Valuation Studies 11(1) 2024: 91–137

Judging by the Rules? The Emergence of Evaluation Practices

Stacy E. Lom

Abstract
How does evaluation work differently, and how do evaluation practices emerge, in different contexts? Drawing on a mixed-methods study of evaluation in figure skating and classical music, I discuss the divergent evaluative cultures in these settings, especially in terms of how formal and standardized they are, to consider how and why evaluation practices change over time and why different settings use different evaluation practices. I emphasize the importance of organizational structure, including context, competition structure, degree of centralization, and governance structure. My findings suggest that highly centralized settings governed by more powerful organizations and where competitions build on each other tend to use more formal and standardized evaluation practices compared to other settings with fewer constraints. Understanding how evaluation practices develop and what they look like in different contexts is important because in addition to influencing the objects of evaluation and perceived fairness and legitimacy, these practices often affect outcomes, which have significant consequences for participants.

Keywords: evaluation; formalization; rule changes; figure skating; classical music

Stacy Lom is Assistant Professor of Sociology at the University of Central Arkansas.

https://doi.org/10.3384/VS.2001-5992.2024.11.1.91-137
Hosted by Linköping University Electronic press
http://valuationstudies.liu.se

Introduction

Evaluation often looks very different in different settings. That statement might seem fairly obvious, but what is much less obvious is what this looks like on the ground. Consider two examples: (1) As American Nathan Chen skates his long program at the 2022 Olympics as the final skater in the men's competition and the overwhelming favorite for the gold medal, the technical panel identifies each technical element he performs, and the judges evaluate each of those elements with a score from -5 to +5. When his program ends, the judges also give him a score for five separate program components. After the judges enter their scores into a computer, Chen scores 218.63 points for his program and wins the gold medal with 332.6 total points, 22.55 points more than Yuma Kagiyama from Japan (ISU 2022a). (2) In contrast, the third and final contestant in a high-level organ competition finishes her performance, and now the judges must decide who will finish first, second, and third. The judges deliberate, arguing about whether to reward technique or artistry. They ultimately reach a consensus that the more artistic performer should win.

In both examples, the goal is to figure out who is going to win a competition, but the methods for achieving that goal are very different. These examples show how evaluation works differently in different settings, which raises the issue of how to explain these differences. In this article, I discuss how organizational structure affects evaluative cultures, especially in terms of the formality and standardization of rules. In particular, I examine how organizational context, competition structure, degree of centralization, and governance structure influence approaches to evaluation practices. This is important because in addition to influencing the objects of evaluation and perceived fairness and legitimacy, evaluation practices often affect outcomes, which can have significant consequences for participants. Based on a mixed-methods study of evaluation in figure skating and classical music and building on discussions of evaluative cultures (Lamont 2009), rules (e.g., Meyer and Rowan 1977; Dobbin and Kelly 2007; Edelman et al. 2011), and objectivity (Porter 1995; Daston and Galison 2007), I focus on how two contexts have developed very different evaluation practices. Based on my main findings, I suggest that settings with high degrees of centralization and shared international governance and where competitions build on each other tend to adopt more formal and standardized evaluation practices compared to other settings with fewer constraints.

Lamont (2009) describes evaluative cultures as cultural scripts surrounding evaluation, including how people think about and practice evaluation, as well as the rule systems and methods they use. In this article, I conceive of evaluative cultures as the meanings, values, and practices surrounding evaluation in particular settings, with an emphasis on rules and how they develop. I focus especially on

variations in evaluative cultures in terms of the formality and standardization of their rules. Based on a study of peer review in academia, Lamont's discussion of evaluative cultures highlights how people deal with disciplinary differences when they come together in an evaluative context where outcomes are based on evaluator consensus. However, she does not address where those original differences come from. This is where my research comes in. While Lamont's work focuses on understanding how evaluative cultures work in general and the process of how decisions are produced in particular, my research examines how evaluative cultures develop and change, with a focus on rules and practices. In addition to a broader emphasis on how perceived fairness and legitimacy affect how rules develop, my work expands Lamont's discussion of differences in evaluative cultures by focusing on why and how different rule systems emerge.

Evaluation, objectivity, and formalization

Valuation is a fundamentally social process, one that is dependent on historical and cultural context and interacts with many other areas of social life (Fourcade 2011; Lamont 2012; Helgesson and Muniesa 2013; Kornberger et al. 2015). What this looks like and how it affects valuation processes is very different in different contexts, which I highlight in this article, focusing on rules and practices surrounding evaluation in competition settings. In general, evaluative rules range from highly formal to highly informal. In many settings in recent years, evaluation has become increasingly formal, often in response to concerns about fairness and legitimacy, but there have been exceptions to this trend. Figure skating is a case of highly formal and standardized evaluation, including a judging system that has been overhauled in recent years following a legitimacy crisis. Classical music operates under a range of evaluative rules and practices, from very informal discussion-based evaluation to more formal and numerical evaluation. Some music competitions use intricate scoring systems, for example. On the whole, however, evaluation in classical music is much less formal and much less standardized than in skating. An evaluative rule system's degree of formality affects both evaluators and the people they evaluate in relation to everything from interactions among evaluators, to evaluators' discretion, to performers' potential creativity. I focus here on distinctions between formal and informal rules and varying degrees of standardization, especially where those differences come from and what they look like on the ground.

At least on the surface, the skating world has emphasized evaluation with formal, standardized rules much more actively than the music world generally has. Why has this been the case, and how has this

happened? The notion of objectivity, which involves neutrality based on following rules and "knowledge that bears no trace of the knower" (Daston and Galison 2007: 17), often plays a key role in how people think about fairness and legitimacy. This affects just about every area of public life, especially settings that rely on overtly subjective evaluation or evaluation that involves a lot of discretion – control, flexibility, or lack of constraint in making decisions (Mastrofski 2004; Shen and Cho 2005; Li and Tang 2010), which applies to both skating and music. The idea is that if judges are objective in their evaluations, participants should determine the outcomes based on what they do rather than what the people judging them do, similar to the popular belief that scientists are supposed to produce knowledge based on objective procedures which are devoid of emotion or political forces (Brown and Malone 2004).

Porter (1995) delineates two distinct types of objectivity: disciplinary objectivity, which revolves around reaching consensus, and mechanical objectivity, which is based on strictly adhering to rules and tends to emerge where reaching consensus would be difficult or would not seem legitimate to external observers or stakeholders. Porter stresses how disciplinary and mechanical objectivity are often at odds with each other and emphasizes that purely mechanical objectivity is impossible because all types of rules have ambiguities, regardless of how clear the people who designed them think they are. While disciplinary objectivity and the discretion that usually goes along with it have mostly remained legitimate in classical music, pressures toward mechanical objectivity and limiting discretion – and the legitimacy that often goes along with them – have had a profound impact on figure skating.

Porter's (1995) discussion of objectivity centers largely around explaining quantification's appeal in the modern world. Other scholars have built on Porter's work, suggesting that institutional legitimacy and accountability based on numbers have become increasingly linked over the past few decades (Power 2003) and that this link has contributed to a proliferation of measurement systems in a wide variety of settings (Espeland and Sauder 2007; Colyvas 2012; Mau 2019). These kinds of legitimacy and accountability have been significant in skating and music, but in different ways, largely due to their distinct organizational structures. Porter (1995) suggests that because people tend to associate quantification with "impersonality, discipline, and rules" (32) and think of it as one of the most credible strategies for achieving pure objectivity, it has been most attractive in fields plagued by outside pressure, suspicion, controversy, and unseemly politics. Especially because of its centralization and international governance, these factors have seemed to affect evaluation practices in skating much more than in music.

In keeping with how perceptions of objectivity tend to be linked to legitimacy and accountability, many organizations have used formalization – implementing written rules and more specific procedures or instructions (Adler and Borys 1996) – as a strategy for improving legitimacy and accountability. Formalization is a key component of Weber's ideal typical bureaucracy (Weber 1968; Adler 2012), where organizations run like machines without any individual discretion (Feldman 1992), and is often geared toward controlling people's actions (Stinchcombe 2001). In general, rules, written documents, and standardization, especially related to jobs and work roles, have played increasingly pivotal roles in organizations in recent years (for instance, see Adler and Borys 1996; Adler 2006 on the software industry; Maccoby 2006 on health care organizations). In particular, a number of researchers have shown how modern organizations tend to emphasize that they have specific rules or codified procedures in place, regardless of what they look like or how they affect practices and outcomes (Jacobsson 2000; Star and Lampland 2009). Based on these trends, one might be surprised to find any organizations that have not engaged in formalization processes. Many people in a wide variety of settings believe that formal, written rules are fairer and more legitimate than informal rules that have not been clearly defined, especially when those formal rules revolve around numbers. Many people also assume that when rules and procedures are written down, they increase transparency and limit discretion, whether or not this happens in practice.

Similar to how Porter (1995) emphasizes that procedures have become increasingly more important than outcomes in many areas of social life, institutional theorists have argued that in many contexts, formal rules exist for purely ceremonial purposes and to enhance legitimacy rather than to improve organizational fairness or efficiency (Meyer and Rowan 1977). Researchers have found that when organizations have the kinds of policies in place that they are "supposed to" have (like anti-discrimination laws or human rights laws, for example), people – including legal professionals – tend to assume that those policies are effective in practice, even though they are often merely symbolic and often perpetuate the issues they were intended to address (Hafner-Burton and Tsutsui 2005; Dobbin and Kelly 2007; Edelman et al. 2011).

In keeping with these trends, according to other research, people's ideas about whether rules are fair or not (rather than how effective they are) affect how likely they are to accept organizational policies in various domains and how they evaluate the people and organizations behind them in terms of how legitimate they are (for example, see Elsbach and Elofson 2000; Tyler 2000; Sunshine and Tyler 2003). One prominent example of this is law, where formal policies play an especially significant role. Tyler (1988), for instance, suggests that

perceptions of procedural justice in legal contexts influence how people understand and interact with the legal system more than the outcomes of those procedures. Tyler (2003) also notes that perceived neutrality and lack of bias among decision-makers play an especially influential role in how people evaluate the fairness of policies.

This link in many people's minds between how fair rules are and their legitimacy should continue to play a significant role in scholarship on evaluation, rules, and organizations more broadly. Unless people believe that the evaluation systems that matter within the institutions that they care about are legitimate, it will be difficult to sustain the institutions that use them (Lamont 2009). This has major implications for competition settings like figure skating and classical music, which will be virtually impossible to maintain unless participants, and outsiders in some cases, think they are legitimate activities with fair rules. One complicating factor here is that a rule system's fairness is often perceived differently in different contexts (Tyler 1988), which helps explain why different fields use different evaluation practices and why so many different ways of evaluating merit could be considered legitimate. While people might think of strict quantitative rules as fairer in one setting, they might view deliberation as more reasonable in another, depending on the evaluators, what they are evaluating, and the larger goal or context.

According to this research, it makes sense that the figure skating world responded to an Olympic judging scandal by implementing a judging system with more explicit, numerically specific rules, especially given the legitimacy crisis and coercive isomorphism surrounding this case. Isomorphism is a constraining process that is generally linked to legitimacy and leads organizations operating under similar environmental conditions to become more similar to each other (Deephouse 1996). Coercive isomorphism in particular arises from indirect pressures from cultural expectations in an organization's environment and direct pressures from powerful organizations (DiMaggio and Powell 1983). Both types of pressures – informal pressures from values emphasizing fair play in sport and a formal mandate from the International Olympic Committee (IOC) – contributed to formalization in figure skating. Judging-related controversy has plagued both figure skating and classical music, however, and theoretically at least, the music world and its outside constituents should also be concerned about its legitimacy.

As Porter (1995) argues, not only do most people believe that standardized measurement helps protect against bias and neutralizes politics; it is also a common method of dealing with distance and crossing "the boundaries of nation, language, experience, and discipline" (1995: 220). This suggests that centralized, internationally organized settings like figure skating tend to be more susceptible to pressures to use more mechanical and quantitative evaluation

practices. Despite how there are international music competitions, a judging system's capacity to cross boundaries is much more important in skating because while international music competitions can use different rules, skating's centralization and competition structure (where competitions often build on each other) require a single evaluation system that all people everywhere will be able to use and understand. The distinct organizational structures surrounding figure skating and classical music, particularly in relation to competition structure, centralization, and governance, have played key roles in how formal and standardized their rules have been. Based on Porter's work, as well as the data I have collected surrounding both settings, these factors, along with embeddedness within the sport world versus the arts world, have emerged as playing key roles in why they have adopted different rules. After a discussion of methods and data, I detail how evaluation works and how it has changed over time in these settings, addressing how these factors have contributed to variations in evaluation practices in skating and music.

Methods and data

The issues I am examining in this article revolve around differences in evaluation practices, which require comparing different evaluation practices. Fourcade (2011: 1725) emphasizes that comparative analysis "affords us precious analytical leverage … and reveals patterns that are not visible otherwise." Comparisons are especially crucial in analyzing distinctions between or among types of systems, as I am analyzing here. The variations I focus on are degrees of formality and standardization. To investigate how evaluative rules and practices develop through a comparison of two contexts, I rely on a mixed-methods approach, drawing on 96 semi-structured interviews with figure skating and classical music insiders, participant-observation, archival materials, and content analysis. These different types of data are useful for addressing evaluative practice development from different angles and allow for triangulation, which helps verify evidence (Jick 1979). Although looking at additional cases could have increased the breadth of my analysis, it would be virtually impossible to gain a full understanding of evaluation practices and how they work within particular settings without conducting in-depth case studies. Other scholars who have studied evaluation, as well as the effects of measures and other types of systems on organizations, have used similar approaches and types of data (e.g., Timmermans and Berg 2003; Stevens 2007; Lamont 2009; Sauder and Espeland 2009).

Figure skating and classical music provide an ideal comparison for addressing differences in evaluation practices. Skating is an extreme case in two ways: its evaluation system is highly formal and

standardized, with very specific numerical rules, and it has overhauled its judging system in recent years, largely in response to a legitimacy crisis, which has completely changed many other aspects of the sport. Many of the music competitions I focus on represent the opposite extreme, with much more informal evaluative rules, but music competitions use evaluation practices that are all over the map. This variation is very different compared to the standardized evaluation in figure skating, despite how the goal in both settings is the same: figuring out how to rank participants and ultimately who should win. Given that their goal is the same, why is evaluation in these settings so different, and what do those differences look like? These are the main issues I address here, focusing especially on the latter question.

For the figure skating component of my research, I interviewed 33 skating insiders in the United States, most of whom I recruited from an official list of about 200 active technical panelists. My sample includes four Olympic-level, four world-level, two international-level, and nine national-level judging officials. My interview request response rate was about 15%, but considering the range of experiences among the skating insiders I did interview and the other data I collected, I do not believe that this relatively low response rate affected my findings. I conducted these interviews over the telephone, mostly in July and August 2006 but also in July and August 2010. Both rounds of interviews took place just a few months after the Olympics, held in February 2006 and 2010, respectively, so the most recent Olympics was fresh in respondents' minds for both rounds. Interviews averaged about 45 minutes long but lasted between 22 minutes and almost two hours, and with interviewees' permission, I recorded and later transcribed them. We covered several broad topics, including the 2002 Olympic judging scandal and ensuing responses, the two judging systems and especially their differences, how skaters and their choreographers construct their competitive programs, and figure skating's relationship with outsiders like skating fans.

I also collected data as a participant–observer in several settings, including two judging seminars for the overhauled judging system – known as the "international judging system" (IJS) – one in August 2006 that took place at a regional-level competition where some participants did trial judging, and a national technical panel training seminar in April 2009. This data is especially useful for showing how judging officials learn how to judge under the IJS and how they interpret the IJS, and it shows how formal the IJS is and its emphasis on numbers and following written rules. In addition, I collected a variety of archival data spanning from 1997–2023, including television coverage and web-based material from a number of sources. The "communications" published by the International Skating Union (ISU) were especially useful because they provide an official record of the ISU's response to the 2002 Olympic scandal and how the judging

changes emerged, and several official documents include guidelines for identifying technical elements and their levels of difficulty, as well as determining technical and artistic scores.

I collected similar types of data for the classical music side of this study. I interviewed 63 music insiders, including 21 music students at the college level or above (many of whom had at least some teaching experience), 40 musicians who had taught at a university-level music school or conservatory (or several) in the United States, and two other professional musicians. Everyone I interviewed was an active musician at the time, and almost everyone had been involved in activities that revolve around evaluating musical performance on a regular basis. These interviews were conducted in person between August 2009 and May 2010 and averaged about one hour in length. Other than one outlier that took only 17 minutes, interviews with professional musicians ranged from 33 minutes to two hours and 20 minutes, averaging about one hour and 15 minutes. Interviews with students were generally a bit shorter, ranging from 23 to 69 minutes and averaging about 45 minutes. We discussed several general topics, including evaluation criteria that music insiders tend to care about, what makes a "great" performance, and evaluation at music competitions.

In addition, I collected data as a participant–observer in a number of settings. The most significant component of this data includes participant-observation at six music competitions, four where I observed judges' deliberations and other meetings: a trumpet competition over two years (March 2009 and 2010), two rounds of a national-level organ competition held in May and July 2010, a conservatory concerto competition held in April 2010, and a conservatory piano competition over two years (April 2009 and 2010). This data illuminates how judges negotiate in a deliberative context and ultimately reach a consensus, as well as how deliberation can affect competition outcomes. It also shows the informality and discretion involved in evaluation in music. Finally, I collected archival data from a variety of sources. This data includes information about rules and judging procedures from many competitions' and related organizations' websites. To supplement this data, I corresponded with staff from several of these competitions about how they are judged and how their evaluation practices have evolved over time. Additional sources included several music schools' and conservatories' websites, music blogs, and other miscellaneous materials. This archival data reveals the wide range of evaluation practices surrounding music competitions, the general informality surrounding music compared to skating, and whether and how particular music competitions' evaluation practices have changed over time.

Data analysis involved mainly open coding and content analysis of interview transcripts, field notes, and archival materials. Mostly

following an inductive, grounded theory approach as outlined by Charmaz (2001), I used my research questions and my prior knowledge of figure skating and classical music as points of departure for developing interview questions and analyzing data. This is also where my initial codes – including "judging," "judging changes," "competitions," "deliberations," "discretion," "evaluating evaluators," and "repertoire" – came from. As I collected and analyzed more and more data, I added many codes that emerged from the data, such as "numbers," "formal/informal," "conformity," "freedom/creativity," and "legitimacy." This two-phase coding process is similar to how several experts in qualitative research methods and data analysis describe how this process often works for qualitative researchers (for example, see Lofland and Lofland 1994; Emerson et al. 1995). I now turn to a discussion of factors influencing the formality of evaluation practices, revolving around organizational structure and focusing on how rules and practices have developed in figure skating and classical music.

Organizational context: Sports vs. the arts

Although figure skating and classical music have different evaluation practices, they are similar on other dimensions. Perhaps most obviously, both contexts have technical and artistic components, which competition evaluators must take into account as they are judging. Given the central role of aesthetic principles in art worlds (Becker 1982) and how often members of these settings need to evaluate quality, people in the arts tend to grapple with evaluation more explicitly than in many other social contexts. This makes the arts in general, and figure skating and classical music competitions in particular, especially useful for studying social processes surrounding evaluation.

Many people describe artistic performance as "elusive" or "ineffable," which contributes to tensions in evaluating performances with artistic features. A lot of experts believe very strongly that using formal criteria to break down and evaluate artistic products reduces them, but competition judges are forced to do this all the time. However, while some settings (like figure skating) have been required to formally codify artistic components in their judging processes, others (like classical music) have not. Although skating is situated within the art world, its position as a sport has had a profound impact on its evaluation practices and its need to rely on formal rules, whereas the music world as a whole has not needed to deal with similar pressures to adopt formal evaluation practices based on its more obvious position within the arts world.

In comparing figure skating and classical music, I focus on competitions in both domains rather than other evaluative settings because the people involved in evaluating competitions must make

excellence explicit. The need to identify one clear "winner," and "losers" by extension, requires competitions to develop clear ways to separate the winner from other competitors. This is different from other evaluative contexts in these arenas, such as skills tests in skating and auditions in music, where in many cases multiple participants can "win." When judges can pick multiple winners rather than just one, they can choose to use a broader definition of excellence and reward participants with a wider variety of strengths rather than one participant who fits a narrower set of criteria. Competition evaluators who must identify a single winner are usually forced to use a less flexible definition of excellence. This makes the evaluation practices surrounding these types of events especially significant.

There are a number of parallels between music competitions and competitive sports, including the complex organizational components that go into them, participants' hard work and preparation leading up to them, and the identification of a clear winner. Music competitions also involve technical aspects, such as sound quality, whether participants play or sing the right notes, and even how long a musician can sing or play without taking a breath, which are comparable to the athletic aspects of figure skating and other competitive sports. Despite these similarities, few people would consider music a "sport," and there are several unique characteristics that distinguish sports from most other areas of social life.

First, competition – where the goal is to win – plays a central role in sport unlike in any other social setting. While competition also plays a role in many other areas, such as education (for instance, being named valedictorian or getting into the "best" college), the workplace (job offers and promotions), politics (winning an election), and even the family (sibling rivalries or being the "best" mother), its role is much more explicit in sport. In fact, for many sports fans and participants, one of its main attractions is that "success" is measured more exactly than in other contexts: by doing well in competition (Chambliss 1989). Sport also involves more clarity than other areas: by the end of any athletic event, the audience and participants should know who won, who lost, by how much, and how they won and lost (Eitzen 2012). Going along with this is "the perceived inherent purity and goodness of sport" (Coakley 2015: 11), the notion that sport encourages "fair play" (Eitzen 2012: 60) and is "played on a level playing field" (16). Despite much evidence to the contrary (Coakley 2015), the idea here is that the outcome of sporting events should be determined solely based on athletic factors like talent, skill, hard work, preparation, and strategy (and perhaps luck), rather than social factors like where participants are from, who their parents are, or who they know. This "obvious meritocratic orientation" (Washington and Karen 2001: 189) necessitates an emphasis on "playing by the rules" (Coakley 2001: 20), which is why many organizations within the sports world have

incorporated bureaucratic practices, including strict rule systems that are in place partly to encourage objectivity (Eitzen 2012).

Beyond the importance of fairness and rules, sport is prevalent as a cultural, political, and economic symbol around the world, to such a degree that some major sporting events, like the Olympic Games, the World Cup, and the Super Bowl, to name a few, attract so much attention and media coverage that they are impossible to ignore, even for people who have no interest in sports otherwise (Coakley 2015). Especially with global sporting events like the Olympics, sport can contribute to nationalism and patriotic sentiment (Frey and Eitzen 1991) through "an emphasis on demonstrating superiority over other countries and other political systems" (Coakley 2001: 31). This is why sport has such strong political implications and plays such an important role in international relations (Guttmann 2003). The importance of the gold medal count in the Olympics is just one example of how success in sport is "interpreted internally and externally as 'proof' of the superiority of a nation's social, economic, and political systems" (Frey and Eitzen 1991: 512). Music also tends to play a cultural, political, and economic role within specific cultures and more globally, but the publicity surrounding even major international music competitions is typically nothing compared to global sporting events, and very few non-experts care about music competitions compared to sporting events.

While figure skating and classical music are very different in some ways, these two cases – one with highly formal, numeric, standardized evaluation, and the other with significant variations in its evaluation practices – provide similarities and differences that are useful for addressing why settings develop different evaluation practices. Figure skating's high profile among Olympic sports and relatively recent judging changes make it a unique case where it is possible to examine a transition between two very different judging systems. Given the similarities between the Olympics and music competitions (McCormick 2009), classical music offers an intriguing comparison. These two cases are different enough on the formal–informal continuum and in relation to how they are organized, which are major foci of this article, but they are similar in other ways. The publicity and visibility surrounding competitions more broadly, especially at the highest levels, means that they must at least appear legitimate to both participants and outsiders. This legitimacy is commonly, but not always, maintained through formal rules. Organizations involved in running competitions need to make sure that participants and outsiders see them as fair and that the evaluators involved can defend their decisions. All of this makes skating and music competitions a useful comparison for examining evaluation. I now turn to a discussion of other factors – competition structure, degree of

centralization, and governance structure, all of which directly come out of organizational differences between these two contexts.

Centralization, governance, and competition structure

Despite the prominent role of international competitions in both figure skating and classical music, these contexts are organized very differently. Figure skating is highly centralized and has an international governing body, the ISU, which also governs speed skating, whereas the music world lacks both of these elements. Based in Lausanne, Switzerland, the ISU was founded in 1892 following the emergence of international competitions in figure skating and speed skating during the second half of the nineteenth century. The ISU's main purpose has been to develop and maintain international standards for these two sports and to organize international competitions (ISU 2023a). In figure skating in particular, the ISU was responsible for overhauling the judging system after the 2002 Olympic judging scandal and is in charge of reviewing and updating the IJS on a regular basis. For individual athletes to compete at events that are linked to the ISU, like the World Championships and the Olympic Games, their countries must be ISU members. As of June 2023, the ISU has 101 member federations representing 80 countries (Australia, Belgium, Canada, Finland, Germany, the United States, and a number of other countries have two organizations, one for figure skating and one for speed skating) (ISU 2023b). Skating's status as an Olympic sport means that the ISU is accountable to a more powerful organization, the IOC, which has the highest authority over anything having to do with the Olympic Games (IOC 2021).

One thing that makes the ISU, and the IOC by extension, so significant in figure skating is its competition structure, where competitions build on each other. Winning an Olympic gold medal is the (rather obvious) pinnacle of the sport, which skating commentators tend to bring up over and over again leading up to and during the Olympics. The following statement by commentator and Olympic gold medalist Scott Hamilton at the beginning of the 2018 Olympics personifies how much of an emphasis skaters place on the Olympics: "It's the BEST!!! … This is the biggest stage these athletes will ever stand on … When they leave here their lives are forever changed" (Sheehan and Michaels 2018).

Especially in countries with skaters who are competitive on the world level, to have any hope of making an Olympic team, skaters must do well in other competitions first. Each year in the United States, for example, the National Championships (which most skaters must qualify for through regional and sectional competitions earlier in the

season) acts as the qualifying competition for many competitions that follow, including the World Championships and the Olympics but also the six-event Grand Prix Series that takes place internationally each autumn and other competitions that receive less publicity but can be very important for skaters to establish themselves. Even before those events, the number of spots that each ISU member nation has at competitions like the World Championships and Olympics depends on their skaters' placements at previous competitions. For instance, in the year before an Olympics, for a member nation to be granted three spots to those Olympics and the following year's World Championships, the combined placement of their top two skaters at that year's World Championships must be no higher than 13. If two Japanese skaters were placed sixth and seventh in the women's competition, Japan would receive three spots for the following year, but if that second skater were placed eighth, they would only receive two spots. Before skaters reach that level, to put their names in the mix for consideration for bigger events, they must do well at other less prestigious competitions.

Especially with this kind of competition structure, it makes sense that the ISU requires a standardized set of rules at all ISU-sanctioned competitions and that the vast majority of other competitions around the world also uses those rules so that athletes competing there are better prepared for ensuing competitions. This competition structure is very different compared to music competitions, which rarely build on each other. Some competitions are limited to a particular instrument, age range (classical music competitions tend to be limited to young musicians at the beginning of their careers), or repertoire, or have other entrance criteria, but as long as musicians meet those requirements, they can enter those competitions. Even for many prestigious international music competitions, any eligible musician can submit an application, usually consisting of a recorded performance.

In the music world, the most similar organization to the ISU is the World Federation of International Music Competitions (WFIMC), a network of many of "the world's most important music competitions," with 125 member competitions as of June 2023 (WFIMC 2021). Founded in 1957 and based in Geneva, Switzerland, one of the WFIMC's chief objectives is helping and advising member competitions, including by helping them communicate with each other and with other organizations, promoting their prizewinners' careers, and presenting international music competitions in a positive light more generally. This organization also requires "that member competitions maintain the highest professional standards and strictest ethics" (WFIMC 2022).

Music competitions can also become members of the Alink-Argerich Foundation (AAF). Founded in 1999, the AAF provides "the most complete details on music competitions ever compiled" (AAF 2023),

with particular emphasis on piano competitions, and works to help and provide information to musicians and competition organizers. While the WFIMC is only open to international competitions, any musician or competition that is willing to make an annual contribution is eligible to join the AAF. In addition to providing "dedicated assistance" to all members, one of the AAF's main goals is to provide member competitions with publicity through its website and inclusion in its annual brochure. As of June 2023, the AAF has 208 member competitions, some of which are also WFIMC members (AAF 2023).

Even though the WFIMC and AAF play significant roles at the international level, neither organization acts as an international governing body for classical music. As long as music competitions do not build on each other, the music world does not have much need for a governing body that would oversee international competitions and enforce standardized rules. In addition, although many international music competitions have applied for WFIMC membership and have paid to join the AAF, hundreds of successful international music competitions are not members of either organization. Without an organization to impose a single set of rules on all competitions, music competitions can use whatever evaluation practices they deem most appropriate. As many of the music insiders I interviewed pointed out, specific competitions' evaluation methods usually depend on the organizations that run them and sometimes even differ from one year to the next. One piano teacher noted, "The guidelines really can be quite different. Sometimes it's just numbers, or you throw out the high and the low, or 'yes, yes, yes, or no,' or I mean it's just, every competition organizer has to make those decisions as to how they will evaluate it" (Pianist and music conservatory faculty member). The music world's lack of centralization and international governance and its disconnected competition structure encourage a wide range of evaluation practices, whereas figure skating's centralization, international governance, and integrated competition structure more or less require shared evaluation practices.

Evaluation practices in figure skating vs. classical music

In the following sections, I provide a detailed account of how evaluation works in figure skating compared to classical music. Building on work on rules (Meyer and Rowan 1977; Dobbin and Kelly 2007; Edelman et al. 2011; Edelman 2016) and evaluative cultures (Lamont 2009, 2012), as well as Porter's (1995) discussion of objectivity, in addition to highlighting the evaluation practices that emerge in different contexts, especially shared versus varied practices, this article enhances our understanding of how the quest for formality

and objectivity on the one hand and the use of deliberation and the discretion that tends to go along with it on the other hand affect evaluation practices on the ground. This matters because in a deliberative context, judges often influence each other's decisions about competitors' merits and how to evaluate them, which can change outcomes, whereas non-deliberative evaluation prevents judges from influencing each other's opinions, at least in theory. The differences discussed here between figure skating and classical music also enhance our understanding of how and why evaluative cultures change.

Figure skating: Shared, rigid evaluation

While skating has been around as a mode of transportation for more than 3,000 years, competitive figure skating did not begin until around the mid-1800s (Hines 2006). Informal international competitions cropped up during the latter half of the century, and the first official International Figure Skating competition was held in Vienna in 1882 (ISU 2012). Skating styles were very different in different parts of the world around this time, with an emphasis on movement across the ice in continental Europe, geometric formations in England, and developing unique designs in North America. These differences created problems for judging international competitions fairly, which led to the need for international judging criteria (Hines 2006).

Following the rise of national skating organizations and international competitions, most prominently the first European Championships in 1891, the ISU was established in 1892 to develop international standards, which involved many compromises but most closely resembled the "international style" of continental Europe. After the first World Championships in St. Petersburg in 1896, in 1908, figure skating was the first winter sport included in the Olympic Games (ISU 2012; Hines 2006, 2011). In current international competitions, skaters in each discipline (men, women, pairs, and ice dance) skate a short program and a long program (also known as a "free skate") (Hines 2006). The ISU first introduced formal rules for figure skating competitions around 1895 with the 6.0 judging system (Hines 2011) and kept using this system, with occasional adjustments over time, until 2005, a few years after a very consequential judging scandal.

The 2002 Olympic judging scandal

Many people inside and outside of skating would identify the judging scandal during the 2002 Olympic Games in Salt Lake City, which preceded the ISU's total overhaul of its judging system, as the

most significant scandal in skating history. The top two teams in the 2002 pairs event were from Canada and Russia. Despite a technically flawless long program by the Canadians, the Russians won the gold medal, even with a noticeable mistake. After the Canadians' marks were unveiled, indicating a second-place finish, NBC commentator Scott Hamilton exclaimed, "How did that happen?! … They won that program! … There's not a doubt in anyone in the place, except for maybe a few judges … That will be debated forever … Debated forever" (Gunts and Michaels 2002). Five of the nine judges involved placed the Russians ahead of the Canadians in the free skate, but the only one who received widespread publicity was Marie-Reine LeGougne of France, who was involved in a vote-swapping deal with Russia. After the medals had been awarded, LeGougne acknowledged that the French Ice Sports Federation had pressured her to favor the Russians so that the Russian ice dance judge would favor the top French ice dancers and essentially assure them a gold medal.

In response to the uproar over this result, the next day, the ISU began an investigation of the event's judging. The day after that, IOC President Jacques Rogge met with ISU President Ottavio Cinquanta "to emphasize that the situation needed to be resolved quickly" (Roberts 2002a). During a "late-night emergency meeting" four days after the event, the ISU decided to throw out the French judge's marks for the Russians, which led to a four-to-four tie and a duplicate gold medal for the Canadians. Only days later, while the scandal was still a big story, the ISU "unveiled plans for a new scoring system designed to limit the threat of collusion between judges." In an attempt to divert people's attention from the scandal, Cinquanta stated, "'This is a total revolution in the history of the International Skating Union … I promise this system will reduce to a minimum the prospect of bloc judging'" (Roberts 2002b). This scandal was so significant in figure skating history partly because "no judge had ever admitted to outright cheating" before (Jackson 2005: 198). There had been many previous scandals involving disputed outcomes at major events like the World Championships and the Olympics, but unlike this 2002 scandal, none of the judges involved in those competitions ever admitted that they had done anything wrong.

Given the ISU's relationship with a more powerful governing body, the IOC, figure skating insiders must answer to a higher power. The power dynamics between these two organizations were critical in shaping the judging changes that followed the 2002 Olympic scandal. This scandal emerged as a central storyline of those Olympics, and in response, the IOC insisted that the ISU take action to restore its public image. This IOC pressure played a fundamental role in accelerating the transition from the 6.0 system to the IJS. An Olympic-level judging official who was involved in designing the IJS explained how this process unfolded:

> It was the International Olympic Committee that went to the ISU and said, "Either you come up with a better way of measuring your sport, or you're out of the Olympic Games." So it was a mandate by the IOC to the International Skating Union to improve the way they measure their athletes. For example, in speed skating you're measured by time. In figure skating in the 6.0 system, you were measured by the impression of someone else, so the IOC said, "That's not acceptable sport. We need a measuring stick" … So the ISU came up with this system of measurement that was acceptable to the IOC and kept the sport in the Olympic Games.
> (Olympic-level judging official and ISU Technical Committee member)

This IOC mandate and the ISU's response to it highlight how influential figure skating's centralization and shared governance have been in shaping its evaluation practices. If the ISU had been self-governing and had not had to worry about satisfying a more powerful organization like the IOC, implementing these judging changes would have probably been a much slower, more deliberate process, and the ISU would have at least had the option of just keeping the old 6.0 system without any changes. The ISU had already been thinking about changing its judging system, and the 2002 Olympic scandal, especially the French judge's admission, gave its leaders a perfect opportunity to go ahead with those changes. In keeping with the link between sport and notions of "fair play," the International Sports Federations that belong to the Olympic Movement, such as the ISU, must ensure that their sports are fair on an international level, including in relation to how competition outcomes are determined. The IOC's commitment to encouraging ethics across all Olympic sports helps explain why it put so much pressure on the ISU to do something to uphold its integrity after the French judge's confession became public.

The judging changes

After this scandal, the ISU completely overhauled its judging system to make it more formal and more numerically specific. The new "IJS" was officially unveiled in late 2003 and gradually implemented leading up to the 2006 Olympics. I outline the main differences between the 6.0 system and the IJS in Table 1.

Table 1: Figure Skating's 6.0 System vs. IJS*
Source: Author's work

	6.0 System	IJS*
Judging Panels	One panel	Two panels: one technical panel and one judging panel
Judging Tasks	All judges identified and evaluated technical elements and artistic presentation	Technical panelists: identify skaters' technical elements and their levels of difficulty Judges: evaluate skaters' technical elements and program components
Anonymity	Judges' nationalities displayed	Judging officials' nationalities hidden
Technical Evaluation	0-6.0 scale for overall technical merit	Pre-determined base values for all technical elements; judges evaluate elements with grade of execution (GOE) scores ranging from -3 to +3**
Deductions	Automatic deductions for specific errors on required elements (short program only)	Deductions for specific errors
Artistic Evaluation	0-6.0 scale for overall artistic presentation	0.25-10.0 scale for five program components: skating skills, transitions, choreography, interpretation, and performance/ execution***
Competition Outcomes	Ordinal rankings combined after short and long programs	Cumulative point totals after short and long programs

*The information here reflects the IJS as it was during much of my data collection and analysis process. Some IJS rules have changed over time, as explained here.

**Since the beginning of the 2018-19 season, judges have been evaluating technical elements with GOE scores ranging from -5 to +5.

***Following the 2021-22 season, the ISU decided to reduce the number of program components from five to three: composition, presentation, and skating skills. The ISU Technical Committees initiated these changes after observing that there were too many criteria at work for the original five components and that those criteria were often unclear. According to the ISU, those criteria "have been simplified and reorganized in a more logical way" to encourage "more objective evaluation" (International Skating Union 2022b).

The most significant difference is that the rules are now much more detailed, with much more specific judging criteria, but the 6.0 system

and the IJS are generally quite different.

Under the 6.0 system, judges' placements determined competition outcomes. Each judge gave skaters two marks for each program: one for "required elements" (in the short program) or "technical merit" (in the long program), based on how well skaters performed the elements in their programs, and one for "presentation" that reflected skaters' overall programs, including composition, style, originality, and musical interpretation. Both marks were combined to determine each skater's total score from each judge, which led to skaters' "ordinals" from each judge. Skaters' placements for a particular phase of a competition were based on comparing ordinals, with the skater who received the most ordinals matching a given placement finishing in that spot. Because the long program was supposed to be worth twice as much as the short program, to determine overall competition results, short program placements were multiplied by 0.5, and free skate placements were multiplied by 1.0.

The IJS is generally much more exact than the 6.0 system, with each facet of skaters' programs broken down and evaluated based on precise guidelines. In contrast to the 6.0 system's ordinal rankings, the IJS is based on cumulative points. Another important distinction is that rather than relying on just one judging panel, the IJS uses two panels with distinct types of judging officials: (1) a three-person technical panel, typically made up of coaches or former skaters, which is responsible for identifying each technical element in a skater's program and its level of difficulty, and (2) a judging panel, usually with nine judges, which is in charge of evaluating each of those elements and several separate program components. Under the 6.0 system, judges needed to both identify elements and evaluate skaters' programs, but with the added technical panel under the IJS, judges can focus on evaluating elements and components. For the remainder of this article, I refer to judges and technical panelists as such and use "judging officials" or "officials" to refer to both groups simultaneously.

In Table 2, I describe how officials evaluate certain facets of skaters' programs under the 6.0 system versus the IJS.

Type of element	6.0 System	IJS*
Jumps **Triple Axel**	Judged as part of the overall required elements mark (short program) or technical merit mark (long program); if the skater made a mistake on this element in the short program, judges took a specified deduction according to what kind of mistake it was (this did not apply to mistakes in the long program); judges may or may not have considered this element and other jumps when determining presentation marks	Technical panel identifies "triple Axel" and determines whether or not the skater completed the required revolutions (if there is any doubt, the panel reviews the jump using video replay after the end of the program); judges evaluate the jump with a GOE score from -3 to +3; the average GOE score is factored and added to or subtracted from the jump's base value to determine the total score for this element**
Spins **Layback Spins**	Judged as part of the overall required elements mark (long program) or technical merit mark (long program); judges may or may not have considered quality of spins when determining presentation marks**	Technical panel identifies the level of difficulty depending on the number of "features" the skater achieves: judges evaluate the spin with a GOE score from -3 to +3; the average GOE score is factored and added to or subtracted from the spin's base value to determine the total score for this element
Footwork **Circular Step Sequence**	Judged as part of the overall required elements mark (short program) or technical merit mark (long program); judges may or may not have considered quality of footwork when determining presentation marks	Technical panel identifies the level of difficulty depending on the number of "features" the skater achieves; judges evaluate the step sequence with a GOE score from -3 to +3; the average GOE score is factored and added to or subtracted from the step sequence's base value to determine the total score for this element
Presentation/ Artistry* **Composition**	Judged as part of the overall presentation mark	Judges evaluate composition on a scale of 0.25 to 10.0 based on specific criteria

*Unless otherwise noted, the information in this table is from ISU Communication No. 1611 (International Skating Union 2010) and reflects the IJS as it was during much of my data collection and analysis process.

**Since the beginning of the 2018-19 season, judges have been evaluating technical elements with GOE scores ranging from -5 to +5.

***Guidelines for determining program component scores are provided in several ISU documents, including "Program Components Overview" (International Skating Union 2004) and "ISU FAQ – Program Components" (International Skating Union 2022b).

Table 2: Element Evaluation in Figure Skating: 6.0 System vs. IJS*
Source: Author's work

Under the IJS, judging officials determine skaters' technical scores as follows. All elements have "base values" that correspond with specific written criteria that the technical panel uses to identify them. Base values for jumps are based on the type of jump and the number of revolutions, with the "more difficult" jumps with more revolutions receiving higher base values.[1] Base values for other elements, such as spins and step sequences, are based on levels of difficulty, which revolve around how many "features" (such as position changes, "difficult" positions, or a particular number of revolutions) they involve. As the technical panel identifies each element and enters it into the computer system, the judges evaluate it with a "grade of execution" (GOE) score from -3 to +3 based on detailed guidelines.[2] Judges can give a GOE of –3 on a jump, for example, based on these criteria: "any of the following individual errors: fall; severe change of edge on take-off of flip or lutz; lesser rotation than required" (ISU 2010). The following guidelines, in contrast, are provided for giving a GOE of +3 on a jump: "Superior in all jump phases (eg. unexpected or difficult entry phase, great height/distance, strong flow in and out and superior extension on landing)" (ISU 2010). Judges must follow similar written criteria for all types of elements. To determine the score for a particular element, the GOEs from all judges for that element are averaged and then added to or subtracted from its base value. The total technical score for a given program is based on adding up all of the element scores from that program.

[1] Of the six types of jumps, the Axel is considered the most difficult (partly because it has an extra half revolution compared to the other jumps), followed by the Lutz, flip, loop, Salchow, and toe loop. The base values of these jumps vary accordingly, as the following examples illustrate:

Double toe loop 1.30 vs. double Lutz 2.10 vs. double Axel 3.30;
Triple toe loop 4.20 vs. triple Lutz 5.90 vs. triple Axel 8.00;
Quad toe loop 9.50 vs. quad Lutz 11.50 vs. quad Axel 12.50 (ISU 2022c).

[2] Since the beginning of the 2018–19 season, judges have been using a GOE scale from -5 to +5.

Under the 6.0 system, judges incorporated all technical elements (including jumps, spins, and step sequences) into skaters' technical marks, whereas each technical element is identified and evaluated separately based on strict guidelines under the IJS. According to the IJS rulebook, judges can award higher GOE marks if skaters show control, good speed, effortlessness, flow, and high-quality positions; well-centered spins; and elements that go along with the music. On the other hand, if skaters under-rotate, two-foot, step out of, or fall on a jump; take too long to prepare for a jump; fail to complete the required number of spin revolutions; or display poor positions, edge quality, or speed, judges are supposed to lower their GOE marks. Even for the best skaters in the world, judges rarely give the highest GOE mark of +3 (ISU 2010).

Similar to technical element judging, the artistic side of the sport is broken down much more specifically under the IJS than under the 6.0 system. Judges used to give skaters a single mark for presentation, which was designed to incorporate all artistic aspects of skaters' programs, often with an emphasis on choreography and interpretation. Some judges, at their discretion, included the quality of skaters' technical elements like spins, footwork, and even jumps in determining their presentation marks for a particular skater. Under the IJS, the artistic side of skaters' programs is judged on five separate "program components": skating skills, transitions, choreography, interpretation, and performance/execution. Judges evaluate each program on each one of these categories on a ten-point scale (in increments of 0.25), and this is supposed to measure a skater's overall technical mastery and presentation. Similar to the GOE marks, the IJS rulebook provides precise criteria for each component, based on the following qualities:

> Skating skills: balance, precision, flow, speed, effortlessness, and edge quality;
>
> Transitions: variety, difficulty, intricacy, and quality;
>
> Choreography: use of space, ice coverage, purpose, originality, and matching the music;
>
> Interpretation: musical expression and nuance;
>
> Performance/Execution: commitment, style, personality, carriage, movement clarity, variety, and projection (and how well a skater does technically).[3]

[3] Following the 2022 Olympic season, the ISU reorganized these five program components into just three: (1) Composition: "how the program is designed in relation to the music;" (2) Presentation: "how the program is performed," including skaters' "expressive abilities" and "musical sensitivity;" (3) Skating skills: technique and movement (ISU 2022b).

In addition to these factors, component marks should be based on the percentage of the program in which the skater meets those guidelines for each component. After the IJS was first introduced, even Olympic champions received many component marks below 8.0, but these marks have generally become much higher over time, with the top skaters at the 2022 Olympics receiving many component scores in the nine-range (ISU 2004). Skaters' total scores for each program are determined by adding their technical score and program component score together, and total competition scores are based on combining skaters' short program score and free skate score. The competition winner is simply the skater with the highest total competition score.

Despite how both the 6.0 system and IJS are generally based on "technical" and "artistic" marks, the IJS is much more rigid than the 6.0 system. Other than requiring judges to take specific deductions for certain types of mistakes on technical elements in the short program, judges' 6.0 system marks were not based on precise written criteria. Judges used the old technical and artistic marks as a tool to rank skaters and based those marks largely on their overall impressions of skaters' programs, which meant that judges had a lot of discretion in determining their marks. Especially following the 2002 Olympic judging scandal, the IJS was intended partly to make cheating and deal-making more difficult and to increase fairness surrounding competitions through stricter rules revolving around evaluating much more specific aspects of skaters' programs.

Judging the judges

Many of the skating insiders in my interview sample applauded the IJS for its relative transparency, including in relation to the training and testing process for judging officials, which has become much more formal since the judging changes. According to my data, assembling international judging panels under the 6.0 system did not involve many formal criteria. One of my informants who had a lot of experience with this on the ISU side explained, "Romania could have just sent someone and said, you know, 'This is our representative and they are qualified to judge at the Olympic Games,' and that's how it was done" (Olympic-level judging official and ISU Technical Committee Member). With the IJS, while each ISU member nation can decide how to train their own judging officials, before they can serve as judges or technical panelists for ISU events like the Olympics or World Championships, all officials must attend judging seminars and pass a series of tests.

To use the United States as an example, anyone who wants to become a judging official needs to pass a written exam just to qualify to attend a judging seminar that will count toward that process. There

are different types of IJS judging seminars, some more focused than others – on the technical panel (like one I attended as an "observer" in April 2009), a particular discipline like singles or ice dance, or certain elements like spins or program components, for instance – and most officials need to attend a number of seminars before they can pass all of the necessary tests and need to continue attending them to keep up with IJS rule changes and to advance to higher levels of officiating. These seminars are taught by experienced judging officials, who also evaluate exams at the end, and were described by some of the officials I interviewed as "unexpectedly grueling" and "very overwhelming" (Technical panelist).

During the technical panel seminar that I attended in April 2009, I heard a lot of discussion revolving around rules and rule changes, how to interpret certain rules, and how to call specific technical elements. The seminar faculty repeatedly emphasized the importance of knowing the rules and applying them correctly, whether or not judging officials agreed with them, and being able to justify their calls. For example, one faculty member said at one point, "If it's not in writing, don't say it," and another followed with, "You have to know the rules … 'it is' or 'it is not;' no 'it might be' … Your job is to apply the rules, not interpret the rules" (Field notes, April 17, 2009).

There was also a lot of discussion of how the tests at the end of the seminar would be evaluated, which gave me and everyone else who was there a great deal of insight into what these tests tend to look like and how they tend to be judged in general. To become a technical official or a judge, applicants must take a written test and act as an official for anywhere from four to six programs. During these "test" programs, technical officials need to call the elements in each program, and judges need to identify and justify GOE and program component scores. The seminar faculty highlighted several important criteria for evaluating these tests: identifying elements correctly and confidently, diction, the review process, pace of calling, focus (not getting hung up by incorrect calls), and teamwork, and they emphasized that fast calls are especially important because they affect everything else that happens during a competitive program, especially judges trying to enter GOE marks after a skater does each element (Field notes, April 17–19, 2009).

These exams are generally very difficult, which was emphasized by not only the faculty at this seminar but also several of my interviewees. One faculty member in particular, the ISU Technical Committee member quoted earlier, described these exams like this:

> [I]t's a very, very high-pressure type of environment, and we do that on purpose to make them nervous, which is what it really feels like when you're a judge or an official at an Olympic Games … If they say +2, they have to specifically identify in the rules and regulations, not by their feelings or, "Gee, I just think it was that good."
> (Olympic-level judging official and ISU Technical Committee member)

Another judging official at this seminar echoed these sentiments during an informal conversation I observed, where she told someone else who was there to take a test at the end that judging actual competitions is much more relaxing than calling programs during a test because there is so much pressure (Field notes, April 17, 2009).

The ISU trains hundreds of judging officials every year, but just a very small percentage of officials who go through the training process passes the exam. Since the IJS was introduced in 2003, the ISU has been collecting a lot of of data, including recordings of all technical panels for all performances at all ISU events. Especially compared to the relative lack of data from the 6.0 system, this comprehensive IJS data has made it much easier to distinguish among judging officials, identifying officials who are consistently quick and accurate and exposing officials who are not. When the ISU Technical Committees are deciding which judging officials to appoint to ISU events, they look at this data and select officials who have done consistently well on exams and at previous competitions, with a particular emphasis on accuracy and speed.

In keeping with all of this, figure skating has also implemented very strict rules for dealing with judging ethics, which I learned about firsthand before and during the technical panel seminar I attended. A few days before it started, all participants received an email asking us to read the "Technical Panel Code of Ethics" ahead of time because everyone would need to sign a compliance agreement during the on-site registration process. The code provides official ethics guidelines for technical officials involved in U.S. Figure Skating (USFSA) events, specifically by enforcing "the highest standards of ethics, fairness, honesty and integrity." One aspect of the code that the instructors at this judging seminar emphasized was that judging officials must "recognize that even the appearance of misconduct, impropriety, insincere attitude or purpose can be damaging." If officials fail to follow any of these rules, the USFSA can suspend their judging appointments or even expel them "from all U.S. Figure Skating events and activities." These guidelines have been designed in part to help participants, including athletes, and outsiders feel more confident that USFSA events will be conducted and judged fairly (USFSA 2009b).

The USFSA's preoccupation with maintaining high ethical standards illustrates how many organizations have become increasingly concerned about ethics over the past several decades

(Weaver et al. 1999; Erwin 2011). Many organizations have implemented corporate ethics programs, for instance, which often include establishing ethics committees and training programs, developing ethics codes and communication systems, and other features geared toward formalizing expectations for employees and promoting ethical organizational cultures (Weaver et al. 1999; Erwin 2011). In figure skating, the IOC's reaction to the 2002 Olympic scandal put pressure on the ISU and its member nations to make sure the sport's ethical conduct improved quickly. Requiring participants to sign an "Agreement to Comply with Code of Ethics" form (USFSA 2009a) before this judging seminar is one of the clearest examples of this from my data.

One result of the judging changes is that many coaches now serve as technical panelists. This means that coaches' and judging officials' positions sometimes overlap, which has complicated the ethical standards within the sport. The faculty at this judging seminar addressed this issue in detail, advising that coaches who have become certified technical panel officials must acknowledge when they are coaching a skater in an event. Whenever this is the case, they are not allowed to judge that event. As one of the instructors pointed out, this is an important rule because "the perception is that you have an inside track … Perception is reality" (Field notes, April 18, 2009). All of this exemplifies skating's highly formal evaluation practices.

Classical music: Varied, flexible evaluation

In contrast to figure skating's formal, standardized evaluation practices, evaluation in classical music is varied and generally much more flexible. While music contests reportedly go back to ancient times, the modern music competition did not become especially prominent until the nineteenth century, throughout which competitions became increasingly popular, especially in Europe. During the twentieth century, many international competitions were created, mostly focused on a particular instrument, especially piano or violin, and many to celebrate a famous musical figure. The International Tchaikovsky Competition, Chopin International Piano Competition, and Paganini International Violin Competition are three prominent examples of competitions commemorating renowned composers (Latham and Spencer 2002).

Table 3 compares evaluation in figure skating and classical music on several significant dimensions.

Figure Skating	Classical Music
Standardized evaluation practices	Varied evaluation practices
Formal, numeric rules	Generally much more informal rules
Technical and artistic components	Technical and artistic components

Table 3: Evaluation in Figure Skating vs. Classical Music Competitions
Source: Author's work

The most significant differences for the purposes of this article are that evaluation in skating is standardized and generally much more formal than in music. While some music competitions use numeric evaluation systems, many do not, and when they do rely on numbers, judges often give a single score (on a scale from zero to ten or zero to 100, for instance) to the performance as a whole, rather than scoring specific components of the performance like judging officials do in skating now.

In Table 4, I compare elements of three well-known international music competitions: the Queen Elisabeth International Music Competition, Chopin International Piano Competition, and Paganini International Violin Competition.

	Queen Elisabeth International Music Competition*	Chopin International Piano Competition**	Paganini International Violin Competition***
Frequency	Annual	Every five years	Annual until 2002 Biennial since 2002
Location	Brussels, Belgium	Warsaw, Poland	Genoa, Italy
Instrument(s)	Violin, Piano, Voice, Composition (competition for one or two instruments/ categories each year)	Piano	Violin
Age Range	18-29	17-30	16-30
Repertoire Guidelines	Guidelines but some choice	Guidelines but some choice; works by Chopin only	Guidelines but some choice
Number of Rounds	Four	Six	Four

Competition Format	*Preselection Round:* Confidential Judges evaluate applicants' DVDs to eliminate participants who do not play well enough for public rounds *Preliminary Round:* Public Unlimited number of participants Participants play their first two or three pieces, then the jury selects one or more etudes from their repertoire lists *Semi-Final Round:* Public 24 participants Recital phase: Participants play a piece written for that year's competition; jury selects recital programs from repertoire lists Concerto phase: Participants play a Mozart concerto *Final Round:* Public 12 participants Prizes awarded	*Screening Round:* Confidential Admissions Committee (10 members) evaluates applicants' documents and DVD recordings Participants choose repertoire order for all rounds *Preliminary Round:* Public No more than 160 participants 20-member jury 12-member Competition jury *Competition Stage I:* Public No more than 80 participants *Competition Stage II:* Public No more than 40 participants *Competition Stage III:* Public No more than 20 participants *Final Round:* Public No more than ten participants Prizes awarded	*Screening Round:* Confidential Judges evaluate applicants' documents and CD recordings Participants choose repertoire order for all rounds *Preliminary Round:* Public Unlimited number of participants *Semi-Final Round:* Public No more than 12 participants *Final Round:* Public No more than six participants Prizes awarded

Number of Judges	Varies from year to year	12	At least seven
Judging Procedures	Secret ballot system for all rounds *Preselection Round:* Yes/no system *Preliminary Round:* Numeric and yes/no system (0-100 points) *Semi-Final Round:* Numeric and yes/no system (50-100 points) *Final Round:* Numeric system (60-100 points) and rankings	*Pre-Final Rounds:* Numeric and yes/no system (1-100 points) Discussion and open ballot voting *Final Round:* Numeric system (1-100 points) Discussion and open ballot voting	*Preliminary and Semi-Final Rounds:* Voting system (yes/no/possible) *Final Round:* Numeric system (70-100 points)
Prizes	1st Prize: 25.000€; concert performances 2nd Prize: 20.000€; concert performances 3rd Prize: 17.000€; concert performances 4th Prize: 12.500€; concert performances 5th Prize: 10.000€; concert performances 6th Prize: 7.000€; concert performances Unranked Finalists: 4.000€; recital	1st Prize: 30.000€ and gold medal 2nd Prize: 25.000€ and silver medal 3rd Prize: 20.000€ and bronze medal 4th Prize: 15.000€ 5th Prize: 10.000€ 6th Prize: 7.000€ Remaining Finalists: 4.000€ each Special prizes (including concert performances, money, and gifts) are also offered independently of the prizes listed above.	1st Prize: 25.000€; concert engagements 2nd Prize: 10.000€ 3rd Prize: 5.000€ Other Finalists: 1.500€ each Additional special prizes

*This information is from the Queen Elisabeth International Music Competition's official website (https://concoursreineelisabeth.be/cgi?lg=en).

**This information is from the Chopin International Piano Competition's official website (https://chopin2020.pl/en).

***This information is from the Paganini International Violin Competition's official website (https://www.premiopaganini.it/) and the website for the 2010 competition (http://www.carlofelice.it/VediFocus.asp?imgPath1=Media&imgPath2=focus&parentZ=5020&itemID=5020&idNews=1247&idSpettacolo=0).

Table 4: A Comparison of International Music Competitions

Source: Author's work

This information illustrates the variation among competitions, including in terms of featured instruments, frequency, numbers of rounds, and how many judges they use. For the purposes of this article, one of the most important distinctions among these three competitions is their different judging practices.

The Queen Elisabeth Competition uses secret ballots across all four rounds, but its procedures for each round are somewhat different. During the first round, before the actual competition starts, judges use a yes/no system, where they simply specify whether they think participants should or should not be admitted to the competition. As the competition progresses, these procedures become increasingly more complicated: numeric systems with various scales added to the yes/no system for the middle two rounds and a numeric system and rankings for the final round. The Chopin and Paganini Competitions also use different procedures depending on the round. The Chopin Competition combines a numeric and yes/no system with discussion and open ballot voting, which allows judges to see how other judges vote, but it drops the yes/no system for the final round. For its first two rounds, the Paganini Competition uses a "yes/no/possible" voting system, but then it uses a 70- to 100-point numeric system in the final round. In contrast, it is not uncommon for competitions to use the same procedures for all rounds, and sometimes a particular competition uses different practices each time it is held depending on who is involved. As these examples highlight, competition judging practices can be wildly different.

While some music competitions have developed more formal evaluation practices over time, many competitions rely heavily on deliberation. This resembles the multidisciplinary peer review panels outlined by Lamont (2009), who shows how participants use their deliberations partly to develop shared evaluation guidelines, which not only contribute to building trust but also leave room for uncertainty and discretion to address it and more complex cases. Music competitions with deliberative evaluation tend to work this way too,

with rules and criteria developing depending on the judges and along with their deliberations.

I observed judges' discussions at several competitions like this, including an annual trumpet competition over two years. During the semi-final rounds of this competition, during breaks between each group of six or seven musicians, the judges typically talked about which performances stood out to them from that group. Then, at the end of the round, they listed the competitors that they thought should advance to the finals – individually – and the competition's judging coordinator and the head judge for that division tallied their lists. This silent ballot process set the starting point for their deliberations by giving the judges a sense of where they agreed and disagreed, with the ensuing discussion focused on performances they disagreed on. At the end of their deliberations, they nominated between three and six musicians for the final round in each division based on where they decided that there was a "natural break" between competitors, but not in any particular order.

Discussions at some competitions relying on deliberation as part of their judging processes work somewhat differently. Judges might start by talking about each performance in detail, noting positive and negative aspects of each performance, and then use that discussion to figure out the results or who should move on to the next round. However, judges often have limited time for their deliberations, so this likely tends to happen more at competitions or after rounds with fewer performances (I observed two discussions where this happened, and both were after rounds with only three or four performances and took 25–30 minutes). More broadly, there is wide variation in how long judges take to discuss competition performances. I observed deliberations ranging from about two minutes to about 30 minutes, but some of the musicians I interviewed reported that they had been involved in longer discussions. Not surprisingly, deliberation length tends to correspond to how much consensus there is when the discussion starts, with longer discussions typically happening when there is a lot of disagreement.

If competitions use numeric or yes/no judging systems, judges typically write a score or "yes" or "no" next to each musician on a list of names (or a list of numbers if judges are not supposed to see performers' names until after they finalize the results), and then they hand in their lists to whomever is responsible for tallying the results. Sometimes judges deliberate after results are tallied, in which case those results might change, but this often does not happen.

Music competition controversies

Classical music competitions have also had to deal with judging-related scandals, perhaps most notably at the International

Tchaikovsky Competition. In the violin portion of the 1974 competition, for example, judges did not award a gold medal because they "decided no one had performed consistently well enough." This result was very controversial, particularly because a North Vietnamese judge reportedly gave American Eugene Fodor, who tied for the silver medal, only five points out of the maximum 25. After this became public, Fodor was described as "the victim of political bias" (Osnos 1974). 20 years later, in 1994, judges failed to award a first prize in all three of the competition's "main categories" (piano, violin, and cello). Observers described this as "the most stunning outcome in the competition's tumultuous history since 1966, when audience members shouted 'Shame!' at Soviet jurors for unjustly favoring the Soviet contestant." Judges tried to defend these outcomes by claiming that none of the competitors had lived up to the standards of past winners, but several disgruntled violin judges boycotted the awards ceremony, and "people gasped and some members of the audience booed" when the results were announced (Stanley 1994).

For decades, the Tchaikovsky Competition, held every four years, was one of the world's most prestigious piano competitions. Some people still think of it as "one of the major events in the international music community" (Peterson 2011), but several issues, including a series of scandals like those described above, have tarnished its reputation (Stults 2010). In an attempt to deal with these issues and restore its reputation, competition organizers made major changes leading up to the 2011 contest. One of the most significant changes surrounded selecting judges. In the past, this competition's judging panels had included "a fair number of teachers, including Moscow Conservatory faculty members notorious for lobbying and voting in favor of their own pupils," which had led to public scrutiny (Stults 2010). For the 2011 competition, organizing committee chair Valery Gergiev (a conductor) tried to prevent this criticism by selecting mostly performing musicians for the judging panels. Gergiev also asked Richard Rodzinski to serve as senior advisor to the 2011 competition. Rodzinski was the Van Cliburn International Piano Competition's executive director for 23 years, and he introduced new rules for the Tchaikovsky Competition that were very similar to the Cliburn Competition's rules (Service 2011).

One of the cornerstones of these new rules was a new evaluation system that was originally developed in 1990 for the Cliburn Competition and other competitions in the United States to address problems in how many music competitions had been judged. Described by Rodzinski as "transparent" and "very sophisticated in its construction" (Stults 2010), rather than relying on judges' raw scores, this new system emphasizes the spacing between judges' scores. As a result of these changes, this system "forces each juror to have an equal impact on the final result … minimizes the impact of 'gaming' by a

juror … permits abstentions while minimizing any resulting bias … [and] removes the effect of strong juror personalities which could impact the results from traditional consensus systems" (International Tchaikovsky Competition 2010). Even with these changes, which were designed to prevent the typical scandals surrounding this competition, controversy tainted the 2011 contest once again, partly because of some widely debated judges' decisions (Morrison 2011).

The Tchaikovsky Competition is an example of a music competition that has at least tried to increase fairness and transparency by making its evaluation practices more formal and mechanical. Some competitions, however, have taken very different approaches, including even changing evaluation procedures in the opposite direction. The Sydney International Piano Competition, for instance, used a numeric evaluation system for its inaugural contest in 1977 but implemented a yes/no system leading up to its 1981 competition (Email correspondence with the Sydney Competition, September 14, 2010; information provided by Warren Thomson, Artistic Director). Despite how this competition has also been surrounded by judging controversies that have led to widespread criticism, its judging process has not changed much in recent years. Porter (1995) would describe this as anomalous based on his discussion of quantification and objectivity, according to which publicity and outside pressure tend to contribute to more reliance on numeric systems.

Many prominent classical music competitions have dealt with judging controversies, but to my knowledge, no one involved has ever admitted to dishonest judging or anything else that might be considered unethical or contribute to legitimacy concerns. The people involved in evaluating music competitions also typically have a great deal of power, especially compared to judging officials in figure skating. One woodwind player and teacher explained that "the people who are usually doing the evaluations are … the ones that are revered" (Woodwind player and music conservatory faculty member). Combined with how music competitions do not need to appease an overarching governing organization that could force them to change their rules, based on my data, this power has played a key role in minimizing pressures on music competitions to mechanize their evaluation procedures. More broadly, the different organizational structures surrounding music and figure skating, with its centralization and international governance, promote different types of evaluation practices.

Judging the judges

The lack of centralization and international governance in music has also had a major impact on how musicians become adjudicators. While figure skating insiders need to go through extensive training and

testing under the IJS, based on my data, musicians simply rely on their experience – their own teaching and evaluating their students, observing other people like their teachers or colleagues evaluating, and listening to performances. One musician explained, "Some of it is very intuitive. You just get an intuitive feeling about a player when you hear them. When you've lived long enough and you've listened long enough, I think you develop that sense of 'this is good'" (Woodwind player and music conservatory faculty member). Another music insider emphasized that musicians don't usually receive any formal training to help them become evaluators: "I never took a course in that … I'm not sure anybody gives one … It's just applying what you know about music" (Vocalist and music conservatory faculty member). This informality and lack of standardization makes sense given the competition structure in music, where competitions usually do not build on each other like they often do in figure skating and where it is not considered important for competitions to use the same rules.

Also, unlike figure skating, the classical music world does not have a formal mechanism for evaluating judges, which makes sense given its lack of centralization and international governance and music evaluators' resulting high degree of power. However, music judges are evaluated more informally all the time. This happens mostly when competition organizers are assembling judging panels before competitions, where musicians' informal peer evaluations tend to play a major role. Competition organizers often call musician friends and ask for judge recommendations, or if organizers are considering inviting specific musicians to serve as judges, they might ask their friends about whether those people would make good judges. Rather than being asked to take a "judging test" before a competition, musicians who are invited to judge competitions are assumed to already have whatever knowledge they need, as one music student described: "It's not like there is a training session beforehand: 'this is what it's supposed to sound like.' You're expected to know what sounds good and what doesn't and all those in-between areas" (Tuba student and adjudicator).

Despite how music competitions do not generally rely on formal procedures for assembling their judging panels, many competitions use strategies to try to ensure as much fairness as possible. The most common examples that came up in my interview data were trying to prevent teachers from evaluating their own students and throwing out the high and low scores in cases involving numeric evaluation systems, which tends to be helpful because judges sometimes give outrageously high or low scores if they want a certain player to win. As most of the musicians I interviewed admitted, however, it would be virtually impossible to "police" the judging at music competitions without interfering with the judging process itself. Regardless of the results,

judges would typically be able to justify their evaluations somehow, as a guitar player emphasized:

> I'm sure there's lots of judging competitions where people really want their students or their style or their clique to win. I am sure that happens all the time. I don't see what you're supposed to do about that. I mean once you hire them, they're gonna say what they like or don't like. I mean, you could say, "Oh, well that was clearly driven by this," but they can always say, "No, I just liked the way they sung." So, that's a really hard one. I don't see how really you could do that.
> (Guitar player and music conservatory faculty member)

Without any formal policing mechanisms, musicians stressed how important it is to find ethical judges with integrity and "good faith" but acknowledged that they cannot always predict whether or not someone will be trustworthy.

These trends highlight the importance of trust within these competition settings. The four music competitions that allowed me to observe judges' deliberations are very clear examples of this trust. For instance, based on the data I collected at the trumpet competition I attended over two years, including informal conversations with board members and judges, this competition revolves around friendship, trust, and camaraderie. One especially illustrative example of this trust, especially among the people involved in the judging process, happened before a semi-final round in 2010. A judge whose student was about to play brought up the issue of whether he should judge his own student. He asked, "Does anyone else have a student playing?" None of the other judges responded, so he continued, "I do, so I'll just excuse myself." Another judge acknowledged that there were three students from his school in this division but that they were not his students. This conversation led to some confusion about the competition's rules surrounding this issue. After the judge whose student was going to play suggested, "I'll take myself out," and joked that he would not be able to be objective, the head judge looked for a board member who would be able to make a recommendation based on the competition's policies. By the time the judging coordinator entered the room a few minutes later, several non-judges had come in, but he stood behind the judges' table and explained quietly, "If your own students are playing, you don't need to take yourself out. We trust you." Moments later, he added, "I've judged my own students before … If you want to take yourself out of the discussion about that student, that's fine … but you don't need to say that it's your student, or that it's not your student" (Field notes, March 12, 2010).

This approach is dramatically different from the Tchaikovsky Competition, where the 2011 organizing committee chair specifically tried to assemble a panel of judges without any students in the

competition. Other high-level international competitions have implemented similar strategies. The Paganini Competition, for instance, has established a formal rule that prohibits judges from evaluating their own students (Paganini International Violin Competition "Criteria for Ranking by the Jury," personal communication, July 7, 2010). Even with this rule, musicians are still allowed to judge competitions that their students are competing in; they just cannot participate in evaluating their own students. More generally, based on my data, competitions with these kinds of rules do not have any way to punish evaluators who break the rules, other than not inviting them back to judge in the future. Without any official policing mechanisms with serious sanctions, music competitions simply need to rely on the trustworthiness of their evaluators. This is acceptable partly because of the organizational and power dynamics surrounding the music world, where competitions do not need to defend their rules or practices to any more powerful organizations.

In general, these trends where music insiders are typically allowed to judge competitions that their students have entered are in stark contrast to figure skating, where coaches cannot judge events where skaters they have recently worked with are competing, even if they have only worked with a competing skater for a few hours. All of these findings illustrate how evaluation practices in classical music tend to be relatively informal, especially compared to figure skating.

Discussion

Based on this discussion of evaluation in figure skating and classical music, organizational structure has a tremendous impact on evaluation practices and their degrees of formality and standardization. In particular, my findings suggest that highly centralized settings governed by more powerful organizations and where competitions build on each other tend to use more formal and standardized evaluation practices compared to other settings with fewer restrictions. In Table 5, I highlight these factors in terms of how they differ in relation to skating and music competitions.

	Figure Skating	Classical Music
Organizational Context	Sport with artistic components	Art
Centralization	Highly centralized	Not centralized
Governance	Internationally governed	Not internationally governed
	International Skating Union (ISU): • Central governing body with power to impose internationally standardized rules • 101 member federations representing different countries	No governing bodies
	International Olympic Committee (IOC): • Governing body for all Olympic sports	Judging officials and other elites: • Not accountable to more powerful organizations
	ISU officials and other elites • Accountable to IOC • Low levels of power	• High levels of power
Competition Structure*	Competitions tend to build on each other.	Competitions usually do not build on each other.
	Need for internationally standardized rules	Less need for standardized rules

*These distinctions reflect my findings on figure skating and classical music competitions and do not necessarily apply to other activities in these arenas.

Table 5: Factors Affecting Evaluation Practices in Figure Skating vs. Classical Music
Source: Author's work

My findings indicate that organizational structure is pivotal to why these settings have developed formal and standardized versus informal and varied evaluation practices. For example, whether or not everyone is working toward the same pinnacle event, such as the Olympics, makes a major difference in whether an evaluative setting is expected to adopt formal, standardized evaluation practices. In contexts where competitions do not build on each other, with international competitions that are not advancing participants toward a pinnacle and are not run by the same organization, there is much less pressure for formal, standardized evaluation. These divergences in competition

structure interact with variations in levels of centralization and governance structure to influence approaches to evaluation practices.

Also based on these findings, when an organization experiences a legitimacy crisis like a judging scandal, that crisis interacts with these other factors to influence the trajectory of the evaluation practices involved. When rules are perceived as fair, it is usually less likely that anyone would want to change them, which tends to contribute to maintaining the status quo. Even if there is a legitimacy crisis, such as a judging scandal where someone does not follow the rules, if the organizations involved in that crisis do not have to worry about a more powerful entity like an international governing body, those organizations would often be able to come up with a solution that contributes to maintaining the status quo as well. This has happened with many music competitions that have dealt with judging scandals. However, when there is a legitimacy crisis in a setting that needs to worry about international governance, especially by a more powerful organization, outsiders would be more likely to demand change. This helps explain why there has been so much more pressure, and ultimately change, in response to judging controversies in figure skating compared to classical music.

This analysis is limited to just two cases, so researchers should conduct further studies in other areas before definitive conclusions can be made about variables that affect evaluation practices and how rules develop, including the factors I have emphasised here as well as others. Other types of power relationships, publicity, and trust dynamics, for example, also come up a lot throughout my data as influencing evaluation practices in skating and music and would be worth including in further analyses. Adding more cases to this discussion would increase our understanding beyond what an in-depth qualitative study like this allows for.

In this article, I have emphasized the significance of organizational structure in shaping evaluation practices, including degrees of reliance on formal, standardized, quantitative rules. Porter (1995) suggests that quantification is such a powerful legitimation device because it appears to separate knowledge from the people who created that knowledge, which many people believe decreases the likelihood of corruption, dishonesty, and otherwise "hazy thinking" (85). While the kind of debate and persuasion that characterizes judges' deliberations at many music competitions is often seen as more legitimate in a context where people in power positions can make decisions about evaluation practices without pressure from external governing bodies and without worrying about affecting other competitions, in a highly centralized, internationally governed context surrounded by a relative lack of power, this type of evaluation would be much more difficult to defend. A common perception is that verbal reasoning allows for too much potential bias for these kinds of settings without providing any "clear

checks against errors of reasoning" (Porter 1995: 52). In many contexts with similarities to figure skating, decision-making based on this kind of deliberation would be considered illegitimate. In this sense, highly formal rule systems act as legitimation devices, especially when they revolve around numbers.

Conclusion

The findings presented here have implications for how we make decisions and distinctions and how we construct quality – based on adding up a series of individual choices about specific parts of some product or performance (as in figure skating) or through reaching consensus based on overall impressions through group discussion (as in some settings in classical music). Variations in valuing parts or wholes can be especially significant in performance settings because in many instances, the sum of a performance's various components might not reflect its overall impact, so this can affect the types of performances that are rewarded and ultimately valued. The contrast between making decisions individually or collectively can also affect outcomes, partly because this affects whether people can influence each other's evaluations. Some people in some settings might think that letting evaluators influence each other is problematic, while others believe that this can lead to a more balanced decision-making process and ultimately fairer outcomes.

More broadly, our evaluation practices often have profound effects on many important distinctions, such as how we think about "good" versus "bad" and "worthy" versus "unworthy," as well as the boundaries between these categories (Lamont 2012). Lamont (2012) emphasizes that one aspect of fighting inequality is broadening how we define social worth, and she suggests that before we can do that, we need a better understanding of evaluation, which she calls a "complex, slippery, and often elusive sociological object" (203). My research contributes to this goal by outlining how evaluation works in two settings and identifying several factors – organizational context, centralization, governance, and competition structure – that other scholars should incorporate into research on different evaluative settings. Looking at how and why evaluation practices change and the effects of those changes would be a useful starting point, one that would also deepen our understanding of organizational responses to rule changes.

Evaluation, especially deciding what kinds of practices to use in various contexts, has been hotly contested in many areas. Debates surrounding how to make decisions about loans, hiring, evaluating teachers, and college admissions, as well as many other issues, illustrate this point. There is so much controversy around these issues and others linked to evaluation because these decisions are so

important for so many people, and many of them contribute to perpetuating existing inequality. When banks decide whether to give someone a loan, organizations decide whether to offer someone a job, or colleges decide whether to admit a student, the outcomes tend to have serious consequences for the people on the other side of those decisions, and those outcomes are often influenced by how formal these organizations' evaluation systems are. Many such organizations use very formal rules to make those decisions because they think this enhances their legitimacy, but formal guidelines usually favor people who look good on paper – people with high credit scores, degrees from prestigious universities, or high SAT scores, for example – which in turn tend to privilege people who have money and other resources. In contrast, more informal criteria typically give organizations more flexibility to reward more subjective indicators (such as work ethic, integrity, creativity, or interpersonal skills), take unique accomplishments into account, and make exceptions.

When organizations adopt more formal rules, in addition to their link to legitimacy, they often do so to limit corruption and encourage fairness, objectivity, and equality. Most people would think of these goals as very honorable. However, although following formal rules makes sense in certain contexts, especially where breaking rules could be very dangerous (take driving, for instance), formalization often leads to unintended consequences, many of which are negative in some way. A number of researchers have emphasized the potential drawbacks of trying to contain discretion (Perrow 1999; Lom 2016) and how many types of rules and measurement devalue qualities that are important but difficult to measure or account for with rules and tend to change what is being measured (for example, see Carruthers and Espeland 1991; Espeland and Sauder 2007; Sauder 2008; Sauder and Espeland 2009; Colyvas 2012; Lom 2016). Among the effects of the judging changes in figure skating has been a devaluation of the artistic side of the sport, in contrast to how many music evaluators continue to stress artistry. This personifies the potential drawbacks of highly formal rule systems and the potential advantages of more informal systems and indicates that in many cases involving evaluation, sometimes less is more.

Acknowledgments

This research was supported by a Research Grant from the University Research Grants Committee at Northwestern University. Earlier versions of parts of this manuscript were presented at the American Sociological Association Annual Meetings and the Arkansas Sociological and Anthropological Association Annual Conference. The

author is grateful for helpful comments and suggestions from Bruce Carruthers, Wendy Espeland, Gary Fine, Carol Heimer, Ed Powers, John Toth, and Claes-Fredrik Helgesson and the anonymous *Valuation Studies* reviewers.

References

Adler, Paul S. 2006. "Beyond Hacker Idiocy: The Changing Nature of Software Community and Identity." In *The Firm as a Collaborative Community: Reconstructing Trust in the Knowledge Economy*, edited by Charles Heckscher and Paul S. Adler, 198–258. New York: Oxford University Press.

Adler, Paul S. 2012. "The Sociological Ambivalence of Bureaucracy: From Weber via Gouldner to Marx." *Organization Science* 23(1): 244–266.

Adler, Paul S., and Bryan Borys. 1996. "Two Types of Bureaucracy: Enabling and Coercive." *Administrative Science Quarterly* 41(1): 61–89.

AAF (Alink-Argerich Foundation). 2023. "Foundation." https://www.alink-argerich.org/foundation, accessed 29 April 2024.

Becker, Howard S. 1982. *Art Worlds*. Berkeley, CA: University of California Press.

Brown, Richard H., and Elizabeth L. Malone. 2004. "Reason, Polities, and the Politics of Truth: How Science is Both Autonomous and Dependent." *Sociological Theory* 22(1): 106–122.

Carruthers, Bruce G., and Wendy N. Espeland. 1991. "Accounting for Rationality: Double-Entry Bookkeeping and the Rhetoric of Economic Rationality." *American Journal of Sociology* 97(1): 31–69.

Chambliss, Daniel F. 1989. "The Mundanity of Excellence: An Ethnographic Report on Stratification and Olympic Swimmers." *Sociological Theory* 7(1): 70–86.

Charmaz, Kathy. 2001. "Grounded Theory." In *Contemporary Field Research: Perspectives and Formulations*, 2nd edn, edited by Robert M. Emerson, 335–352. Long Grove, IL: Waveland Press.

Coakley, Jay J. 2001. "Sport in Society: An Inspiration or an Opiate?" In *Sport in Contemporary Society*, 6th edn, edited by D. Stanley Eitzen, 20–36. New York: Worth.

Coakley, Jay. 2015. *Sports in Society: Issues and Controversies*, 11th edn. New York: McGraw-Hill.

Colyvas, Jeannette A. 2012. "Performance Metrics as Formal Structures and through the Lens of Social Mechanisms: When Do They Work and How Do They Influence?" *American Journal of Education* 118(2): 167–197.

Daston, Lorraine, and Peter Galison. 2007. *Objectivity*. New York: Zone Books.

Deephouse, David L. 1996. "Does Isomorphism Legitimate?" *Academy of Management Journal* 39(4): 1024–1039.

DiMaggio, Paul J., and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48(2): 147–160.

Dobbin, Frank, and Erin L. Kelly. 2007. "How to Stop Harassment: Professional Construction of Legal Compliance in Organizations." *American Journal of Sociology* 112(4): 1203–1243.

Edelman, Lauren B. 2016. *Working Law: Courts, Corporations, and Symbolic Civil Rights*. Chicago, IL: University of Chicago Press.

Edelman, Lauren B., Linda H. Krieger, Scott R. Eliason, Catherine R. Albiston, and Virginia Mellema. 2011. "When Organizations Rule: Judicial Deference to Institutionalized Employment Structures." *American Journal of Sociology* 117(3): 888–954.

Eitzen, D. Stanley. 2012. *Fair and Foul: Beyond the Myths and Paradoxes of Sport,* 5th edn. Lanham, MD: Rowman and Littlefield.

Elsbach, Kimberly D., and Greg Elofson. 2000. "How the Packaging of Decision Explanations Affects Perceptions of Trustworthiness." *Academy of Management Journal* 43(1): 80–89.

Emerson, Robert M., Rachel I. Fretz, and Linda L. Shaw. 1995. *Writing Ethnographic Fieldnotes*. Chicago, IL: University of Chicago Press.

Erwin, Patrick M. 2011. "Corporate Codes of Conduct: The Effects of Code Content and Quality on Ethical Performance." *Journal of Business Ethics* 99: 535–548.

Espeland, Wendy N., and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113(1): 1–40.

Feldman, Martha S. 1992. "Social Limits to Discretion: An Organizational Perspective." In *The Uses of Discretion*, edited by Keith Hawkins, 163–184. Oxford: Clarendon Press.

Fourcade, Marion. 2011. "Cents and Sensibility: Economic Valuation and the Nature of 'Nature.'" *American Journal of Sociology* 116(6): 1721–1777.

Frey, James H., and D. Stanley Eitzen. 1991. "Sport and Society." *Annual Review of Sociology* 17: 503–522.

Gunts, Bucky, and David Michaels. 2002. "Figure Skating." XIX Olympic Winter Games NBC. February 11, 2002.

Guttmann, Allen. 2003. "Sport, Politics and the Engaged Historian." *Journal of Contemporary History* 38(3): 363–375.

Hafner-Burton, Emilie M., and Kiyoteru Tsutsui. 2005. "Human Rights in a Globalizing World: The Paradoxes of Empty Promises." *American Journal of Sociology* 110(5): 1373–1411.

Helgesson, Claes-Fredrik, and Fabian Muniesa. 2013. "For What It's Worth: An Introduction to Valuation Studies." *Valuation Studies* 1(1): 1–10.

Hines, James R. 2006. *Figure Skating: A History.* Urbana, IL: University of Illinois Press; Colorado Springs, CO: World Figure Skating Museum and Hall of Fame.

Hines, James R. 2011. *Historical Dictionary of Figure Skating*. Lanham, MD: Scarecrow Press.

IOC (International Olympic Committee). 2021. "International Olympic Committee – History, Principles & Financing." https://olympics.com/ioc/overview.
ISU (International Skating Union). 2004. "Program Components Overview."
ISU (International Skating Union). 2010. "Communication No. 1611: Scale of Values, Levels of Difficulty and Guidelines for Marking Grade of Execution for Single and Pair Skating."
ISU (International Skating Union). 2012. "Some Key Dates in ISU History."
ISU (International Skating Union). 2022a. "ISU Results: XXIV Olympic Winter Games 2022: Men." http://results.isu.org/events/owg2022_Men.htm.
ISU (International Skating Union). 2022b. "ISU FAQ – Program Components." https://isu.org/media-centre/press-releases/2022-4/29705-isu-faq-components/file.
ISU (International Skating Union). 2022c. "Communication No. 2475: Single & Pair Skating Scale of Values season 2022/23."
ISU (International Skating Union). 2023a. "ISU History." https://isu.org/inside-isu/about/history.
ISU (International Skating Union). 2023b. "Member Federations." https://isu.org/inside-isu/about/member-federations.
International Tchaikovsky Competition. 2010. "The Scores of Harmony Processing System."
Jackson, Jon. 2005. *On Edge: Backroom Dealing, Cocktail Scheming, Triple Axels, and How Top Skaters Get Screwed*. New York: Thunder's Mouth Press.
Jacobsson, Bengt. 2000. "Standardization and Expert Knowledge." In *A World of Standards*, edited by Nils Brunsson and Bengt Jacobsson, 40–49. New York: Oxford University Press.
Jick, Todd D. 1979. "Mixing Qualitative and Quantitative Methods: Triangulation in Action." *Administrative Science Quarterly* 24(4): 602–611.
Kornberger, Martin, Lise Justesen, Jan Mouritsen, and Anders K. Madsen (eds.) 2015. *Making Things Valuable*. Oxford: Oxford University Press.
Lamont, Michèle. 2009. *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.
Lamont, Michèle. 2012. "Toward a Comparative Sociology of Valuation and Evaluation." *Annual Review of Sociology* 38: 201–221.
Latham, Alison, and Piers Spencer. 2002. "Competitions in Music." In *The Oxford Companion to Music*, edited by Alison Latham. Oxford Music Online. http://www.oxfordmusiconline.com/subscriber/article/opr/t114/e1527.
Li, Jiatao, and Yi Tang. 2010. "CEO Hubris and Firm Risk Taking in China: The Moderating Role of Managerial Discretion." *Academy of Management Journal* 53(1): 45–68.

Lofland, John, and Lyn H. Lofland. 1994. *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*, 3rd edn. Belmont, CA: Wadsworth.

Lom, Stacy E. 2016. "Changing Rules, Changing Practices: The Direct and Indirect Effects of Tight Coupling in Figure Skating." *Organization Science* 27(1): 36–52.

Maccoby, Michael. 2006. "Health Care Organizations as Collaborative Learning Communities." In *The Firm as a Collaborative Community*, edited by Charles Heckscher and Paul S. Adler, 259–280. New York: Oxford University Press.

McCormick, Lisa. 2009. "Higher, Faster, Louder: Representations of the International Music Competition." *Cultural Sociology* 3(1): 5–30.

Mastrofski, Stephen D. 2004. "Controlling Street-Level Police Discretion." *Annals of the American Academy of Political and Social Science* 593: 100–118.

Mau, Steffen. 2019. *The Metric Society: On the Quantification of the Social*. Medford, MA: Polity Press.

Meyer, John W., and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83(2): 340–363.

Morrison, Richard. 2011. "Off-Key Antics Failed to Derail the Tchaikovsky Competition." *The London Times* July 6: Section T2, p. 11.

Osnos, Peter. 1974. "Fodor: Tied for Silver, Not Gold." *The Washington Post* July 5.

Perrow, Charles. 1999. *Normal Accidents: Living with High-Risk Technologies*, with a New Afterword and a Postscript on the Y2K Problem. Princeton, NJ: Princeton University Press.

Peterson, Diane. 2011. "Sonoma Violinist Advances to Tchaikovsky Finals." *The Press Democrat* June 25.

Porter, Theodore M. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.

Power, Michael. 2003. "Auditing and the Production of Legitimacy." *Accounting, Organizations and Society* 28(4): 379–394.

Roberts, Selena. 2002a. "Canada to Appeal on Skating; Judge Pressured, French Say." *The New York Times* February 14: Section A, Column 1: 1.

Roberts, Selena. 2002b. "Skating Group Proposes a New System of Judging." *The New York Times* February 19: Section A, Column 2: 1.

Sauder, Michael. 2008. "Interlopers and Field Change: The Entry of U.S. News into the Field of Legal Education." *Administrative Science Quarterly* 53(2): 209–234.

Sauder, Michael, and Wendy N. Espeland. 2009. "The Discipline of Rankings: Tight Coupling and Organizational Change." *American Sociological Review* 74(1): 63–82.

Service, Tom. 2011. "Everything to Play for at the Tchaikovsky Competition." *The Guardian* September 20.

Sheehan, Mike, and David Michaels. 2018. "Figure Skating." XXIII Olympic Winter Games NBC. February 8, 2018.
Shen, Wei, and Theresa S. Cho. 2005. "Exploring Involuntary Executive Turnover through a Managerial Discretion Framework." *Academy of Management Review* 30(4): 843–854.
Stanley, Alessandra. 1994. "Musical Tradition of Acrimony." *The New York Times* July 2.
Star, Susan L., and Martha Lampland. 2009. "Reckoning with Standards." In *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, edited by Martha Lampland and Susan L. Star, 3–24. Ithaca, NY: Cornell University Press.
Stevens, Mitchell L. 2007. *Creating a Class: College Admissions and the Education of Elites*. Cambridge, MA: Harvard University Press.
Stinchcombe, Arthur L. 2001. *When Formality Works: Authority and Abstraction in Law and Organizations*. Chicago, IL: University of Chicago Press.
Stults, Raymond. 2010. "Rodzinski to Revive Tchaikovsky Competition." *The Moscow Times* January 29.
Sunshine, Jason, and Tom R. Tyler. 2003. "The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing." *Law and Society Review* 37(3): 513–548.
Timmermans, Stefan, and Marc Berg. 2003. *The Gold Standard: The Challenge of Evidence-Based Medicine and Standardization in Health Care*. Philadelphia, PA: Temple University Press.
Tyler, Tom R. 1988. "What is Procedural Justice? Criteria Used by Citizens to Assess the Fairness of Legal Procedures." *Law and Society Review* 22(1): 103–136.
Tyler, Tom R. 2000. "Multiculturalism and the Willingness of Citizens to Defer to Law and to Legal Authorities." *Law and Social Inquiry* 25(4): 983–1019.
Tyler, Tom R. 2003. "Procedural Justice, Legitimacy, and the Effective Rule of Law." *Crime and Justice* 30: 283–357.
USFSA (United States Figure Skating Association). 2009a. "Agreement to Comply with Code of Ethics."
USFSA (United States Figure Skating Association). 2009b. "Technical Panel Code of Ethics."
Washington, Robert E., and David Karen. 2001. "Sport and Society." *Annual Review of Sociology* 27: 187–212.
Weaver, Gary R., Linda K. Trevino, and Philip L. Cochran. 1999. "Corporate Ethics Programs as Control Systems: Influences of Executive Commitment and Environmental Factors." *Academy of Management Journal* 42(1): 41–57.
Weber, Max. 1968. *Economy and Society: An Outline of Interpretive Sociology*. Edited by Guenther Roth and Claus Wittich. New York: Bedminster.

WFIMC (World Federation of International Music Competitions). 2021. "The Federation." https://www.wfimc.org/federation.
WFIMC (World Federation of International Music Competitions). 2022. "WFIMC Statutes 2022." https://www.wfimc.org/federation.

Stacy E. Lom is Assistant Professor of Sociology at the University of Central Arkansas. Her research interests include evaluative cultures, particularly how they develop and their effects in different contexts; rules and formalization; quantification; organizational decision making; and the politics of knowledge production and consumption. Her research focuses on rule systems and evaluation in sport, the arts, and education.