Do superstitious beliefs have permanent effects on education? An analysis of the year of the Dragon and its implications.

Erasmus University Rotterdam: Erasmus School of Economics
Bachelor Thesis
Victor Dănescu (504613)
Thesis supervisor: Dr. Dinand Webbink

July 2021

Contents

Abstract

In Chinese folklore it is commonly believed that being born in the year of the Dragon predestines one to success and fortune. Many parents follow the Chinese zodiac, preferring their children to be born in the year of the Dragon. The main question asked in this article is whether or not believing in superstitions by means of being born in a Dragon year makes one achieve better education in the future. This is theorized to be due to a different parental treatment and self-perception of the individual. Harmonized microdata collected by the fifth National Population Census of China, contained in the IPUMS-International database, is primarily used along with OLS and logistic regressions to determine the effects of five different birth cohorts on educational outcomes of years of education and literacy of Chinese citizens. The findings suggest positive and significant effects for the birth cohorts of 1976 and 1964, with no effects for the period between 1928 and 1956. The results are supported by socio-economic and historical circumstances of The People's Republic of China. The paper concludes by suggesting highly specified research into the topic of superstition.

1.Introduction

In the modern world, superstitious beliefs such as misfortune-related happenings and the horoscope are widely acknowledged as innocent when it comes to success-related consequences. This study will test and analyse the impact of superstition, measured by children born in the Chinese year of the Dragon, on education prospects. The focus is on Mainland China and its afferent regions, but there is an additional emphasis on effects in other Asian countries, such as Japan, Korea and Thailand. The attempt is to investigate whether or not belief in a certain characteristics of the Chinese horoscope is self-fulfilling. It is a widespread belief in Asian cultures that people born in the year of the Dragon are destined to become good leaders and are prone to good fortune, compared to people born in other years. Can this lead to parents to parents raising their children in a more dignified matter? Perhaps the children themselves display an artificially inflated sense of self-esteem due to their perceived “gift”. In this paper, I hypothesise that being born in the Chinese year of the Dragon has a positive effect on future educational outcomes.

Beliefs lead to behaviour and behaviour leads to decisions. Superstitious beliefs can be easily classified as cultural biases, which have been shown to influence decision-making and behavioural patterns in economic trading (Guiso et al. 2009). Limited participation in the stock market is justified by a general inability to trust (Guiso et al. 2008). Belief in conspiracy theories was found to be correlated to a lack of trustworthiness, low self-esteem and hostility (Abalakina-Paap et al. 2002), as well positively associated with a young age and insecurity about employment prospects (Goertzel, 1994). An analysis based on a series of interviews discovered a positive link between suicidality and superstitious beliefs in Chinese women (Zhang and Xu, 2007).

Despite the seemingly evident fact that one’s time of birth does not come with any biological advantages, pseudo-scientific beliefs were found amongst life science and biology teachers (Raymond and Dunn, 1990). How could something so reasonably false persist for so long? Bénabou and Tirole (2016) found that the rigidity of some beliefs come as a sort of implicit trade-off between reason and hope. The resistance to evidence manifests itself either as a sunk cost fallacy or some sort of reasoned irrationality, in the sense that people willingly deny the truth to allow themselves to dream. Previous evidence has shown that planting false beliefs in subjects’ minds leads to direct consequences in their behaviour in the long term

(Laney et al. 2008). Ko and Huang (2012) found a link between loss aversion and falling back on past beliefs in the investment market

This study will make use of two datasets, the most focused on being the year 2000 census from the Chinese Bureau of Statistics of the IPUMS-International database. This database relates to the main goal of this study. Focusing on 32 provinces of China, this dataset provides the necessary variables for the following computations. Two educational outcomes will be analysed, the first being a variable indicating educational attainment by means of years of completed education, and the second a binary variable that indicates literacy or a lack thereof. A variable indicating whether or not an individual was born a Dragon will be used as an independent covariate, as well as five birth cohorts for the Dragon-years of 1928, 1940,1952,1964 and 1976, restricted around a tight interval of 5 years. [1]

An ordinary least squares regression was applied for each outcome and each predictor respectively, as well as a Regression Discontinuity regression with a stricter interval of only two months. Because of the large number of observations and wide time period used, a logistic regression was used for the binary outcome of literacy and a generalized ordered logistic regression (or gologit) was chosen for the categorical and ordinal outcome of education in years. The effects on years of completed education, after being confirmed by a placebo analysis, were significantly positive for the birth cohorts of 64 and 76, arguably insignificant for the 28,40 and 52 cohorts, and ambiguous for the general effect. This can be potentially justified by the sudden change in the Chinese socio-economic climate brought by the creation of the People's Republic of China in 1949, as well by the rapid urbanization that came after the interwar period.[2] Similar results were found for the literacy outcome.

The Program for International Student Assessment (PISA) is a worldwide study done by the OECD. It evaluates the development of pupils aged 15 and 16, by conducting a series of tests for the subjects of reading, science and mathematics. The countries analyzed are Japan, South Korea, Thailand and the Hong Kong Special Administrative Region of the People's Republic of China[3.] Dependent variables take the form of standardized average scores for the three subjects. Independent variables take the form of interaction effects ran within a

---

[1] The cohorts will be sometimes referred to as 28,40,52,64 and 76 respectively
[2] See Section 7
[3] Referred to as "Hong Kong" in this study

difference-in-difference regression. A general and a per-country analysis were done, both concluding in insignificant results.

This paper expands on previous research into the topic of superstition by especially focusing on Mainland China, relating the results back to historical circumstances and including other Asian cultures into the study. Section 2 provides background information and previous literature regarding superstitious beliefs in China. Section 3 describes the two datasets used in detail. Section 4 explains the main empirical strategies employed in this study, as well as their shortcomings. Sections 5 and 6 show the results of the analysis and a robustness test respectively. Section 7 is a discussion on the results that were found and section 8 concludes this article.

2.Background information and previous literature

Figure 1 depicts the twelve Chinese zodiacs, in order. They are Rat, Ox, Tiger, Rabbit, Dragon, Snake, Horse, Goat, Monkey, Rooster, Dog and Pig. The Korean zodiac has the exact same creatures, whilst the Japanese calendar replaces Pig with Boar. The Thai zodiac replaces the Dragon symbol with the Naga, but its cultural significance is still that of prosperity and fertility (Jiang, updated in 2021). The zodiacs are cycled in a lunar calendar that lasts approximately twelve years, with the year of the Dragon starting around February and ending roughly at the end of January of the next calendar year (see figure 2)[4].

The Dragon symbol is especially different from the other eleven. Characteristics such as wealth, leadership qualities, ambition and general charisma have been associated with this zodiac since ancient times. Many Asian cultures believe that children born in Dragon years will end up wealthy and successful. The rabbit and snake years will be marginally relevant to the PISA analysis, so a brief description of their canon is warranted. The rabbit is a symbol of pureness and auspiciousness, it represents amicability, tenderness and elegance. Individuals born in the year of the snake are characterised by thoughtfulness, calmness, loyalty and rationality.

---

[4] Figure 2 also contains the 5 elements associated with the zodiac. They are of marginal interest to this study. For more information, see https://www.purewow.com/wellness/chinese-zodiac-elements#:~:text=There%20are%20five%20elements%20of,and%20interdependence%20between%20all%20things.

Superstition in the form of numerology was shown to be quite important in determining human decision-making. Second-hand apartments in Chengdu, China with floors ending with the lucky number of “8” were shown to sell for higher prices than other apartments. The same study showed that newly built apartments on the same floor tend to sell much faster than others and tend to be bought by people with more 8’s in their phone number (Shum et al. 2014). Another study on Chinese residents in North America showed similar results. Houses with their address number ending in the traditionally unlucky number of “4” sold at a discount whilst houses with addresses ending in “8” sold at a premium (Fortin et al. 2014). Parents of children born in the year of the Dragon were shown to have higher expectations for their kids, compared to other parents (Mocan and Yu, 2020). What all this past evidence seems to show is that superstition has a significant effect on the behaviour of Chinese individuals, no matter where in the world they are located.

The perceptions that run in a family can have impactful ramifications on a child’s development. Alesina and Giuliano (2011) found that strong family ties lead to decreased political participation and civic engagement. The self-fulfilling prophecy that is theorized in this article acts in the same manner. Despite an evident downward trend of the fertility rate for the past 20 years in Hong Kong, short spikes in fertility were observed for the Dragon years of 1988 and 2000(Yip et al. 2002). This study also found no influences of the zodiac of fertility before 1976. Daniel M. Goodkind (1991) found a tendency of Chinese couples outside of mainland China to aim the birth of their children in the Dragon years of 1976 and 1988, but curiously enough this propensity was never present before 1976. These studies are partly in line with what was discussed in the introduction. The study justifies this anomaly by implying that this phenomenon is a result of a general relaxation provided by the modern era.

Finally, the superstition of Taian-Butsumetsu, relating to lucky and unlucky days of the week, is common amongst Japanese citizens. In a study done at a hospital in Kyoto, Japan, the mean number of hospital discharges was the highest in the lucky day of Taian and the lowest in the unlucky day of Butsumetsu. Evidence also showed that individuals were even willing to extend their hospitalization to ensure they were discharged on a lucky day (Hira et al. 1998).

3.Data

3.1 IPUMS-International Data

The IPUMS-International is a harmonized international census data for research and social science. It is entirely dedicated to distributing and collecting census microdata from around the world, covering 102 countries, 473 surveys and censuses and over 1 billion person records. The source data for IPUMS is provided by national statistical offices in countries of interest. For this research, the country of interest is China and the census used is the fourth National Population Census, conducted in 2000. The source data is once again provided by the Chinese National Bureau of statistics and it includes all individuals aged 6 or more who have Chinese nationality and reside within the country's borders.

The census day was November 1st, 2000. The entire field work period ran between November 1st and November 10th of 2000. Data collection was split in four forms: A short census form, a long census form, a population mortality survey form and a census appendix form. The fieldwork involved direct house-to-house enumeration which means that surveyors verify and complete forms at the residence based upon responses given by the members of the household. The sample design chosen was systematic sampling, based on households. This method selects sample members from a large population based on a random starting point, given a periodic and fixed interval. The interval is calculated by dividing the population size by the intended sample size. The sample size is equal to 11,804,344 households. Domestic households include individuals who live in the same place due to familial relationships and single individuals living alone. Collective dwellings include schools, churches and any institution that involves a large number of individuals engaging in cohabitation.

Variables of interest are indicators of demographic factors and educational attainment. (See Appendix B for a detailed description). Harmonised variables will be used in order to enable comparisons between the two censuses. The process of harmonization makes variables from different censuses and countries comparable by employing traditional methods such as the lowest common denominator or certain coding schemes in order to not lose information. After removing missing variables in years of birth and education outcomes and controlling for province fixed effects, 7,619,783 observations are left to work with. The fourth census, conducted in 1990 was also considered for this research. Because data for the province of birth is only available for the 2000 census, the 1990 census was not used. Controlling for province fixed effects will be shown to be very relevant to this study. There are 9.61%

Dragons in the database, with 90.39% of individuals being born in other years (see table 9, Appendix C).

3.2 PISA Data

The Program for International Student Assessment(PISA) for the years 2003,2006,2009,2012,2015 and 2018 will additionally be used to test whether or not parents tend to predispose their children to better education outcomes in other Eastern-Asian countries. Students from schools all around the world take standardized tests and answer detailed questionnaires every three years. The database includes mathematics, reading and science test scores for students aged 15 and 16 from a maximum of 80 different countries (both OECD and non-OECD; depending on the database used) and a maximum of 19811 different schools (for 2018). The students are clustered by different schools. The database includes multiple questionnaires, out of which the relevant one for this inquiry is the one given to students. The student dataset includes data on student performance, parent assessment and personal questions at an individual student level.

Each dataset, from 2003 to 2018, includes students born either 15 or 16 years prior to the testing. The 2003 and 2015 databases are especially important, as they include students born in 1988 and 2000, which are Dragon-years. The data is filtered to only include relevant variables and remove missing observations. The countries of interest are the Chinese Republic of Hong Kong, Japan, South Korea and Thailand. These four countries satisfy all the required restrictions:

- All four are Asian countries, with a general cultural belief in the Chinese horoscope, with only a few slight variations(see Section 2).
- They are common to all 6 databases used.
- They have a reliable number of observations, despite the seemingly low proportions of individuals born in the year of the Dragon(see Appendix D)

The datasets were appended to allow for the difference-in-difference approach used in this study. This method essentially compares the effect of being younger by means of being born in the later year across different datasets conducted in different years. After accounting for all these restrictions, we are left with a database that has 139,897 individual observations, out of which 12,387 are born in the year of the Dragon (1988 and 2000 for the dataset 2003 and 2015). Younger individuals comprise 25.65% of the observations, numbering 35,888. Out of the younger students, 34.52% are Dragons.

A stratified two-stage sampling design is used for most countries. The strata can be defined by the school's size and type and the gender of the individuals involved. A probability proportional-to-size sample of schools was used. The size is defined by the number of 15-year olds enrolled in the school. At least 150 schools were selected in each country (if less than 150, all schools were used). Following this, a sample of 35 students was drafted from the required age group for each of the sampled schools. This would indeed suggest that individuals in the sample do not have an equal probability of being selected and thus they are assigned sampling weights. The weights are adjusted to indicate the probability of inclusion in the sample and to account for schools and students that did not participate. This method is primarily concerned with maintaining external validity.

The students' scores in mathematics, reading and science are not directly observed in the study and have to be deduced from responses. PISA uses plausible values to input the scores. Plausible values are a representation of reasonable expectations from students' abilities[5]. The databases contain ten sets of plausible values for each subject (for 2015 and 2018) or five sets respectively (for 2003 up to and including 2012). These sets were averaged for each subject after which they were normalized to have a mean of 0 and a standard deviation of 1. This permits the estimated effect to be interpreted as a percentage of standard deviation. A more uniform interpretation is possible this way, as the scores are calculated on an arbitrary scale (0 to 700) and the regular coefficients did not provide much insight. The averages obtained were compared to each of the sets of standardized plausible values and the results obtained were almost identical.

[5] For a detailed description of the imputation of plausible values, see Appendix D.

4.Empirical Strategy

The main focus of this paper is the data from the IPUMS-International database. In order to determine the effect of being born in a Dragon year on educational achievement, we estimate three different methods: two for the IPUMS-International data and one for the PISA data.

4.1 Ordinary least squares regression

The OLS regression will be used in conjunction with the IPUMS-International data. The following models are estimated:

1) $EducationYears_i = \alpha_1 + \beta_1 D_i + \gamma X_i + \rho_p + \varepsilon_i$
2) $Literacy_i = \alpha_1 + \beta_1 D_i + \gamma X_i + \rho_p + \varepsilon_i$

-where $EducationYears_i$ is a categorical variable indicating the attained years of education for individual i. It can take values 0,3,6,9,12,14 and 16[6]. $Literacy_i$ is a binary variable that takes value 0 if the individual is illiterate and value 1 otherwise. A general comprehension of both reading and writing makes an individual literate. D_i is an interchangeable variable that takes value 1 for a certain year of the Dragon(spanning from 1928 to 1976) and value 0 for the previous two and next two years. It can also represent a binary variable that takes value 1 for Dragons and value 0 for non-Dragons; α_1 is the constant term. X_i is a vector of control variables including gender and age, ρ_p stands for province fixed effects[7].(see Appendix B for a more detailed explanation of the variables).The error term is represented by ε_i Each equation involves running five separate regressions.

The treatment and control groups might differ in observable characteristics. To account for this potential source of bias, one can use an OLS regression. This method mainly relies on the Conditional Mean Independence assumption(CIA) holding. This would imply that the difference between people born in the middle year of the interval used(ex.1976) and individuals born in the lower and upper bounds (1974,1975,1977,1978) only differ in observable characteristics. If it is the case that this assumption holds, one can add the observable controls to the regression to ensure that the correlation between the variable of interest and the error term is equal or close to zero. The gender of the individual and province

---

[6] See Appendix C, table 10 for frequencies and proportions for each level of education.
[7] See table 11 of appendix C for name, proportion and frequencies of the 32 provinces

fixed effects are used as controls, with the occasional use of age. I argue that due the restrictive nature of the independent variable, only spanning 5 years, and the nature of the analysis being a study on birth years, there is hardly any unobserved bias that can influence the coefficients. Teachman and Schollaert (1989) explore the possibility of gender being an influence on the timing of one's birth. They find a tendency of women to speed up their third births not based on wanting more boys or girl, but rather based on a balancing act of the sexes of the children. They also find a tendency of women to time their births faster if they already have a boy, justified by the higher likelihood of parents with boys still being married, thus less likely to disrupt fertility careers. This isn't to say that the gender of the individual by itself can have an influence on his or her birth date, and especially the birth year. It is safe to assume that many parents know their child's gender in advance, but to time a birth in such a way that it takes place in another year would only be possible around the very beginning and very end of the year. Table 12 (Appendix C) shows the effect of gender on the Dragon variables, starting with a significant negative effect and diminishing as we go back in time. The general effect on being born a Dragon seems to be positively significant, so the variable will be used.

I will concede the fact that variables such as parental income and parental education could have a plausibly significant effect on the child's birth year. Schooling was shown to have a largely negative effect on self-reported religiosity and on the propensity to believe in horoscopes or superstition of any kind (Mocan and Pogorelova, 2014). This would further imply that parents would not time their child's birth specifically if they have no concern for horoscopes. Immigration is not of much concern, as the IPUMS-International data only concerns individuals with both Chinese nationality and residence. In 2016, 1576 permanent residency cards were issued by China, a number over 750 lower than the U. S's 1.2 million (Zwetsloot and Peterson, 2019). Because residency does not guarantee nationality, it is safe to assume that this variable is not a confounder for this analysis. My argument is that all these potential confounders are captured by the province fixed effects used in the analyses. Figure 3 of the appendix shows the income distribution across China's regions, with the highest per capita incomes being concentrated on the right side of the country. Parental wealth and education are well known influencers of the place of residence and especially the place of birth. Figure 4 shows the mean age of the respondents scattered with the number of education years. Since the older individuals appear to be the least educated, the variable of age will only be used conservatively as a control. This will later be justified in the Discussion section.

A Regression Discontinuity Design(RDD) will be employed in order to further control for potential unobserved bias. The temporal interval will be restricted to a period of two months, at the beginning and at the end of the year of the Dragon. January of the year of the Rabbit will be compared to February of the year of the Dragon as the start of the Dragon year, followed by comparing January of the Dragon year with February of the Snake Year[8]. The equations look the same as the OLS equations used previously, but they imply different interpretations due to the restrictions that were imposed. Because the dataset does not provide exact birth days, it is not possible to pinpoint an actual moment of transition.This approach trades a reduction in bias for a potential increase in the variance of the estimates.

I. $EducationYears_i = \alpha_2 + \beta_1 D_i + \gamma X_i + \rho_p + \varepsilon_i$

II. $Literacy_i = \alpha_2 + \beta_1 D_i + \gamma X_i + \rho_p + \varepsilon_i$

In the interval of January/Rabbit-February/Dragon, and:

III. $EducationYears_i = \alpha_2 + \beta_1 D_i + \gamma X_i + \rho_p + \varepsilon_i$

IV. $Literacy_i = \alpha_2 + \beta_1 D_i + \gamma X_i + \rho_p + \varepsilon_i$

In the interval of January/Dragon-February/Snake

The variables represent the same properties as with models 1) and 2), but this time the implied values of D_i are 1 for being born in February of the Dragon-Year(28,40,52,64,76) and 0 for being born in January of the Rabbit Year(28,40,52,64,76), in the case of the first interval. The second interval implies that D_i takes value 1 for being born in January of the Dragon Year(29,41,53,65,77) and 0 for being born in February of the Snake Year(29,41,53,65,77).

4.2 Generalized ordered logistic regression/Logistic Regression

Logistic regression is used in scenarios where predicting the presence or absence of an outcome is possible. This model is best suited when the dependent variable is binary, as it is mainly used to predict probabilities. As it will be seen, this is not the only case. An important difference between the logit model and the OLS regressions is that the latter may offer probabilities that are not captured between 0 and 1. Another is that the logit model thinks in

[8] Meaning the first comparison involves the last month of the rabbit Year and the first month of the Dragon year and the second comparison involves the last month of the Dragon year and the last month of the snake year.

odds instead of probabilities. This part of the analysis will also focus on the IPUMS-International database.

Let y stand for a value of the literacy outcome for individual i. Name p the probability that y is equal to 1. The odds that y is equal to 1 are given by: $\frac{p}{1-p}$. The logit or log odds of p is: $\ln \frac{p}{1-p}$ .The logit is estimated as a linear combination of the given independent variables. Given that the literacy variable is binomial, a regular logistic regression will be used to assess it. The following logistic function is estimated:

a) $L_i = \tau_1 + \beta_1 D_i + \gamma X_i + \rho_p + \varepsilon_i$

-where L_i is the log-odds function for the outcome of literacy and the other variables hold the same meaning as with the OLS approach.

When it comes to the other outcome variable used, education in years, a more appropriate method to use would be the ordered logistic regression. This method is especially useful when the dependent variable has a meaningful order and more than two levels or categories. In our case, the variable has increasing levels of education with 7 different levels. This method essentially groups the different categories as if they were binomial(for example category 1 vs categories 2-7) as many times as necessary in an increasing order. The categories of the outcome variable are essentially collapsed into two subsets.

An ordered logit implies proportionality between the categories. For example, the coefficients that describe the relationship between the 0 years of education category(the first) versus all higher categories of education are identical to those that show the relationship between the next level and all higher levels and so on. This is the proportional odds assumption, also known as the parallel regression assumption. If this holds, the relationship between all pairs of groups is the same, so only one model is needed. There are two ways of testing this assumption. The first is a likelihood ratio test with the null hypothesis that coefficients are identical between models. The Brant test, coined by Rollin Brant(1990), is also a reliable method of testing this assumption(Williams,2016). With a similar null hypothesis, and not unlike the first test, this test hopes for an insignificant coefficient in order to prove that the assumption holds. Unsurprisingly, given the nature of the variables used, the assumption is violated in every case, no matter the test, as highlighted in table 1. This is unfortunate news as it means that the ordered logit approach cannot be used.

Table 1: Likelihood ratio test and Brant Test for the proportional odds assumption

	(1)	(2)	
Variables	Likelihood test chi2	Brant Chi2	Observations
Dragon1976	21086.82***	19887***	887,148
Dragon1964	18502.36***	16900.47***	1,117,290
Dragon1952	8110.23***	7396.11***	805,870
Dragon1940	4596.35***	4361.17**	425,697
Dragon1928	958.08***	955.40***	255,978
All Dragons	149861.99***	135065.11***	7,619,783

*Note: Column 1 indicates the chi-squared of the approximate likelihood-ratio test of proportionality across response categories. A chi-squared that is significant implies that the proportionality condition does not hold. Column 2 shows the chi-squared of the Brant test of the parallel regression assumption. A significant test statistic proves that the assumption has been violated. The first 5 rows are indicated by the 5 birth cohorts used, with the year indicated as the treatment group and the 2 years before and after as the control group. For example: Dragon1976 takes value 1 if person is born in 1976 and 0 if born in 1974,1975,1977 or 1978. The 6th row indicates the use of the general variable that indicates being born in a Dragon year, for all birth years provided in the database. The above table involves two separate test; each ran six times. Other variables tested at the same time are gender and province. They were also shown to be individually significant. ***P-value<0,01, **P-value<0,05, *P-value<0,1, no asterisk: P-value >0,1.*

4.2.1 The generalized ordered logistic regression

Encapsulated in the module "Gologit"(Fu, 1997), this method can replicate the ologit coefficients, estimate models completely unconstrained by the proportional odds assumption, or partially constrain models by only having certain variables meet the assumption (Williams,2006). This method fits less restrictive criteria compared to an ordered logistic regression, while resulting in more parsimonious and interpretable results than a multinomial logistic regression. The values taken by the outcome variable are irrelevant apart from the assumption that higher values are associated with higher outcomes. The gologit will be used to analyze the dependent variable education-years. Multiple panels will be estimated in order to account for the lack of proportionality between the categories. The following model is estimated:

b) $GOL_i = \tau_1 + \beta_1 D_i + \gamma X_i + \rho_p + \varepsilon_i$

-where GOL_i is the generalized ordered logistic function for the outcome of education in completed years of schooling, for individual i. The other variables hold the same meaning as with the OLS approach.

4.3 Difference-in-Difference approach

This method will be applied to the PISA dataset. This method tests 2 different groups in two or more different regions or time periods. In our case, the two groups are individuals born in the later year, or younger students(treatment) and older students, or individuals born in the earlier year(control). Do keep in mind that every one of the 6 PISA datasets involves students aged 15 to 16 born in either of two consecutive years(see Data section and Appendix D). The different time periods are the years of the databases, spanning from 2003 to 2018, every three years. By interacting the effect of being younger with the year of the database, 2003 and 2015 would yield comparisons between differences of Dragons and rabbits(1988-1987 and 2000-2001) and placebo differences for 2006,2009, 2012 and 2018. One of the highlighted effects would be, for instance, the difference between the effect of being younger in 2003, compared with the effect of being younger in 2006. The following model is estimated:

i. $S_i = c_i + \beta_1 Young_i + \beta_2 Year_i + \beta_3 (Young_i * Year_i) + e_i$
ii. $S_{ic} = c_i + \beta_1 Young_{ic} + \beta_2 Year_{ic} + \beta_3 (Young_{ic} * Year_{ic}) + e_{ic}$

-where S_i is either the standardized average score across all three disciplines or the standardized average score for mathematics, reading ability or science for individual *i* (see Appendix D and section 3). This already implies four separate regressions, depending on the outcome used. The variable $Young_i$ takes value 1 if the individual is born in the later year and value 0 otherwise. $Year_i$ is a categorical variable that takes the value of the year the database was published in. β_3 is the difference between the effect of being younger in either 2006,2009,2012,2015 and 2018 and the effect of being younger in 2003. The baseline year is 2003. e_i is the error term and c_i is the constant term. The subscript *ic* refers to individual *i* in country *c*. The second regression is similar to the first one, only this time instead of running it for the entire sample, it will be ran separately for each of the four countries chosen[9].

[9] See table 13 of Appendix C for proportions and frequencies for Japan, South Korea, Thailand and Hong Kong.

The DiD approach allows controlling for time-invariant bias and time-varying constant bias. Students born in the Dragon years also happen to be the youngest in the dataset. It would be rather difficult for a regular OLS regression to distinguish between the effect of being younger[10] and the Dragon effect. The DiD method isolates the effect of the Dragon by differencing out the effect of a younger age and its implications. The treated and non-treated are allowed to evolve over time, provided that other potentially unseen outcomes evolve in the same way in the absence of treatment. This is the parralel trends assumption, which this study takes as a given because of the lack of proper ways of testing it and its rather low perceived likelihood of being violated due to the nature of the variables used.

5.Results

5.1 OLS Regression Results: IPUMS-International

Table 2 showcases the models described by equations 1) and 2). There is a general positive effect of being a Dragon on the education outcome, with Dragons having, on average, 0.086 additional years of schooling, or around 1 month of extra education, compared to non-Dragons. The same can be said about the literacy outcome, where being a Dragon increases the probability of being literate by 0.3%, compared to not being one. When it comes to the five birth cohorts, a curious trend can be observed. In the case of education, positive and significant effects were observed for the birth cohorts of 76 and 64, seeming to be generally more predisposed to higher education than their close seniors and juniors. The birth cohorts of 52,40 and 28 present themselves with negative and significant coefficients, implying either a bias in estimation or perhaps parents actively raising their children in worse conditions due to them being Dragons. The effects on literacy present the same way, with the exception of a zero effect for the birth cohort of 1976. For now, this is rather inconclusive. One possible interpretation would be that the rapid urbanization and economic development of China brought along the required relaxation needed to develop superstitious beliefs.

[10] So presumably less experienced, in a lower grade-year, etc.

Table 2: OLS Regression estimates of being born in a Dragon year on educational outcomes.

Variables	(1) Education in Years	(2) Literacy	Observations
All Dragons	0.086***	0.003***	7,619,783
Dragon1976	0.029***	0	887,148
Dragon1964	0.042***	0.001***	1,117,290
Dragon1952	-0.125***	-0.002*	805,870
Dragon1940	-0.030***	-0.005**	425,697
Dragon1928	-0.065**	-0.009**	255,978
Controls	Yes	Yes	

*Note: Columns 1 and 2 show the same OLS regression for two different dependent variables: education in completed school years and a binary variable indicating literacy. The first row indicates the use of the general variable that indicates being born in a Dragon year, for all birth years provided in the database. The next 5 rows indicate the 5 birth cohorts used, with the year indicated as the treatment group and the 2 years before and after as the control group. For example: Dragon1976 takes value 1 if person is born in 1976 and 0 if born in 1974,1975,1977 or 1978. Controls used are the gender and age of the individual and province fixed effects. ***P-value<0,01, **P-value<0,05, *P-value<0,1, no asterisk: P-value >0,1. Numbers rounded to the third decimal.*

5.2 Regression discontinuity results

The results of the same regressions with the applied restrictions for stricter intervals can be seen in tables 3 and 4. They illustrate the models highlighted by equations I-IV. Looking at education first, rather inconsistent changes seem to take place between the start and end of the Dragon year. For the 1976 cohort, being a Dragon in February results, on average, in 0.08 fewer years of education, or 29 days, compared to being a rabbit in just the previous month of January. This result completely flips for the end-year, with Dragons in the last month of the Dragon year having an average of 0.09 more years of education, or 33 days, compared to snakes in the next month. The 1964 cohorts present itself with positive results in both directions, with Dragons achieving on average 2 months of extra education at the start of the year and approximately one year of additional completed schooling at the end of the year. The 1952 and 1940 cohorts only have significant positive results for the start and end respectively, with the 1928 cohort seemingly not affecting the education level. This could be attributed to the lower amount of observations, the further in time the study goes, or it may fall in line with the theory discussed above, that Chinese individuals only started acting on their superstitious beliefs in the 60's.

Table 3: Regression discontinuity estimates of the effect of being born in a Dragon year on years of completed education.

Variables	(1)	(1*)	(2)	(2*)
Education/Years	Start	Observations	End	Observations
Dragon1976	-0.080**	28,369	0.090***	24,987
Dragon1964	0.173***	37,455	1.049***	37,278
Dragon1952	0.111***	26,009	-0.015	25,128
Dragon1940	-0.033	14,607	0.136**	13,557
Dragon1928	-0.015	9,339	0.046	7,918
Controls	Yes		Yes	

Note: The first model from table 2 is used here, with an additional restriction of only including two months' worth of observations. Column 1 depicts the comparison between the last month of the previous Rabbit Year(January) and the first month of the Dragon year (February. Column 2 illustrates the comparison between the last month of the Dragon year (January of the next calendar year) and the first month of the next Snake Year(February). Columns 1 and 2* show the number of observations for their respective column and predictor. The first 5 rows indicate the 5 birth cohorts used, with the year indicated as the treatment group and the 2 years before and after as the control group. For example: Dragon1976 takes value 1 if person is born in 1976 and 0 if born in 1974,1975,1977 or 1978. Row 1, column 1 shows the coefficient for the interval between January 1976 and February 1976. Controls used are the gender of the individual and province fixed effects. ***P-value<0,01, **P-value<0,05, *P-value<0,1, no asterisk: P-value >0,1. Numbers rounded to the third decimal.*

Moving on to the literacy outcome (table 4), the regressions ran presented themselves with insignificant results all across the board. The number of observations is the same as before. If this OLS regression is assumed to be free of bias, it can be safely concluded that being born a Dragon does not affect literacy in the future, at least when compared to being born a rabbit or a snake. This study was unable to control for minute differences between each symbol in the Chinese horoscope. Presumably, if these differences are significant enough, a Dragon effect on literacy might still be present.

Table 4: Regression discontinuity estimates of the effect of being born in a Dragon year on literacy.

Variables	(1)	(1*)	(2)	(2*)
Literacy	Start	Observations	End	Observations
Dragon1976	-0.002	28,369	-0.081	24,987
Dragon1964	- 0.001	37,455	0.002	37,278
Dragon1952	0.002	26,009	0	25,128
Dragon1940	-0.002	14,607	0.002	13,557
Dragon1928	0.008	9,339	0.005	7,918
Controls	Yes		Yes	

Note: The second model from table 2 is used here, with an additional restriction of only including two months' worth of observations. Column 1 depicts the comparison between the last month of the previous Rabbit Year(January) and the first month of the Dragon year (February. Column 2 illustrates the comparison between the last month of the Dragon year (January of the next calendar year) and the first month of the next Snake Year(February). Columns 1 and 2* show the number of observations for their respective column and predictor. The first 5 rows indicate the 5 birth cohorts used, with the year indicated as the treatment group and the 2 years before and after as the control group. For example: Dragon1976 takes value 1 if person is born in 1976 and 0 if born in 1974,1975,1977 or 1978. Row 1, column 1 shows the coefficient for the interval between January 1976 and February 1976. Controls used are the gender of the individual and province fixed effects. ***P-value<0,01, **P-value<0,05, *P-value<0,1, no asterisk: P-value >0,1. Numbers rounded to the third decimal.*

5.3 Logit and Generalized ordered logit results: IPUMS-International

The logit regression ran with for the literacy outcome yielded mostly significant results, with the exception of the 1976 birth cohort. The second column of table 5 illustrates the odds ratios, or the probability of literacy divided by the probability of illiteracy, for the 5 birth cohorts and the general effect of being a Dragon. Being born in a Dragon year makes being literate 1.039 times as likely as being illiterate, on average. Given the large amount of observations for this category, it is quite difficult to dismiss this as being too low of an effect.

The 1964 birth cohort has a comparable and significant odds ratio of 1.049, whilst the 52,40 and 28 birth cohorts highlight negative effects of being a Dragon on literacy. All of the effects discussed are considerably small, but highly significant nonetheless.

Table 5: Logistic regression estimates of the effect of being born a Dragon on literacy.

Variables	(1) Literacy	(2) Odds Ratio	Observations
All Dragons	0.038***	1.039***	7,619,783
Dragon1976	-0.033	0.975	887,148
Dragon1964	0.049***	1.049***	1,117,290
Dragon1952	-0.029***	0.974**	805,870
Dragon1940	-0.027***	0.966***	425,697
Dragon1928	-0.026**	0.973**	255,978
Controls	Yes	Yes	

*Note: Column 1 shows the logistic regression for the binary literacy level, in log odds. Column 2 shows the proportional odds ratios. This ratio is equal to e raised to the power of the coefficient. The first row indicates the use of the general variable that indicates being born in a Dragon year, for all birth years provided in the database. The next 5 rows indicate the 5 birth cohorts used, with the year indicated as the treatment group and the 2 years before and after as the control group. For example: Dragon1976 takes value 1 if person is born in 1976 and 0 if born in 1974,1975,1977 or 1978. Controls used are the gender and age of the individual and province fixed effects. ***P-value<0,01, **P-value<0,05, *P-value<0,1, no asterisk: P-value >0,1. Numbers rounded to the third decimal.*

Six generalized ordered logistic regressions were run for the general Dragon identifier and the five birth cohorts. The results of the regression are shown in table 7. Positive coefficients imply a positive influence of being a Dragon on education years completed. Negative coefficients imply the opposite. If the coefficients take both positive and negative values from left to right, the effect is ambiguous. Looking at the temporally unconstrained Dragon covariate first, after transforming log odds into odds, and turning odds into probabilities, the following results were obtained:

- Dragons are 3.9% **less** likely to find themselves in the higher education categories, compared to having 0 years of schooling (being uneducated)
- Dragons are 1.7% **less** likely to find themselves in the 6 years to 16 years categories compared to the 0 and 3 years categories
- Dragons are 4.6% **more** likely to find themselves in the 9-16 categories compared to the 0,3 and 6 categories
- Dragons are 2.1% **less** likely to find themselves in the 12 to 16 years categories compared to the 0 to 9 categories
- Dragons are 4.9% **more** likely to find themselves in the 14 and 16 years categories compared to the 0 to 12 years categories
- Dragons are 7.8% **more** likely to find themselves in the 16 years category compared to the 0 to 14 years categories

This implies a general tendency of Dragons to either be poorly educated (0 and 3 years of education) or highly educated (9, 14 and 16 years of education) with the odd tendencies of having less Dragons have 6 years or 12 years of education. For now, this looks discouragingly inconclusive.

The 1976 birth cohort presents itself with all positive and significant coefficients, apart from two instances of insignificance in the first and fourth columns. It would seem that Dragons born in the 1976 birth cohort tend to be better educated than their non-Dragon seniors and juniors. Coefficients for the 1964 cohort are also positive and this time they are all significant. The same conclusion can be reached for this cohort as for 1976. Apart from a few zero coefficients, the 52,40 and 28 birth cohorts have negative and significant coefficients. Parents of Dragons born in these time periods do not only seem to be unconcerned by the horoscope, but the results imply that they actively opt for a poorer development of their children if they are Dragons.

Table 6: Generalized ordered logistic regression estimates of the Dragon-effect on education in completed years of schooling.

Education	(1)	(2)	(3)	(4)	(5)	(6)
Log Odds	0	3	6	9	12	14
All Dragons	-0.039***	-0.017***	0.045***	-0.021***	0.048***	0.075***
Dragon1976	0.012	1.010***	0.098***	-0.006	0.093***	0.178***
Dragon1964	0.043***	0.045***	0.044***	0.014***	0.084***	0.149***
Dragon1952	-0.065***	-0.051***	-0.105***	-0.180***	-0.115***	-0.126***
Dragon1940	-0.008	0.002	-0.002	0.003	-0.037**	-0.044**
Dragon1928	-0.034***	-0.032***	-0.036***	-0.046***	-0.037	-0.096**
Controls	Yes	Yes	Yes	Yes	Yes	Yes

*Note: Columns 1-6 show the generalized ordered logistic regressions for education in school years, classified by the number of years one completed in education (see Appendix B). The first row indicates the use of the general variable that indicates being born in a Dragon year, for all birth years provided in the database. The next 5 rows indicate the 5 birth cohorts used, with the year indicated as the treatment group and the 2 years before and after as the control group. For example: Dragon1976 takes value 1 if person is born in 1976 and 0 if born in 1974,1975,1977 or 1978. Coefficients represent the log-odds as with any logistic regression. Interpretation goes from left to right. Controls used are the gender and age of the individual and province fixed effects. ***P-value<0,01, **P-value<0,05, *P-value<0,1, no asterisk: P-value >0, 1. Numbers rounded to the third decimal.*

5.4 Difference-in-difference results: PISA data

This section shows the tests ran to assess the influence of being born in the year of the Dragon on test scores for Japan, South Korea, Hong Kong and Thailand. Tables 14 and 15(Appendix C) show the difference-in-difference regressions ran for the entire sample and for each country separately. The general of effect of being younger is expectedly negative, with the interaction effects being generally insignificant between periods that include Dragons and periods that do not. I expect this to be due to a combination between the low number of observations and the generally lower propensity to believe in superstition found in non-Chinese Asian cultures.

5.5 Final remarks

For the Chinese sample analysed, the conclusion reached is that being born a Dragon in 1964 and 1976 brings forth better conditions in one's own upbringing that in turn lead to a higher level of education achieved during one's lifetime. For Dragons born in 1928, 1940 and 1952, a negative effect on educational attainment is found, for reasons this study is not properly equipped to explain. It is worth noting that each different birth cohort used fits into a different zodiac element, being Earth, Metal, Water, Wood, and Fire. The possibilities of inherent superstitious beliefs associated with these elements might explain away some of the previous inconsistencies. After conducting a placebo analysis (see section 6), it is found that this negative effect is not caused by being a Dragon, so the overall conclusion is that Dragons in this period have no additional educational advantage. As far as the literacy outcome is concerned, the two-month interval regressions showed overall insignificance whilst the regular OLS regression and the logit implied insignificance for the 1976 cohort, a positive effect for the 1964 cohort and negative effects for the 1928,1940 and 1952 cohorts. This is the same conclusion as with the education outcome, with the exception of the irregularity in 1976. No real effects were found for other Asian cultures in the PISA datasets, so the conclusion for this side-quest of this study is insignificance. Do note that the setting for this analysis was quite different from the IPUMS inquiry, having a different dependent variable and Dragons born much later than 1976.

6.Robustness

A placebo analysis was conducted for the OLS method concerning both educational outcomes and for the logistic regression regarding the literacy result. Five placebo birth cohorts were coded in the same manner as the cohorts used so far, taking place in periods where there are no Dragons born (see Appendix B). The following equations were estimated:

1. $EducationYears_i = \alpha_1 + \beta_1 P_i + \gamma X_i + \rho_p + \varepsilon_i$
2. $Literacy_i = \alpha_1 + \beta_1 P_i + \gamma X_i + \rho_p + \varepsilon_i$
3. $L_i = \tau_1 + \beta_1 P_i + \gamma X_i + \rho_p + \varepsilon_i$

-where equations 1. and 2. are similar to equations 1) and 2), with the exception that the variable *D* is replaced by P_i ,which signifies Dragons in the placebo cohorts. Each of the two regressions is ran 5 times, each time for a different birth cohort. Equation 3. is alike model a), it is a regular logistic function for the literacy outcome, with the independent variables being the five placebos.

Table 7 shows the modelling of equations 1. and 2. An(almost) insignificant effect is found for the 1973 and 1961 cohorts, which falls in line with the conclusion found earlier regarding these time periods. Negative significant effects are found for the 1949,1937 and 1925 cohorts, aligning with the negative results found for their respective Dragon cohorts. This leads to the conclusion that the negative effect found is not caused by the individual being born a Dragon, with the true cause being somewhere else.

Table 7: Ordinary Least Squares regression estimates of the placebo-effect on educational outcomes.

	(1)	(2)	
Variables	Education Years	Literacy	Observations
Placebo1973	0.004	0	1,048,631
Placebo1961	0.00745*	-0.001*	856,865
Placebo1949	-0.026**	0	661,757
Placebo1937	-0.042***	- 0.004*	398,329
Placebo1925	-0.045**	-0.008***	194,456
Controls	Yes	Yes	

*Note: Columns 1 and 2 show the same OLS regression for two different dependent variables: education in completed school years and a binary variable indicating literacy. The first 5 rows indicate the 5 placebo birth cohorts used, with the year indicated as the treatment group and the 2 years before and after as the control group. For example: Placebo1973 takes value 1 if person is born in 1973 and 0 if born in 1971,1972,1974 or 1975.The main highlight of this method is using a period close to the periods including the Dragon years, coded in the same manner while not including a year of the Dragon. Controls used are the gender and age of the individual and province fixed effects. ***P-value<0,01, **P-value<0,05, *P-value<0,1, no asterisk: P-value >0,1. Numbers rounded to the third decimal.*

The following table shows the results found by running regression 3. The results found align with those found in the previous table.

Table 8: Logistic regression estimates of the placebo-effect on literacy.

	(1)	
Variables	Literacy	Observations
Placebo1973	- 0.008	1,048,631
Placebo1961	0.005	856,865
Placebo1949	- 0.008*	661,757
Placebo1937	- 0.032***	398,329
Placebo1925	-0.041***	194,456
Controls	Yes	

*Note: Column 1 shows the logistic regression for the binary literacy level, in log odds, for the outcome variable education in years. Column 2 illustrates the same regression for the outcome literacy. The first 5 rows indicate the 5 placebo birth cohorts used, with the year indicated as the treatment group and the 2 years before and after as the control group. For example: Placebo1973 takes value 1 if person is born in 1973 and 0 if born in 1971,1972,1974 or 1975.The main highlight of this method is using a period close to the periods including the Dragon years, coded in the same manner while not including a year of the Dragon. Controls used are the gender of the individual and province fixed effects. ***P-value<0,01, **P-value<0,05, *P-value<0,1, no asterisk: P-value >0,1. Numbers rounded to the third decimal.*

7.Discussion

The results found harken back to the conclusions reached by Goodkind and Yip et al. in their studies on the influence of the Dragon year on fertility. The difference from their conclusion and the one reached in this study is that they believe that the influence of the Dragon is null before 1976. I argue that there are quite a few notable differences that have to be accounted for. The differences justify the same effect for the birth cohort of 1964. Goodkind found that the increasing adherence to superstition was justified by an amalgamation of quick urbanization, consumerism and historical circumstances. Yip, Lee and Cheung found that contraceptives became increasingly more available after 1964 for the Chinese population. One important difference is that these studies were focused on Chinese samples outside of the People's Republic of China, and that they focus on fertility rather than educational outcomes. While their findings are in line with Dragons influencing educational outcomes after 1976, I believe that the main historical circumstance that primarily affected both development in Mainland China and superstitious belief was the foundation of the People's Republic of China in 1949 during the aftermath of the Chinese civil war.

An analysis on prices, wages and living standards in China for 18th, 19th and 20th centuries shows that in the 18th century Chinese incomes were much lower compared to developed economies such as England and Japan. The wages of Chinese workers continued to stagnate throughout the 18th and 19th century until they finally started growing in the 1900's (Allen et al. 2011). The comparison with the rapid industrialization of Japan also highlights a potential reason for not finding significant results for other Asian cultures. The communist regime brought along severe censorship of external information, poor human rights, lacking infrastructure and secularism(Smith,2015). The arrival of the Chinese Communist Party was also marked by the emergence of "Mass Superstitious Incidents" as early as 1953 when over 50,000 people gathered to experience a prophetic curing of diseases. The communist party actively tried to supress these types of incidents, but in a manner similar to the rampant black market of the Prohibition, it led to the appearance of many other incidents of this kind(Jiao,2020). The combination between the rapid urbanization, development of consumerist behaviour, and the degradation brought along by the CCP created a perfect boiling pot for a return to superstition. For the reasons stated above, this study claims proper external validity for the Chinese population of Mainland China. It is worth noting that Hong Kong is not considered as a part of Mainland China.

Studies such as this one are important in distinguishing between calculated, rational behaviour, and actions driven by long-lasting beliefs of past generations. In a conventionally secularist modern era where the emergence of the field of behavioural economics is an increasingly important factor in policy-making, analysing superstitious behaviour can highlight flaws in active and subconscious decision-making. Casual readers may find it useful to self-reflect on their habits of making important decisions based on circumstantial hearsay. Paternalistic social engineers and policy makers can use information on Dragon birth years to either drive people away from this behaviour and towards rational decision-making, or integrate the tendency to follow the Chinese zodiac into effective nudges. An ethicist might consider these developments worrying and strive towards more effective education of the populous while a business owner can tap into a goldmine of irrationality by employing specific sales and discounts around time periods of particular interest.

This analysis is limited in its approach by a few factors. First of all, the unavailability of more detailed datasets, especially for the case of other Asian cultures, makes the conclusion derived from the PISA datasets rather circumstantial. Secondly, the nature of the data and the predictors used makes it irresponsible to quantify the positive effects obtained, even considering that the significance and positive sign of the effects were properly argued to be valid. This study showed that the negative effect experienced before 1964 is not due to the year of the Dragon, but it was unable to pinpoint its actual cause. Thirdly, it cannot be said that the results found here apply for Chinese populations outside of the Chinese Mainland, nor can they be used to predict effects for individuals living in other countries believing in versions of the Chinese zodiac. Finally, despite the rather robust nature of the treatment variables uses, being coded from birth years, there are still potential confounders that may not have been accounted for by the provincial analysis. These include other superstitious beliefs associated with the year of the Dragon, such as specific differences between each of the 12 zodiacs, instead of just the Dragon year (especially highlighted in the strict interval analysis), the heavenly branch of the zodiac, and belief in other superstitions such as numerology associated with the year one was born in.

A suggestion for future data gatherers is to construct datasets specifically around superstitious beliefs, with more detailed predictors such as numerology, belief in conspiracy theories and religious background. Future research can expand on this study by analysing the 18th and 19th centuries as well as very recent Dragon years in order to determine the existence of a trend. Relations between the birth years of the parents and the birth years of their children may be useful in establishing whether there is a persistence of folklore across generations. Looking into the entire universe of the zodiacs might help find effects of other years that were previously unaccounted for. This would help create a hierarchy of zodiacs in terms of future education prospects and wealth. The heavenly branch and the 5 elements of earth, wood, water, metal and fire should also be delved into to establish differences between cycles of the same zodiac sign. Japanese, Korean and Thai populations should be approached separately in a manner similar to this study. Relating studies like this to Western astrological signs may help identify similar effects in non-Asian populations.

8.Conclusion

This paper aimed to estimate the effect of superstition, by means of the Chinese zodiac, on educational outcomes of residents of Mainland China and international students. The initial hypothesis was that merely being born in the year of the Dragon results in circumstances that provide higher educational attainment in the future. Data from Chinese residents was primarily used to assess this question. Evaluation was approached via three empirical methods: An OLS regression with an added asterisk of a restrictive interval, logistic and generalized ordered logistic regressions, and difference-in-difference models. The paper underlines the main strenghts and weaknesses of the methods used after which it finds the theorized positive and significant effect of Dragon-years on educational outcomes for the cohorts of 1976 and 1964, with no effect before that. The effect is justified by the historical and economic circumstances of China. A placebo analysis was conducted to strenghten the belief in the previously mentioned results. An effect for populations outside of the People's Republic of China cannot be confirmed or infirmed. Suggestions for future research involve the creation and use of datasets specifically tailored around behavioural analysis on superstition. Comparisons within the Chinese zodiac and with other forms of superstition should also provide insightful results in the future.

References

Abalakina-Paap, M., Stephan, W. G., Craig, T., & Gregory, W. L. (1999). Beliefs in conspiracies. *Political Psychology*, *20*(3), 637-647.

Alesina, A., & Giuliano, P. (2011). Family ties and political participation. *Journal of the European Economic Association*, *9*(5), 817-839.

Allen, R. C., Bassino, J. P., Ma, D., Moll-Murata, C., & Van Zanden, J. L. (2011). Wages, prices, and living standards in China, 1738–1925: in comparison with Europe, Japan, and India. *The Economic History Review*, *64*, 8-38.

Asmita, S. (2019, July 26). *The 12 Chinese Zodiac Signs Can Reveal Hidden Depths of Your Personality*. https://her.womenworking.com. https://her.womenworking.com/12-chinese-zodiac-signs-reveal-personality.

Bénabou, R., & Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, *30*(3), 141-64.

Brant, R. (1990). Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression. *Biometrics, 46*(4), 1171-1178. doi:10.2307/2532457

Christine. (2019, December 18). *What does your Chinese zodiac sign say about your travel style?* Travels of a Madman. http://airportshuttlecapetown.blogspot.com/2019/12/what-does-your-chinese-zodiac-sign-say.html.

Data - PISA. OECD. (n.d.). https://www.oecd.org/pisa/data/.

Dragon (zodiac). (2005, June 8). In *Wikipedia*. https://en.wikipedia.org/wiki/Dragon_(zodiac)

Eve, R., & Dunn, D. (1990). Psychic Powers, Astrology & Creationism in the Classroom? Evidence of Pseudoscientific Beliefs among High School Biology & Life Science Teachers. *The American Biology Teacher, 52*(1), 10-21. doi:10.2307/4449018

Fortin, N. M., Hill, A. J., & Huang, J. (2014). Superstition in the housing market. *Economic Inquiry*, *52*(3), 974-993.

Fu, V. (2002). GOLOGIT: Stata module to estimate generalized ordered logit models.

Goertzel, T. (1994). Belief in conspiracy theories. *Political psychology*, 731-742.

Goodkind, D. M. (1991). Creating new traditions in modern Chinese populations: Aiming for birth in the year of the Dragon. *Population and Development review*, 663-686.

Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. *the Journal of Finance*, *63*(6), 2557-2600.

Guiso, L., Sapienza, P., & Zingales, L. (2009). Cultural biases in economic exchange? *The quarterly journal of economics*, *124*(3), 1095-1131.

Hira, K., Fukui, T., Endoh, A., Rahman, M., & Maekawa, M. (1998). Influence of superstition on the date of hospital discharge and medical cost in Japan: retrospective and descriptive study. *Bmj*, *317*(7174), 1680-1683.

IPUMS-*International*: Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 7.3 [dataset]. Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/D020.V7.3

Jiang, F. (n.d.). *The Japanese / Vietnamese / Burmese Zodiac*. The Japanese Zodiac, Vietnamese Zodiac and Burmese Zodiac. https://www.chinahighlights.com/travelguide/chinese-zodiac/chinese-zodiac-in-other-asian-countries.htm.

Jiao, Y. (2019). The People's Living Guanyin Bodhisattva: Superstition, Entrepreneurship, Healthcare, Rural Economic Control, and Huidaomen in the Early PRC. *The Chinese Historical Review*, *26*(2), 175-196.

Ko, K. J., & Huang, Z. (2012). Persistence of beliefs in an investment experiment. *The Quarterly Journal of Finance*, *2*(01), 1250005.

Laney, C., Fowler, N. B., Nelson, K. J., Bernstein, D. M., & Loftus, E. F. (2008). The persistence of false beliefs. *Acta Psychologica*, *129*(1), 190-197.

Mocan, N., & Pogorelova, L. (2017). Compulsory schooling laws and formation of beliefs: Education, religion and superstition. *Journal of Economic Behavior & Organization*, *142*, 509-539.

Mocan, N., & Yu, H. (2020). Can superstition create a self-fulfilling prophecy? school outcomes of Dragon children in China. *Journal of Human Capital*, *14*(4), 485-534.

Schiavenza, M. (2013). Mapping China's Income Inequality. *The Atlantic*, *13*.

Shum, M., Sun, W., & Ye, G. (2014). Superstition and "lucky" apartments: Evidence from transaction-level data. *Journal of Comparative Economics*, *42*(1), 109-117.

Smith, S. A. (2015). 13. Redemptive Religious Societies and the Communist State, 1949 to the 1980s. In *Maoism at the Grassroots* (pp. 340-364). Harvard University Press.

Statistical data. National Bureau of Statistics of China. (n.d.). http://www.stats.gov.cn/english/.

Teachman JD, Schollaert PT. Gender of children and birth timing. Demography. 1989 Aug;26(3):411-23. PMID: 2792477.

Williams, R. (2005, July). Gologit2: Generalized Logistic Regression Models for Ordinal Dependent Variables. In *North American Stata Users' Group Meetings 2005* (No. 21). Stata Users Group.

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *The Stata Journal*, *6*(1), 58-82.

Williams, R. (2016). Understanding and interpreting generalized ordered logit models. *The Journal of Mathematical Sociology*, *40*(1), 7-20.

Wright, J. (2020, November 13). *Chinese Zodiac Elements: How to Know What Yours Is and What It Means*.PureWow.https://www.purewow.com/wellness/chinese-zodiac-elements#:~:text=There%20are%20five%20elements%20of,and%20interdependence%20between%20all%20things.

Yip, P. S., Lee, J., & Cheung, Y. B. (2002). The influence of the Chinese zodiac on fertility in Hong Kong SAR. *Social science & medicine*, 55(10), 1803-1812.

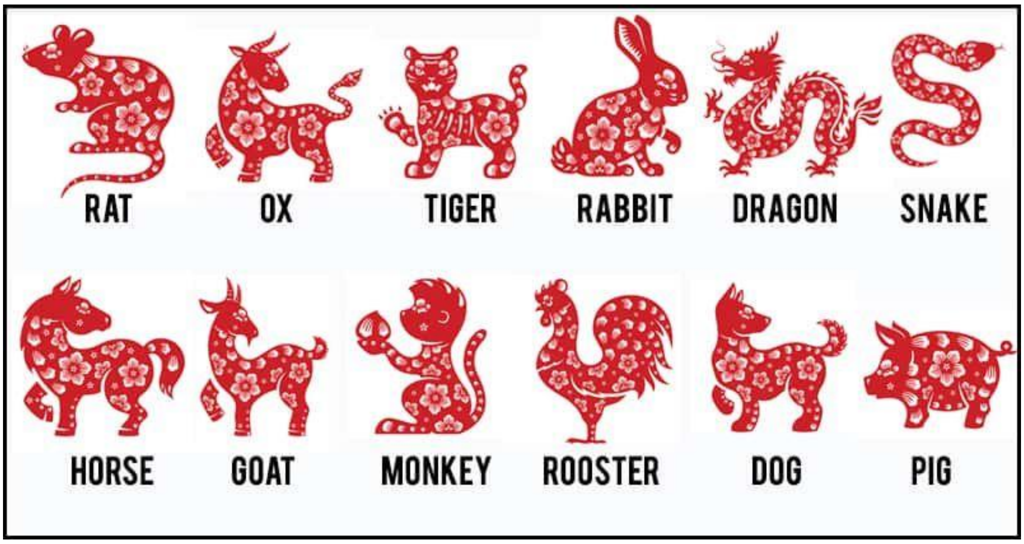
Zhang, J., & Xu, H. (2007). The effects of religion, superstition, and perceived gender inequality on the degree of suicide intent: a study of serious attempters in China. *OMEGA-Journal of Death and Dying*, *55*(3), 185-197.

Zwetsloot, R., & Peterson, D. (2019, December 31). *The US-China Tech Wars: China's Immigration Disadvantage*. – The Diplomat. https://thediplomat.com/2019/12/the-us-china-tech-wars-chinas-immigration-disadvantage/.

Appendix

Appendix A: Figures

Figure 1: The twelve Chinese zodiacs, in order



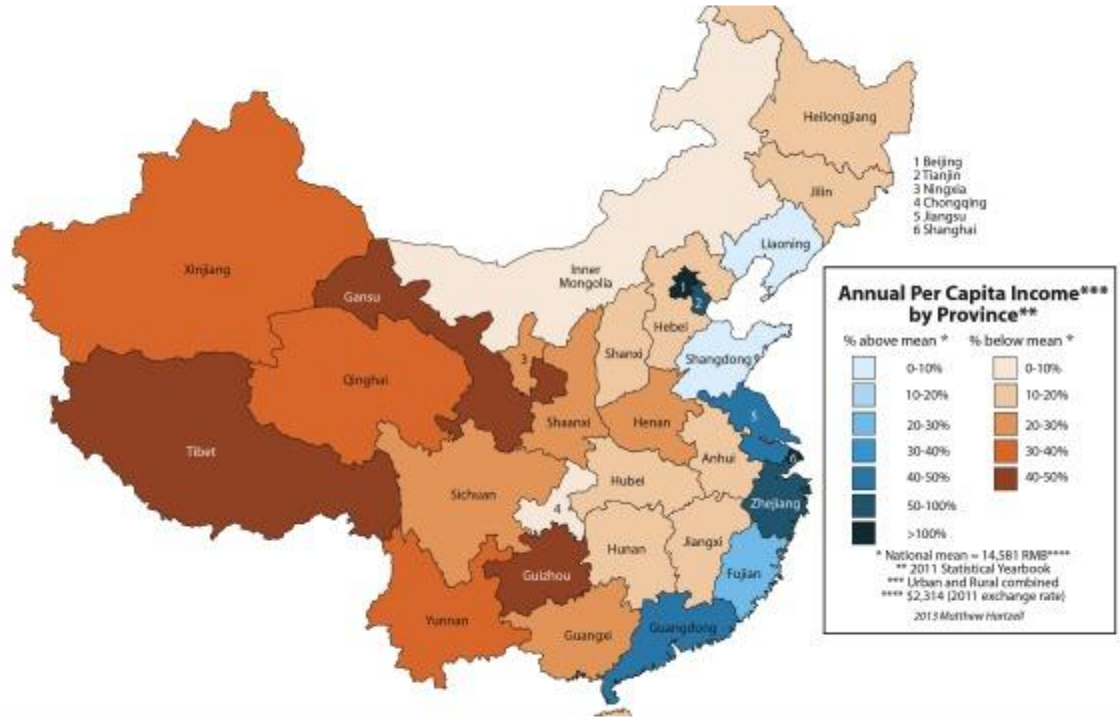
Source: Asmita S. (2019)

Figure 2: The start and end of the 5 Dragon-years used in this research

23 January 1928	9 February 1929	Earth Dragon
8 February 1940	26 January 1941	Metal Dragon
27 January 1952	13 February 1953	Water Dragon
13 February 1964	1 February 1965	Wood Dragon
31 January 1976	17 February 1977	Fire Dragon

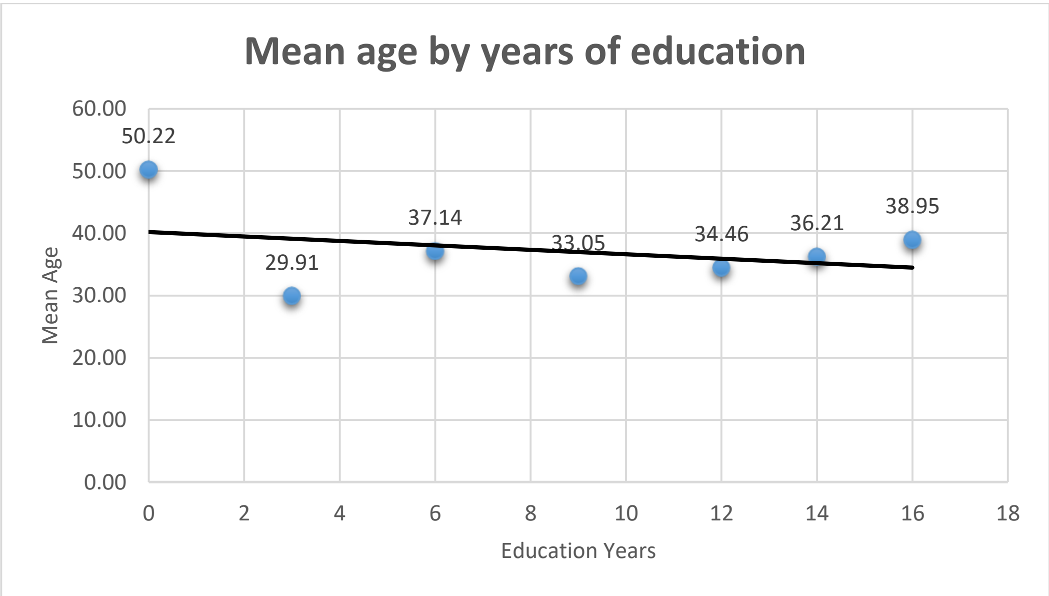
Note: The last column represents the heavenly branch compatibility. Source: Wikipedia (2005)

Figure 3: Regional income distribution in China



Source: (Schiavenza, 2013)

Figure 4: Scatterplot of the mean age and years of education of individuals in the sample



Note: Y-Axis represents the average age per-year category. X-axis represents the amount of completed education years.

Appendix B: Variables Used

Outcome Variables:

Education in years: Coded to years of education from the educational attainment for China Variable. If the initial variable indicated uncertainty regarding a level of schooling, the lower bound was chosen (i.e. secondary school not finished= 6 years of schooling). The levels are described below:

-0 years of schooling: Illiterate individuals; aged 6 and over knowing less than 1500 Chinese characters; unable to read basic literature or write informal notes.

-3 years of schooling: Individuals confirmed to be attending primary school and individuals confirmed to have not graduated primary school.

-6 years of schooling: People who have completed primary school and people confirmed to either be attending or have not graduated junior high-school

-9 years of schooling: Individuals that completed junior high-school and people confirmed to either be attending or have not graduated senior high-school, whether it is academic or technical

-12 years of schooling: Individuals that have completed senior high-school, whether it is academic or technical, and people confirmed to be attending or have not graduated college/university.

-14 years of schooling: People that completed a two-year college degree.

-16 years of schooling: Individuals with a completed bachelor's degree or higher. Bachelor degrees typically last 4 years in China.

Literacy level: Takes value 0 for individuals classified as illiterate and value 1 for individuals classified as literate. A person is considered literate if capable of both reading and writing. Illiterate people comprise 10.12% of the observations, with literate people representing 89.88% respectively.

Treatment variables: five different variables were constructed from the initial database containing years of birth: Dragon1976, Dragon1964, Dragon1952, Dragon1940 and Dragon1928. They revolve around the year of reference and a period of two years before and after the fact. Dragon1976 takes value 1 for people born in the year of the Dragon 1976 and value 0 otherwise (for 1974,1975,1977 and 1978). The same rule applies to the other 4 variables.

Placebo Variables: five different variables were constructed from the initial database containing years of birth: Placebo1973, Placebo1961, Placebo1949, Placebo1937 and Placebo1925. They revolve around the year of reference and a period of two years before and after the fact. Placebo1973 takes value 1 for people born in the year 1973 and value 0 otherwise (for 1971,1972,1974 and 1975). The same rule applies to the other 4 variables.

Other Variables:

Gender: Takes the value 1 for females and 0 for males. The sample is almost evenly distributed, with 50.71% males and 49.29% females.

Age in years: Can take values between 21 and 77

Birth year: The year the individual was born in, ranging from 1923 to 1978 in the kept data.

Birth time in months: Since data is available for birth years as well as for birth months, this variable was used to indicate the time of birth for the individual, in months.

Age in months: Difference between the time of the census, in months, and the birth time in months

Province: Categorical. Indicates one’s province of birth, within China and its afferent territories such as Hong Kong. Contains 32 provinces in total. The province treated as a baseline in all regressions conducted is Beijing.

Appendix C: Tables

Table 9: Frequencies and proportions of Dragons per cohort: IPUMS-International

	(1)	(2)	(3)
Variables	Dragons	Non-Dragons	Total
Dragon1976	175,506(19.78%)	711,642(80.22%)	887,148
Dragon1964	229,252(19.47%)	948,038(80.53%)	1,117,290
Dragon1952	171,742(21.31%)	634,128(78.69%)	805,870
Dragon1940	90,230(21.20%)	335,467(78.80%)	425,697
Dragon1928	53,037(20.72%)	202,941(79.28%)	255,978
Total	721,688(9.61%)	6,791,089(90.39%)	7,512,777

Note: Table displays the number of observations for the 5 birth cohorts used, with percentages in parentheses. Columns 1 and 2 signify the dragon and non-dragon classifications. The first 5 rows indicate the 5 birth cohorts used, with the year indicated as the treatment group and the 2 years before and after as the control group. For example: Dragon1976 takes value 1 if person is born in 1976 and 0 if born in 1974,1975,1977 or 1978.

Table 10: Frequencies and proportions for each level of schooling.

Years of Schooling	Observations
0	936,037(12.28%)
3	313,280(4.11%)
6	2,342,969(30.75%)
9	2,712,776(35.60%)
12	964,194(12.65%)
14	241,959(3.18%)
16	108,568(1.42%)

Note: The amount of education years in the left column (coding explained in Appendix B). The number of observations in the sample on the right, with percentages in parentheses

Table 11: Frequencies and proportions for each province.

Province	Observations	Province	Observations
Beijing	62,960(0.83%)	Hubei	340,088(4.46%)
Tianjin	58,420(0.77%)	Hunan	407,031(5.34%)
Hebei	437,789(5.75%)	Guangdong	383,756(5.04%)
Shanxi	196,798(2.58%)	Guangxi	259,425(3.40%)
Inner Mongolia	136,506(1.79%)	Hainan	38,063(0.50%)
Liaoning	291,967(3.83%)	Chongqing	187,928(2.47%)
Jilin	175,624(2.30%)	Sichuan	559,694(7.35%)
Heilongjiang	215,323(2.83%)	Guizhou	203,614(2.67%)
Shanghai	91,833(1.21%)	Yunnan	244,610(3.21%)
Jiangsu	488,562(6.41%)	Tibet	12,642(0.17%)
Zhejiang	304,894(4.00%)	Shaanxi	213,604(2.80%)
Anhui	379,174(4.98%)	Gansu	156,400(2.05%)
Fujian	190,445(2.50%)	Qinghai	26,882(0.35%)
Jiangxi	232,679(3.05%)	Ningxia	28,536(0.37%)
Shandong	629,185(8.26%)	Xinjiang	82,756(1.09%)
Henan	579,935(7.61%)	Hong Kong, Macau, Taiwan or foreign regions	2,690(0.04%)
		Total	7,619,783(100%)

Note: The 32 provinces of China are highlighted in the first and third columns. The second and fourth columns display the number of observations for the province of interest, with percentages in parentheses. Regions such as Hong Kong, Macau, Taiwan and Tibet are included in this analysis.

Table 12: Effect of gender on the treatment variables.

Variables	Dragon1976	Dragon1964	Dragon1952	Dragon1940	Dragon1928	All Dragons
Gender	-0.002***	0.002**	0	0	0.001	0.001***

*Note: The variables for the five birth cohorts and the variable indicating all individuals born in Dragon-years are regressed on the binary control variable gender. The first 5 columns indicate the 5 birth cohorts used, with the year indicated as the treatment group and the 2 years before and after as the control group. For example: Dragon1976 takes value 1 if person is born in 1976 and 0 if born in 1974,1975,1977 or 1978. ***P-value<0,01, **P-value<0,05, *P-value<0,1, no asterisk: P-value >0,1. Numbers are rounded to the third decimal.*

Table 13: Frequencies and proportions of Dragons per country: PISA.

	(1)	(2)	(3)
Average Score	Dragons	Non-Dragons	Total
Hong Kong	1,177(3.92%)	28,852(96.08%)	30,029
South Korea	2,241(6.82%)	30,632(93.18%)	32,873
Japan	2,752(7.68%)	33,102(92.32%)	35,854
Thailand	6,217(15.11%)	34,924(84.89%)	41,141
Total	12,387(8.85%)	127,510(91.15%)	139,897

Note: The rows imply the country classification; the columns differentiate between individuals born in Dragon years and individual born in non-Dragon years (in this case the year of the rabbit, or the earlier year). The number of observations is shown first, with the percentage out of the total in parentheses.

Table 14: Difference in difference estimates of the Dragon-effect on the three score categories and the average score: all four countries included.

	(1)	(2)	(3)	(4)
Variables	Average Score	Mathematics	Reading	Science
Young	-0.386***	-0.409***	-0.335***	-0.374***
2003v2006	-0.095***	- 0.053**	-0.136***	-0.088***
2003v2009	-0.015	-0.002	-0.031	-0.012
2003v2012	0.012	0.006	0.008	0.022
2003v2015	0.003	0.040*	-0.065**	0.032
2003v2018	0.008	0.022	-0.086***	0.087***
Observations	139,897	139,897	139,897	139,897

*Note: Row 1 depicts the general effect of being younger on the score category. Rows 2-6 illustrate the difference in difference estimates with Dragons in 2003 as a baseline. The only two relevant years for Dragons are 2015 and 2003, as they include individuals born in 1988 and 2000. The 2006,2009,2012 and 2018 comparisons are used as potential control groups due to the fact that they include people born in consecutive years, none of which are Dragons. Essentially, the coefficient for 2003vs2006 signifies the effect of being born in the later year (1991) compared to being born in the earlier year (1990) in 2006 compared to the baseline: the comparison between being born in the later year (1988) and the earlier year (1987) for 2003. Since 1990 is not a Dragon year and 1988 is, a positive effect of being a Dragon on test scores would imply a negative and significant coefficient for 2003v2006. The outcomes used are the standardized test scores for mathematics, reading and science (see Appendix D and Data) and a standardized average score. ***P-value<0,01, **P-value<0,05, *P-value<0,1, no asterisk: P-value >0,1. Numbers rounded to the third decimal.*

Table 15: Difference in difference estimates of the Dragon-effect on the standardized average score: four countries separately.

	(1)	(2)	(3)	(4)
Average Score	Hong Kong	South Korea	Japan	Thailand
Young	-0.070**	-0.052*	-0.114***	-0.113***
2003v2006	-0.079	- 0.046	0.044	-0.084***
2003v2009	-0.089	-0.064	0.032	0.027
2003v2012	-0.016	0.029	-0.004	0.032
2003v2015	-0.019	0.013	0.058	0.015
2003v2018	-0.053	0.051	0.063	0.035
Observations	30,029	32,873	35,854	41,141

*Note: Row 1 depicts the general effect of being younger on the standardized average score, for the 4 different countries in the database. This involves running the same regression 4 times, with a different country restriction. Rows 2-6 illustrate the difference in difference estimates with Dragons in 2003 as a baseline. The only two relevant years for Dragons are 2015 and 2003, as they include individuals born in 1988 and 2000. The 2006,2009,2012 and 2018 comparisons are used as potential control groups due to the fact that they include people born in consecutive years, none of which are Dragons. Essentially, the coefficient for 2003vs2006 signifies the effect of being born in the later year (1991) compared to being born in the earlier year (1990) in 2006 compared to the baseline: the comparison between being born in the later year (1988) and the earlier year (1987) for 2003. Since 1990 is not a Dragon year and 1988 is, a positive effect of being a Dragon on test scores would imply a negative and significant coefficient for 2003v2006. The outcome used is the standardized average score (see Appendix D and Data). ***P-value<0,01, **P-value<0,05, *P-value<0,1, no asterisk: P-value >0,1. Numbers rounded to the third decimal.*

Appendix D: PISA Databases

Imputation of plausible values in PISA

Let ρ_i be the conditional distribution of the unobserved proficiency score of pupil i, given covariates x_i representing background characteristics for the student and n known responses: r_i = (r_{1i}, r_{2i}, ..., r_{ni}). The multiple imputation method draws at random from the conditional distribution, expressed as:

f($\rho_i|r_i, x_i$) = $\prod_{i=1}^{n}$[Pr($R_{ji} = 1|a_j, b_j, p_i$)]$^{r_{ji}}$ X Pr($R_{ji} = 0|a_j, b_j, p_i$)]$^{(1-r_{ji})}$ f($\rho_i|x_i$, λ, σ)

where Pr($R_{ji} = 1|a_j, b_j, p_i$)=$[1 + \exp\left(-a_i(p_i - b_i)\right)]^{-1}$ and f($\rho_i|x_i$, λ, σ) is the normal distribution with standard deviation σ and mean x_i'λ. a_j shows how question j distinguishes between students and b_i depicts the perceived level of difficulty of the question. The answers are assumed independent given(a_j, b_j, p_i). Ten or five imputed values for ρ_i are drawn for every pupil i and added to the respective PISA database.

Outcome variables:

Mathematics, Reading, Science: The outcome variables are the standardized average scores for mathematics, reading and science. For each specific dataset, the 5 or 10 plausible values for each discipline are averaged and then standardized to have a mean of 0 and a standard deviation of 1. (see Appendix D for a more information regarding plausible values).

Average Score: The average scores of the three disciplines are added together and divided by 3. The number obtained is standardized. This variable is used to showcase general ability, irrespective of discipline.

Potential Independent Variables

Dragon: Takes value 1 for people born in the year of the Dragon and 0 otherwise, used in the 2003 and 2015 databases.

Fake-Dragon: Takes value 1 for people born in “false” year of the Dragon (later year) and 0 otherwise. Used in the 2006,2009,2012 and 2018 datasets.

Young: Equal to Dragon+ Fake-Dragon. Takes value 1 if individual is born in the later year of the two years included in each dataset, and 0 otherwise.