# (Not Finding a) Sequential Order Bias in Elite Level Gymnastics

Kurt W Rotthoff*
Seton Hall University
Stillman School of Business

Summer 2014

**Abstract**
Sequential order bias is often used to refer to timing biases in sequential order judging. However, there are two distinct biases within this structure: overall order bias, a bias throughout the event, and a sequential order bias, a judgment biased by the immediately preceding performance. I independently test these forms of bias using uniquely suitable data from elite level gymnastics. After modeling overall order bias, I find evidence this bias exists; scores escalate throughout the competition. However, I find no evidence of a sequential order bias; scores are independent of the immediately preceding performer.

JEL: L10, L83, D81, J70, Z1

Keywords: Sequential Order Judging; Sequential Order Bias; Judging Bias; Overall Order Bias

* Kurt Rotthoff at: Kurt.Rotthoff@shu.edu, Seton Hall University, JH 674, 400 South Orange Ave, South Orange, NJ 07079. A special thanks to Hillary Morgan for helpful comments. Any mistakes are my own.

## I. Introduction

Bias of social comparison has been found in sequential order competitions, primarily in two main forms: an overall order bias which occurs throughout the event, referred to as primacy and recency, and a sequential order bias, where one person's score is influenced by the information gathered by the person who immediately preceded them. Often these two types of biases are lumped together as a sequential order bias. In this study, I measure these two forms of bias separately, and find that an overall order bias (primacy and recency) does exist, but find no evidence of a sequential order bias. Although the sample is near ideal, the semi-randomly assigned order has an upward bias in the overall order bias. This can overstate the true impact, or existence, of overall order bias.

Damisch, Mussweiler, and Plessner (2006) find that gymnasts' scores are influenced by the previous performance when using data from the 2004 Olympics. However, there are three main concerns with this data: 1) there is a team competition, 2) they use finals data, and 3) there was an apparent judging controversy in this Olympics which caused a major overhaul of the gymnastics scoring system. When a team competition exists, each team is given multiple spots in which to place their athletes. It is commonly known that coaches choose to order athletes to optimize team performance; this means coaches place their best athletes last. Given this information, one would expect to find a sequential order bias, as it is driven by the coach's selection of athlete placement.

The second issue is the use of finals data in the 2006 study. This data is biased in-and-of-itself because the ordering in the finals competition is based off the scores in the

preliminary competition. The athlete who scores the highest in the preliminary competition goes last, with the second highest score going second to last, etc. Thus, if the preliminary scores are any prediction of the athletes' ability to complete the task, the structure of the competition is set-up to be correlated. Damisch, Mussweiler, and Plessner (2006) find a contestant positive and significant sequential order bias, which is expected when using finals data (with a team competition). Although they admit this bias in the data structure, their ability to control for it is limited by statistical tools. In this study I use a sample that limits these structurally imbedded biases to get a truer understanding of sequential order bias.

The third issue, the judging controversy, led the FIG (Federation Internationale de Gymnastique) to separate the scores into a 'difficulty' score and an 'execution' score. The difficulty and execution scores are now judged by completely separate panels of judges. The difficulty score assesses the complexity of each routine the athlete attempts, and the execution score measures the performance of each routine the athlete attempts. This separation of scoring allows for a proxy to better understand the impact of both overall order bias and sequential order bias. Because the difficulty of each routine is decided by the athletes themselves, no order bias should exist. However, because the execution score is decided by the judge, both forms of order bias are testable in this portion of the scoring system.

With this unique dataset I am able to separately test for an overall order bias and sequential order bias. Although Damisch, Mussweiler, and Plessner (2006) claim to find a sequential order bias, I find no evidence that a sequential order bias exists. Either the bias in their sample was driving the results, or the gymnastics world has since found a

better way to eliminate this form of bias from the judging process. I also test if this bias exists for athletes performing immediately following a superstar athlete, continuing to find no evidence this form of bias exists. I do find evidence of an overall order bias, however, which is expected. The model and literature on different forms of bias are covered in the next section. Section three discusses the data and focuses specifically on how it uniquely allows for an accurate measure of sequential order bias. Section four describes the methodology followed by the results, separately for overall order bias and sequential order bias, and the last section concludes.

## II. Forms of Bias

Wilson (1977) points out that 'order affects' matter. He finds that synchronized swimmers who appear in the first grouping receive significantly lower scores than those competing in the second or third groupings. Flôres and Ginsburgh (1996) find that order matters in "The Queen Elizabeth Musical Competition," as the day a contestant competes impacts their overall ranking. Glejser and Heyndels (2001) confirm their findings, again using "The Queen Elizabeth Musical Competition," finding that it is better to go later in the week, or later in the day, *ceteris paribus*. Page and Page (2010) classify a found order bias in the "Idol" song contest as a form of sequential order bias. They find this ordering bias in the form of a J-shaped function indicating that it is better to go first than second or third, but it is even better to go last than first. This confirms the idea of primacy and recency found in the psychology literature which states that it is better to go first (primacy) or last (recency) when being judged, but not in the middle; which gives a U-shaped function (Gershberg and Shimamura 1994, Burgess and Hitch 1999, and

Mussweiler 2003). Bruine de Bruin (2005) separate out judgments made at the end of the competition in the "Eurovision" song contest and judgments made after each contestant in figure skating. They find that both forms of contest have an overall order bias.

These order bias issues occur differently depending on the structure of the competition: if the judgment is made after each contestant, like in this data set, or at the end of all contestants. Thus, it is possible that a perceived overall order bias, as found in the literature, is an efficient response to the set-up of sequential order contests. When judging occurs throughout a competition, it is not the memory response to primacy and recency that matter; it is the unexpected outcomes of future contestants. In a contest with 100 people, if order is assigned randomly and judging is conducted sequentially, the probability that the first person is the best (worst) is one percent. Therefore, the probability that this person deserves the highest (lowest) score is very low, and the judges are aware of this. Even if the first person is good, the odds are high that a later competitor will be even better. By withholding the highest score from the first competitor, judges hedge their bets by suppressing high scores. When scores are submitted at the end of each performance, rather than at the conclusion of the competition, the suppression of high scores will lead to an increase in scores over time. This will differ from the U-shaped function found in the psychology literature.

This simple example can be approached from a more mathematical standpoint. Assuming talent is randomly distributed, or unknown, the probability, $p$, of any given individual in the population, $n$, being the best is:

$$p = \frac{1}{n} \tag{1}$$

As *n* increases, the probability that the first person is the best (worst) decreases. Given that scoring has an upper (lower) bound, the use of the highest (lowest) score for the first person restricts the ability of the judge to give this high (low) score for someone better (worse) later in the competition. Although we worry about these scores being limited at both the high and low ends, given the high level of this competition, the relevant margin will be the top scores.[1]

Given that these judges are experts, and conditional on observing a performance, they form a posterior belief about the probability a given performance is the best, which may be significantly different from 1/n. It is clear that if the gymnast falls, their probability of being highly ranked is low, and if that gymnast competes without fault a routine with a high level of difficulty, their probability of being highly ranked is high. While judges will be able to estimate an approximate rank of the performance given their experience of the distribution of past performances, they still face an uncertainty about the distribution of performances in the present sample of athletes.

Given the remaining uncertainty, I conjecture that judges still recognize that the odds of the first person being the best are low, and thus may withhold the top scores from early contestants. If this is true, it is expected that the judges will reserve high scores for later in the competition.[2] This suppression of high scores is a function of the number of participants left and the expected distribution of the quality of future participants. As a judge observes more participants, this score ceiling effect diminishes. With the right type

[1] Although it is possible that the judge could withhold both high and low scores, due to the fact that the average execution score is approximately 8.0, out of 10.0, a judge's ability to give an athlete a low score later in the competition for a terrible performance is not a real constraint. In this data, the main constraint is moving an athlete's score up for a better performance later in the competition.

[2] Because the athletes are known, this affect may be diminished. For this reason I control for 'superstar' athletes.

of data, this description of judges' behavior can be observed. Although this conjecture explains an overall order bias, it does not explain a sequential order bias. The scores can escalate throughout the competition, but given a large enough *n*, these increasing scores will not impact one individual's performance relative to the subsequent individual's score.[3] The use of gymnastics data allows an accurate measure of these two distinct forms of bias.

Other forms of bias have been found in the literature: such as a racial bias, gender bias, nationalistic bias, and difficulty bias. Racial preference has been shown by referees in basketball (Price and Wolfers 2010) and baseball (Parsons, Sulaeman, Yates, and Hamermesh 2011). Glejser and Heyndels (2001) find that women receive lower scores in piano in "The Queen Elizabeth Musical Competition," and that prior to 1990, contestants from the Soviet Union received higher scores than contestants of other nationalities. Other studies have found evidence of a nationalistic bias in figure skating (Seltzer and Glass 1991, Campbell and Galbraith 1996, Sala, Scott, and Spriggs 2007, and Zitzewitz 2006). A difficulty bias, which occurs when a competitor attempting a more difficult routine yields an artificially higher execution score, is found in gymnastics by Morgan and Rotthoff (2014). Because these forms of bias exist, I control for a nationalistic bias and difficulty bias to accurately measure the existence of an overall order bias and sequential order bias.

[3] It is assumed that these events, all with an *n* over 99, are large enough.

## III. Data

Once every four years (in the year after the Olympics) the top gymnastics competition in the world occurs without a team competition. Using data from the 2009 World (Artistic) Gymnastic Championships, held in London, England, allows an accurate measure of both overall order bias and sequential order bias because it is the first competition, and only at this point in time, that has both a new scoring system and no team competition. I analyze all four women's events (vault, uneven bars, beam, and floor) and all six men's events (vault, floor, pommel horse, rings, high bar, and parallel bars). Although athletes typically compete in many events, there is enough recovery time between events for each athlete to recover for the next event. For each of the events there are between 106 and 134 performers, each judged by the two distinct judging panels: the difficulty panel and the execution panel.

The 2009 World Gymnastics Championships consisted of two rounds of competition, the preliminary round and finals. The final round is done in traditional gymnastics meet fashion where the top talent performs last. Thus, the final round of data cannot be used because of the inherent bias in the ordering structure. The preliminary round does not have this bias. Before the meet begins, each participating country is randomly assigned between one and three spots. This random assignment occurs at three levels: which session each athlete will compete in, which event they will start on (their rotation), and in what order they will appear for that event. Each country, or the countries' gymnastics governing bodies (in the U.S. this is USA Gymnastics), places their athletes in these spots.

This semi-random order assignment increases the accuracy of measuring order biases relative to Damisch, Mussweiler, and Plessner's (2006) findings. Although each country can strategically place their athletes, each athlete's relative position is independent of the immediately preceding athlete (countries were not given two back-to-back spots). Also, given that each country could have a maximum of three athletes in the competition, any one country's overall impact is very small. However, because this is semi-random, but not completely random, the overall order bias measure is biased upward. Still, the measurement of sequential order bias is unaffected by the semi-random assignment due to the limited spots each country receives.

Table 1 – Summary statistics for the women's events.

| Summary Statistics (women) | | | | |
|---|---|---|---|---|
| Variable | Vault | Uneven Bars | Balance Beam | Floor |
| Participants | 107 | 113 | 118 | 113 |
| Mean Difficulty Score | 4.94 | 4.89 | 4.99 | 4.92 |
| Standard Deviation of Difficulty Score | 0.706 | 1.194 | 0.650 | 0.564 |
| Mean Execution Score | 8.24 | 6.91 | 7.21 | 7.37 |
| Standard Deviation of Execution Score | 0.904 | 1.517 | 1.161 | 0.778 |

Table 2 – Summary statistics for the men's events.

| Summary Statistics (men) | | | | | | |
|---|---|---|---|---|---|---|
| Variable | Parallel Bars | High Bar | Rings | Floor | Vault | Pommel Horse |
| Participants | 127 | 127 | 126 | 134 | 122 | 132 |
| Mean Difficulty Score | 5.31 | 5.31 | 5.43 | 5.51 | 5.31 | 5.14 |
| Standard Deviation of Difficulty Score | 0.88 | 1.00 | 0.91 | 0.79 | 0.88 | 0.90 |
| Mean Execution Score | 8.07 | 7.80 | 7.94 | 8.16 | 8.07 | 7.68 |
| Standard Deviation of Execution Score | 0.78 | 0.85 | 0.66 | 0.96 | 0.78 | 1.17 |

As gymnastics went to the new scoring system, they also went to two completely separate judging panels, one judging the difficulty of the routine and one judging the execution of the routine. The difficulty score is theoretically infinite and is limited by the

routine attempted by each athlete. The execution score has a maximum score of 10.0, keeping the traditional 'perfect 10' of the sport alive. The average and standard deviation of scores are shown in table 1 (women) and 2 (men).

To accurately measure both the overall order bias and the sequential order bias I normalize the data following Morgan and Rotthoff (2014). Each event's data is normalized with a mean of zero and standard deviation of one to measure the overall order bias. Sequential order bias is dependent on the person performing immediately before each athlete; thus, it is necessary to analyze each event separately. Each panel of judges remains on the same panel for each event throughout the competition. Any measure of a sequential order bias from one athlete to another will exist on each event. The summary statistics for the normalized data are shown in Table 3.

Table 3 – Normalized data, mean zero and standard deviation one, of all events.

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Order | 1219 | 63.40689 | 36.59816 | 1 | 135 |
| Order-squared | 1219 | 5358.76 | 4849.31 | 1 | 18225 |
| Normalized Difficulty Score | 1220 | 6.90E-09 | 0.996302 | -7.009 | 2.208469 |
| Normalized Execution Score | 1220 | 4.35E-09 | 0.996302 | -9.11125 | 1.75576 |
| Superstar | 1220 | 0.053279 | 0.224681 | 0 | 1 |
| Same Judge | 1220 | 0.101639 | 0.302297 | 0 | 1 |
| Male | 1220 | 0.630328 | 0.482914 | 0 | 1 |

Given the different forms of bias found in the data, I control for both a nationalistic bias and a difficulty bias. From GymnasticsResults.com I have the country of each judge on the execution panel.[4] These countries are presented in Table 4 for the women's events and Table 5 for the men's events.

[4] Data on the difficulty panel's countries do not publicly exist.

Table 4 – Country of the execution judges, by event.

| Country of Execution Judges (women) | | | |
|---|---|---|---|
| Vault | Uneven Bars | Balance Beam | Floor |
| Mexico | N. Korea | India | Slovenia |
| Bulgaria | Egypt | Ireland | Germany |
| S. Korea | Norway | Portugal | Venezuela |
| Italy | Canada | Argentina | Lithuania |
| Romania | Brazil | France | China |
| Ukraine | Germany | Israel | Russia |

Table 5 – Country of the judges, by event.

| Country of Execution Judges (men) | | | | | |
|---|---|---|---|---|---|
| Parallel Bars | High Bar | Rings | Floor | Vault | Pommel Horse |
| Netherland | Algeria | Bulgaria | Japan | Mexico | Slovenia |
| S. Korea | Portugal | France | Venezuela | New Zealand | Russia |
| Lithuania | Austria | Germany | Luxemburg | Belarus | Portugal |
| Argentina | Ukraine | Qatar | Romania | Germany | Brazil |
| Czech Republic | Hungry | Jordan | Egypt | Canada | N. Korea |
| Poland | Great Britain | South Africa | Italy | Israel | Denmark |

This allows a dummy variable to represent if the country of a judge on the execution panel matches the country an athlete represents. In order to control for a few very talented individuals driving the results, I also control for athletes that come from superstar countries.[5] Superstar countries are defined as countries that are shown to be top performers on a given event over the previous nine years. This means that they have won at least three medals in the last three Olympics (2000, 2004, and 2008) and/or in the last six world's competitions (2001-2003 and 2005-2007). Superstar countries are shown in Tables 6 (women) and 7 (men).

[5] In gymnastics, different countries have measurably different talent levels. Given that it is possible for some of these 'powerhouse' countries to drive the results, I use a 'superstar' variable as a proxy for any self- selection issues. The gymnastics governing body (FIG) has a world ranking system based on the previous year's performance, but these rankings are uninformative the year following an Olympics. This data is from the coming out of the next group of elite level gymnasts, which occurs for each event's world championship in the year following the Olympics.

Table 6 – Countries that are considered 'superstar' countries for women's events.

| Super Star Countries (women) | | | |
|---|---|---|---|
| Vault | Uneven Bars | Balance Beam | Floor |
| USA | USA | USA | USA |
| Russia | Russia | Russia | Romania |
| China | China | Romania | |
| Germany | | China | |

Table 7 – Countries that are considered 'superstar' countries for men's events.

| Super Star Countries (men) | | | | | |
|---|---|---|---|---|---|
| Parallel Bars | High Bar | Rings | Floor | Vault | Pommel Horse |
| China | Germany | China | Canada | China | China |
| S. Korea | Slovakia | Bulgaria | Romania | Romania | Romania |
| | | Italy | | Poland | Japan |

## IV. Methodology

There are two forms of order bias that need to be tested: overall order bias and sequential order bias. Overall order bias will reveal if there are benefits to going at specific different points in the competition and will capture the primacy and recency effects found in the psychology literature. Sequential order bias measures the impact of the performance immediately preceding the performance in question; this bias is the perceived bias that worries presenters, or those taking oral exams, that the person presenting immediately before you impacts your score. I do two robustness tests for each form of bias.

*Overall Order Bias (Primacy and Recency)*

In 2006 elite level gymnastics moved away from the perfect 10 system that made Nadia Comaneci famous during the 1976 Olympics. Gymnastics' scoring system now separates out the difficulty score and the execution score. I estimate the following for each athlete, *i*, aggregating all events, for both men and women, together:

$$Score_i = \beta_0 + \beta_1 OverallOrder_i + \beta_2 OverallOrder_i^2$$

$$+ \beta_3 SuperStar_i + \beta_4 SameJudge_i + \delta E + \varepsilon \qquad (2)$$

Individual athlete's scores, *Score*, is used to measure whether performing at a different point in the contest has an impact on the score received. Difficulty and execution scores are separately estimated. This separation reveals vital information on the biased nature of the results. If there is a bias, it will only exist in the execution score; as the difficulty score is randomly distributed throughout the competition. Finding an order impact on the execution score, and not the difficulty score, provides the strongest evidence for an overall order bias.

Overall order (*OverallOrder*) and overall order squared capture each athlete's relative order in the competition, with a squared term to capture any non-linear relationship possibilities. To control for a few very talented individuals driving the results, I include a variable for athletes from *SuperStar* countries. In order to control for a potential nationalistic bias, I include *SameJudge* (for the execution score only) to capture whether the athlete and a judge on the execution panel are from the same country. To control for the potential for different results from each of the events, the E vector is an event-specific fixed effect. [6]

I also add a difficulty control to the execution results to test if the overall order bias still exists with this control (equation 3). Morgan and Rotthoff (2014), specifically at this meet, argue that a difficulty bias exists. Although this could be a self-selection issue,

[6] The excluded event is women's vault. These results are not included as no important results are found. The same judge bias can only be controlled for on the execution panel because the countries are not known for the judges on the difficulty panel.

which Morgan and Rotthoff admit, controlling for *Difficulty Bias* will not bias the results in this study.

$$ExecutionScore_i = \beta_0 + \beta_1 OverallOrder_i + \beta_2 OverallOrder_i^2 + \beta_3 SuperStar_i + \beta_4 SameJudge_i + \beta_5 DifficultyScore_i + \delta\ E + \varepsilon \quad (3)$$

*Robustness - Overall Order Bias*

I run two robustness tests to measure the potential impacts of these forms of bias. First, I test for an overall order bias in the execution score by separating the athletes into four different groups. The first 30 people to go are in the first group, the next 30 are in the second group, the following 30 are in the third group, and the remaining people are in the fourth group. This test will confirm an overall order bias, but also separates out whether the impact is consistent throughout the competition or focused on one particular group.

Second, I measure, individually, the first and last person in each rotation, if that person was in the first or last rotation, and if that person was in the first or last session.

$$Score_{ij} = \beta_0 + \beta_1 FirstInRotation_{ij} + \beta_2 LastInRotation_{ij} + \beta_3 FirstRotation_{ij} + \beta_4 LastRotation_{ij} + \beta_5 FirstSession_{ij} + \beta_6 LastSession_{ij} + \beta_7 SuperStar_{ij} + \beta_8 SameJudge_{ij} + \varepsilon \quad (4)$$

These scores, *Score*, will be used individually to measure if performing at a different point in the contest has an impact on the score received. Score will be measure for each athlete, *i*, for each event, *j*, with a dummy for if they go first, *First*, or last, *Last*, in their rotation (*In Rotation*), in the first or last rotation of a session (*Rotation*), or in the first or last session (*Session*) of the day. I also control for a superstars (*SuperStar*) and judges

that come from the same country as the athlete (*Same Judge*). This structure allows for the measurement of any bias that occurs right after the judge has a short break between rotations, or a longer break between sessions. Given that Danziger et al. (2011) find that judges issue different ruling after the return from a break, this robustness measure tests for any 'break effect' on the judging in this competition.

*Sequential Order Bias*

A sequential order bias occurs when the score of one particular judgment is dependent on the previous judgment. Scores are then related to the most recently performing person. Using data for each event, I match each performer with his or her predecessor, estimating equation 5 for each athlete, *i*, for each event, *j*.

$$Score_{ij} = \beta_0 + \alpha\ Score_{(i-1)j} + \beta_1 OverallOrder_{ij} + \beta_2 OverallOrder_{ij}^{\ 2} + \beta_3 SuperStar_{ij} + \beta_4 SameJudge_{ij} + \varepsilon \quad (5)$$

This tests whether an athlete's score is independent of the previous athlete's performance, *i*-1. I am focused only on the alpha coefficient on the lagged score variable, thus the other results are suppressed for brevity. There are three outcomes to look for in this model. The scores could be positively related, meaning that the previous athlete's score positively impacts the next athlete's score. If the previous athlete does well, that helps the next athlete's score; however, if the previous athlete does poorly, the next athlete's score is artificially lower because of this. The second possibility is that the scores are negatively correlated, thus the subsequent athlete's score is inversely affected by the preceding athlete's score. Both of these outcomes support the finding of a sequential bias. The third possibility is that α is zero, which would imply that scores are

completely independent of each other. Finding a coefficient of zero would be the strongest evidence that a sequential order bias does not exist.

*Robustness - Sequential Order Bias*

Although the sequential order bias has been measured in other studies (Wilson 1977, Flôres and Ginsburgh 1996, Bruine de Bruin 2005, Damisch, Mussweiler, and Plessner 2006, and Page and Page 2010), I extend the measure of sequential order bias to include the potential bias in sequential order judgment following superstar athletes. It is possible that there is a separate bias of judgment that is made after a judge views an abnormally high quality athlete. To measure this impact, in equation 6, I include a lagged Superstar measure, for each athlete, *i*, for each event, *j*.

$$Score_{ij} = \beta_0 + \alpha_1 Score_{(i-1)j} + \beta_1 OverallOrder_{ij} + \beta_2 OverallOrder_{ij}^2 + \beta_3 SuperStar_{ij} + \alpha_2 SuperStar_{(i-1)j} + \beta_4 SameJudge_{ij} + \varepsilon \quad (6)$$

This tests whether an athlete's score is independent on the previous athlete's performance, *i*-1 and including if that previous athlete was a superstar, *i*-1. The focus is now expanded to both alpha coefficients on the lagged score variables. Again, the other results are suppressed for brevity.

As a second robustness check on sequential order bias I separate out the execution scores for each event by the judges on the panel. For each event there are six judges on the execution panel who judge the same event throughout the preliminary round of the competition. This test reveals if there is a sequential order bias for a given judge, even if one is not found for the average score given by the panel of judges.

## IV. Results – Overall Order Bias

To test if the difficulty score is truly exogenous from the timing in the competition, I look at the timing measures to determine if they have any impact on the difficulty of the routines the gymnasts complete. Table 8 displays results from equation 2, where the dependent variable is the difficulty score, controlling for the order in which each athlete competes and whether the athlete is from a super star country.

Table 8 – Regression of difficulty score controlling for time bias, super stars, and individual events.

| Difficulty Score | |
|---|---|
| Order | -0.000005<br>(0.003) |
| Order Squared | 0.000030<br>(0.000) |
| Super Star | 1.189910***<br>(0.122) |
| Constant | -0.233673*<br>(0.122) |
| Event Level Fixed Effects | Yes |
| Observations | 1,219 |
| R-squared | 0.093 |
| Standard errors in parentheses;<br>Women's vault is excluded event | |
| *** p<0.01, ** p<0.05, * p<0.1 | |

The variables for order are insignificant, suggesting that the performance order is independent of the athletes' chosen level of difficulty for their routines. The expected positive sign on superstar is found, meaning that athletes from powerhouse countries are more likely to have harder routines.

Table 9 includes a control for same judge to determine how the predictors are related to the execution score. I hypothesize that if any bias exists, it will be found in the execution portion of the scoring, which has a maximum value of 10.0. Significant coefficients on the time variables indicate an overall order bias.

Table 9 – Regression of execution score controlling for time bias, super stars, same judge, and individual events.

| Execution Score | |
|---|---|
| Order | 0.008110***<br>(0.003) |
| Order Squared | -0.000043*<br>(0.000) |
| Super Star | 0.730150***<br>(0.126) |
| Same Judge | 0.021926<br>(0.093) |
| Constant | -0.348287***<br>(0.126) |
| Event Level Fixed Effects | Yes |
| Observations | 1,219 |
| R-squared | 0.039 |
| Standard errors in parentheses;<br>Women's vault is excluded event | |
| *** p<0.01, ** p<0.05, * p<0.1 | |

Going first is detrimental to a gymnast's score, while going later in the competition yields a statistically significantly higher score. This pattern is compatible with the fact that judges could be withholding top scores early in order to discriminate between later performances when they have a more complete knowledge of the level of competition they will see throughout the day. This is consistent with previous research on overall order bias and primacy and recency (Wilson 1977, Flôres and Ginsburgh 1996, Glejser and Heyndels 2001, and Page and Page 2010), and supports the idea that it is bad to go first (primacy) and simultaneously good to go last (recency)). I also find that athletes from superstar countries receive better scores. There is no statistical impact of having an execution judge from the same country as the athlete.

Following equation 3 (presented in Table 10), I control for the difficulty bias found in Morgan and Rotthoff (2014). Overall order bias continues to be present when controlling for this difficulty bias.

Table 10: Regression of execution score controlling for time bias, super stars, same judge, difficulty bias, and individual events.

| Execution Score | |
|---|---|
| Order | 0.008104*** (0.003) |
| Order Squared | -0.000060*** (0.000) |
| Super Star | 0.046359 (0.108) |
| Same Judge | -0.036960 (0.077) |
| Normalized Difficulty Score | 0.576618*** (0.025) |
| Constant | -0.206036** (0.105) |
| Event Level Fixed Effects | Yes |
| Observations | 1,219 |
| R-squared | 0.340 |
| Standard errors in parentheses; Women's vault is excluded event | |
| *** p<0.01, ** p<0.05, * p<0.1 | |

*Robustness - Overall Order Bias*

When splitting the order into four groups where the first 30 people to go are in the first group, the next 30 are in the second group, the following 30 are in the third group, and the remaining people are in the fourth group, I continue to find it is more valuable to go later in the competition.

Table 11 – Regression of execution score controlling for time bias, super stars, same judge, and individual events. Splitting time by one of four relative positions.

| Execution Score | |
|---|---|
| Second | 0.276946***<br>(0.081) |
| Third | 0.239626***<br>(0.081) |
| Fourth | 0.301424***<br>(0.078) |
| Super Star | 0.720233***<br>(0.126) |
| Same Judge | 0.018517<br>(0.093) |
| Constant | -0.264690**<br>(0.108) |
| Event Level Fixed Effects | Yes |
| Observations | 1,219 |
| R-squared | 0.041 |
| Standard errors in parentheses;<br>The excluded group is the first 30 athletes. | |
| *** p<0.01, ** p<0.05, * p<0.1 | |

Table 11 continues to show evidence that judges withhold top scores early in the competition. Thus, given this structure of the model, I find it is optimal not to go in the first group of 30 athletes. However, I find no significant difference in being in any other the other groups in the sample; second, third, or fourth.

The second robustness test on overall order bias measures the first and last person in each rotation, if that person was in the first or last rotation, and if that person was in the first or last session. The first regression is on the difficulty score and is presented in table 12.

Table 12 – Regression of difficulty score with dummy variables for the first and last person in the rotation, in the first and last rotation of the session, and in the first or last session. Controls for super stars, same judge, and individual events are also included.

| Difficulty | | | | |
|---|---|---|---|---|
| | Vault | Uneven Bars | Balance Beam | Floor |
| First in Rotation | -0.272<br>(1.58) | 0.108<br>(0.36) | 0.12<br>(0.77) | -0.157<br>(1.16) |
| Last in Rotation | -0.227<br>(1.32) | -0.063<br>(0.21) | 0.102<br>(0.67) | -0.217<br>(1.56) |
| First Rotation | -0.228<br>(1.53) | 0.338<br>(1.23) | -0.028<br>(0.20) | -0.08<br>(0.65) |
| Last Rotation | -0.059<br>(0.37) | 0.188<br>(0.68) | 0.116<br>(0.86) | -0.173<br>(1.44) |
| First Session | -0.329<br>(2.16)* | -0.433<br>(1.64) | -0.063<br>(0.47) | -0.072<br>(0.59) |
| Last Session | 0.334<br>(1.77) | 0.025<br>(0.08) | 0.228<br>(1.41) | 0.303<br>(2.14)* |
| Super Star | 0.673<br>(3.01)** | 1.311<br>(3.05)** | 0.873<br>(4.72)** | 0.65<br>(2.85)** |
| Constant | 5.078<br>(43.18)** | 4.771<br>(25.74)** | 4.829<br>(50.01)** | 4.989<br>(53.21)** |
| Observations | 106 | 113 | 118 | 113 |
| R-squared | 0.22 | 0.13 | 0.20 | 0.17 |

Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%

The variables for time are all insignificant except for a positive coefficient on the last session on the floor and a negative coefficient on the first session on the vault. Following the expected result that difficulty is not dependent on when a gymnast performs, twenty-two of the twenty-four time coefficients are not significant. I conclude that the difficulty score is not dependent on the timing of the performance.[7]

In Table 13 the same variables are used to determine how they are related to the Execution Score. If any judging bias occurs immediately following a break time

[7] Of the two that are significant, it is possible that the first group of vaulters were trying lower levels of difficulty. It is also possible that the group in the last rotation of floor increased their difficulty in response to earlier performers. Further analysis of this is encouraged, including the possibility that these results do follow the expected result of overall order bias.

(Danziger et al. 2011), it would appear on the dummy variables for the people performing first in a rotation, or in the first rotation.

Table 13 – Regression of difficulty score with dummy variables for the first and last person in the rotation, in the first and last rotation of the session, and in the first or last session. Controls for super stars, same judge, and individual events are also included.

| Execution | | | | |
|---|---|---|---|---|
| | Vault | Uneven Bars | Balance Beam | Floor |
| First in Rotation | -0.449<br>(1.93) | -0.213<br>(0.58) | -0.008<br>(0.03) | -0.067<br>(0.37) |
| Last in Rotation | 0.07<br>(0.30) | 0.103<br>(0.28) | 0.507<br>(1.88) | -0.366<br>(1.93) |
| First Rotation | -0.002<br>(0.01) | 0.248<br>(0.74) | -0.142<br>(0.58) | -0.029<br>(0.17) |
| Last Rotation | -0.338<br>(1.56) | 0.161<br>(0.48) | 0.393<br>(1.65) | -0.148<br>(0.89) |
| First Session | -0.505<br>(2.46)* | -1.182<br>(3.67)** | -0.214<br>(0.90) | -0.436<br>(2.62)* |
| Last Session | 0.026<br>(0.10) | 0.678<br>(1.78) | 0.039<br>(0.14) | 0.274<br>(1.43) |
| Super Star | 0.426<br>(1.40) | 1.384<br>(2.63)** | 1.132<br>(3.44)** | 0.795<br>(2.56)* |
| Same Judge | 0.178<br>(0.69) | 0.426<br>(1.10) | -1.08<br>(3.06)** | 0.435<br>(2.02)* |
| Constant | 8.441<br>(50.79)** | 6.843<br>(29.24)** | 7.084<br>(40.60)** | 7.451<br>(55.13)** |
| Observations | 106 | 113 | 118 | 113 |
| R-squared | 0.14 | 0.22 | 0.23 | 0.21 |

Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%

When looking at the time variables on the execution score, there is evidence that going in the first session of the day is detrimental to one's score, which is consistent with previous research on primacy and recency (Glejser and Heyndels's 2001). However, there is no evidence that being the first person to go after a short break, between routines, after a long break, or in the first rotation of a session, lead to higher (or lower) scores, as would be predicted from the results in Danziger et al. (2011).

## V. Results – Sequential Order Bias

Sequential bias occurs when one person's results are related the person that goes immediately before them. Tables 14-17 show separately for difficulty and execution whether the score received can be predicted by the score awarded to the preceding athlete. This data also tests each event separately for both the women and men to measure the effect of the previous person in that event. For each result I control for overall order, overall order squared, superstar, and same judge (when appropriate), but I only report the coefficient on the lagged term because this is where the sequential order bias is observed if it exists.

Table 14 – Testing for sequential bias on difficulty score in women's events. Controlling for Order, Order Squared, Super Star, and Same Judge; only the lagged effect is reported.

Difficulty Score (women)

| | Vault | Uneven Bars | Balance Beam | Floor |
|---|---|---|---|---|
| L.Event | 0.067 | -0.128 | 0.070 | -0.041 |
| | (0.10) | (0.09) | (0.09) | (0.09) |

Standard errors in parentheses;
*** p<0.01, ** p<0.05, * p<0.1

Table 15 – Testing for sequential bias on difficulty score in men's events. Controlling for Order, Order Squared, Super Star, and Same Judge; only the lagged effect is reported.

Difficulty Score (men)

| | Parallel Bars | High Bar | Rings | Floor | Vault | Pommel Horse |
|---|---|---|---|---|---|---|
| L.Event | -0.001 | -0.056 | -0.139 | 0.015 | -0.045 | -0.067 |
| | (0.09) | (0.09) | (0.09) | (0.097) | (0.10) | (0.09) |

Standard errors in parentheses;
*** p<0.01, ** p<0.05, * p<0.1

Table 16 – Testing for sequential bias on execution score in women's events. Controlling for Order, Order Squared, Super Star, and Same Judge; only the lagged effect is reported.

| | Execution Score (women) | | | |
|---|---|---|---|---|
| | Vault | Uneven Bars | Balance Beam | Floor |
| L.Event | 0.004 | -0.137 | 0.111 | 0.021 |
| | (0.10) | (0.09) | (0.08) | (0.09) |

Standard errors in parentheses;
*** p<0.01, ** p<0.05, * p<0.1

Table 17 – Testing for sequential bias on execution score in men's events. Controlling for Order, Order Squared, Super Star, and Same Judge; only the lagged effect is reported.

| | Execution Score (men) | | | | | |
|---|---|---|---|---|---|---|
| | Parallel Bars | High Bar | Rings | Floor | Vault | Pommel Horse |
| L.Event | -0.017 | -0.088 | -0.117 | 0.017 | -0.075 | 0.007 |
| | (0.09) | (0.09) | (0.09) | (0.09) | (0.10) | (0.09) |

Standard errors in parentheses;
*** p<0.01, ** p<0.05, * p<0.1

None of the results show evidence of a sequential bias for difficulty scores or execution scores. Because none of the results are significantly different from zero, this indicates that judges award each score independently of the previous score. Although other studies have found a sequential order bias (i.e. Damisch, Mussweiler, and Plessner 2006) it may be that their data was not an accurate measure of this form of bias or there has been a conscientious effort to eliminate this bias in the judging community.

*Robustness - Sequential Order Bias*

Although there is no evidence of a sequential bias, there is the potential that there is a separate bias for the person immediately following a superstar athlete. Tables 18-21 show, separately for difficulty and execution, whether the score received can be predicted by the score awarded to the preceding athlete and if the score is predictable for an athlete following a superstar.

This data tests each event separately for both the women and men separately on both difficulty and execution score. For each result I control for overall order, overall order squared, superstar, and same judge (when appropriate) but I only report the coefficient on the lagged terms because these are the coefficients of interest.

Table 18 – Testing for sequential bias on difficulty score in women's events. Controlling for Order, Order Squared, Super Star, and Same Judge; only the lagged effect is reported.

Difficulty Score (women)

| | Vault | Uneven Bars | Balance Beam | Floor |
|---|---|---|---|---|
| L.Event | 0.042 | -0.133 | 0.079 | -0.044 |
| | (0.10) | (0.10) | (0.10) | (0.10) |
| L.Superstar | 0.159 | 0.073 | -0.042 | 0.027 |
| | (0.24) | (0.46) | (0.20) | (0.25) |

Standard errors in parentheses;
*** p<0.01, ** p<0.05, * p<0.1

Table 19 – Testing for sequential bias on difficulty score in men's events. Controlling for Order, Order Squared, Super Star, and Same Judge; only the lagged effect is reported.

Difficulty Score (men)

| | Parallel Bars | High Bar | Rings | Floor | Vault | Pommel Horse |
|---|---|---|---|---|---|---|
| L.Event | 0.009 | -0.070 | -0.129 | 0.017 | -0.046 | -0.077 |
| | (0.09) | (0.09) | (0.10) | (0.09) | (0.10) | (0.09) |
| L.Superstar | -0.175 | 0.579 | -0.140 | -0.053 | 0.985 | 0.172 |
| | (0.44) | (0.63) | (0.39) | (0.36) | (0.62) | (0.36) |

Standard errors in parentheses;
*** p<0.01, ** p<0.05, * p<0.1

Table 20 – Testing for sequential bias on execution score in women's events. Controlling for Order, Order Squared, Super Star, and Same Judge; only the lagged effect is reported.

Execution Score (women)

| | Vault | Uneven Bars | Balance Beam | Floor |
|---|---|---|---|---|
| L.Event | -0.008 | -0.155 | 0.109 | -0.016 |
| | (0.10) | (0.10) | (0.09) | (0.10) |
| L.Superstar | 0.272 | 0.344 | 0.026 | 0.483 |
| | (0.32) | (0.54) | (0.32) | (0.34) |

Standard errors in parentheses;
*** p<0.01, ** p<0.05, * p<0.1

Table 21 – Testing for sequential bias on execution score in men's events. Controlling for Order, Order Squared, Super Star, and Same Judge; only the lagged effect is reported.

Execution Score (men)

| | Parallel Bars | High Bar | Rings | Floor | Vault | Pommel Horse |
|---|---|---|---|---|---|---|
| L.Event | -0.013 | -0.088 | -0.092 | -0.018 | -0.072 | -0.006 |
| | (0.09) | (0.09) | (0.09) | (0.09) | (0.10) | (0.09) |
| L.Superstar | -0.087 | -0.272 | -0.389 | 0.277 | -0.331 | 0.037 |
| | (0.39) | (0.53) | (0.27) | (0.45) | (0.82) | (0.48) |

Standard errors in parentheses;
*** p<0.01, ** p<0.05, * p<0.1

These results continue to show no evidence of a sequential bias for difficulty scores or execution scores. There is also no evidence that following a superstar athlete biases the scores of the preceding athlete.

It is also possible that, although the judges on average do not show any evidence of a sequential order bias, individual judges on the panel still display this bias. Each panel of judges has multiple judges scoring each event. Each judge judges the same event for the entire preliminary competition, so the second judge on the floor panel will remain the second judge on the floor panel throughout the competition. I have data on each individual execution judge's score throughout the meet. In tables 22 for the men and 23 for the women, the sequential order measure given for each individual judge.

Only one of the sixty judges shows a statically significant impact on the sequential order nature of the data. Judge one on the women's uneven bars shows a negative and significant impact at the ten percent level. The remaining judges continue to show no significant relationship between an athlete's score and the previous athlete's performance.

Table 22 - Testing for sequential bias on execution score in men's events. Controlling for Order, Order Squared, Super Star, and Same Judge; only the lagged effect is reported.

Execution Score (men)

| | Judge One | Judge Two | Judge Three | Judge Four | Judge Five | Judge Six | Observations |
|---|---|---|---|---|---|---|---|
| L.Vault | -0.066 | -0.076 | -0.099 | -0.071 | -0.091 | -0.070 | 115 |
| | (0.09) | (0.10) | (0.10) | (0.10) | (0.10) | (0.09) | |
| L.Floor | 0.000 | 0.023 | -0.044 | -0.012 | -0.021 | -0.055 | 132 |
| | (0.09) | (0.09) | (0.09) | (0.09) | (0.09) | (0.09) | |
| L.Pommel Horse | 0.031 | -0.016 | 0.002 | -0.005 | 0.018 | 0.026 | 130 |
| | (0.03) | (0.09) | (0.09) | (0.09) | (0.09) | (0.09) | |
| L.High Bar | -0.117 | -0.101 | -0.080 | -0.093 | -0.059 | -0.060 | 124 |
| | (0.09) | (0.09) | (0.09) | (0.09) | (0.09) | (0.09) | |
| L.Parrellel Bars | -0.004 | 0.022 | -0.051 | -0.041 | -0.041 | 0.002 | 121 |
| | (0.09) | (0.09) | (0.09) | (0.09) | (0.09) | (0.09) | |
| L.Rings | -0.123 | -0.090 | -0.140 | -0.142 | -0.008 | -0.131 | 123 |
| | (0.09) | (0.09) | (0.09) | (0.09) | (0.09) | (0.09) | |

Standard errors in parentheses;
*** p<0.01, ** p<0.05, * p<0.1

Table 23 - Testing for sequential bias on execution score in women's events. Controlling for Order, Order Squared, Super Star, and Same Judge; only the lagged effect is reported.

Execution Score (women)

| | Judge One | Judge Two | Judge Three | Judge Four | Judge Five | Judge Six | Observations |
|---|---|---|---|---|---|---|---|
| L.Vault | -0.028 | -0.009 | 0.012 | 0.007 | 0.018 | 0.015 | 99 |
| | (0.10) | (0.10) | (0.01) | (0.10) | (0.10) | (0.10) | |
| L.Floor | 0.004 | -0.013 | 0.049 | -0.042 | 0.074 | 0.038 | 109 |
| | (0.09) | (0.09) | (0.09) | (0.10) | (0.09) | (0.09) | |
| L.Beam | 0.094 | 0.099 | 0.089 | 0.132 | 0.135 | 0.091 | 111 |
| | (0.09) | (0.08) | (0.08) | (0.08) | (0.08) | (0.08) | |
| L.Uneven Bars | -0.162* | -0.149 | -0.152 | -0.105 | -0.123 | -0.072 | 107 |
| | (0.09) | (0.09) | (0.09) | (0.09) | (0.09) | (0.10) | |

Standard errors in parentheses;
*** p<0.01, ** p<0.05, * p<0.1

## V. Conclusion

Sequential order events are a regular part of life. Many people try to strategically place themselves in sequential order for events such as oral exams, business or classroom presentations, debates, meetings, interviews, or releasing vital statistics. People engage in strategic behavior because of two types of bias that occur in sequential order contests: overall order bias and sequential order bias. Sequential order bias occurs when people believe they need a strategic placement because the information presented or released immediately before them will impact their relative performance or score. However, the results of this study find no evidence that time spent choosing a strategic sequential order in this manner is not useful; the information released immediately before you is not found to impact your score. However, there is evidence of an overall bias. In general, going later in the competition is valuable. Thus, using resources to enable oneself go later in a competition, especially when each score is finalized before the next person goes, is beneficial.

Using a unique data set where elite level gymnasts are randomly assigned starting positions and there is no team competition, which is known to bias the data, I find evidence of an overall bias in the 2009 Worlds Gymnastics Championships. In this meet there is no evidence of a sequential order bias. This means that either the judging in elite level gymnastics has increased its ability to mitigate this bias or this sample, without the team competition and finals competition, allows for a more accurate estimation of this form of bias (relative to Damisch, Mussweiler, and Plessner's 2006 findings). Either way, understanding that the sequential order bias does not exist in the data is useful when judging gymnasts, song contestants, job candidates, employees, students, (presidential)

debates, movie awards, stock analyst estimations, or any other sequential order competition.

However, the finding of an overall bias continues to mean that there are strategic advantages to the timing of appearance in any of these arenas. Knowing that the placement of a contestant, job interviewer, employee, or debater can influence the overall fairness of the competition also means there are strategic responses to this bias. Continued research on both of these forms of bias, as well as all forms of bias, is encouraged.

**Works Cited**


Apesteguia, Jose and Ignacio Palacios-Huerta, 2010. Psychological Pressure in Competitive Environments: Evidence from a Randomized Natural Experiment *Forthcoming, American Economic Review*

Bhaskar, V., 2009. Rational Adversaries? Evidence from Randomised Trials in one Day Cricket *The Economic Journal*, 119, 1–23.

Bruine de Bruin, W., 2005. Save the Last Dance for Me: Unwanted Serial Position Effects in Jury Evaluations. *Acta Psycologica* 118, 245-260

Burgess, N. and G. Hitch, 1999. Memory for serial order: a network model of the phonological loop and its timing *Psychological Review* 106, 551–581

Campbell, Bryan and John Galbraith, 1996. Nonparametric Tests of the Unbiasedness of Olympic Figure-Skating Judgments *The Statistician* 45(4), 521-526.

Chiappori, P.A., Levitt, S. and Groseclose, T., 2002. Testing Mixed Strategy Equilibrium When Players are Heterogeneous: The Case of Penalty Kicks *American Economic Review* 92, 1138–1151.

Damisch, L, T. Mussweiler, and H. Plessner, 2006. Olympic Medals as Fruits of Comparison? Assimilation and Contrast in Sequential Performance Judgments *Journal of Experimental Psychology Applied* 12, 166

Danziger, S., J. Levav, and L. Avnaim-Pesso (2011) Extraneous Factors in Judicial Decisions *Proceedings of the National Academy of Sciences of the United States of America* Vol 108, No. 17, 6889–6892

Emerson, John, W. Seltzer, and D. Lin, 2009. Assessing Judging Bias: An Example from the 2000 Olympic Games *The American Statistician* 63, 124-131

Flôres, Jr., R.G. and V. A. Ginsburgh, 1996. The Queen Elisabeth Musical Competition: How Fair Is the Final Ranking? *The Statistician* 45 (1): 97–104

Garicano, Luis, Palacios-Huerta, Ignacio and Canice Prendergast, 2005. Favoritism Under Social Pressure *Review of Economics and Statistics* 87, 208-216

Gershberg, F. and A. Shimamura, 1994. Serial position effects in implicit and explicit tests of memory *Journal of Experimental Psychology: Learning, Memory and Cognition* 20, 1370–1378

Glejser, H. and B.Heyndels, 2001. Efficiency and inefficiency in the ranking in competitions: the case of the Queen Elisabeth music contest *Journal of Cultural Economics* 25, 109–129.

Goldin, Claudia and Cecilia Rouse, 2000. Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians *The American Economic Review,* Vol. 90, No. 4. pp. 715-741.

Morgan, Hillary N. and Rotthoff, Kurt W., (2014) The Harder the Task, the Higher the Score: Findings of a Difficulty Bias *forthcoming in Economic Inquiry*

Mussweiler, T., 2003. Comparison Processes in Social Judgments: Mechanisms and Consequences *Psychological Review* 110 (3), 472-489

Neilson, W., 1998 Reference Wealth Effects in Sequential Choice. *Journal of Risk and Uncertainty* 17, 27-48

Novemsky, N. and R. Dhar, 2005. Goal Fulfillment and Goal Targets in Sequential Choice *Journal of Consumer Research* 32, 396-404

Page, Lionel and Katie Page, 2010. Last shall be first: A field study of biases in sequential performance evaluation on the Idol series *Journal of Economic Behavior & Organization* 73, 186–198

Parsons, Christopher A.; Sulaeman, Johan; Yates, Michael C.; Hamermesh, Daniel S., 2011 Strike Three: Discrimination, Incentives, and Evaluation *The American Economic Review* Vol. 101, No. 4, June, pp. 1410-1435

Price, Joseph and Justin Wolfers, 2010. Racial Discrimination Among NBA Referees *Forthcoming, Quarterly Journal of Economics*

Romer, D., 2006. Do Firms Maximize? Evidence From Professional Football *Journal of Political Economy* 114, 340–365.

Sala, Brian, Scott, John, and James Spriggs, 2007. The Cold War on Ice: Constructivism and the Politics of Olympic Skating Judging *Perspectives on Politics* 5(1), 17-29

Sarafidis, Y., 2007. What Have you Done for me Lately? Release of Information and Strategic Manipulation of Memories *The Economic Journal* 117, 307-326

Segrest Purkiss, S., P. Perrewe, T. Gillespie, B. Mayes, and G. Ferris, 2006. Implicit Sources of Bias in Employment Interview Judgments and Decisions *Organizational Behavior and Human Decision Processes* 101, 152-167

Seltzer, Richard and Wayne Glass, 1991. International Politics and Judging in Olympic Skating Events: 1968-1988 *Journal of Sports Behavior* 14, 189-200

Wilson, V., 1977. Objectivity and Effect of Order of Appearance in Judging of Synchronized Swimming Meets. *Perceptual and Motor Skills* 44, 295–298

Zitzewitz, Eric, 2006. Nationalism in Winter Sports Judging and its Lessons for Organizational Decision Making *Journal of Economics and Management Strategy*, Spring, 67-99

Zitzewitz, Eric, 2010. Does Transparency Really Increase Corruption? Evidence from the 'Reform' of Figure Skating Judging *Working Paper*