Understanding Schoolchildren Test Anxiety through Online Writing Analysis

Xuetao Tian[1] , **Liping Jing**[2,*] , **Fang Luo**[3] , **Enrui Cui**[3] and **Sheng Zhang**[3]

[1]School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China
[2]Beijing Key Lab of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing, China
[3]Collaborative Innovation Center of Assessment toward Basic Education Quality, Beijing Normal University, Beijing, China

{xttian, llpjing}@bjtu.edu.cn, {luof, zhangsheng}@bnu.edu.cn, 295114542@qq.com

Abstract

Test anxiety has a great influence on academic performance. How to scientifically and effectively identify whether a student has test anxiety, thus give timely help to relieve the effect of test anxiety is a matter worthy of attention. Compared with high-cost self-report scales, our research firstly tries to understand the performance of test anxiety and predict the degree of individual test anxiety through analyzing online writing. Based on word frequency and word type extracted from our compiled dictionary, the study finds that test anxiety is related to some writing habits, such as frequency of using words in *Ease* and type of using words in *Jealousy*. By applying machine learning techniques, models for test anxiety prediction based on online writing data are established and evaluated. Random forest regressor achieves the best performance for now to provide a baseline on test anxiety prediction, where over 20% high level of test anxiety can be predicted with precision being 50%.

1 Introduction

Test anxiety is a personality with which students will have tension or other inner experiences and stimulate negative emotions when they realize that a test they are going to face has a potential threat to themselves. Some early works [Bruch, 1981; Paulman and Kennelly, 1984] evidence that students with high or moderate levels of test anxiety adopt significantly less effective problem-solving strategies than low-anxiety students. And some other previous works [Tian and Guo, 2004; Fayegh *et al.*, 2010] demonstrates that there is a strong correlation between test anxiety and poor test scores. Therefore, how to scientifically and effectively identify whether a student has test anxiety, thus give timely help to alleviate or solve the effect of test anxiety is a matter worthy of attention.

Currently, test anxiety is measured and observed using questionnaires, that is self-report scales. The investigated subject is required to complete a standardized scale developed by professional psychologists. However, it is always high-cost and time-consuming. Inspired by some researches [Mairesse *et al.*, 2007; Majumder *et al.*, 2017] on mining personality traits from text data, our research firstly tries to understand the performance of test anxiety and predict the degree of an individual test anxiety through analyzing online writing data.

A dataset was established to support our research. We collected online writing data of 1, 231 schoolchildren from TeachGrid[1] and measured their degrees of test anxiety by using Student Common Part Questionnaire in PISA 2015[2]. Reference to LIWC-C [Huang *et al.*, 2012] and TextMind [Gao *et al.*, 2013], a dictionary applicable to schoolchildren is compiled, which contains 6, 000 high-frequency words in schoolchildren writing and consists of 118 parts of speech. Based on it, word frequency and word type are respectively extracted from each individual online writing data, where we look forward to discovering some writing habits related to test anxiety and predicting individual degree of test anxiety by introducing some machine learning techniques. Overall, the contributions in this paper mainly include:

a) a dataset established for schoolchildren test anxiety research;

b) analysis of the relationship between writing habits and test anxiety;

c) firstly trying to predict the degree of test anxiety through text data and machine learning techniques.

The following sections are structured as follows: In Section 2, we summarize some relevant theories on the definition and measurement of test anxiety. Section 3 illustrates the analysis results of the relationship between test anxiety and writing habits, where some details about the dataset are also described. The study on predicting the degree of test anxiety by machine learning techniques will be elaborated in Section 4. Section 5 presents some discussions on test anxiety before we conclude our research in Section 6.

2 Relevant Theories of Learning

Before illustrating our research, some relevant theories about the definition and measurement of test anxiety are introduced to clarify the background in this section. Some measurement results in our dataset are displayed incidentally.

[1]This is a website (https://www.jiaokee.com) for schoolchildren to write compositions and communicate with each other.

[2]https://nces.ed.gov/surveys/pisa/questionnaire.asp

Reliability	Items	Content Validity
	I often worry that taking a test will be difficult for me.	0.806**
	I worry that I will get poor grades at school.	0.828**
0.864	Even if I am well-prepared for a test, I feel very anxious.	0.835**
	I get very tense when I study for a test.	0.805**
	I get nervous when I don't know how to solve a task at school.	0.750**

Table 1: Test anxiety measurement by PISA 2015 (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

2.1 Definition of Test Anxiety

Scholars at home and abroad do not have a clear and unified definition of test anxiety. Initially, Mandler and Sarason [1952] considers test anxiety as an emotion in a state of helplessness and disorder. Spielberger [1974] believes that test anxiety refers to the disorder, tension and psychological arousal in evaluation scenarios, caused by the individual concerns about a wide range of situations. After that, Sarason [1990] refines previous theories and proposes that test anxiety is threat perception, low self-efficacy, self-depreciation or strong emotional reaction, which are easily produced or awakened in evaluation scenarios. At home, Tian and Guo [2004] absorb international research results and analyzes the actual situation to define test anxiety as a personality with which a student will produce complex emotional reactions, such as anxiety, low self-efficacy and other cognition, in the face of examination situations. Although the definition is always updated, the commonality among different statements from different scholars is to emphasize the destructiveness of test anxiety on psychological status in a specific situation.

2.2 Measurement of Test Anxiety

Since many studies have shown that test anxiety has a great influence on students' academic performance [Bruch, 1981; Paulman and Kennelly, 1984; Cassady and Johnson, 2002; Fayegh *et al.*, 2010; Carey *et al.*, 2017], the measurement of test anxiety has always been concerned. In the field of psychology, there are some developed questionnaires to measure test anxiety, such as The Scale of Test Anxiety Behavior [Suinn, 1969], State-Trait Anxiety Inventory [Spielberger, 1970], Test Anxiety Inventory [Spielberger, 1974] and Worry-Emotionality Scale [Morris and Engle, 1981]. Accordingly, some localization studies have been carried out. For example, Wang [2003] from China adapts Spielberger Test Anxiety Inventory and achieves a good reliability (Cronbach's α = 0.9) in domestic experiments.

Student Common Part Questionnaire in PISA 2015 increases the measurement of test anxiety, which is widely adopted in recent years [Qiao and Liu, 2017; Zhao and Huang, 2018]. In our research, the ground truth of test anxiety measurement is obtained by using it as well. This questionnaire consists of 5 items rated on 4-point scale of 1 (strongly disagree) to 4 (strongly agree) as shown in Table 1. Based on our 1, 321 subjects, the reliability and content validity are tested and perform well, where Cronbach's α is 0.864 and corrected item-total correlations are 0.806, 0.828, 0.835, 0.805 and 0.750.

3 Test Anxiety and Writing Habits

In this section, the relationship between individual test anxiety and writing habits is explored. We use word frequency and word type to express writing habits and find some performances that are related to individual test anxiety.

3.1 Data Preparation

Student Common Part Questionnaire in PISA 2015 quantifies individual test anxiety as a value of 5 to 20. The overall distribution of the measurements of all subjects is shown in Figure 1, where the mean is 11.73 and the standard deviation is 3.92. All subjects are divided into three groups, with low level of test anxiety being 5 to 9, high level being 16 to 20 and middle level being the rest. The number of subjects is respectively 336, 192 and 703.

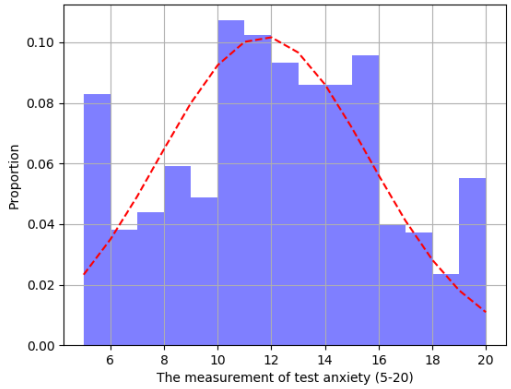

Figure 1: Overall distribution of test anxiety measurements

We select 6, 000 high-frequency words in schoolchildren writing to compile a dictionary. Each word is mapped to a subset of 109 parts of speech referring to LIWC-C and TextMind and the size of a subset is not more than 8. Another 9 parts of speech like WordCount, and WordPerSentence will be generated separately. Based on the dictionary, word frequency and word type are extracted, where word frequency is "what the proportion of the used words is in total words for each part of speech" and word type is the proportion of "how many kinds of words are used for each part of speech".

3.2 Correlation Analysis

To explore the difference between writing habit in each part of speech and individual degree of test anxiety, independent sample *t*-test [Abbott, 2016] is applied to each pair of groups

Group	Part of speech	Level	*Mean*	*SD*	*t*	*df*	*p*
1. High-low test anxiety	Ease	High	0.160	1.474	-2.058	239.445	0.041
		Low	-0.072	0.688			
2. High-middle test anxiety	Doubt	High	0.200	1.537	-2.109	222.277	0.036
		Low	-0.043	0.830			
3. Middle-low test anxiety				None			

Table 2: Independent sample *t*-test with word frequency between any two groups

Group	Part of speech	Level	*Mean*	*SD*	*t*	*df*	*p*
1. High-low test anxiety	Number	High	0.136	1.102	-2.073	526	0.039
		Low	-0.058	0.992			
	Family	High	0.124	1.136	-2.021	335.258	0.044
		Low	-0.070	0.923			
	Joy	High	0.117	1.138	-2.037	526	0.042
		Low	-0.073	0.967			
	Ease	High	0.165	1.152	-2.485	526	0.013
		Low	-0.061	0.911			
	Wish	High	0.165	1.152	-2.499	526	0.013
		Low	-0.061	0.911			
	Jealousy	High	0.100	1.144	-2.035	313.133	0.043
		Low	-0.093	0.851			
2. High-middle test anxiety	PastM	High	0.144	1.097	-2.035	893	0.042
		Low	-0.021	0.970			
	ProgM	High	0.154	1.010	-2.431	893	0.015
		Low	-0.044	0.999			
	Swear	High	0.131	1.035	-2.053	893	0.04
		Low	-0.034	0.970			
	Ease	High	0.165	1.152	-2.149	893	0.032
		Low	-0.016	0.994			
3. Middle-low test anxiety				None			

Table 3: Independent sample *t*-test with word type between any two groups

of subjects. It should be noted that the data were normalized before do this test. The results with word frequency and word type of the front 109 parts of speech are displayed respectively in Table 2 and Table 3. From the results, test anxiety shows a correlation with some writing habits in both word frequency and word type.

Before explaining the results, the meanings or some examples of the mentioned parts of speech are presented in Table 4 to make it clear. Under the representation of word frequency, schoolchildren with high level of test anxiety tend to use more words in *Ease* and *Doubt*. We would like to explain it that high level of test anxiety leads to self-doubt, with which schoolchildren will express self-comfort by writing. Under the representation of word types, more phenomena were discovered. *Ease* and *Joy* would be explained as self-comfort as similar as the above. Schoolchildren with high level of test anxiety use more words in *Wish*, which may indicate their expectations for the outcome of many things and *Family* expresses their desires to depend on somebody. More seriously, there is a significant correlation between test anxiety and *Jealousy* or *Swear*, which supports that test anxiety can produce strong emotional reaction [Sarason *et al.*, 1990].

Part of speech	Meaning (m) or examples (e)
Number	(e) one, hundred, thousand
Family	(e) family, mother, brother
Joy	(e) happy, laugh, easy
Ease	(e) well, I can, contented
Wish	(e) strive, heartfelt, I hope
Jealousy	(e) refuse, envy, disdain
PastM	(m) adverbs of past time
ProgM	(m) adverbs of continuity time
Swear	(e) low-down, shameless, bah
Doubt	(e) unexpectedly, disbelieve

Table 4: Meaning or examples of mentioned parts of speech

As for *Number*, *PastM* and *ProgM*, we think this is because schoolchildren with test anxiety should be more concerned about time than others.

4 Prediction of test anxiety

Among the existing personality prediction researches, the problem is formalized into a supervised classification [Wei

Model	Methods/Params	Training set			Testing set		
		RMSE	MAE	R^2	RMSE	MAE	R^2
LR	Basic LR	3.7650	3.0895	0.0633	4.0319	3.2018	0.0387
	Ridge	3.7650	3.0895	0.0633	4.0319	3.2019	0.0387
	Lasso	3.8079	3.1284	0.0418	4.0160	3.2395	0.0462
	ElasticNet	3.7937	3.1126	0.0489	4.0501	3.2617	0.0299
SVR	Linear kernel	3.7850	3.0751	0.0533	4.0228	3.1935	0.0430
	RBF kernel	3.7377	3.0099	0.0768	4.1054	3.2789	0.0032
	Poly(2) kernel	3.7930	3.0467	0.0493	4.0876	3.3194	0.0119
TR	CART	3.7425	3.0559	0.0744	3.9928	3.2003	0.0572
	RFR	**3.3988**	**2.7613**	**0.2367**	**3.9456**	3.1374	**0.0794**
MLP	Layers -16-	3.6913	3.0172	0.0996	4.0517	3.1973	0.0292
	Layers -16-4-	3.5640	2.8797	0.1606	3.9878	3.1978	0.0596
	Layers -64-	3.6361	2.9662	0.1264	3.9943	3.1528	0.0565
	Layers -64-16-	3.4408	2.7729	0.2177	3.9488	**3.1023**	0.0778
OWF + RFR		3.7146	3.0354	0.0882	4.0064	3.1897	0.0508
OWT + RFR		3.7129	3.0413	0.0890	4.0212	3.1937	0.0437

Table 5: Regression results with CWFT for test anxiety

et al., 2017] or regression [Majumder *et al.*, 2017] issue. In our research on test anxiety, all current studies indicate that high level of test anxiety should be paid more attention to. However, we think that low level of test anxiety is also noteworthy, which will be discussed in the next section. Thus, we try to predict test anxiety with the idea of regression or three-category classification.

4.1 Settings

In our research, we try to use various models to regress the degree of test anxiety, including linear regression (LR), support vector regression (SVR), tree regression (TR) and multi-layer perceptron (MLP). For each regressor, we test three different inputs: only word frequency in the 118 parts of speech (OWF), only word type in the 118 parts of speech (OWT) and correlative word frequency or word type mentioned in subsection 3.2 and the values in the last 9 parts of speech (CWFT). For the classification issue, we adopt two strategies: various machine learning classifiers and directly mapping the regression results to three categories as subsection 3.1. We select evaluation criteria RMSE, MAE and R^2 for regression and Precision, Recall and F1-score for classification.

4.2 Experimental Results

To concurrently observe the accuracy and generalization of a regressor, the dataset is split into training set and testing set with a ratio of 9:1 considering the small size of our dataset. For each type of regression model, we choose various methods, parameters or structures and select appropriate hyperparameters by grid search. Experimental results indicates any regressor with CWFT performs better than it with OWF or OWT. So we only display the results with CWFT and the best results with OWF and OWT because of space limitations. From the results in Table 5, random forest regressor (RFR) with max-depth being 5 achieves a good performance for the regression issue and shows a better generalization based on our dataset.

For the three-category classification, various machine learning classifiers are applied and fine-tuned, among which the best result is obtained by random forest classifier (RFC) with max-depth being 10 using CWFT and displayed in Table 6 and 7. For directly mapping the regression result of RFR to classification result, there is a special operation to illustrate. The original value of the degree of test anxiety is 5 to 20 with the mean being 11.73 and the standard deviation being 3.92. The mean and standard deviation of the prediction value are 11.70 and 1.23 respectively. As we know, any nonstandard normal distribution can be transformed into a standard normal distribution. According to it, it is checked that 10.84 and 13.04 are the boundary of three groups of test anxiety instead of 9 and 16. The classification performance is shown in Table 6, of which confusion matrices are displayed in Table 7. RFR has a poor performance in training set, but is shows a better generalization than RFC. Based on the fact that the error rate is very low for misjudgement between high and low level of test anxiety, we believe that both RFC and RFR to predict test anxiety are feasible.

5 Discussions

There are some issues that we would like to point out and discuss. The first one is whether only high level of test anxiety should be paid attention to. Of course we believe that high level of test anxiety has a great influence on students' academic performance as illustrated above, but low level of test anxiety may lead to low grades sometimes. As illustrated as Yerkes-Dodson Law [Broadhurst, 1957], appropriate tension is conducive to bring normal performance into test. In that way, low level of test anxiety should also be concerned, at least before some important examinations. Thus, what we want to emphasize is that it is correct to formalize test anxiety prediction into a regression or three-category classification issue.

The next one is whether it is possible to design a scale of writing style rather than multi-choice questions for test anxiety measurement. Our research has shown that there is a

Method	Groups	Training set			Testing set		
		Precision	Recall	F1-score	Precision	Recall	F1-score
RFC	Low	**0.9871**	**0.7459**	**0.8497**	0.3333	0.2069	0.2553
	Middle	**0.8177**	**0.9984**	**0.8991**	0.6200	**0.8611**	**0.7209**
	High	**1.0000**	**0.6331**	**0.7753**	**0.6667**	0.1739	0.2758
RFR	Low	0.6316	0.3518	0.4519	**0.3684**	**0.2414**	**0.2917**
	Middle	0.6516	0.8653	0.7434	**0.6211**	0.8194	0.7066
	High	0.6939	0.4024	0.5094	0.5000	**0.2174**	**0.3030**

Table 6: Three-category classification results for test anxiety

Method	Groups	Prediction in training set			Prediction in testing set		
		Low	Middle	High	Low	Middle	High
RFC	Low	229	78	0	6	23	0
	Middle	3	628	0	8	62	2
	High	0	62	107	4	15	4
RFR	Low	108	193	6	7	22	0
	Middle	61	546	24	8	59	5
	High	2	99	68	4	14	5

Table 7: Confusion matrices of the classification results

relationship between test anxiety and some writing habits. If the context of schoolchildren writing is designed, some appointed habits may be better displayed. We expect that relevant research is carried out by some psychologists.

The last one is to explore better ways to predict test anxiety. As shown in Table 5, although RFR achieves a good performance in our dataset, it can be observed that the performance of MLP regressor gets better as its structure gets deeper and wider. We believe that deep model can better predict the degree of test anxiety after more data has been collected. Some deep learning techniques for natural language processing may promote this research as well.

6 Conclusions

Test anxiety has a great influence on students' academic performance. The current measurements of test anxiety rely on self-report scales, which are always time-consuming and high-cost. Our research firstly tries to understand the performance of test anxiety and predict the degree of an individual test anxiety through analyzing online writing data. A dataset from 1, 231 subjects is established to support it. The study indicates test anxiety is related to some writing habits with the representation of word frequency and word type. To predict the degree of test anxiety, we applied various machine learning techniques. From experiments, random forest regressor shows the best performance for now to provide a baseline on test anxiety prediction, where over 20% high level of test anxiety can be predicted with precision being 50%.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61822601, 61773050, and 61632004; the Beijing Natural Science Foundation under Grant Z180006; the Beijing Municipal Science & Technology Commission under Grant Z181100008918012; National Key Research and Development Program (2017YFC1703506).

References

[Abbott, 2016] Martin Lee Abbott. *Independent Sample T Test*. Using Statistics in the Social and Health Sciences with SPSS® and Excel®, John Wiley & Sons, Inc, 2016.

[Broadhurst, 1957] P. L. Broadhurst. Emotionality and the yerkes-dodson law. *Journal of Experimental Psychology*, 54(5):245–352, 1957.

[Bruch, 1981] Monroe A. Bruch. Relationship of test-taking strategies to test anxiety and performance: toward a task analysis of examination be-havior. *Cognitive Therapy & Research*, 5(1):41–56, 1981.

[Carey *et al.*, 2017] Emma Carey, Amy Devine, Francesca Hill, and Denes Szucs. Differentiating anxiety forms and their role in academic performance from primary to secondary school. *Plos One*, 12(3), 2017.

[Cassady and Johnson, 2002] Jerrell C. Cassady and Ronald E. Johnson. Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, 27(2):270–295, 2002.

[Fayegh *et al.*, 2010] Yousefi Fayegh, Mansor Abu Talib, Mariani Mansor, and Juhari Rumaya. The relationship between test-anxiety and academic achievement among iranian adolescents. *Asian Social Science*, 6(5):100–105, 2010.

[Gao *et al.*, 2013] Rui Gao, Bibo Hao, He Li, Yusong Gao, and Tingshao Zhu. Developing simplified chinese psychological linguistic analysis dictionary for microblog. In *International Conference on Brain & Health Informatics,*, pages 119–128, Maebashi, Japan, Oct 2013. Springer.

[Huang *et al.*, 2012] Chin-Lan Huang, Cindy K. Chung, N. Hui, Yi-Cheng Lin, Yi-Tai Seih, BC Lam, and James W. Pennebaker. Development of the chinese linguistic inquiry and word count dictionary. *Chinese Journal of Psychology*, 54(2):185–201, 2012.

[Mairesse *et al.*, 2007] Francois Mairesse, Marilyn A. Walker, Matthias R. Mehl, and Roger K. Moore. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30:457–500, 2007.

[Majumder *et al.*, 2017] Navonil Majumder, Soujanya Poria, Alexander Gelbukh, and Erik Cambria. Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, 32(2):74–79, 2017.

[Mandler and Sarason, 1952] George Mandler and Seymour B. Sarason. A study of anxiety and learning. *Journal of Abnormal Psychology*, 47(2):166–173, 1952.

[Morris and Engle, 1981] Larry W. Morris and Benjamin W. Engle. Assessing various coping strategies and their effects on test performance and anxiety. *Journal of Clinical Psychology*, 37(1):165–171, 1981.

[Paulman and Kennelly, 1984] Ronald G. Paulman and Kevin J. Kennelly. Test anxiety and ineffective test taking: different names, same construct? *Journal of Educational Psychology*, 76(2):279–288, 1984.

[Qiao and Liu, 2017] Naxi Qiao and Xiaoyu Liu. The present situation of campus bullying and its influence on life satisfaction and test anxiety. *The Guide of Science & Education*, 31:191–192, 2017.

[Sarason *et al.*, 1990] Irwin G. Sarason, Barbara R. Sarason, and Gregory R. Pierce. Anxiety, cognitive interference, and performance. *Journal of Social Behavior & Personality*, 5(2):1–18, 1990.

[Spielberger, 1970] Charles D. Spielberger. *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press, Inc.: Palo Alto, 1970.

[Spielberger, 1974] Charles D. Spielberger. Case seminar of group mental health consultation. *Professional Psychology*, 5(3):303–308, 1974.

[Suinn, 1969] Richard M. Suinn. The relationship between fears and anxiety: A further study. *Behaviour Research and Therapy*, 7(3):317–318, 1969.

[Tian and Guo, 2004] Bao Tian and Dejun Guo. The basic model of test anxiety influencing test performance. *Psychological Science*, 27(6):1360–1364, 2004.

[Wang, 2003] Caikang Wang. The reliability and validation of chinese version test anxiety inventory in college students. *Chinese Journal of Clinical Psychology*, 11(1):69–70, 2003.

[Wei *et al.*, 2017] Honghao Wei, Fuzheng Zhang, Nicholas Jing Yuan, Chuan Cao, Hao Fu, Xing Xie, Yong Rui, and Wei-Ying Ma. Beyond the words: Predicting user personality from heterogeneous information. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining,*, Cambridge, UK, Feb 2017. ACM.

[Zhao and Huang, 2018] Decheng Zhao and Liang Huang. B-s-j-g (china) and singapore students' science literacy and its influencing factors: An analysis on pisa 2015 data. *Journal of Beijing Normal University*, 2:23–31, 2018.