The deadly effect of high-stakes testing on teenagers with reference-dependent preferences

Liang Choon Wang[*]

Abstract

This paper explains why suicidal tendency and test performance of teenagers may not be inversely related when individuals have reference-dependent preferences. Using panel survey data of South Korean secondary school students, I show that the relationship between suicidal ideation and test performance is consistent with reference-dependent preferences. When a student's rank in the high-stakes College Scholastic Ability Test falls below her average ranks in prior national examinations, she exhibits greater suicidal tendency. The reference dependent effects, however, are absent for low-stakes in-school academic performance. The findings highlight the potential adverse consequences of disappointment in high-stakes testing.

Keywords: High-Stakes Testing, Reference-Dependent Preference, Suicide, Suicidal Ideation, Korea.
JEL codes: I12, I21, I31.

[*]Affiliation: Monash University, Department of Economics. Postal Address: Monash University, Department of Economics, Clayton, VIC 3800, Australia. Phone: + 61 3 9905 5448. Email: liang.c.wang@monash.edu.

1. Introduction

Suicide is now one of the leading causes of death for young adults aged 10 to 24 (Patton et al. 2009). According to Cutler et al. (2001), youth suicide rates in the U.S. tripled between 1950 and 1990, while suicide rates for adults and elderly were falling during the same period of time. On the other hand, examination, especially a high-stakes one, is increasingly common as a mean to quantify learning, improve achievement, and hold schools accountable. Although the evidence regarding the effects of testing on learning outcomes is mixed, some studies have shown that achievement is higher when high-stakes testing (Roderick et al. 2002), low-stakes diagnostic testing (Betts et al. 2011), or curriculum based examination (Bishop 1998) is in place, suggesting the usefulness of testing in improving outcomes.[1] However, a cross-country correlation also reveals that teenagers living in countries with high international standardized test performance are more likely to commit suicide (figure 1). The pattern raises the questions of why suicidal tendency and test score are positively correlated and whether testing may have adverse consequences on young individuals.

Existing economic models of suicide are useful in explaining the relationships between suicidal tendency and social, economic, or demographic factors, but they tend to suggest a negative relationship between test score and suicide or suicide attempt.[2] For example, given the positive economic return to higher achievement, Hamermesh and Soss's (1974) rational choice model predicts lower suicide rates among high-achieving students as better test score implies greater permanent income and utility. Similarly, Rosenthal's (1993) strategic-suicide model also does not suggest a positive relationship between suicide attempt and achievement, as one would not expect high-achieving students to be unhappy and attempt suicide to signal their unhappiness for help. Other theories of suicides, such as the contagion theory and the instrumentality hypothesis that Cutler et al. (2001) highlight, are less readily applicable to the relationship between suicidal tendency and test performance. Contagion theory of suicide cannot offer a clear reason for why high-achieving students would first commit suicides and then trigger other high-achieving students to also commit suicides. The instrumentality hypothesis also does not

---

[1] Some other papers show no or negative effect of testing on a range of outcome measures. For example, Muralidharan and Sundararaman (2010) find no evidence that low-stakes diagnostic testing improves students' academic performance in India. Glewwe et al. (2003), Figlio and Winicki (2005), and Jacob and Levitt (2003) show that the presence of high-stakes testing leads to perverse practices among teachers and schools.

[2] The study of the demographic, social, and economic determinants of suicides dated back to Emile Durkheim's *Suicide* in 1897.

1

immediately provide an explanation for why students in high-achieving places would have greater access to lethal means.

This paper explains why suicidal tendency may not be inversely related with students' test performance when students have reference-dependent preferences (Tversky and Kahneman 1991). When students display Kahneman and Tversky's (1979) form of loss aversion, a worse-than-expected performance in high-stakes examination will lead to greater suicidal tendency. If students have high degree of loss aversion, the negative effect of performing worse than expected can potentially undermine the positive effect of good performance on subjective wellbeing that standard economics of suicide model predicts. The reference-dependent explanation is agnostic as to whether teenagers attempt suicides because they are determined to die or they do so to signal unhappiness in order to influence resources provided to them. This reference-dependent effect on suicide is closely related to Baumeister's (1990) escape theory of suicide in the psychology literature, which hypothesizes that individuals experience increased suicide risks when they fail to attain an important standard.

Using panel survey data of South Korean secondary school students, I test whether the relationship between suicidal ideation and test performance is consistent with reference-dependent preferences. South Korea (hereafter Korea) provides a unique opportunity to examine how deviation of youths' outcomes away from their reference points can increase the risk of suicide. Statistics Korea (2010) reports that suicide was the leading cause of death of Korean individuals aged 15-24 in 2008. In addition to taking a number of relatively low-stakes national and in-school tests, secondary school students intending to attend college in Korea must also undertake the college entrance examination, the College Scholastic Ability Test (CSAT). CSAT is a high-stakes examination, as it is administered once every year and an individual's CSAT performance determines her likelihood of entering into her choice of university and major. This feature of the Korean schooling system facilitates an investigation of the reference-dependent effects of testing on suicidal ideation of teenagers.

The results show that students' performance in low-stakes examination is not negatively correlated with self-reported suicidal ideation even after controlling for individual and parental characteristics, other potential influences of suicidal ideation, and past suicidal ideation. Specifically, the expected sign of test performance is ambiguous because someone performing well in the low-stakes examination may also do well in the high-stakes examination, lowering

2

suicide risks, but the person may also have higher expectation, increasing the suicide risks when the performance in the high-stakes examination falls short of her expectation. To examine the predictions of the reference-dependent explanation for suicidal risks, I use students' average past national ranks in the relatively low-stakes National Assessment of Educational Achievement (NAEA) examinations before they took CSAT to proxy their expected ranks in CSAT and the difference between the realized and expected ranks in CSAT to capture the deviation from the reference point to test for the reference-dependent effect. The constructed reference point is hence fixed as in Tversky and Kahneman (1991). The estimates show that deviation of test performance below the reference point leads to greater suicidal ideation. The effects are robust to controlling for past suicidal ideation, past academic performance, and other correlates of academic performance and suicidal tendency. The findings are consistent with the reference-dependent model of suicide. In contrast, the deviation of current rank from past rank of in-school academic performance does not predict suicidal ideation of students in grade eight. Although actual suicide is not observed, the strong correlation between suicidal ideation and suicide attempts that Dubow et al. (1989) documented suggests the results are useful for understanding suicidal risks. The results imply that the combination of reference-dependent preferences and high-stakes examination poses significant suicide risks for young individuals.

This paper is closely related to the literature on the relationship between academic performance and suicide risks of youths, where evidence to date has been mixed. For example, studies by Braaten and Darling (1962), Hendin (1982), and Alaraisanen et al. (2006) show that suicidal individuals have better grades, whereas Park et al. (2006), Dubow et al. (1989), Gunell et al. (2005), and Bjarnason and Thorlindsson (1994) show that suicidal tendency and academic performance are inversely related. This paper uses reference-dependent preferences to demonstrate why the empirical relationship between academic performance and suicide risks is mixed and links the economics literature with the escape theory of suicide by Baumeister (1990) in the psychology literature. The results highlight why it is important to identify reference points and distinguish between high-stakes and low-stakes testing. Despite the reliance on survey data that are susceptible to measurement errors and misreporting, the results demonstrate the usefulness of panel data with rich information on academic performance and suicidal ideation of individuals to examine reference-dependent effects of testing on suicide risks. The findings call for more attention to the potential unintended consequences of high-stakes testing and possible

policy remedies to safeguard students against depression and suicide risks. Finally, this study also extends the growing literature on the importance of reference-dependent effects, which has centered on labor supply decisions (Camerer et al. 1997; Farber 2005, 2008; Crawford and Meng 2012) and violence behaviors (Card and Dahl 2011).

2. A Reference-Dependent Explanation for Suicide Risks and Test Performance

In this section, I incorporate reference-dependent preferences into a simple model of suicide to highlight why test performance and suicidal tendency may not necessarily be inversely related as standard economic models of suicide imply. This simple model guides the empirical specifications to examine whether changes in suicidal ideation are consistent with reference-dependent preferences.

Assume student i who has reference-dependent preference and maximizes the following single-period welfare function:

$$W_i = \beta U(c_i, l_i) + \alpha G(c_i - \tilde{c}_i),$$

subject to the time constraint $T = l_i + h_i$, where l_i denotes leisure hours and h_i denotes study hours. $U(\cdot)$ stands for the traditional utility function and $G(\cdot)$ represents the Kahneman and Tversky's (1979) behavioral gain-loss function. β and α are weights attached to traditional utility $U(\cdot)$ and behavioural gain/loss $G(\cdot)$, respectively. c_i denotes the student's test performance. Higher test score brings greater welfare to the student through higher likelihood of college attendance. To focus attention on the effect of test performance on suicidal tendency, I assume away the decisions to attend college and the subsequent labor market choices. The higher the stakes of the test, the greater the marginal return of c_i to W_i. $\tilde{c}_i$ captures the student's expectation of her performance in the exam. For simplicity, assume that expected performance in the high-stakes college entrance exam is formed on the basis of prior average national test performance:

$$\tilde{c}_i = x_i$$

Note that the value of x_i does not enter into $U(\cdot)$ as the prior test performance does not bring direct benefit to the student. The student does not anticipate what the reference point will be, but form a fixed reference point that is entirely based on her performance relative to every other

student's.[3] This fixed reference-point assumption is consistent with Tversky and Kahneman's (1991) model.

The student's problem is to choose leisure hours l_i and in turn her study hours h_i. Her study hours affect how well she does in the high-stakes college entrance exam. Since the student cannot perfectly control how well she does in the high-stakes college entrance exam, we have:

$$c_i = F_i(h_i) + \epsilon_i$$

$F_i(h_i)$ is the student's test score production function and ϵ_i captures some idiosyncratic random factors that are outside the student's control. The test score production function differs across individuals and is increasing in time spent studying, $F_i'(h_i) > 0$. Individuals with higher innate ability and better family influences do better for a given study hour.

For simplicity and in the spirit Hamermesh and Soss (1974), assume that the student will enter into depression and commit suicide when $W_i < W^*$, where W^* denotes some threshold level of welfare for an individual to sustain life. Alternatively, we may allow for the possibility of strategic suicide as in Rosenthal's (1993) model and assume that the student will attempt suicide when $W_i < W^*$ so as to signal unhappiness to their parents, for example, and to change how family resources are distributed. The student's welfare is increasing in test score when $\alpha = 0$. With reference-dependent preferences, the welfare enhancing effect of test performance coming from $\beta U(\cdot)$ may be more than offset by the feeling of loss driven by poorer than expected test performance $\alpha G(\cdot)$. Thus, differences in suicide risks between two students are driven by differences between the realized test scores and the reference points, as well as the differences in the levels of realized test scores.

The model offers several testable predictions that I will examine empirically in the next few sections. First, the higher is c_i, the realized performance, the lower is the student's suicidal tendency, holding all else constant. Second, if a negative shock ϵ_i leads to the realized performance c_i falling below the reference point $\tilde{c}_i$, the student will experience higher suicide risks. Likewise, a positive shock will diminish a student's suicide risks. Finally, if an examination does not provide information about c_i and $\tilde{c}_i$, then it will have no effect on the student's suicidal tendency.

---

[3] See Koszegi and Rabin (2006) for a discussion on anticipated reference point. I do not model anticipated reference point because the data used do not have high frequency and detailed data on time use and other effort variables that may be used to estimate the anticipated reference points.

3. Background and Data

The empirical focus of the paper is the secondary school students in metropolitan cities and provinces in Korea. Major metropolitan school districts in Korea have gradually implemented the so-called "equalization policy" since the 1970s. The equalization policy essentially abolished all entrance or exit exams at the primary and secondary levels and introduced randomization of students into secondary schools (middle and high schools). Although there is no longer any entrance and exit exams, students in grade 6, grade 9, grade 10 and grade 11 are administered the National Assessment of Educational Achievement (NAEA) tests each year to inform school district offices and public about the distribution of students at different proficiency levels in each school (Kim et al. 2003). In addition, students intending to apply for college admission must take the College Scholastic Ability Test (CSAT) on one day of November during grade 12 (Kim et al. 2010). CSAT has five sections, covering verbal ability, quantitative ability, foreign language ability, and other electives, such as social and physical sciences. As all sections must be completed within the same day, the test usually takes a whole day to complete. Official CSAT scores are typically released in December. Students' CSAT scores play a major role in determining whether they are admitted into their desired universities.

This paper uses the 2003 Korean Youth Panel Survey (KYPS) – Junior High School Sample and the 2004 KYPS – Elementary School Sample.[4] In 2003, KYPS sampled roughly 3500 second-grade (grade 8) junior high school students across 12 metropolitan cities and provinces and followed them over a six-year period. In 2004, KYPS sampled roughly 3000 fourth-grade elementary school students across 15 metropolitan cities and provinces and followed them over a five-year period. Questions regarding personal and family characteristics, career plan and preparation, academic performance, school activities, social deviances, self-identity, attitudes, and so forth were asked in several waves of the surveys. I use the junior high school sample to perform the analyses related to the effects of high-stakes testing, and use the elementary school sample to perform analyses related to the effects of low-stakes testing.

Beginning in 2004, the junior high school and elementary school respondents answered questions related to suicidal ideation and other measures of mental wellbeing, such as sadness, loneliness, satisfaction with life, and so on. I code suicidal ideation as a binary variable, which

---

[4] The data are available at the Nation Youth Policy Institute's (NYPI) Youth and Children data archive: http://e-archive.nypi.re.kr/

takes the value of 1 when the student agreed or strongly agreed with the question: "do you sometimes feel suicidal without apparent reason?" There are more individuals with suicidal ideation than individuals who actually attempt suicide and successfully commit suicide. On the other hand, individuals having suicidal thoughts do not always attempt suicide and even if they attempt, many do not succeed. For example, it takes roughly 100-200 suicide attempts for every one completed suicide by an American 15-24 years-old (Goldsmith et al. 2002). Amongst twelfth grade students in the U.S., roughly 13.5% seriously considered suicide in 2007, but only 5.4% actually attempted suicide (Centers for Disease Control and Prevention 2008). Table 1 provides summary statistics of key variables drawn from the junior high school sample.

Right after completing CSAT in 2007, roughly 8.6% of the respondents indicated that they sometimes felt suicidal with no apparent reason, comparing to 10.2% in the prior year. The fraction of respondents exhibiting suicidal ideation fell to 5.7% one year after they took CSAT. These figures are much lower than the suicide ideation rate of 13.5% reported for American twelfth graders in 2007 (Centers for Disease Control and Prevention 2008). Given that Korea and the United States have similar suicide rates for 15-19 years-olds (9 per 100,000 persons vs. 7 per 100,000 persons), the suicidal tendency measure used in this paper is probably a conservative proxy for suicidal ideation.[5]

[Table 1]

Table 1 also reports summary statistics of the various test performance used. KYPS asked students how they ranked nationally in the most recent national examination (NAEA) in each wave of the survey. Since students were randomly sampled to take NAEAs prior to 2008, some students have more than one NAEA scores over the years, while others have only one or none. Hence, I average each student's ranks in all recent national examinations as the proxy for the student's reference performance. Using the average ranks maximizes the sample size and minimizes the extent of measurement error due to misreporting. On average, the students ranked nationally in the 60.6 percentile. When KYPS interviewed the students in grade 12, those who intended to attend college had already taken CSAT, but they had not learned their exact CSAT

---

[5] The suicide rate of 15-19 years-olds for the U.S. is computed using 2007 causes of mortality data combining with 2007 population estimates.

scores. The answers to CSAT become publicly available the day after the examination, but official results are released within a month. Therefore, students would only react to their perceived performance in CSAT at the time of the survey, even though there is some uncertainty regarding their exact performance. The students reported their CSAT scores to KYPS one year later in the final wave of the survey.[6] I use the actual distribution of CSAT scores to construct the corresponding national rank for each score.[7] On average, the respondents ranked nationally in the 68.2 percentile. The average CSAT rank is not only considerably higher than the 50 percentile, but also much higher than the average ranks in national examinations. This means that the reference points constructed using average ranks in national examinations are likely understated relative to the CSAT ranks inferred from the self-reported CSAT scores. Since test performance measures are used as explanatory variables, the measurement errors will likely lead to attenuation bias in coefficient estimates. Thus, the estimated effects are likely lower bounds.

4. Empirical Specifications

Before examining the predictions of the reference-dependent suicide model, I estimate the relationship between suicidal tendency and test performance in the following Probit regression to illustrate the problem with the standard economic model of suicide:

$$Pr(y_i = 1) = \Phi(\beta_0 + \beta_1 x_i + \delta' z_i + \varepsilon_i > 0) \qquad (1)$$

The dependent variable takes the value of one if student i reports to have suicidal ideation as defined in the previous section, and zero otherwise. x_i measures the average national ranks of student i's performance in previous rounds of NAEAs. z_i is a set of observed characteristics that are potentially correlated with mental wellbeing and test performance, such as household income, parental education, gender, whether the student easily gets stressed about poor academic performance or by parental pressure to perform academically, whether the student mostly spends time alone. Φ denotes the cumulative normal density function. Standard economic model of suicide predicts $\beta_1 < 0$, regardless of whether the test is high-stakes or low-stakes. That is, students who do well in test should have lower tendency to have suicidal ideation.

---

[6] The students only reported their CSAT digit scores that range from 1 to 9, where 1 is the highest.
[7] Korea Institute for Curriculum and Evaluation (KICE), which administers CSAT, kindly provided the data.

To examine whether the relationship between test performance and suicidal tendency of teenagers is consistent with reference-dependent preferences, I estimate the following Probit regression separately for students performing below expectation and above expectation:

$$Pr(y_i = 1) = \Phi[\alpha_0 + \alpha_1|c_i - x_i| + \delta'z_i + u_i > 0] \qquad (2)$$

The explanatory variable c_i measures the rank of student i's performance in CSAT, while the expected national rank in CSAT is captured by x_i, proxied by the mean of the student's national ranks in previous rounds of NAEAs. For students with below-reference-point performance $(c_i < x_i)$, we expect $\alpha_1 > 0$. For students with above-reference-point performance $(c_i > x_i)$, we expect $\alpha_1 < 0$.

It is possible that students may have unobserved traits that jointly affect mental wellbeing and test performance, leading to a relationship between test score and poor mental wellbeing when equation (1) or (2) is estimated at the cross-sectional level. To address this concern, I add the lagged dependent variable and past average ranks of in-school academic performance as additional control variables in some specifications.

To further examine whether reference-dependent effects are present for relatively low-stakes academic performance that does not inform much about the national rank of a student's CSAT performance, equation (2) is also estimated using data of junior high school students and their in-school academic performance.

5. Results

5.1 Non-Reference-Dependent Specifications

Table 2 reports the estimated relationship between suicidal ideation in grade 12 and average ranks in past NAEA examinations for students who planned to attend college. The coefficient of interest is mostly positive, and none of them is statistically significant. Adding control variables tend to make the coefficient of interest more positive. In particular, when past suicidal tendency and past academic performance are controlled for in column (5) in table 2, the estimated effect of test performance on suicidal tendency becomes fairly large and positive, although it is still not statistically significant. The lack of relationship is not too surprising in the reference-dependent framework since students performing well in general are also likely to perform well in CSAT, reducing their likelihood of having suicidal ideation, while high-achieving students may also have higher reference point that put them at risk of not meeting their expectation and raises their

probability of having suicidal thought. Thus, similar to the cross-country correlation in figure 1, the results based on individual survey data do not support the standard economic model of suicide.

[Table 2]

5.2 Reference-Dependent Specifications

Table 3 shows the estimates based on the reference-dependent specification (2). Columns 1 and 2 in table 3 report the estimated effect of the reference-dependent effect of test performance on suicidal ideation in grade 12 based on students having realized CSAT ranks below the reference ranks. Columns 3 and 4 in table 3 report the estimated reference-dependent effect for students having realized CSAT ranks above the reference ranks. The estimates are broadly in line with the reference-dependent explanation. The significant positive coefficients of the deviation term $|CSAT - NAEA|$ in columns 1 and 2 indicate that students performing worse than expected exhibit increasing suicidal tendency as the difference between their CSAT rank and their reference point grows. On the other hand, columns 3 and 4 show that students performing better than expected display lower suicide risks as their test performance improve, but the estimate is statistically insignificant. The estimates may suffer from attenuation bias, as the reference points are likely understated for many individuals given the self-reporting bias discussed in Section III. Moreover, because students did not know their exact CSAT scores when they were surveyed about their suicidal ideation in grade 12, some students may be incorrectly classified as having above or below reference-point performance.

[Table 3]

The labeling of students into the below-expectation and above-expectation groups is more likely correct among those whose CSAT scores are considerably away from their reference points. To assess whether the reference-dependent effects are stronger for them, observations with test score deviation greater than the top 10 percentile and the top 20 percentile among those with above-expectation performance and below expectation performance are used to re-estimate

10

the model.[8] Note that using these restricted samples may improve the size of the coefficient estimates, but will also lead to larger standard errors. The two columns in Table 4 show that the estimated positive effect becomes larger but noisier for students performing below the reference point after dropping roughly 80% of the sample. The last two column shows that the estimated negative effect becomes statistical significant for students performing above the reference point after dropping roughly 90% of the sample. Thus, the main results are not too sensitive to measurement errors in test performance. The reference-dependent effects are stronger for students deviating considerably from the reference point, especially those performing better than expected.

[Table 4]

5.3 Low-Stakes Academic Performance

If a low-stakes test informs little about a student's likely performance in the high-stakes college entrance examination, the model predicts that the test is likely to have little effect on suicidal ideation. To examine this possibility, I re-estimate equation (2) using panel survey data from the elementary school sample, which followed a different set of students from their elementary school (grade 4) into junior high school (grade 8). Ideally, I would like to focus on the within-school rank of academic performance in elementary school, as it has low stakes, does not determine the type of secondary schools students attend in Korean metropolitan cities, and provides little information about how well a student will rank nationally in CSAT.[9] However, because the elementary school sample only contains information about students' academic performance in grades 7 and 8, I measure the reference point using the within-school percentile rank in grade 7 and the realized performance using the within-school percentile rank in grade 8.

Table 5 shows the estimates of reference-dependent effects of in-school academic performance on suicidal ideation in grade 8. For students performing worse in grade 8 than in grade 7, the likelihood of having suicidal ideation is higher but not statistically significant. Furthermore, the magnitude of the coefficient is also much smaller than that reported in Table 3.

---

[8] The sample is restricted to the top 20 percentile, rather than the top 10 percentile, the sample of students having below-expectation performance is smaller than that having above-expectation performance.
[9] Metropolitan cities in South Korea implement "equalization policy" which randomizes students into schools within school districts.

Similarly, students performing better in grade 8 than in grade 7 have lower likelihood of suicidal ideation but the estimate is not statistically significant. Thus, students' suicidal ideation is not significantly related to their low-stakes in-school performance that provides little information about how well they may perform nationally.

[Table 5]

6. Robustness

6.1 Functional form assumptions

Table 6 reports the estimates based on Logit (columns 1-2), linear probability (columns 3-4), and ordered Logit regressions (columns 5-6). The ordered logit model uses five categories of suicidal ideation as the dependent variable (strongly disagree, disagree, neutral, agree, and strongly agree). The estimated marginal effects of deviation of CSAT from reference point for Logit and linear probability specifications are similar to that for Probit specification in magnitude and significance as previously reported. However, the coefficient estimates of interest for ordered Logit specification are statistically insignificant, although the signs are consistent with those based on the binary suicidal ideation measure.

[Table 6]

6.2 Life satisfaction

Although the main outcome of interest in this paper is suicidal tendency, we may also consider life satisfaction as the dependent variable, as it also captures an individual's mental wellbeing. Table 7 presents the marginal effects of Probit specifications using the dummy dependent variable taking the value of one if students agreed or strongly agreed that they were satisfied with their life. Column 1 of table 7 reports the estimates for students having below-expectation performance and column 2 reports the estimates for students having above-expectation performance. Students performing worse than their reference points are statistically less likely to be satisfied with their life, whereas those performing better than their reference points are more satisfied with their life.

12

[Table 7]

6.3 Suicidal ideation one year later

The reference-dependent effects attenuate over time. Columns 3 and 4 of Table 7 report the reference-dependent effect of test performance on suicidal ideation one year after the students took CSAT. The marginal effect of performing worse than expected on post grade-12 suicidal ideation falls by roughly half and is still statistically significant. On the other hand, the marginal effect of performing better than expected on suicidal ideation is positive but remains statistically insignificant. Thus, the reference-dependent effects subside one year later.

7. Conclusions

This paper uses reference-dependent preferences to explain why student's test performance and suicidal tendency are not necessarily inversely related. Students experience increased suicide risks when their test performance falls short of their expectation, and suicide risks are higher when the examination has higher stakes. Using panel survey data of secondary school students in Korea, I show that students whose national ranks in the high-stakes college entrance examination fell short of the self-reported average ranks in previous national examinations exhibit higher likelihood of having suicidal ideation than those whose performance was better than expected. The effects of test score on suicidal ideation fade one year after the examination, but remain somewhat statistically significant. On the other hand, students in lower grade level do not experience such reference-dependent effects of academic performance on suicidal ideation when using the ranks in low-stakes in-school academic performance to measure performance deviation from reference point.

Several caveats with respect to the model and findings are worth noting. First, I assume that reference point is fixed and not endogenously determined as in Koszegi and Rabin (2006). Although the fixed reference point assumption can be relaxed, we will need high frequency data with multiple realizations of high-stakes test scores and detailed data on effort and input into test score production to estimate anticipated reference points. It is not possible to have a richer analysis given the data in hand, as information on study effort and diagnostic test performance is lacking. Second, the empirical analysis relies on self-reported information on test performance and suicidal ideation and the estimates are plagued with measurement errors. It possibly explains

why some of the estimates are noisy. Third, suicidal ideation is not the same as suicide attempts and the effects of test performance on suicide risks should be interpreted with this difference in mind.

Despite the data limitation, the results support the argument that high-stakes testing may pose suicide risks to individuals with reference-dependent preferences when they perform worse than expected, while low-stakes testing is less likely to pose such risks. However, there are ways for policymakers to minimize the suicide risks associated with high-stakes testing. Recent field experimental evidence by List (2003) indicates that with (market) experience, the endowment effects associated with loss aversion (in trading) tend to disappear. Taking the field experimental evidence together with the current findings, more tests and reduction in the stakes involved with a test can potentially reduce the suicide risks that testing imposes on students. It is important for policy makers to consider the potential risks they may introduce when pushing for greater use of high-stakes testing.

References

Alaraisanen, A., Miettunen, J., Lauronen, E., Rasanen, P., Isohanni, M., 2006. Good school performance is a risk factor of suicide in psychoses: A 35-year follow up of the Northern Finland 1966 birth cohort. Acta Psychiatrica Scandinavica 114 (5), 357-62.

Baumeister, R. F., 1990. Suicide as escape from self. Psychological Review 97 (1), 90-113.

Betts, J. R., Hahn, Y., Zau, A. C., 2011. Does Diagnostic Math Testing Improve Student Learning? Public Policy Institute of California, San Francisco.

Bishop, J. H., 1998. The effect of curriculum-based external exit exam systems on student achievement. Journal of Economic Education 29 (2), 171-182.

Braaten, L. J., Darling, C. D., 1962. Suicidal tendencies among college students. Psychiatric Quarterly 36 (1-4), 665-692.

Camerer, C., Babcock, L., Loewenstein, G., Thaler, R., 1997. Labor supply of New York City cabdrivers: One day at a time. Quarterly Journal of Economics 112 (2), 407-441.

Card, D., Dahl, G., 2011. Family violence and football: The effect of unexpected emotional cues on violent behavior. Quarterly Journal of Economics 126 (1), 103-143.

Centers for Disease Control and Prevention. *Youth Risk Behavior Surveillance —United States*, 2007. Surveillance Summaries, June 6, 2008. MMWR 2008; 57(No. SS-4).

Crawford, V. P., Meng, J., 2011. New York City cab drivers' labor supply revisited: Reference-dependent preferences with rational expectations targets for hours and income. American Economic Review 101 (5), 1912-1932.

Cutler, D. M., Glaeser, E. L., Norberg, K. E., 2001. Explaining the rise in youth suicide, NBER chapters. In: Risky Behavior Among Youths: An Economic Analysis. National Bureau of Economic Research, 219-270.

Dubow, E. F., Kausch, D. F., Blum, M. C., Reed, J., Bush, E., 1989. Correlates of suicidal ideation and attempts in a community sample of junior high and high school students. Journal of Clinical Child Psychology 18 (2), 158-166.

Durkheim, E., (1897) 1966. Suicide. Macmillan, New York.

Farber, H. S., 2005. Is tomorrow another day? The labor supply of New York City cabdrivers." Journal of Political Economy 113 (1), 46-82.

Farber, H. S. 2008. Reference-dependent preferences and labor supply: The case of New York City taxi drivers. American Economic Review 98 (3): 1069-1082.

Goldsmith, S. K., Pellmar, T. C., Kleinman, A. M., Bunney, W. E., 2002. Reducing Suicide: A National Imperative. National Academies Press, Washington, DC.

Gunnell, D., Magnusson, P. K. E., Rasmussen, F., 2005. Low intelligence test scores in 18 year old men and risk of suicide: Cohort study. British Medical Journal 330 (7484), 167-170.

Hamermesh, D. S., Soss, M. N., 1974. An economic theory of suicide. Journal of Political Economy 82, 83-98.

Hendin, Herbert. 1982. *Suicide in America*. New York: Norton.

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica 47 (2), 263-291.

Kim, K., Kim, G., Kim, S., Kim, J., Kim J., Park, J., 2010. OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes – Country Background Report for Korea, Korean Educational Development Institute.

Kim, T., Lee, J., Lee, Y., 2003. Mixing versus Sorting in Schooling: Evidence from the Equalization Policy in South Korea. KDI School Working Paper 03-07.

Koszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. Quarterly Journal of Economics 121 (4), 1133-1165.

List, J. A., 2003. Does market experience eliminate market anomalies? Quarterly Journal of Economics 118 (1), 41-71.

Muralidharan, K., Sundararaman, V., 2010. The impact of diagnostic feedback to teachers on student learning: Experimental evidence from India. Economic Journal 120 (546), F187–F203.

Roderick, M., Jacob, B. A., Bryk, A. S., 2002. The impact of high-stakes testing in Chicago on student achievement in promotional gate grades. Educational Evaluation and Policy Analysis 24 (4), 333-357.

Rosenthal, R. W., 1993. Suicide attempts and signaling games. Mathematical Social Sciences 26, 25-33.

Tversky, A., Kahneman, D., 1991. Loss aversion in riskless choice: A reference-dependent model. Quarterly Journal of Economics 106 (4), 1039-1061.

Figure 1: Suicides per 100,000 15-19 year-olds versus Mean Test Score
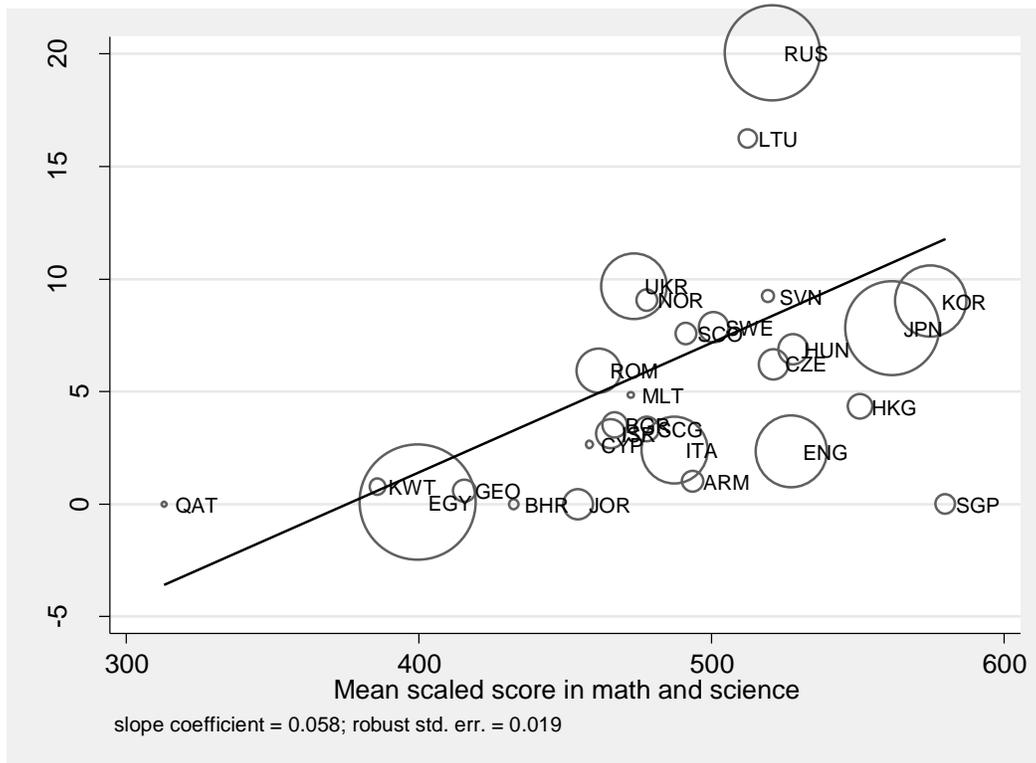
Sources: World Health Organization's mortality database (ICD 10) and Trends in International Mathematics and Science Study 2007.

Notes: Suicide rates of 15-19 year-olds (per 100,000 persons) are averages of 2008 to 2010 figures. Mean scaled score in math and science is the mean of average math and science scaled scores of 14 year olds in 2007. Fitted line is computed based on a weighted least squares regression of suicide rates against mean scaled scores. The weights used are the average number of 15-19 year-olds between 2008 and 2010.

Table 1: Summary Statistics

	Obs.	Weighted mean	Mean	Std. Dev.	Min	Max
Suicidal ideation - grade 12	1166	0.086	0.082	0.275	0	1
Suicidal ideation - grade 11	1166	0.102	0.103	0.304	0	1
Suicidal ideation - post-graduation	1166	0.057	0.060	0.238	0	1
CSAT percentile rank	1166	0.682	0.694	0.250	0	1.00
NAEA average percentile rank	1166	0.606	0.611	0.222	0.01	0.99
Absolute deviation from reference point	1166	0.189	0.189	0.182	0.0001	0.986
Deviation from reference point	1166	0.076	0.084	0.025	-0.975	0.986
Easily stress - grade 11	1166	0.664	0.672	0.470	0	1
Loner in grade 12	1166	0.143	0.148	0.355	0	1
Household income in grade 12 ('000)	1166	0.358	0.357	0.169	0	1.5
Parent attended college	1166	0.522	0.523	0.500	0	1
Male	1166	0.561	0.531	0.499	0	1
Average percentile ranks in school	1166	0.665	0.670	0.189	0.1	0.99
Life satisfaction – grade 12	1166	0.613	0.618	0.486	0	1
Life satisfaction – grade 11	1166	0.552	0.556	0.497	0	1

Notes: Sample includes only students who intended to attend college and have non-missing values for all variables used. Weighted means are weighted by the panel sampling weight in wave 5.

Table 2: The Effect of Test Performance on Suicidal Ideation

	(1)	(2)	(3)	(4)	(5)
Average NAEA ranks	0.002	0.002	0.004	0.018	0.051
	(0.039)	(0.039)	(0.039)	(0.038)	(0.050)
Household income		0.052	0.048	0.040	0.041
		(0.053)	(0.053)	(0.050)	(0.050)
Parent attended college		-0.019	-0.022	-0.018	-0.016
		(0.016)	(0.016)	(0.015)	(0.015)
Male		0.018	0.021	0.028	0.027
		(0.017)	(0.017)	(0.017)*	(0.017)
Easily stressed in grade 11			0.031	-0.002	-0.002
			(0.017)*	(0.016)	(0.016)
Loner			0.028	0.014	0.016
			(0.022)	(0.021)	(0.021)
Suicidal ideation in grade 11				0.161	0.159
				(0.020)***	(0.020)***
Average percentile ranks in school					-0.065
					(0.058)
Observations	1166	1166	1166	1166	1166

Notes: Sample includes only students who intended to attend college and have all included variables available. Estimates are the marginal effects computed from Probit regressions. Suicidal ideation was reported right after the students took CSAT in grade 12. Panel sampling weight in wave 5 is used. Robust standard errors clustered at the (wave 5) school level reported in parentheses. *** p<0.01 **p<0.05 *p<0.1

Table 3: The Reference-Dependent Effects of High-Stakes Testing on Suicidal Ideation

	(1)	(2)	(3)	(4)
	Below Reference Point		Above Reference Point	
| CSAT – NAEA |	0.135	0.159	0.002	-0.017
	(0.071)*	(0.069)**	(0.055)	(0.052)
Household income		0.036		0.044
		(0.104)		(0.049)
Parent attended college		0.052		-0.032
		(0.034)		(0.018)*
Male		0.016		0.028
		(0.034)		(0.019)
Easily stress in grade 11		-0.043		0.018
		(0.033)		(0.019)
Loner		0.040		0.009
		(0.044)		(0.025)
Suicidal ideation in grade 11		0.175		0.154
		(0.041)***		(0.022)***
Average percentile ranks in school		-0.055		-0.012
		(0.077)		(0.055)
Observations	332	332	834	834

Notes: Sample includes only students who intended to attend college and have all included variables available. Estimates are the marginal effects computed from Probit regressions. Suicidal ideation was reported right after the students took CSAT in grade 12. CSAT scores reported one year after grade 12 were converted to CSAT ranks using the actual test score distribution. Panel sampling weight in wave 5 is used. Robust standard errors clustered at the (wave 5) school level reported in parentheses. *** $p<0.01$ **$p<0.05$ *$p<0.1$

Table 4: The Reference-Dependent Effects of High-Stakes Testing on Suicidal Ideation

	(1)	(2)	(3)	(4)
	Below Reference Point		Above Reference Point	
| CSAT – NAEA |	0.025	0.185	-0.281	-0.646
	(0.241)	(0.184)	(0.232)	(0.329)**
Household income		-0.075		0.399
		(0.186)		(0.186)**
Parent attended college		0.035		-0.167
		(0.090)		(0.057)***
Male		-0.070		0.075
		(0.091)		(0.059)
Easily stress in grade 11		-0.078		-0.106
		(0.065)		(0.076)
Loner		-0.007		-0.157
		(0.123)		(0.091)*
Suicidal ideation in grade 11		0.334		0.371
		(0.099)***		(0.096)***
Average percentile ranks in school		-0.438		0.026
		(0.195)**		(0.142)
Observations	67	67	81	81

Notes: Samples restricted to students having performance deviation from the reference point in the top 20 percentile if the students have below expectation performance and the top 10 percentile if the students have above expectation performance. Estimates are the marginal effects computed from Probit regressions. Suicidal ideation was reported right after the students took CSAT in grade 12. CSAT scores reported one year after grade 12 were converted to CSAT ranks using the actual test score distribution. Panel sampling weight in wave 5 is used. Robust standard errors clustered at the (wave 5) school level reported in parentheses. *** $p<0.01$ **$p<0.05$ *$p<0.1$

Table 5: The Reference-Dependent Effects of Low-Stakes Testing on Suicidal Ideation

	(1)	(2)	(3)	(4)
	Below Reference Point		Above Reference Point	
| Grade 8 rank – Grade 7 rank |	0.065	0.057	-0.018	-0.053
	(0.071)	(0.070)	(0.081)	(0.074)
Household income		-0.056		-0.086
		(0.052)		(0.065)
Parent attended college		-0.038		-0.046
		(0.021)*		(0.016)***
Male		-0.022		-0.054
		(0.021)		(0.017)***
Easily stress in grade 7		0.049		0.046
		(0.016)***		(0.018)***
Loner		-0.031		0.049
		(0.026)		(0.032)
Suicidal ideation in grade 7		0.155		0.232
		(0.054)***		(0.050)***
Observations	706	706	708	708

Notes: Sample includes respondents in KYPS elementary school sample. Estimates are the marginal effects computed from Probit regressions. Suicidal ideation measured at the end of grade. Control variables used are those reported in column 4 of Table 2, except past academic performance because the information was not collected prior to grade 7. Panel sampling weight in wave 5 is used. Robust standard errors clustered at the (wave 5) school level reported in parentheses. *** $p<0.01$ **$p<0.05$ *$p<0.1$

Table 6: Robustness – Functional Form

	(1)	(2)	(3)	(4)	(5)	(6)
	------- Logit -------		Linear Probability		Ordered Logit	
	CSAT – NAEA		0.159	-0.011	0.187	-0.006	0.583	-0.271
	(0.066)**	(0.054)	(0.097)*	(0.049)	(0.536)	(0.380)
Performance relative to ref. point	Below	Above	Below	Above	Below	Above
Observations	332	834	332	834	332	834

Notes: Columns 1-4 report the marginal effects, whereas columns 5 and 6 report the Ordered Logit estimates. Sample includes only students who intended to attend college and have all included variables available. All specifications include a set of control variables as in columns 2 and 4 of Table 3. Suicidal ideation was reported right after the students took CSAT in grade 12. CSAT scores reported one year after grade 12 were converted to CSAT ranks using the actual test score distribution. Panel sampling weight in wave 5 is used. Robust standard errors clustered at the (wave 5) school level reported in parentheses. *** $p<0.01$ **$p<0.05$ *$p<0.1$

Table 7: Robustness – Alternative Outcome Measures

	(1)	(2)	(3)	(4)
	Life satisfaction		Suicidal ideation a year later	
	CSAT – NAEA		-0.249	0.055	0.088	0.055
	(0.122)**	(0.089)	(0.047)*	(0.040)
Performance relative to ref. point	Below	Above	Below	Above
Observations	332	834	332	834

Notes: Sample includes only students who intended to attend college and have all included variables available. Estimates are the marginal effects computed from Probit regressions. Life satisfaction was reported right after the students took CSAT in grade 12. Suicidal ideation was reported one year after students took their CSAT. CSAT scores were reported one year after grade 12 and were converted to CSAT ranks using the actual test score distribution. Panel sampling weight in wave 5 is used. Robust standard errors clustered at the (wave 5) school level reported in parentheses. *** $p<0.01$ **$p<0.05$ *$p<0.1$