GCOE Discussion Paper Series

Global COE Program
Human Behavior and Socioeconomic Dynamics

Peer effects of swimmers

Shoko Yamane and Ryohei Hayashi

January 2012

Peer effects of swimmers

Shoko Yamane[†]

Ryohei Hayashi[‡]

January 2012

Abstract

In this study, we purely and directly revealed peer effects by using a large dataset from the official Internet site of the Japanese Swimming Federation. First, we completely excluded the endogeneity of peer assignment and found that the performance of adjacent peers positively influences swimmers' performance; swimmers can swim faster with fast, high-ability peers. We also found that swimmers are aware of their peer who has a lower best record than theirs. Being chased improves swimmers' performance. Second, using absent-peer data, we directly compared the performance of individual swimmers with and without an adjacent peer. We found that the existence of adjacent peers enhances swimmers' performance. Furthermore, when we compared the records for freestyle and backstroke competitors, we found that the ability to observe peers affects the emergence of peer effects.

† Graduate School of Economics, Osaka University, Research Fellow of the Japan Society for the Promotion of Science, E-mail: evidence8money@gmail.com
  Address: 1-7-504 Machikaneyama, Toyonaka, Osaka 560-0043, Japan.
‡ Graduate School of Economics, Osaka University, Email: pey@ryohei.info

1. Introduction

   Do employees work hard when their coworkers are working hard? In a workplace, this is a very important question because the workplace is a social place where workers must share and cooperate. This question is related to designing the optimal workplace environment and incentives. Guryan et al. (2009) suggested three pathways of social effects that influence us in the workplace. The first is called the "learning effect," which describes the process that the workers learn from their coworkers about how to perform a given task in the best way. The second is the "motivation effect," whereby workers are motivated when they see their coworkers working hard and performing well. The third is the "mechanical effect." In some production processes such as assembly lines in automobile factories, coworkers' productivity mechanically influences a worker's productivity. Guryan et al. (2009) defined learning effects and motivation effects as "peer effects." Peer effects have been thoroughly studied in economics, and also in many psychological studies under a different label.

   In psychology, researchers seem to have reached consensus on the existence of peer effects. Their main issue is the direction of peer effects, that is, whether a person is facilitated or inhibited by others. The earliest study Tripllet (1898) found positive peer effects among cyclists, whereas Pessin (1993) found negative effects. Zajonc (1965) suggested that the direction of peer effects depends on the characteristics of tasks, indicating that there are positive peer effects in well-learned tasks and negative peer effects in complicated tasks.

   In contrast, economics has focused on the existence of peer effects. Economics research has assumed that peer effect implicitly means a positive effect, and no regard is given to negative peer effects. Peer effects in economics were first studied in crime rates and education. Falk and Ichino (2006) studied peer effects in the workplace with a laboratory experiment, whereas Mas and Moretti (2009) and Bandiera, Barankay, and Rasul (2010) measured peer effects using actual workplace data. Mas and Moretti (2009) investigated the productivity of cashiers in supermarkets and suggested their optimal placement. Bandiera, Barankay, and Rasul (2010) focused on social ties, indicating the importance of the presence of friends. Most economic studies conclude that peer effects exist, but Guryan et al. (2009) by using golf tournament data found no peer influence. The golf tournament data are well suited to random assignment, which are always important in avoiding the statistical problem of common shock inherent in estimating peer effects. They found that in golf tournaments, performance arises from strong financial incentives, but they did not consider the nonlinearity of peer effects. If negative peer effects are balanced by positive peer effects, the result can be no peer effects. Such mutual cancellation may explain their inability to observe peer effects among golfers.

   The purpose of the present paper is to examine peer effects using swimming data. Swimming data offer many advantages for testing peer effects so as to reveal the peer effect purely and directly. The primary advantage of swimming data is that the rule for peer assignment is completely observable, and we can eliminate endogeneity in peer assignment. Previous studies found the optimal condition

in which the assignment of peers is considered to be random (e.g., cashiers or golf tournament). In swimming, however, the adjacent peer is assigned mechanically using each player's best record according to the International Rules. Because we have each swimmer's best record in our dataset, we can perfectly control the endogeneity of peer assignment using the lane assignment rule. Thus, we can extract the pure peer effect.

The second advantage is that, with absent-peer data, we can also test the effect of peer existence by comparing the performance of individual swimmers with and without an adjacent competitor. Swimmers are aware only of competitors in adjacent lanes, and hence, if both adjacent competitors are absent, swimmers perceive themselves to be swimming alone. Using this absent-peer data, we can obtain and compare an individual's records both with and without adjacent peers, allowing us to identify the effects of peer existence. Third, we can use the differences between stroke styles. There are four styles in swimming: freestyle, breaststroke, backstroke, and butterfly. As noted below in detail, we test the existence of peer observability using the differences between freestyle and backstroke. Because swimmers cannot see adjacent peers in backstroke, we can assume that the difference between freestyle and backstroke results from the ability to observe peers.

The other advantage of swimming data is that they are free from common shocks that hamper the estimation of peer effects, such as the weather condition in Guryan et al. (2009). In addition, because our data are extracted from an Internet site, the large dataset of swimming records is sufficient to include swimmers with low to high performance levels. This allows us to examine the relationship between peer effects and skill levels. On the other hand, Mas and Moretti (2009) investigated peer effects only in low-skilled workers, and Guryan, Kroft, and Notowidigdo (2009) studied professional, highly skilled workers.

There is no need to consider learning in swimming because the very brief time (only about one minute) during which swimmers swim is too short for learning to occur. Therefore, our definition of peer effects is limited to the effect of being motivated by the performance of their adjacent swimmers. Our first contribution is revealing purely the peer effect and showing the nonlinearity of its influence by excluding peer assignment endogeneity. Our secondary contribution is showing directly the effect of peer existence.

Although swimming data have many advantages in testing peer effect, they may also seem to have some weakness, but are not. There is no direct monetary reward in swimming data, but there are the qualifying standard and the qualification rank provided by Japanese Swimming Federation[1] (JSF), which motivate swimmers to achieve better scores. If a swimmer obtains a higher qualifying standard and qualification rank, he can participate in more numerous and higher rank meets. They

---

[1] To be exact, the qualifying standards are provided by the organizer of each meet. Some meets are organized by the Japanese Swimming Federation, while others are organized by the Japan Sports Association, the swimming federation of each prefecture, the All Japan High School Athletic Federation, and so on.

also gain fame and subsidies for national meets or training camps if they achieve a very high standard or rank. Thus, we might say that the rewards that motivate swimmers include primarily honor (nonmonetary), in addition to a few monetary rewards. Furthermore, there seem to be mechanical effects from waves. It is possible that the waves caused by peers affect swimmers' performance. As noted below in detail, we reject this doubt in representing the results of backstroke.

The rest of this paper is organized as follows. Section 2 describes the international rules of swimming. Section 3 explains the estimation model and dataset. Section 4 describes the results, and Section 5 presents the conclusion and discussion.

2. Rules

The International Rules specify how swimmers are assigned their lanes. Our explanation focuses on eight-lane pools because most competitions have eight lanes[2]. Each swimmer applies to a competition by submitting his best records to date. Swimmers are classified into groups of eight. As shown in Figure 1, lane 1 is on the right extremity of the pool as seen when facing the pool from the starting line. The swimmer with the fastest record is placed in lane 4, and the next fastest is assigned to his immediate left (lane 5). The assignment of the other swimmers alternates to the right and left in accordance with their submitted records.

	Pool							
Lane number	lane 8	lane 7	lane 6	lane 5	lane 4	lane 3	lane 2	lane 1
Order of best record	8th	6th	4th	2nd	1st	3rd	5th	7th

Figure 1: Lane numbers and order

3. Data and Model

3.1 Data Detail

We used two online datasets, both from "Swim-Record dot com" (http://www.swim-record.com/index.html), the official Internet search site of the Japanese Swimming Federation (JSF). The site includes approximately 1,500 official competition records per year from competitions held by each JSF member organization (mainly in each prefecture).

---

[2] The method of assignment in a pool with more or fewer lanes than eight is provided in Appendix 1, the International Rules.

Members are obliged to reveal all records of official JSF competitions to the public. The official record contains only JSF-registered swimmers in competitions governed by International Rules. Therefore, unofficial records (for example, of citizen's competitions) are excluded from our data. The number of these unofficial records is insufficient to affect our results.

The athletic events comprising our data from both datasets were 100 meter freestyle short course competitions for men from 2007 to 2010 for the following reasons: (1) Short course competition is held year-round, providing extensive and easily mined data. (2) The 100 meter freestyle is a general event, whereas the 200 meter freestyle is a specialty event. The general event includes competitors with a range of skills and therefore provides the necessary range of skill-level data. (3) The time span of a 50 meter freestyle event is approximately 30 seconds, and the competitor's success often depends on the quality of his entry into water. Therefore, it provides less suitable data for testing peer effects. (4) Swimmers customarily breathe with their heads turned consistently to the left or right. Since the 50 meter event involves swimming one length of the pool, competitors can see only the competitor on their left or right side during the competition. On the other hand, the 100 meter event involves swimming two lengths, enabling athletes to see both adjacent competitors during the race if he breathes only with right or left side. We used data only from an event's qualifying heats, not its final competition. With the slower competitors eliminated, only the fastest swimmers compete in the final match, which would have skewed our statistical results.

There are two types of races. One is finals; only swimmers who survive the heats swim in the final match. The Olympic Games adopt this system. In the final match, all the fastest competitors swim simultaneously. Therefore, we cannot distinguish the peer effects and order effect using finals data because it is natural for the swimmers to swim faster with fast peers, all of whom want to win the cup. The other type is timed finals, which has no final match. Swimmers are classified into groups of approximately eight, and each swimmer's time in the heats determines the final rank. Here a swimmer cannot necessarily win the cup if he swims faster than adjacent swimmers, because even faster swimmers may be competing in different heats. Thus, each swimmer's optimal strategy is to give his best performance regardless of his peers' performance. We use timed finals data to examine peer effects.

3.2   Estimation Models and Data Setting

In this paper, we examine peer effects and their attributes from three viewpoints. We test peer effects by creating three datasets, one suitable for each effect.

First, we test the peer effect that has been thoroughly examined in previous studies. We examine this effect more purely than have previous studies by using the rules of lane assignment. Let R_{it}

denote the records of an individual i in competition t, and R_{si} denote the records of an individual i in competition s. We denote peer status by a dummy variable P_i, which has value 1 with R_{1i} and 0 with R_{0i}. We use this notation to distinguish it from the other peer effect noted below, and can describe this effect as a conditional expectation of R_i given P_i. The peer effect is expressed as follows:

$$E[R_{ti} - R_{si} \mid P_i = 1] \tag{1}$$

By this estimator, we can observe how the performance of the peers improves one's performance. We test this effect by using our first dataset, the "Peer Dataset," which consists of the panel data of the same individual in the same lane; the panel identifier consists of the individual and lane number. Thus, when the swimmer is assigned the same lane, the peers assigned on his left and right are considered random. For each swimmer and his competitors on the left and right, we compute the best records from all past competitions in our data. We created this elaborate dataset because we faced two problems in estimating the peer effect. One is the reflection problem (Manski 1993). A swimmer influenced by his adjacent peer also influences that peer, so we cannot use the peer records for the current competition as peer productivity. The other is the assignment problem. As shown in Section 2, lane assignment occurs according to the rule and each lane has its unique characteristics. Therefore, we cannot use the panel data of only swimmers' names as the panel identifier. To control for this characteristic of the lane, we estimate the fixed effects on the panel identifier consisting of individual and lane number. For the Peer Dataset, we specify the estimation model as follows:

$$R_{ilt} = \mu_{il} + \gamma_1 B_{it} + \gamma_2 B_{jt} + \gamma_3 X_{it} + \varepsilon_{ilt} \tag{2}$$

R_{ilt} denotes the record of swimmer i in lane l of competition t. B_{it} is swimmer i's best record at t. In this model, subscript j implies i's peer, so γ_2 captures the peer effect. X_i is i's personal variable such as school age. We estimate these parameters with fixed effects on i and l. Each swimmer's optimal strategy is to give his best performance regardless of his peers' performance, whether faster or slower. Then, the coefficient of the peer's performance equals 0. Thus, we can say that there is a positive peer effect if swimmers swim faster when adjacent peers swim fast and a negative peer effect if swimmers swim slower when adjacent peers swim fast.

We use the records of only swimmers who swim more than three times and have better records than both their left and right swimmers. We also exclude the records of the endmost lanes, such as lanes 1 and 8, in the eight-lane pool. The Peer Dataset contains 11486 records of 5373 swimmers.

Second, we examine the peer existence effect directly. Let R_{1i} denote the records of an individual i if he had peers and R_{0i} denote the records of an individual i if he had no peer by the abstention of peers. The peer existence effects are expressed as follows:

$$E\left[R_{1i} - R_{0i}\right] \tag{3}$$

In the previous test, it is impossible to obtain both records R_{1i} and R_{0i} for the same individual. However, we can observe both records R_1 and R_0 for each individual in swimming data because of the abstention of peers. We have the data of individual swimmers' performances with and without a competitor in the adjacent lane. Swimmers can watch only the peer competitor swimming in the immediately adjacent lanes of the pool; hence, if no peer swimmer is present, there is no peer effect. We directly compare the performance in both situations. Since the absence of an adjacent swimmer is an exogenous shock for a swimmer, we can consider the data as resulting from a natural experiment. Using the estimator of equation (3), we investigate whether the existence of adjacent peers enhances or diminishes performance.

In our second dataset to test the effect of peer existence, the "Abstention Dataset," we have two records: R_{0i} denotes the records when the swimmer has no peer and R_{1i} denotes the existence of a peer in both his right and left adjacent lanes (i.e., two peers). We choose R_{1i} from all records for each swimmer i as follows:

$$R_{1i} = \arg\min\{D_{1i} - D_{0i}\} \tag{4}$$

D_{0i} is the date of a competition in which i had no peer. The nearest data of all competitions for i are chosen as a target for comparison so that the progressive athletic development of i would not distort the results. To estimate the peer existence effects, we specify the model as follows:

$$R_i = \mu_i + \beta_1 P_i + \beta_2 D_i + \varepsilon_i \tag{5}$$
$$R_i \in R_{0i}, R_{1i}, \quad D_i \in D_{0i}, D_{1i}$$

P_i is a dummy variable that has value 1 with R_{1i} and 0 with R_{0i}. Hence, β_1 captures the peer existence effects. Swimmers' performance is enhanced by adjacent peers if β_1 is positive and inhibited if negative. This dataset contains R_{1i} that are older than R_{0i}; other data have more recent R_{1i}. Swimmers become faster as they develop. To take that into account when comparing data with and without peers, we use the date of record as an independent variable. Thus β_2 captures the effect of swimmers' development.

We primarily use the records when both-sided peers are present as R_{1i}, but also have records when only a one-sided peer is present for some individuals. To examine whether the number of peers influences the impact of the peer existence, we also compare these one-sided presence records with R_{0i}. If the number of peers linearly influences the impact of peer existence, the presence of a one-sided peer has half the impact of the presence of both-sided peers. However, there might be some nonlinearity between the number of peers and the impact of the peer existence.

We count the swimmers who did not participate in the race at all for the abstentions. We do not regard abstention players who swam until the middle of the race. Similarly, we exclude disqualified

7

players from the Abstention Dataset because we do not know when a competitor left the race. We also exclude the records of the endmost lanes. The Abstention Dataset contains 21954 records of 5187 unique swimmers and 3924 one-sided abstention records of 1267 unique swimmers.

Finally, we examine the influence of observable peers by comparing the records of backstroke and freestyle swimmers. In backstroke, swimmers cannot see adjacent peers because they swim on their back. Hence, the difference between freestyle and backstroke results from the observability of peers. To reveal the existence of observable peers, we created a third dataset, the "Backstroke Dataset," structured the same as the Peer Dataset, with the panel data of the same swimmer in the same lane. We can thus reveal how the existence of observable peers influences swimmers' performance. We perform the estimation with model (2) using backstroke records, and compare two γ_2, one from freestyle estimation and the other from backstroke estimation. The difference between these two γ_2 represents the effect of observability. The Backstroke Dataset contains 1096 backstroke (100 meter in short course) records of 602 swimmers[3].

Now we can examine all three effects: the pure peer effect, the peer existence effect, and the observability effect. This examination demonstrates the influence of peers and its attributes more clearly and minutely.

4. Results

4.1 Descriptive Statistics

We begin with the descriptive statistics of key variables, as shown in Table 1. Panels A, B, and C show descriptive statistics of the Peer Dataset, Abstention Dataset, and Backstroke Dataset, respectively. A smaller value of "record" indicates higher performance. We represent the date by a serial value in which January 1, 1900 takes numeric value 1; January 2, 1900 takes 2; and so on. Definitions of all variables appear in Appendix 2. In the Peer Dataset, the average of a swimmer's own record, the swimmer's own best record, and peers' best records are approximately 62 seconds and 70 seconds, respectively, in the Backstroke Dataset. The mean records are approximately 64 s regardless of the existence or the number of peers in the Abstention Dataset. There is no difference between the average record of the competition with and without peer (t(21964) = −1.64, n.s.). The quality of competitions is same regardless the existence of abstention swimmers.

---

[3] The sample size is too small compared with the above two datasets. It is obtained from the number of swimmers because backstroke is a specialized event compared with 100 meter free style, in which the most swimmers participate. Later, we create the Freestyle Dataset, which has the same sample size as the Backstroke Dataset, and denote its results as well.

Table 1 Descriptive statistics

Panel A: Peer Dataset

Variable	N	Mean	SD	Min	Max
record	11486	62.00	6.27	48.10	104.99
bestrecord	11486	62.42	6.64	48.88	106.49
bestrecord_side	11486	62.65	6.73	49.06	116.90
bestrecord_left	11486	62.67	6.88	48.88	116.67
bestrecord_right	11486	62.64	6.94	48.88	126.41
schoolage	11486	7.58	2.16	1	16

Panel B: Abstention Dataset

Data	Records type	Variable	N	Mean	SD	Min	Max
both-sided abstention data	with peer records	record	10977	64.52	7.97	49.52	121.58
		date	10977	17809.88	353.51	17257	18342
	without peer records	record	10977	64.34	7.91	49.63	121.67
		date	10977	17848.10	347.99	17257	18343
one-sided abstention data	with peer records	record	1962	64.88	8.69	50.13	118.85
		date	1962	17840.95	304.27	17257	18342
	without peer records	record	1962	64.75	8.21	50.18	118.13
		date	1962	17842.34	352.53	17264	18343

Panel C: Backstroke Dataset

Variable	N	Mean	SD	Min	Max
record	2534	68.45	6.61	53.34	95.58
bestrecord	2534	68.93	6.97	52.63	109.33
bestrecord _side	2534	69.55	6.89	53.375	97.28
bestrecord _left	2534	69.61	7.42	52.63	102.64
bestrecord _right	2534	69.49	7.21	52.63	98.88
schoolage	2534	7.53	2.11	2	17

4.2 Pure Peer Effect

In this section, we identify the peer speed effect using the Peer Dataset. We examine the estimated results shown in Table 2. In column 1, the coefficient of the average of left and right peers' best records is highly significant and the signs are positive. Thus, the faster the peers swim, the faster the swimmer swims. In average neighborhood, the performance of a swimmer improves by 0.18 seconds when his peer swims faster by 1 second. Column 2 represents the result of regression testing using each of the right and left best records regardless of the average of best records. Both left and right best records are highly significant, with both-sided peers equally influencing the swimmers.

In column 3, we present the estimation results considering the land characteristics. As shown in section 3.1, some swimmers have the peer with the faster best records than theirs in their left lane, and others have the faster peer in their right lane. For example, in an eight-lane pool, the swimmers in lane numbers 1, 2, and 3 have the faster peer in their left lane and the slower peer in their right lane. In contrast, the swimmers in lanes 5, 6, 7, and 8 have the faster peer in their right lane and the slower peer in their left lane. In the regression of column 3, we omit the fastest lane (lane 4 in an eight-lane pool) because they have only slower peers. Column 3 of Table 2 shows that the coefficient of the slower lane is positively significant but that of faster lane is not. We found that the faster the slower lane peer swims, the faster the swimmer can swim. Thus, the performance of a swimmer is improved by the slow-lane peer. This result shows that there is a "chased effect," with a swimmer aware only of his peer in the slower lane who has the slower best records[4]. We obtain the same result using the swimmers who swam more than five times or ten times.

We can also use the current as well as best records of peers. Then, we perform the three-stage estimation. The first and second stages consist of OLS estimation of the peer record at the current competition; the records of the right or left peer are treated as endogenous variables and his best record and school age are used as instrumental variables. In the third stage, we conduct OLS estimation with estimation model (2) using peers' current records instead of their best records. Table 3 represents the estimation result with instrumental variables. The coefficients of the records of the right and left peers remain positively significant. The coefficient of the slow lane peer's record is also positively significant; however, the fast lane peer's record is also positively significant, unlike that shown in Table 2[5].

---

[4] We divide the sample by the relative position in the current competition as in Section 4.3, and obtain the same result as shown in Table 2 in all divisions.
[5] We cannot, however, control the individual fixed effect in IV regression because of insufficient sample.

Table 2: Regression result of peer effects

The dependent variable is the swimmer's own record, with individual and course fixed effects.

	(1) with average of left and right best records		(2) with each of left and right best records		(3) with adjacent lane characteristics	
	Coef.	p value	Coef.	p value	Coef.	p value
Constant	20.535	[0.000]**	20.515	[0.000]**	18.997	[0.000]**
best record	0.553	[0.000]**	0.553	[0.000]**	0.554	[0.000]**
schoolage	-0.548	[0.000]**	-0.547	[0.000]**	-0.476	[0.000]**
best record _side	0.177	[0.000]**				
best record _right			0.093	[0.000]**		
best record _left			0.084	[0.000]**		
best record _fastlane					0.036	[0.070]
best record _slowlane					0.156	[0.000]**
Obs		11486		11486		7541
R squared		0.834		0.834		0.832

Note: ** 1%, * 5% significance.

Table 3: IV result of peer effects

Model	Dependent var.	Independent var.	Coef.	p value
(1) with each of left and right best records	recprd_right	Constant	4.985	[0.000]***
		besttime_right	0.915	[0.000]***
		schoolage_right	-0.013	[0.000]***
	record_left	Constant	4.758	[0.000]***
		besttime_left	0.918	[0.000]***
		schoolage_left	-0.011	[0.000]***
	record	Constant	4.039	[0.000]***
		besttime	0.665	[0.000]***
		time_right	0.134	[0.000]***
		time_left	0.136	[0.000]***
		schoolage	-0.043	[0.000]***
(2) with adjacent lane characteristics	record_fastlane	Constant	5.428	[0.000]***
		besttime_fastlane	0.907	[0.000]***
		schoolage_fastlane	-0.012	[0.000]***
	record_slowlane	Constant	5.365	[0.000]***
		besttime_slowlane	0.910	[0.000]***
		schoolage_slowlane	-0.015	[0.000]***
	record	Constant	5.083	[0.000]***
		besttime	0.668	[0.000]***
		time_fastlane	0.090	[0.000]***
		time_slowlane	0.161	[0.000]***
		schoolage	-0.063	[0.000]***

Note: ** 1%, * 5% significance.

Next, we examine whether the impact of peer effect is influenced by skill level. Because skill level and age can be highly correlated, we use quantile regression by school level: elementary school, junior high school, senior high school and adult. We have five quantiles: 0.1 quantile (very low skill), 0.25 quantile (low skill), 0.5 quantile (medium skill), 0.75 quantile (high skill), and 0.9 quantile (very high skill). Table 4 reports the regression result at each school level. Panel A of Table 4 shows the elementary school students (sample size 4036), Panel B shows the junior high school (sample size 5192), and Panel C shows high school and older (sample size 2258). Figure 2 shows the size of the coefficient of "besttime_side" in each quantile and each school level. The results show that the relationship of the size of the peer effect and skill level is nonlinear. Among the elementary school swimmers, low-skilled swimmers feel the strongest peer effects and the impact of the peer effect decreases with rising skill level. However, the highest-skilled swimmers are more influenced by peers. Although the junior high school swimmers exhibit almost same results as the elementary school swimmers, the coefficients of peers' records are smaller than those for elementary school swimmers in all quantiles. In contrast, the high school and adult swimmers seem to have a very different pattern. That is, the 0.25 quantile is least affected and the highest-skilled swimmers are most affected. However, we found that the coefficient of peer speed effect for the 0.25 quantile of high school is similar to that for the 0.75 quantile of junior high school, and that of the 0.5 quantile for high school is as large as the 0.9 quantile for junior high school, as shown in Figure 2. We can consider that the 0.25 quantile of high school overlaps with the 0.75 quantile of junior high school, and the 0.5 quantile of high school overlaps with the 0.9 quantile of junior high school. Therefore, we can draw a U-shaped line whose lowest point is approximately the median[6]. On the whole, the lowest-skilled swimmers (beginners) are most affected and the second-most affected are the top swimmers.

---

[6] We think this idea is feasible because the average records of the two resemble each other; the 0.25 quantile of high school is 57.52 seonds, and 0.75 quantile of junior high school is 57.77 seconds. Similarly, the average record of the 0.5 quantile of high school is 55.35 seconds, and that of the 0.9 quantile of junior high school is 55.81 seconds.

Table 4: Result of quantile regression for peer speed effect in school

The dependent variable is the swimmer's own record.

Panel A: Elementary school students

	0.1 quantile		0.25 quantile		0.5 quantile		0.75 quantile		0.9 quantile	
	Coef.	p value	Coef.	p value	Coef.	p value	Coef.	p value	Coef.	p value
Constant	11.98	[0.000]**	8.481	[0.000]**	6.26	[0.000]**	3.924	[0.000]**	0.707	[0.264]
bestrecord	0.584	[0.000]**	0.705	[0.000]**	0.789	[0.000]**	0.840	[0.000]**	0.861	[0.000]**
bestrecord_side	0.232	[0.000]**	0.170	[0.000]**	0.129	[0.000]**	0.122	[0.000]**	0.156	[0.000]**
schoolage	−0.382	[0.000]**	−0.283	[0.000]**	−0.259	[0.000]**	−0.223	[0.000]**	−0.169	[0.001]**

Panel B: Junior high school students

	0.1 quantile		0.25 quantile		0.5 quantile		0.75 quantile		0.9 quantile	
	Coef.	p value	Coef.	p value	Coef.	p value	Coef.	p value	Coef.	p value
Constant	10.988	[0.000]**	7.228	[0.000]**	5.097	[0.000]**	2.98	[0.000]**	1.427	[0.031]*
best record	0.645	[0.000]**	0.724	[0.000]**	0.793	[0.000]**	0.847	[0.000]**	0.850	[0.000]**
bestrecord_side	0.170	[0.000]**	0.146	[0.000]**	0.121	[0.000]**	0.109	[0.000]**	0.132	[0.000]**
schoolage	−0.197	[0.000]**	−0.052	[0.070]	−0.027	[0.284]	−0.006	[0.836]	0.071	[0.090]

Panel C: High school student and adults

|	0.1 quantile		0.25 quantile		0.5 quantile		0.75 quantile		0.9 quantile	
	Coef.	p value	Coef.	p value	Coef.	p value	Coef.	p value	Coef.	p value
Constant	6.308	[0.000]**	4.791	[0.000]**	2.423	[0.000]**	0.926	[0.127]	0.387	[0.598]
bestrecord	0.725	[0.000]**	0.793	[0.000]**	0.810	[0.000]**	0.807	[0.000]**	0.820	[0.000]**
bestrecord_side	0.152	[0.000]**	0.122	[0.000]**	0.151	[0.000]**	0.186	[0.000]**	0.194	[0.000]**
schoolage	−0.051	[0.096]	0.048	[0.020]*	−0.021	[0.170]	0.004	[0.858]	−0.005	[0.860]

Note: ** 1%, * 5% significance.
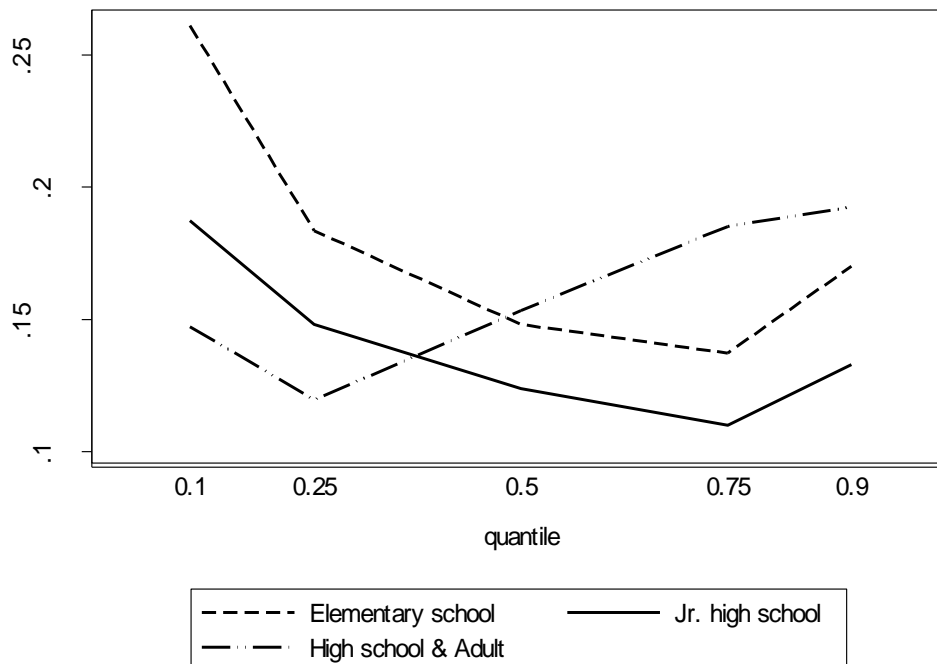
Figure 2: Coefficient of peer's best records

4.3 Peer Existence Effects

   In this section, we test the peer existence effects by the regression of the dummy variable for having peers using the Abstention Dataset. Table 5 represents the results of the regression using the both-sided abstention and presence records. In the first column, which shows the results of all samples, the coefficient of the peer dummy is highly significant, and its sign is negative; swimmers swim faster against competitors than alone. Then, we divide the sample by the relative position with the peers, and regression (2) in Table 5 shows the result of swimmers who swim faster than both peers. Here we find that the coefficient of the peer dummy is also negatively significant. However, in regression (3), the result of the swimmers who swim slower than peers, the coefficient of the peer dummy is positively significant. Thus, we find that the relative position to peers changes the direction of the peer existence effect. Because the swimmer who is behind both adjacent peers is at least rank 3, he might lose his motivation.

   Regressions (4) and (5) show the results of the swimmers who swim faster or slower than at least one of two peers. The coefficient of the peer dummy is also negatively significant in swimmers faster than at least one peer, but the size of its coefficient is smaller than regression (2). In the group slower than at least one peer, the coefficient of the peer dummy is not significant.

   Next, to confirm whether the number of peers influences the impact of peer existence, we compare

the records when the one-sided peer is present and the other-sided peer is absent with those when both-sided peers are absent. Column 1 of Table 6 shows the results of all samples, and the peer dummy is not significant. However, if we consider the relative position to peers, we obtain the same results as those in Table 2: a swimmer can swim faster with a slower peer than alone, but he swims slower with a faster peer than alone.

From Tables 5 and 6, we found that a swimmer can swim faster than alone when he is pursued from behind, but slower when his peer swims ahead of him. Thus, the existence of a pursuer improves swimmers' performance. This is similar to the chased effect noted in Section 4.2. On average, peer existence improves a swimmer's performance by 0.64 seconds when the peers are slower than him and diminishes it by 0.48 seconds when the peers are faster than him[7]. In all regressions of Tables 5 and 6, the coefficient of date is negatively significant, which appropriately captures the effects of development.

The performance of swimmers far from peers might be same as that of the swimmers swimming alone, and there could be stronger peer effects in a close match. Therefore, we divide the samples by the difference between the swimmer's own and peers' records and perform the same regression for each group. Our first group is within 0.5 seconds faster or slower than the peers. The 0.5 seconds difference is nearly equal to half the body-length's distance, so this case is considered a close match. The other group has 2 seconds difference, representing approximately a 3 meter difference, which is more than a two-body-lengths' distance. Table 7 shows the regression results. We found no difference between a swimmer's own and peers' speeds (lead/lag distance).Thus, Swimmers are more influenced by peers' relative position.

---

[7] Naturally, a 1-s improvement from 45 s and 70 s should be considered to be different. Therefore, we can interpret the coefficient of peer existence "in average neighborhood".

Table 5: Result of regression with fixed effects of swimmers' names

(both-sided peer existence)

Dependent variables: swimmers' own records

	(1) Default		(2) Faster than both peers		(3) Slower than both peers	
	Coef.	p value	Coef.	p value	Coef.	p value
Constant	221.365	[0.000]**	226.14	[0.000]**	221.878	[0.000]**
peer	−0.151	[0.000]**	−0.641	[0.000]**	0.475	[0.000]**
date	−0.009	[0.000]**	−0.009	[0.000]**	−0.009	[0.000]**
Obs	21954		6878		5746	
R squared	0.589		0.611		0.566	

	(4) Faster than one peer at least		(5) Slower than one peer at least	
	Coef.	p value	Coef.	p value
Constant	222.028	[0.000]**	220.269	[0.000]**
peer	−0.375	[0.000]**	0.067	[0.077]
date	−0.009	[0.000]**	−0.009	[0.000]**
Obs	16208		15076	
R squared	0.598		0.578	

Note: "Peer" is a dummy variable that equals 1 if the record includes data when both adjacent swimmers are present and 0 otherwise. ** 1%, * 5% significance.

Table 6: Result of regression with fixed effects of swimmers' names

(one-sided peer existence)

Dependent variables: swimmers' own records

	(1) Default		(2) Faster than peer		(3) Slower than peer	
	Coef.	p value	Coef.	p value	Coef.	p value
Constant	248.933	[0.000]**	242.011	[0.000]**	248.889	[0.000]**
peer	0.124	[0.111]	−0.258	[0.011]*	0.489	[0.000]**
date	−0.01	[0.000]**	−0.01	[0.000]**	−0.01	[0.000]**
Obs	3924		1980		1944	
R squared	0.575		0.602		0.555	

Note: "Peer" is a dummy variable that equals 1 if the record includes data when an adjacent swimmer is present and 0 otherwise. ** 1%, * 5% significance.

Table 7: Close match or not, with fixed effects of swimmers' names

	Faster by more than 2 s than both peers		Slower by more than 2 s than both peers		Faster within 0.5 s than both peers		Slower within 0.5 s than both peers	
Constant	281.294	[0.000]**	297.781	[0.000]**	216.928	[0.000]**	174.583	[0.000]**
peer	-0.671	[0.000]**	1.695	[0.000]**	-0.827	[0.003]**	-0.288	[0.210]
date	-0.012	[0.000]**	-0.013	[0.000]**	-0.009	[0.000]**	-0.006	[0.000]**
Obs	1714		1154		282		268	
R squared	0.663		0.623		0.574		0.515	

	Faster by more than 2 s than one peer		Slower by more than 2 s than one peer		Faster within 0.5 s than one peer		Slower within 0.5 s than one peer	
Constant	207.597	[0.000]**	247.343	[0.000]**	155.933	[0.000]**	154.47	[0.000]**
peer	-0.389	[0.013]*	1.006	[0.000]**	-0.329	[0.045]*	-0.309	[0.044]*
date	-0.008	[0.000]**	-0.01	[0.000]**	-0.005	[0.000]**	-0.005	[0.000]**
Obs	820		802		386		416	
R squared	0.49		0.47		0.43		0.447	

Note: ** 1%, * 5% significance.

4.4 Peer Observability

Finally, we examine the effect of peer observability using the Backstroke Dataset. Columns 1 and 2 of Table 8 show the estimated results. The coefficient of peers' performance is not significant. We know that the observability of peers is important; however, unlike freestyle, backstroke swimmers are not aware of the existence of the adjacent peer[8]. This result also shows that there is no wave effect. If the wave effect existed, peer performance would be also significant in the backstroke.

To examine the effect of peer observability again, we estimate using the next two lanes. The result is presented in columns 3 and 4 of Table 8. Although we might assume that swimmers are actually aware only of swimmers in adjacent lanes, we test this assumption. If swimmers are aware of not only their adjacent peers but also those to the right and left past the next lane peers, the coefficient of peers' best records is significant in this estimation. We have 3757 short course freestyle records (100

---

[8] To exclude the doubt that this result comes from a small sample, we choose randomly a freestyle sample equaling the backstroke sample (about 1000). We repeat this choice 10 times, and the coefficient of peer performance is positively significant in all trials, two at 10% significance, one at 5% significance, and the rest at 1% significance.

m) of 2485 swimmers with the best records of their two adjacent peers. The coefficients of the next two peers' best records are not significant. Thus, we find that swimmers are aware of only their immediately adjacent peers.

Table 8: Regression result using backstroke and two adjacent peers

The dependent variable is the swimmer's own record, with individual and course fixed effects.

	Backstroke				Two adjacent peers			
	(1) with average of left and right best records		(2) with each of left and right best records		(3) with average of left and right best records		(4) with each of left and right best records	
	Coef.	p value	Coef.	p value	Coef.	p value	Coef.	p value
Constant	29.481	[0.000]**	30.129	[0.000]**	22.165	[0.000]**	22.21	[0.000]**
bestrecord	0.632	[0.000]**	0.629	[0.000]**	0.649	[0.000]**	0.651	[0.000]**
schoolage	−1.100	[0.000]**	−1.116	[0.000]**	−0.551	[0.000]**	−0.554	[0.000]**
bestrecord_side	0.054	[0.414]						
bestrecord_left			0.061	[0.256]				
bestrecord_right			−0.012	[0.840]				
bestrecord_2side					0.055	[0.088]		
bestrecord_2left							0.042	[0.106]
bestrecord_2right							0.011	[0.713]
Obs	1094		1094		3757		3757	
R squared	0.814		0.815		0.827		0.828	

Note: ** 1%, * 5% significance.

5. Conclusion

In this study, we investigated peer effects on swimmers. First, we found that the performance of adjacent peers positively influences swimmers' performance. In particular, swimmers are aware of and influenced by their slower-lane peer with the lower best record than theirs. Second, we revealed a peer's existence effect using abstention data. These data are nearly natural experiment data that enable us to observe the pure impact of peer existence. We found that swimmers swim faster than when alone if their peer is behind them, but they swim slower if their peer swims ahead. Our findings showed that being chased by others enhances individual performance. There are similar situations to our findings in daily life. For example, the existence of a brilliant senior member does not improve one's performance, but that of excellent younger members increases it. We also found that the relationship between the skill level and the impact of peer speed effect is nonlinear and takes

19

a U shape. The lowest-skilled swimmers and top-level swimmers are more influenced than medium level swimmers. This result might be related to their development stage. It is considered that the low-skill swimmers are sensitive to the exogenous environment including peer effects. This sensitivity fades with experience, and higher-level swimmers become aware of competitors because they want to obtain the relative information about their ability. Additional research is, however, needed.

   Our study is the first empirical economic study to purely and directly reveal the effects of peer existence and their nonlinear aspects. The result of this study suggests that competition improves performance. To construct the optimal environment or incentive design in a workplace, we should take into account the existence and nonlinearity of peer effects.

6.   References

Bandiera, O., Barankay, I. and Rasul, I. 2010 Social Incentives in the Workplace. Review of Economic Studies, 77, pp. 417–458.

Falk, A. and Ichino, A. 2006 Clean Evidence on Peer Effects. Journal of Labor Economics, 24(1), pp. 39–57.

Guryan, J., Kroft, K. and Notowidigdo, M. J. 2009 Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments. American Economic Journal: Applied Economics, 1(4), pp. 34–68.

Manski, C. 1993 Identification of Endogenous Social Effects: The Reflection Problem. Review of Economic Studies, 60 (3), pp. 531–542.

Mas, A. and Moretti, E. 2009 Peers at Work. American Economic Review, 99(1), pp. 112–145.

Pessin, J. 1933 The Comparative Effects of Social and Mechanical Stimulation on memorizing. American Journal of Psychology, 45(2), pp. 263–270.

Prendergast, C. 1992 Career Development and Specific Human Capital Collection. Journal of the Japanese and International Economics, 6, pp. 207–227.

Triplett, N. 1898 The Dynamogenic Factors in Pacemaking and Competition. American Journal of Psychology, 9(4), pp. 507–33.

Zajonc, R. B. 1965 Social Facilitation. Science, 149(3681), pp. 269–274.

Appendix 1: International Swimming Rules about heats by Federation Internationale de Natation (FINA).

SW 3.1 Heats

SW 3.1.1 The best competitive times of all entrants for the preceding twelve (12) months prior to the entry deadline of the competition, shall be submitted on entry forms and listed in order of time by the Management Committee. Swimmers who do not submit official recorded times shall be considered the slowest and shall be placed at the end of the list with a no time. Placement of swimmers with identical times or of more than one swimmer without times shall be determined by draw. Swimmers shall be placed in lanes according to the procedures set forth in SW 3.1.2. below. Swimmers shall be placed in trial heats according to submitted times in the following manner.

SW 3.1.1.1 If one heat, it shall be seeded as a final and swum only during the final session.

SW 3.1.1.2 If two heats, the fastest swimmer shall be seeded in the second heat, next fastest in the first heat, next fastest in the second heat, next in the first heat, etc.

SW 3.1.1.3 If three heats, the fastest swimmer shall be placed in the third heat, next fastest in the second, next fastest in the first. The fourth fastest swimmer shall be placed in the third heat, the fifth in the second heat, and the sixth fastest in the first heat, the seventh fastest in the third heat,etc.

SW 3.1.1.4 If four or more heats, the last three heats of the event shall be seeded in accordance with SW 3.1.1.3 above. The heat preceding the last three heats shall consist of the next fastest swimmers; the heat preceding the last four heats shall consist of the next fastest swimmers, etc. Lanes shall be assigned in descending order of submitted times within each heat, in accordance with the pattern outlined in SW 3.1.2 below.

SW 3.1.1.5 Exception: When there are two or more heats in an event, there shall be a minimum of three swimmers seeded into any one preliminary heat, but subsequent scratches may reduce the number of swimmers in such heat to less than three.

SW 3.1.1.6 Where a 10 lane pool is available and equal times are established for the 8th place in the heats of 800m and 1500m Freestyle events, lane 9 will be used with a draw for lane 8 and lane 9. In case of three (3) equal times for 8th place, lane 9 and 0 will be used with a draw for lane 8, 9 and 0.

SW 3.1.1.7 Where a 10 lane pool is not available SW 3.2.3 will apply.

SW 3.1.2 Except for 50 metre events in 50 metre pools, assignment of lanes shall be (number 1 lane

being on the right side of the pool when facing the course from the starting end) by placing the fastest swimmer or team in the centre lane in pool with an odd number of lanes, or in lane 3 or 4 respectively in pools having 6 or 8 lanes. The swimmer having the next fastest time is to be placed on his left, then alternating the others to right and left in accordance with the submitted times. Swimmers with identical times shall be assigned their lane positions by draw within the aforesaid pattern.

SW 3.1.3 When 50 metre events are contested in 50 metre pools, the races may be swum, at the discretion of the Management Committee, either from the regular starting end to the turning end or from the turning end to the starting end, depending up suchfactors as existence of adequate Automatic Equipment, starter's position, etc. The Management Committee should advise swimmers of their determination well before the start of the competition. Regardless of which way the race is swum, the swimmers shall be seeded in the same lanes in which they would be seeded if they were both starting and finishing at the starting end.

Appendix 2: Definition of the variables

Table A1: Definition of the variables

Variables	Explanation
record	Own record
peer	Dummy variable: 1 if the record includes data when both adjacent swimmers are present and 0 otherwise
date	Date by serial value in which January 1, 1900 takes 1; January 2, 1900 takes 2; and so on
bestrecord	Own fastest records at that point
bestrecord_side	Average of right and left swimmers' best record
bestrecord_left	Record of left swimmer
bestrecord_right	Record of right swimmer
bestrecord_fastlane	Record of peer who is assigned fast lane
bestrecord_slowlane	Record of peer who is assigned slow lane
bestrecord_2side	Average of right and left after the next swimmers' best record
bestrecord_2left	Record of left after the next swimmer
bestrecord_2right	Record of right after the next swimmer
schoolage	School age; 7 corresponds to the first year of junior high school, 10 to the first year of high school, and 13 to the first year of college.