37), 14818. doi:10.1073/pnas.1313540110

Dweck, C. S. (2014). Mindsets and math/science achievement.

Dweck, C. S., & Bempechat, J. (1983). Children's theories of intelligence: Consequences for learning. In S. G. Paris, G. M. O., & H. W. S. (Eds.), *Learning and motivation in the classroom* (pp. 239-256). New Jersey: Lawrence Erlbaum Associates.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256-273.

Dweck, C. S., & Yeager, D. S. (2019). Mindsets: A view from two eras. *Perspectives on Psychological Science, 14*(3), 481-496. doi:10.1177/1745691618804166

Ebenstein, A. (2010). The "missing girls" of China and the unintended consequences of the one child policy. *Journal of Human Resources, 45*(1), 87-115. doi:10.3368/jhr.45.1.87

Egan, V., Deary, I. J., & Austin, E. (2000). The NEO-FFI: Emerging British norms and an item-level analysis suggest N, A and C are more reliable than O and E. *Personality and Individual Differences, 29*(5), 907-920. doi:10.1016/s0191-8869(99)00242-1

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54*(1), 5-12. doi:10.1037/0022-3514.54.1.5

Ellis, L. (2011). Identifying and explaining apparent universal sex differences in cognition and behavior. *Personality and Individual Differences, 51*(5), 552-561. doi:10.1016/j.paid.2011.04.004

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103-127. doi:10.1037/a0018053

Eysenck, H. J. (1953). *The structure of human personality*. London: Methuen.

Eysenck, H. J. (1990). *Biological dimensions of personality*. New York: The Guilford Press.

Eysenck, H. J. (1992). Four ways five factors are not basic. *Personality and Individual Differences, 13*(6), 667-673. doi:10.1016/0191-8869(92)90237-j

Fagot, B. I., Leinbach, M. D., & O'Boyle, C. (1992). Gender labeling, gender stereotyping, and parenting behaviors. *Developmental Psychology, 28*(2), 225-230. doi:10.1037/0012-1649.28.2.225

Fergusson, D. M., & Horwood, L. J. (1997). Gender differences in educational achievement in a New Zealand birth cohort. *New Zealand Journal of Educational Studies, 32*(1), 83-96.

Findley, M. J., & Cooper, H. M. (1983). Locus of control and academic achievement: A literature review. *Journal of Personality and Social Psychology, 44*(2), 419-427. doi:10.1037/0022-3514.44.2.419

Finn, A. S., Kraft, M. A., West, M. R., Leonard, J. A., Bish, C. E., Martin, R. E., . . . Gabrieli, J. D. (2014). Cognitive skills, student achievement tests, and schools. *Psychological Science, 25*(3), 736-744. doi:10.1177/0956797613516008

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology, 44*(3), 329-344.

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology, 53*(1), 25-44. doi:10.1016/j.jsp.2014.10.002

Furnham, A., Chamorro-Premuzic, T., & McDougall, F. (2003). Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance. *Learning and Individual Differences, 14*(1), 47-64. doi:10.1016/j.lindif.2003.08.002

Furnham, A., Reeves, E., & Budhani, S. (2002). Parents think their sons are brighter than their daughters: Sex differences in parental self-estimations and estimations of their children's multiple intelligences. *The Journal of Genetic Psychology, 163*(1), 24-39. doi:10.1080/00221320209597966

Gates, W. (2015). Mindset over matter: What you believe affects what you achieve. Retrieved from https://www.gatesnotes.com/Books/Mindset-The-New-Psychology-of-Success

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin, 114*(2), 345-362. doi:10.1037/0033-2909.114.2.345

Geary, D. C., Fan, L., & Bow-Thomas, C. C. (1992). Numerical cognition, loci of ability differences comparing children from China and the US. *Psychological Science, 3*(3), 180-185.

Gignac, G. E. (2018). A moderate financial incentive can increase effort, but not intelligence test performance in adult volunteers. *British Journal of Psychology, 109*(3), 500-516. doi:10.1111/bjop.12288

Goldberg, L. R. (1990). An alternative" description of personality": The big-five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216 - 1228.

Goldberg, L. R. (1993). The structure of personality traits: Vertical and horizontal aspects. In D. C. Funder, R. Parke, C. Tomlinson-Keasey, & K. Widaman (Eds.), *Studying lives through time: Personality and development*. Washington, DC: American Psychological Association.

Goldhaber, D. (2002). The mystery of good teaching. *Education Next, 2*(1), 50-55.

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. doi:https://doi.org/10.1016/j.appdev.2003.09.002

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology, 102*(4), 700-717. doi:10.1037/a0026659

Gunderson, E. A., Gripshover, S. J., Romero, C., Dweck, C. S., Goldin-Meadow, S., & Levine, S. C. (2013). Parent praise to 1- to 3-year-olds predicts children's motivational frameworks 5 years later. *Child Development, 84*(5), 1526-1541. doi:10.1111/cdev.12064

Gunderson, E. A., Sorhagen, N. S., Gripshover, S. J., Dweck, C. S., Goldin-Meadow, S., & Levine, S. C. (2018). Parent praise to toddlers predicts fourth grade academic achievement via children's incremental mindsets. *Developmental Psychology, 54*(3), 397-409. doi:10.1037/dev0000444

Haimovitz, K., & Dweck, C. S. (2016). What predicts children's fixed and growth intelligence mind-sets? Not their parents' views of intelligence but their parents' views of failure. *Psychological Science, 27*(6), 859-869. doi:10.1177/0956797616639727

Haimovitz, K., Wormington, S. V., & Corpus, J. H. (2011). Dangerous mindsets: How beliefs about intelligence predict motivational change. *Learning and*

Individual Differences, 21(6), 747-752. doi:https://doi.org/10.1016/j.lindif.2011.09.002

Hannum, E., Kong, P., & Zhang, Y. (2009). Family sources of educational gender inequality in rural china: A critical assessment. *International Journal of Educational Development, 29*(5), 474-486. doi:10.1016/j.ijedudev.2009.04.007

Hashim, N. M. H. N., Alam, S. S., & Yusoff, N. M. (2014). Relationship between teacher's personality, monitoring, learning environment, and students' EFL performance. *GEMA Online® Journal of Language Studies, 14*(1), 101-116

Hau, K., & Salili, F. (1996). Motivational effects of teachers' ability versus effort feedback on Chinese students' learning. *Social Psychology of Education, 1*(1), 69-85.

Heine, S. J., Kitayama, S., Lehman, D. R., Takata, T., Ide, E., Leung, C., & Matsumoto, H. (2001). Divergent consequences of success and failure in Japan and North America: An investigation of self-improving motivations and malleable selves. *Journal of Personality and Social Psychology, 81*(4), 599-615. doi:10.1037//0022-3514.81.4.599

HESA. (2019). Non-continuation: UK performance indicators 2017/18. Retrieved from https://www.hesa.ac.uk/news/07-03-2019/non-continuation-tables

Hogan, R., Johnson, J., & Briggs, S. (1997). *Handbook of personality psychology*. USA: Academic Press.

Hogan, R., & Sherman, R. A. (2020). Personality theory and the nature of human nature. *Personality and Individual Differences, 152*. doi:10.1016/j.paid.2019.109561

Hojnoski, R. L., Caskie, G. I. L., & Miller Young, R. (2017). Early numeracy trajectories: Baseline performance levels and growth rates in young children by disability status. *Topics in Early Childhood Special Education, 37*(4), 206-218. doi:10.1177/0271121417735901

Hong, Y., Chiu, C., & Dweck, C. S. (1995). Implicit theories of intelligence: Reconsidering the role of confidence in achievement motivation. In M. H. Kernis (Ed.), *Efficacy, agency, and self-esteem* (pp. 197-216). New York: Plenum Press.

Hong, Y., Chiu, C., Dweck, C. S., Lin, D. M.-S., & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology, 77*(3), 588-599.

Hong, Y., Chiu, C., Dweck, C. S., & Sacks, R. (1997). Implicit theories and evaluative processes in person cognition. *Journal of Experimental Social Psychology, 33*(3), 296-323. doi:10.1006/jesp.1996.1324

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85*(6), 869-879. doi:10.1037/0021-9010.85.6.869

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*(6), 581-592. doi:10.1037/0003-066X.60.6.581

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science, 321*(5888), 494-495. doi:10.1126/science.1160364

Ivcevic, Z., & Brackett, M. (2014). Predicting school success: Comparing conscientiousness, grit, and emotion regulation ability. *Journal of Research in Personality, 52*, 29-36. doi:10.1016/j.jrp.2014.06.005

Jackson, J. J., Wood, D., Bogg, T., Walton, K. E., Harms, P. D., & Roberts, B. W. (2010). What do conscientious people do? Development and validation of the Behavioral Indicators of Conscientiousness (BIC). *Journal of Research in Personality, 44*(4), 501-511. doi:10.1016/j.jrp.2010.06.005

Job, V., Walton, G. M., Bernecker, K., & Dweck, C. S. (2013). Beliefs about willpower determine the impact of glucose on self-control. *Proceedings of the National Academy of Science USA, 110*(37), 14837-14842. doi:10.1073/pnas.1313475110

John, O. P. (1989). *Towards a taxonomy of personality descriptors*. New York, NY: Springer.

John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality, 2*(3), 171-203. doi:10.1002/per.2410020302

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory - versions 4a and 54. Retrieved from https://www.ocf.berkeley.edu/~johnlab/bfi.htm

John, O. P., Hampson, S. E., & Goldberg, L. R. (1991). The basic level in personality-trait hierarchies: Studies of trait use and accessibility in different contexts. *Journal of Personality and Social Psychology, 60*(3), 348-361.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big-Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins , & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114-158). New York, NY: Guilford Press.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of Personality: Theory and Research* (Vol. 2). New York: Guilford Press.

Johnston, I. (2017). Creating a growth mindset. *Strategic HR Review, 16*(4), 155-160. doi:10.1108/SHR-04-2017-0022

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology, 45*(3), 850-867. doi:10.1037/a0014939

Jussim, L., Crawford, J. T., Anglin, S. M., Chambers, J. R., Stevens, S. T., & Cohen, F. (2016). Stereotype accuracy: One of the largest and most replicable effects in all of social psychology. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 31-63). U.K.: Psychology Press.

Jussim, L., Harber, K. D., Crawford, J. T., Cain, T. R., & Cohen, F. (2005). Social reality makes the social mind: Self-fulfilling prophecy, stereotypes, bias, and accuracy. *Interaction Studies, 6*(1), 85-102. doi:10.1075/is.6.1.07jus

Jussim, L., Nelson, T. E., Manis, M., & Soffin, S. (1995). Prejudice, stereotypes, and labeling effects: Sources of bias in person perception. *Journal of Personality and Social Psychology, 68*(2), 228-246. doi:10.1037/0022-3514.68.2.228

Kamins, M. L., & Dweck, C. S. (1999). Person versus process praise and criticism: Implications for contingent self-worth and coping. *Developmental Psychology, 35*(3), 835-847.

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *The Journal of Abnormal and Social Psychology, 28*(3), 280-290. doi:10.1037/h0074049

Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M., & Patrick, H. (2006). Sex differences in math performance: The role of children's approach to schoolwork. *Developmental Psychology, 42*(1), 11-26. doi:10.1037/0012-1649.42.1.11

Khwaileh, F. M., & Zaza, H. I. (2011). Gender differences in academic performance among undergraduates at the University of Jordan: Are they real or stereotyping. *College Student Journal, 45*(3), 633-648.

Kim, L. E., Dar-Nimrod, I., & MacCann, C. (2018). Teacher personality and teacher effectiveness in secondary school: Personality predicts teacher support and student self-efficacy but not academic achievement. *Journal of Educational Psychology, 110*(3), 309-323. doi:10.1037/edu0000217

Kim, L. E., Jorg, V., & Klassen, R. M. (2019). A meta-analysis of the effects of teacher personality on teacher effectiveness and burnout. *Educational Psychology Review, 31*(1), 163-195. doi:10.1007/s10648-018-9458-2

Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M., & Hedges, L. V. (2006). Preschool children's mathematical knowledge: The effect of teacher "math talk.". *Developmental Psychology, 42*(1), 59-69. doi:10.1037/0012-1649.42.1.59

Knee, C. R., & Petty, K. N. (2013). Implicit theories of relationships: Destiny and growth beliefs. In J. A. Simpson & L. Campbell (Eds.), *The Oxford handbook of close elationships* (pp. 183-198). New York: Oxford University Press.

Komarraju, M., Karau, S. J., & Schmeck, R. R. (2009). Role of the Big Five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences, 19*(1), 47-52. doi:10.1016/j.lindif.2008.07.001

Koo, R. D., Kam, M. C. K., & Choi, B. C. (2003). Education and schooling in Hong Kong under one country, two systems. *Childhood Education, 79*(3), 137-144. doi:10.1080/00094056.2003.10522213

Kuncel, N. R., Credé, M., & Thomas, L. L. (2016). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75*(1), 63-82. doi:10.3102/00346543075001063

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*(1), 148-161. doi:10.1037/0022-3514.86.1.148

Lachance, J. A., & Mazzocco, M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learning and Individual Differernces, 16*(3), 195-216. doi:10.1016/j.lindif.2005.12.001

Lai, F. (2010). Are boys left behind? The evolution of the gender achievement gap in Beijing's middle schools. *Economics of Education Review, 29*(3), 383-399. doi:10.1016/j.econedurev.2009.07.009

Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences, 42*(3), 441-451. doi:10.1016/j.paid.2006.08.001

Lam, S. F., Yim, P. S., & Ng, Y. L. (2008). Is effort praise motivational? The role of beliefs in the effort–ability relationship. *Contemporary Educational Psychology, 33*(4), 694-710. doi:https://doi.org/10.1016/j.cedpsych.2008.01.005

Laski, E. V., & Siegler, R. S. (2014). Learning from number board games: You learn what you encode. *Developmental Psychology, 50*(3), 853-864. doi:10.1037/a0034321

Lauzen, M. M., Dozier, D. M., & Horan, N. (2008). Constructing gender stereotypes through social roles in prime-time television. *Journal of Broadcasting & Electronic Media, 52*(2), 200-214. doi:10.1080/08838150801991971

Lazarus, R. S., & Opton, E., M. Jr. (1967). *Personality*. England: Penguin Books.

Leary, M. R., & Baumeister, R. F. (2000). The nature and function of self-esteem: Sociometer theory. *Advances in Experimental Social Psychology, 32*, 1-62. doi:10.1016/S0065-2601(00)80003-9

Lee, W. Y. (1995). Women's education in traditional and modern China. *Women's History Review, 4*(3), 345-367.

Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science, 347*(6219), 262-265. doi:10.1126/science.1261375

Li, C. (2000). Confucianism and feminist concerns: Overcoming the confucian "gender complex". *Journal of Chinese Philosophy, 27*(2), 187-199. doi:10.1111/0301-8121.00012

Li, J. (2002). A cultural model of learning: Chinese “heart and mind for wanting to learn”. *Journal of Cross-Cultural Psychology, 33*(3), 248-269. doi:10.1177/0022022102033003003

Li, P., Zhou, N., Zhang, Y., Xiong, Q., Nie, R., & Fang, X. (2017). Incremental theory of intelligence moderated the relationship between prior achievement and school engagement in Chinese high school students. *Frontiers in Psychology, 8*, 1703. doi:10.3389/fpsyg.2017.01703

Li, Y., & Bates, T. C. (2019). You can't change your basic ability, but you work at things, and that's how we get hard things done: Testing the role of growth mindset on response to setbacks, educational attainment, and cognitive ability. *Journal of Experimental Psychology: General, 148*(9), 1640-1655. doi:10.1037/xge0000669

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*(6), 1123-1135. doi:10.1037/a0021276

Lindholm, K. (2017). Handling stereotypes of religious professionals: Strategies hospice chaplains use when interacting with patients and families. *Journal of Pastoral Care & Counseling, 71*(4), 284-290. doi:10.1177/1542305017744493

Lindsey, R. V., Shroyer, J. D., Pashler, H., & Mozer, M. C. (2014). Improving students' long-term knowledge retention through personalized review. *Psychological Science, 25*(3), 639-647. doi:10.1177/0956797613504302

Lippa, R. A. (2010). Gender differences in personality and interests: When, where, and why? *Social and Personality Psychology Compass, 4*(11), 1098-1110. doi:10.1111/j.1751-9004.2010.00320.x

Lippmann, W. (1922). *Public opinion*. London: Allen & Unwin.

Liu, F. (2006). Boys as only-children and girls as only-children—parental gendered expectations of the only-child in the nuclear Chinese family in present-day China. *Gender and Education, 18*(5), 491-505. doi:10.1080/09540250600881626

Liu, J., & Carpenter, M. (2005). Trends and issues of women's education in China. *The Clearing House, 78*(6), 277-281.

Liu, S. S., Comer, L. B., & Dubinsky, A. J. (2001). Gender differences in attitudes toward women as sales managers in the People's Republic of China. *Journal of Personal Selling & Sales Management, 21*(4), 303-311

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705-717. doi:10.1037//0003-066x.57.9.705

Löckenhoff, C. E., Chan, W., McCrae, R. R., De Fruyt, F., Jussim, L., De Bolle, M., . . . Terracciano, A. (2014). Gender stereotypes of personality. *Journal of Cross-Cultural Psychology, 45*(5), 675-694. doi:10.1177/0022022113520075

Lounsbury, J. W., & Gibson, L. W. (1998). *Personal Style Inventory: A work-based personality measurement system*. Knoxville, TN: Resource Associates.

Lynn, R., & Mikk, J. (2009). Sex differences in reading achievement. *TRAMES: A Journal of the Humanities & Social Sciences, 13*(1), 3-13. doi:10.3176/tr.2009.1.01

Ma, X. (2008). Within-school gender gaps in reading, mathematics, and science literacy. *Comparative Education Review, 52*(3), 437-460. doi:10.1086/588762

Machin, S., & McNally, S. (2005). Gender and student achievement in English schools. *Oxford Review of Economic Policy, 21*(3), 357-372. doi:10.1093/oxrep/gri021

Macnamara, B. N., & Rupani, N. S. (2017). The relationship between intelligence and mindset. *Intelligence, 64*, 52-59. doi:https://doi.org/10.1016/j.intell.2017.07.003

Madon, S., Guyll, M., Aboufadel, K., Montiel, E., Smith, A., Palumbo, P., & Jussim, L. (2001). Ethnic and national stereotypes: The Princeton trilogy revisited and revised *Personality and Social Psychology Bulletin, 27*(8), 996-1010. doi:10.1177/0146167201278007

Martel, J., McKelvie, S. J., & Standing, L. (1987). Validity of an intuitive personality scale: Personal responsibility as a predictor of academic achievement. *Educational and Psychological Measurement, 47*(4), 1153-1163. doi:10.1177/0013164487474033

Matheny, A. P., Wachs, T. D., Ludwig, J. L., & Phillips, K. (1995). Bringing order out of chaos: Psychometric characteristics of the confusion, hubbub, and order scale. *Journal of Applied Developmental Psychology, 16*(3), 429-444. doi:10.1016/0193-3973(95)90028-4

Matthews, G., Deary, I., & Whiteman, M. (2009). *The trait concept and personality theory*. Cambridge: Cambridge University Press.

Matthews, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology, 101*(3), 689-704. doi:10.1037/a0014240

Mau, W., & Lynn, R. (2001). Gender differences on the scholastic aptitude test, the American college test and college grades. *Educational Psychology, 21*(2), 133-136. doi:10.1080/01443410020043832

McAdams, D. P., & Pals, J. L. (2006). A new Big Five: Fundamental principles for an integrative science of personality. *American Psychologist, 61*(3), 204-217. doi:10.1037/0003-066X.61.3.204

McCauley, C., Stitt, C. L., & Segal, M. (1980). Stereotyping: From prejudice to prediction. *Psychological Bulletin, 87*(1), 195-208. doi:10.1037/0033-2909.87.1.195

McCrae, R. R. (1990). Traits and trait names: How well is Openness represented in natural languages? *European Journal of Personality, 4*(2), 119-129. doi:10.1002/per.2410040205

McCrae, R. R. (2002). NEO-PI-R data from 36 cultures. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures*. London: London Kluwer Academic/Plenum.

McCrae, R. R., Costa, P. T., Hřebíčková, M., Urbánek, T., Martin, T. A., Oryol, V. E., . . . Senin, I. G. (2004). Age differences in personality traits across cultures: Self-report and observer perspectives. *European Journal of Personality, 18*(2), 143-157. doi:10.1002/per.510

McCrae, R. R., & Costa, P. T., Jr. (1983). Joint factors in self-reports and ratings: Neuroticism, extraversion and openness to experience. *Personality and Individual Differences, 4*(3), 245-255. doi:10.1016/0191-8869(83)90146-0

McCrae, R. R., & Costa, P. T., Jr. (1985). Updating Norman's" adequacy taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology, 49* (3), 710-721.

McCrae, R. R., & Costa, P. T., Jr. (1990). *Personality in adulthood*. New York: Guilford.

McCrae, R. R., & Costa, P. T., Jr. (2008). *The five-factor theory of personality*: The Guilford Press.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175-215. doi:10.1111/j.1467-6494.1992.tb00970.x

McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review, 15*(1), 28-50. doi:10.1177/1088868310366253

McCrae, R. R., Terracciano, A., & 79 Members of the Personality Profiles of Cultures Project. (2005). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology, 89*(3), 407-425. doi:10.1037/0022-3514.89.3.407

McCrae, R. R., Terracciano, A., & Personality Profiles of Cultures, P. (2005). Universal features of personality traits from the observer's perspective: data from 50 cultures. *Journal of Personality and Social Psychology, 88*(3), 547-561. doi:10.1037/0022-3514.88.3.547

McNabb, R., Pal, S., & Sloane, P. (2002). Gender differences in educational attainment: The case of university students in England and Wales. *Economica, 69*(275), 481-503. doi:10.1111/1468-0335.00295

Meyer, M., Cimpian, A., & Leslie, S. J. (2015). Women are underrepresented in fields where success is believed to require brilliance. *Frontiers in Psychology, 6*, 235. doi:10.3389/fpsyg.2015.00235

Miller, R. L., Brickman, P., & Bolen, D. (1975). Attribution versus persuasion as a means for modifying behavior. *Journal of Personality and Social Psychology, 31*(3), 430-441. doi:10.1037/h0076539

Mojavezi, A., & Tamiz, M. P. (2012). The impact of teacher self-efficacy on the students' motivation and chievement. *Theory and Practice in Language Studies, 2*(3). doi:10.4304/tpls.2.3.483-491

Mottus, R. (2017). Five-Factor Theory and personality development. In *Personality development across the lifespan* (pp. 87-100). Specht, Jule: Academic Press.

Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology, 75*(1), 33-52. doi:10.1037/0022-3514.75.1.33

Muldoon, K., Simms, V., Towse, J., Menzies, V., & Yue, G. (2011). Cross-cultural comparisons of 5-year-olds' estimating and mathematical ability. *Journal of Cross-Cultural Psychology, 42*(4), 669-681. doi:10.1177/0022022111406035

Munafo, M. R., Zammit, S., & Flint, J. (2014). Practitioner review: A critical perspective on gene-environment interaction models--what impact should they have on clinical perceptions and practice? *Journal of Child Psychology and Psychiatry, 55*(10), 1092-1101. doi:10.1111/jcpp.12261

Nallapothula, D., Lozano, J. B., Han, S., Herrera, C., Sayson, H. W., Levis-Fitzgerald, M., & Maloy, J. (2020). M-LoCUS: A scalable intervention enhances growth mindset and internal locus of control in undergraduate students in STEM. *Journal of Microbiology & Biology Education, 21*(2). doi:10.1128/jmbe.v21i2.1987

National Bureau of Statistics of the People's Republic of China. (1998). *China statistical year book*. China Statistics Press Retrieved from http://www.stats.gov.cn/yearbook/indexC.htm

National Bureau of Statistics of the People's Republic of China. (2015). *China statistical year book*. China Statistics Press. Retrieved from http://www.stats.gov.cn/tjsj/ndsj/2015/indexeh.htm

National Bureau of Statistics of the People's Republic of China. (2017). *China statistical year book*. Retrieved from http://www.stats.gov.cn/tjsj/ndsj/2016/html/0411EN.jpg

National Mathematics Advisory Panel. (2008). Foundations for success: The final report of the National Mathematics Advisory Panel. U.S.: Washington, DC.

Niederle, M., & Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives, 24*(2), 129-144. doi:10.1257/jep.24.2.129

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology, 93*(1), 116-130. doi:10.1037/0022-3514.93.1.116

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology, 66*(6), 574-583. doi:10.1037/h0040291

Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal, 48*(5), 1125-1156. doi:10.3102/0002831211410683

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 Natlonal Academy of Sciences of the United States of America, 106*(26), 10593-10597. doi:10.1073/pnas.0809921106

Núñez, J. C., Suárez, N., Rosário, P., Vallejo, G., Cerezo, R., & Valle, A. (2014). Teachers' feedback on homework, homework-related behaviors, and academic achievement. *The Journal of Educational Research, 108*(3), 204-216. doi:10.1080/00220671.2013.878298

O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences, 43*(5), 971-990. doi:10.1016/j.paid.2007.03.017

O'Neill, S. A. (2011). Developing a young musician's growth mindset: The role of motivation, self-theories, and resiliency. In I. Deliège & J. Davidson (Eds.), *Music and the mind: Essays in honour of John Sloboda* (pp. 31-46). New York: Oxford University Press.

OECD. (2008). Growing unequal? Income distribution and poverty in OECD countries. Paris, France.

Okimoto, T. G., & Heilman, M. E. (2012). The "bad parent" assumption: How gender stereotypes affect reactions to working mothers. *Journal of Social Issues, 68*(4), 704-724. doi:10.1111/j.1540-4560.2012.01772.x

Ommundsen, Y. (2003). Implicit theories of ability and self-regulation strategies in physical education classes. *Educational Psychology, 23*(2), 141-157. doi:https://doi.org/10.1080/01443410303224

Open Science Collaboration. (2015). Psychology. Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. doi:10.1126/science.aac4716

Ostergren, R., & Traff, U. (2013). Early number knowledge and cognitive ability affect early arithmetic ability. *Journal of Experimental Child Psychology, 115*(3), 405-421. doi:10.1016/j.jecp.2013.03.007

Park, D., Gunderson, E. A., Tsukayama, E., Levine, S. C., & Beilock, S. L. (2016). Young children's motivational frameworks and math achievement: Relation to teacher-reported instructional practices, but not teacher theory of intelligence. *Journal of Educational Psychology, 108*(3), 300-313. doi:10.1037/edu0000064

Paulsen, K., & Johnson, M. (1983). Sex role attitudes and mathematical ability in 4th-, 8th-, and 11th-grade students from a high socioeconomic area. *Developmental Psychology, 19*(2), 210-214. doi:10.1037/0012-1649.19.2.210

Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S., & Dweck, C. S. (2015). Mind-set interventions are a scalable treatment for academic underachievement. *Psychological Science, 26*(6), 784-793. doi:10.1177/0956797615571017

Peabody, D., & Goldberg, L. R. (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology, 57* (3), 552-567.

Pervin, L. A., & John, O. P. (1996). *Personality: Theory and research*. USA: John Wiley & Sons, Inc.

Pianta, R. C., la Paro, K. M., Payne, C., Cox, M. J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *The Elementary School Journal, 102*(3), 225-238. doi:10.1086/499701

Platt, J. R. (1964). Strong inference. *Science, 146*(3642), 347-353.

Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance and internal distress. *Journal of Educational Psychology, 94*(2), 396-404. doi:10.1037/0022-0663.94.2.396

Popper, K. (1963). *Conjectures and refutations: The growth of scientific knowledge*. UK: Routledge.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*(2), 322-338. doi:10.1037/a0014996

R Core Team. (2019). R: A language and environment for statistical computing (Version 3.5.2). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development, 79*(2), 375-394. doi:10.1111/j.1467-8624.2007.01131.x

Ramani, G. B., & Siegler, R. S. (2011). Reducing the gap in numerical knowledge between low- and middle-income preschoolers. *Journal of Applied Developmental Psychology, 32*(3), 146-159. doi:10.1016/j.appdev.2011.02.005

Ramani, G. B., Siegler, R. S., & Hitti, A. (2012). Taking it to the classroom: Number board games as a small group learning activity. *Journal of Educational Psychology, 104*(3), 661-672. doi:10.1037/a0028995

Raphals, L. A. (1998). Chapter six yin and yang. In L. A. Raphals (Ed.), *Sharing the light: Representations of women and virtue in early China* (pp. 139-167). N.Y.: State University of New York Press.

Rattan, A., Good, C., & Dweck, C. S. (2012). "It's ok—Not everyone can be good at math": Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology, 48*(3), 731-737. doi:https://doi.org/10.1016/j.jesp.2011.12.012

Rattan, A., Savani, K., Chugh, D., & Dweck, C. S. (2015). Leveraging mindsets to promote academic achievement: Policy recommendations. *Perspectives on Psychological Science, 10*(6), 721-726. doi:10.1177/1745691615599383

Rattan, A., Savani, K., Naidu, N. V. R., & Dweck, C. S. (2012). Can everyone become highly intelligent? Cultural differences in and societal consequences of beliefs about the universal potential for intelligence. *Journal of Personality and Social Psychology, 103*(5), 787-803. doi:10.1037/a0029263

Raven, J., Raven, J. C., & Court, J. H. (2000). *Raven manual: Standard progressive matrices, including the parallel and plus versions (2000 edition ed.)*. Oxford, UK: Oxford Psychologists Press.

Reilly, D., Neumann, D. L., & Andrews, G. (2015). Sex differences in mathematics and science achievement: A meta-analysis of National Assessment of Educational Progress assessments. *Journal of Educational Psychology, 107*(3), 645-662. doi:10.1037/edu0000012

Rienzo, C., Rolfe, H., & Wilkinson, D. (2015). Changing mindsets. Retrieved from https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/changing-mindsets/

Rimm-Kaufman, S. E., & Hamre, B. K. (2010). The role of psychological and developmental science in efforts to improve teacher quality. *Teachers College Record, 112*(12), 2988-3023.

Ritchie, S. J., Bates, T. C., & Deary, I. J. (2015). Is education associated with improvements in general cognitive ability, or in specific skills? *Developmental Psychology, 51*(5), 573-582. doi:10.1037/a0038981

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science, 2*(4), 313-345. doi:10.1111/j.1745-6916.2007.00047.x

Roberts, B. W., Walton, K. E., & Bogg, T. (2005). Conscientiousness and health across the life course. *Review of General Psychology, 9*(2), 156-168. doi:10.1037/1089-2680.9.2.156

Robins, R. W., & Pals, J. L. (2002). Implicit self-theories in the academic domain: Implications for goal orientation, attributions, affect, and self-esteem change. *Self and Identity, 1*(4), 313-336. doi:10.1080/15298860290106805

Romero, C., Master, A., Paunesku, D., Dweck, C. S., & Gross, J. J. (2014). Academic and emotional functioning in middle school: The role of implicit theories. *Emotion, 14*(2), 227-234. doi:10.1037/a0035490

Rong, X. L., & Shi, T. (2001). Inequality in Chinese education. *Journal of Contemporary China, 10*(26), 107-124. doi:10.1080/10670560124330

Rosander, P., & Backstrom, M. (2014). Personality traits measured at baseline can predict academic performance in upper secondary school three years late. *Scandinavian Journal of Psychology, 55*(6), 611-618. doi:10.1111/sjop.12165

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton Univ. Press.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80*(1), 1-28. doi:10.1037/h0092976

Ryoo, J. H., Molfese, V. J., Heaton, R., Zhou, X., Brown, E. T., Prokasky, A., & Davis, E. (2014). Early mathematics skills from prekindergarten to first grade. *Journal of Advanced Academics, 25*(3), 162-188. doi:10.1177/1932202x14538975

Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology, 82*(1), 30-43.

Salili, F., & Hau, K. (1994). The effect of teachers' evaluative feedback on Chinese students' perception of ability: a cultural and situational analysis. *Educational Studies, 20*(2), 223-236. doi:10.1080/0305569940200206

Saucier, G., & Goldberg, L. R. (1998). What is beyond the Big Five? *Journal of Personality, 66* (4), 495-524.

Scheiber, C., Reynolds, M. R., Hajovsky, D. B., & Kaufman, A. S. (2015). Gender differences in achievement in a large, nationally representative sample of children and adolescents. *Psychology in the Schools, 52*(4), 335-348. doi:10.1002/pits.21827

Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2016). The geographic distribution of Big Five personality traits. *Journal of Cross-Cultural Psychology, 38*(2), 173-212. doi:10.1177/0022022106297299

Schunk, D. H. (1983). Ability versus effort attributional feedback: Differential effects on self-efficacy and achievement. *Journal of Educational Psychology, 75*(6), 848-856. doi:http://dx.doi.org/10.1037/0022-0663.75.6.848

Schweinhart, L. J., & Weikart, D. P. (2016). Effects of the Perry Preschool Program on youths through age 15. *Journal of the Division for Early Childhood, 4*(1), 29-39. doi:10.1177/105381518100400105

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*(2), 428-444.

Siegler, R. S., & Mu, Y. (2008). Chinese children excel on novel mathematics problems even before elementary school. *Psychological Science, 19*(8), 759-763.

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games -but not circular ones- improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology, 101*(3), 545-560.

Siewert, L. (2011). The effects of written teacher feedback on the academic achievement of fifth-grade students with learning challenges. *Preventing School Failure: Alternative Education for Children and Youth, 55*(1), 17-27. doi:10.1080/10459880903286771

Sirin, S. R. (2016). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417-453. doi:10.3102/00346543075003417

Sisk, V. F., Burgoyne, A. P., Sun, J., Butler, J. L., & Macnamara, B. N. (2018). To what extent and under which circumstances are growth mind-sets important to academic achievement? Two meta-analyses. *Psychological Science, 29*(4), 549-571. doi:10.1177/0956797617739704

Smith, T. W. (2016). Personality as risk and resilience in physical health. *Current Directions in Psychological Science, 15*(5), 227-231. doi:10.1111/j.1467-8721.2006.00441.x

Song, J., Zuo, B., & Yan, L. (2016). Effects of gender stereotypes on performance in mathematics: A serial multivariable mediation model. *Social Behavior and Personality: An international journal, 44*(6), 943-952. doi:10.2224/sbp.2016.44.6.943

Song, L., Appleton, S., & Knight, J. (2006). Why do girls in rural China have lower school enrollment? *World Development, 34*(9), 1639-1653. doi:10.1016/j.worlddev.2005.12.009

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science: A critical review. *American Psychologist, 60*(9), 950-958. doi:10.1037/0003-066X.60.9.950

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. doi:10.1006/jesp.1998.1373

Speybroeck, S., Kuppens, S., Van Damme, J., Van Petegem, P., Lamote, C., Boonen, T., & de Bilde, J. (2012). The role of teachers' expectations in the association between children's SES and performance in kindergarten: A moderated mediation analysis. *PLoS One, 7*(4), e34502. doi:10.1371/journal.pone.0034502

Sriram, R. (2014). Rethinking intelligence: The role of mindset in promoting success for academically high-risk students. *Journal of College Student Retention: Research, Theory & Practice, 15*(4), 515-536. doi:10.2190/CS.15.4.c

Stevenson, H. W., Lee, S., Chen, C., Lummis, M., Stigler, J., Fan, L., & Ge, F. (1990). Mathematics achievement of children in China and the US. *Child Development, 61*(4), 1053-1066.

Stevenson, H. W., Lee, S. Y., & Stigler, J. W. (1986). Mathematics achievement of Chinese, Japanese, and American children. *Science, 231*(4739), 693-699. doi:10.1126/science.3945803

Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science, 29*(4), 581-593. doi:10.1177/0956797617741719

Stronge, J. H., Tucker, P. D., & Hindman, J. L. (2004). *Handbook for qualities of effective teachers*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Styles, I., Raven, M., & Raven, J. C. (1998). *Raven's Standard Progressive Matrices: SPM Parallel sets A-E*.

Su, R., & Rounds, J. (2015). All STEM fields are not created equal: People and things interests explain gender disparities across STEM fields. *Frontiers in Psychology, 6*, 189. doi:10.3389/fpsyg.2015.00189

Szirmák, Z., & De Raad, B. (1994). Taxonomy and structure of Hungarian personality traits. *European Journal of Personality, 8*(2), 95-117. doi:10.1002/per.2410080203

Tahir, A. A., & Shah, A. F. (2012). Relationship of students' perceptions about teacher's personality with academic achievement of students. *Journal of Educational Research, 15*(2), 5-13

Tatli, A., Ozturk, M. B., & Woo, H. S. (2017). Individualization and marketization of responsibility for gender equality: The case of female managers in China. *Human Resource Management, 56*(3), 407-430. doi:10.1002/hrm.21776

Tellegen, A. (1991). Personality traits: Issues of definition, evidence, and assessment. In W. M. Grove & D. Cicchetti (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl* (Vol. 2). Minneapolis: University of Minnesota Press.

Terracciano, A., McCrae, R. R., Brant, L. J., & Costa, P. T., Jr. (2005). Hierarchical linear modeling analyses of the NEO-PI-R scales in the Baltimore Longitudinal Study of Aging. *Psychology and Aging, 20*(3), 493-506. doi:10.1037/0882-7974.20.3.493

Tiedemann, J. (2000). Gender-related beliefs of teachers in elementary school mathematics. *Educational Studies in Mathematics, 41*(2), 191-207. doi:10.1023/a:1003953801526

Tsui, M. (2007). Gender and mathematics achievement in China and the United States. *Gender Issues, 24*(3), 1-11. doi:10.1007/s12147-007-9044-2

Tsui, M., Xu, X., Venator, E., & Wang, Y. (2016). Stereotype threat and gender: Math performance in Chinese college students. *Chinese Sociological Review, 48*(4), 297-316. doi:10.1080/21620555.2016.1166339

Tsui, M., Xu, X. Y., & Venator, E. (2011). Gender, stereotype threat and mathematics test scores. *Journal of Social Sciences, 7*(4), 538-549.

Tucker-Drob, E. M., & Bates, T. C. (2016). Large cross-national differences in gene x socioeconomic status interaction on intelligence. *Psychological Science, 27*(2), 138-149. doi:10.1177/0956797615612727

Tupes, E. C., & Christal, R. E. (1961). Recurrent personality factors based on trait ratings. *Journal of Personality, 60*(2), 225-251. doi:10.1111/j.1467-6494.1992.tb00973.x

Tweed, R. G., & Lehman, D. R. (2002). Learning considered within a cultural context: Confucian and Socratic approaches. *American Psychologist, 57*(2), 89-99. doi:10.1037/0003-066X.57.2.89

UCAS. (2019). UCAS Tariff tables. Retrieved from https://www.ucas.com/files/new-tariff-tables

Ulug, M., Ozden, M. S., & Eryilmaz, A. (2011). The effects of teachers' attitudes on students' personality and performance. *Procedia - Social and Behavioral Sciences, 30*, 738-742. doi:10.1016/j.sbspro.2011.10.144

Wagerman, S. A., & Funder, D. C. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality, 41*(1), 221-229. doi:10.1016/j.jrp.2006.03.001

Walker, S. O., & Plomin, R. (2006). Nature, nurture, and perceptions of the classroom environment as they relate to teacher-assessed academic achievement: A twin study of nine-year-olds. *Educational Psychology, 26*(4), 541-561. doi:10.1080/01443410500342500

Wang, J., & Lin, E. (2009). A meta-analysis of comparative studies on Chinese and US students' mathematics performance: Implications for mathematics education reform and research. *Educational Research Review, 4*(3), 177-195. doi:10.1016/j.edurev.2009.06.003

Wang, M. T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review, 29*(1), 119-140. doi:10.1007/s10648-015-9355-x

Wang, Q., & Ng, F. F. Y. (2012). Chinese students' implicit theories of intelligence and school performance: Implications for their approach to schoolwork. *Personality and Individual Differences, 52*(8), 930-935. doi:https://doi.org/10.1016/j.paid.2012.01.024

Weisberg, Y. J., Deyoung, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the Big Five. *Frontiers in Psychology, 2*, 178-188. doi:10.3176/tr.2009.1.01

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology, 102*(1), 43-53. doi:10.1037/a0016738

Whyte, J. C., & Bull, R. (2008). Number games, magnitude representation, and basic number skills in preschoolers. *Developmental Psychology, 44*(2), 588-596. doi:10.1037/0012-1649.44.2.588

Wilberg, S., & Lynn, R. (1999). Sex differences in historical knowledge and school grades: A 26 nation study. *Personality and Individual Differences, 27*(6), 1221-1229. doi:10.1016/s0191-8869(99)00066-5

Williams, J. E., Satterwhite, R. C., & Best, D. L. (1999). Pancultural gender stereotypes revisited: The Five Factor Model. *Sex Roles, 40*(7-8), 513-525. doi:Doi 10.1023/A:1018831928829

World Bank. (2017). Sex ratio at birth (male births per female births). Retrieved from https://data.worldbank.org/indicator/SP.POP.BRTH.MF?end=2017&locations=CN&start=1962&type=shaded&view=chart

World Economic Forum. (2018). The global gender gap report 2018. Retrieved from https://www.weforum.org/reports/the-global-gender-gap-report-2018

Wu, H., Kung, F. Y. H., Chen, H., & Kim, Y. (2016). Academic success of "tiger cubs". *Social Psychological and Personality Science, 8*(6), 698-705. doi:10.1177/1948550616675667

Yeager, D. S., & Dweck, C. S. (2012). Mindsets that pomote resilience: When students believe that personal characteristics can be developed. *Educational Psychologist, 47*(4), 302-314. doi:10.1080/00461520.2012.722805

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., . . . Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature, 573*(7774), 364-369. doi:10.1038/s41586-019-1466-y

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., . . . Dweck, C. S. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology, 108*(3), 374-391. doi:10.1037/edu0000098

Yettick, H., Lloyd, S., Harwin, A., Riemer, A., & Swanson, C. B. (2016). Mindset in the classroom: A national survey of K-12 teachers. Retrieved from http://www.edweek.org/media/ewrc_mindsetintheclassroom_sept2016.pdf

Yun, S. H. (2012). An analysis of Confucianism's yin-yang harmony with nature and the traditional oppression of women: Implications for social work practice. *Journal of Social Work, 13*(6), 582-598. doi:10.1177/1468017312436445

Zeng, G., Hou, H., & Peng, K. (2016). Effect of growth mindset on school engagement and psychological well-being of Chinese primary and middle school students: The mediating role of resilience. *Frontiers in Psychology, 7*, 1873. doi:10.3389/fpsyg.2016.01873

Zentall, S. R., & Morris, B. J. (2010). "Good job, you're so smart": The effects of inconsistency of praise type on young children's motivation. *Journal of Experimental Child Psychology, 107*(2), 155-163. doi:10.1016/j.jecp.2010.04.015

Zhang, J., & Ziegler, M. (2016). How do the big five influence scholastic performance? A big five-narrow traits model or a double mediation model. *Learning and Individual Differences, 50*, 93-102. doi:10.1016/j.lindif.2016.07.011

Zhao, J., & Wang, M. (2014). Mothers' academic involvement and children's achievement: Children's theory of intelligence as a mediator. *Learning and Individual Differences, 35*, 130-136. doi:10.1016/j.lindif.2014.06.006

Zhou, M. (2015). Moderating effect of self-determination in the relationship between Big Five personality and academic performance. *Personality and Individual Differences, 86*, 385-389. doi:10.1016/j.paid.2015.07.005