Organization Science
Vol. 21, No. 2, March–April 2010, pp. 331–346
ISSN 1047-7039 | EISSN 1526-5455 | 10 | 2102 | 0331
informs®
DOI 10.1287/orsc.1080.0400

Certifications and Reputation: Determining the Standard of Desirability Amidst Uncertainty

Scott D. Graffin
Terry College of Business, University of Georgia, Athens, Georgia 30602, sgraffin@terry.uga.edu

Andrew J. Ward
Department of Management, College of Business and Economics, Lehigh University,
Bethlehem, Pennsylvania 18015, award@lehigh.edu

We develop a theory that examines how a particular type of third-party quality signal, certifications, influences an actor's long-term reputation by addressing two different types of uncertainty. The first type deals with the degree to which the capabilities of an actor can be inferred over time based on known performance dimensions. We label this technical uncertainty. The second type deals with assessing whether the perceived capabilities of the actor meet the standard of desirability, which we call performance standard uncertainty. We propose and test that certifications will positively influence the long-term reputation of actors in situations that involve minimal technical uncertainty, and that, across levels of technical performance, certifications will have an inverted U-shaped relationship with assessments of actors such that certifications will have the greatest impact on assessments of actors who are close to the uncertain standard of desirability. We test our hypotheses in the context of the voting for Major League Baseball's Hall of Fame, an environment where comprehensive technical performance measures leave little technical uncertainty. Our results support our hypotheses and suggest that certifications can influence an actor's reputation by reducing performance standard uncertainty rather than just technical uncertainty, as previously presumed.

Key words: reputation; certification; uncertainty; standard of desirability; assessment
History: Published online in *Articles in Advance* March 6, 2009; revised July 2, 2009.

Introduction

The reputation of actors has been shown to influence outcomes such as performance at the organization level (e.g., Deephouse 2000, Fombrun and Shanley 1990) and executive compensation at the individual level (Malmendier and Tate 2005, Graffin et al. 2008, Wade et al. 2006). We define reputation as the collective judgment of observers regarding the quality or capabilities of a focal actor within a specific domain that is earned over time (Fombrun 1996, Podolny 2005, Rindova et al. 2005, Washington and Zajac 2005). As such, reputation is the assessed quality or capabilities of an actor, which is informed over time by the continued appraisal of the actor's performance (Carter and Ruefli 2006, Fombrun 1996). The general consensus in the literature is that the more uncertain the assessment of the actor's capabilities becomes, the more likely it is that the actor's reputation will be influenced by third-party quality signals such as certifications, accreditations, coverage by the media, or the opinions of other informed parties such as analysts (Festinger 1954, Podolny 2005, Rao 1994, Rindova et al. 2005, Zuckerman 1999). By third-party quality signals, we mean information generated by reputable observers that provides evaluations and endorsements or repudiations of actors within a given domain (Rao 1994, Scott 1994, Wade et al. 2006).

In this paper, we seek to demonstrate that third-party quality signals help build an actor's reputation even in a context where the performance of an actor is easily observable on a comprehensive range of factors, as we conceive of reputation as being influenced by two separate types of assessments. The first type of assessment deals with the degree to which the quality or capabilities of an actor can be inferred over time based on known performance dimensions. To the extent that the direct observation of performance is not possible, or if there is a perceived loose coupling between an actor's performance and his or her underlying capabilities, this assessment is uncertain. We refer to this type of uncertainty as *technical uncertainty*. The second type of assessment deals with the uncertainty surrounding what Thompson (1967, p. 84) termed "the standards of desirability." That is, even if an actor's capabilities are perceived to be tightly coupled with his or her observable performance, there may remain uncertainty surrounding the standards or yardstick against which the actor's capabilities are to be judged in order for them to be considered acceptable. We refer to this type of uncertainty as *performance standard uncertainty*.

These separate assessments reflect the idea that uncertainty is a multifaceted construct. However, most prior reputation studies have treated uncertainty as a unidimensional construct that only exists when the capabilities of actors are difficult to assess, i.e., when technical uncertainty is present. This unidimensional conceptualization mandates the presence of technical uncertainty as a necessary condition for third-party quality signals to be influential, and unnecessarily restricts the role such signals are thought to play in building an actor's reputation. We recognize an additional dimension of uncertainty, performance standard uncertainty, which is critical to the overall evaluation of an actor. In explicitly separating these two dimensions of uncertainty, we aim to demonstrate that certifications are not only valuable in building reputation by reducing the technical uncertainty observers face in assessing the quality or capabilities of an actor, but also that certifications can influence an actor's reputation by reducing the performance standard uncertainty as to whether the actor's perceived capabilities meet or exceed the standard of desirability. The recognition of this second dimension of uncertainty unbounds the role that third-party quality signals play in influencing the overall evaluation of actors, as technical uncertainty is no longer a necessary condition for such quality signals to be influential.

Thus, the central tenet of this paper is that the study of certifications and reputations may each be advanced through a theoretical clarification regarding the different types of uncertainty that may be addressed by certifications, as well as a clearer understanding of the process by which these third-party quality signals influence the overall assessments of actors. By doing so, we make two related contributions to the reputation literature. First, we theorize and test the idea that, even in the absence of technical uncertainty, the existence of performance standard uncertainty will continue to allow third-party quality signals, in this case certifications, to influence an actor's reputation. Second, we suggest that the role certifications play in influencing the evaluation of an actor may be more or less influential, depending on the level of the actor's technical performance. Previous studies have implicitly assumed that certifications have a homogeneous impact on actors' reputations within a given domain (e.g., Rao 1994, Wade et al. 2006) while we assert that this influence may vary across actors.

We test our hypotheses in a unique setting that allows us to examine the impact of one form of a third-party quality signal, certifications, on the reputation of actors—voting for Major League Baseball's (MLB) Hall of Fame (the Hall)—which is an environment where comprehensive measures of technical performance leave little uncertainty regarding the capabilities of an actor. Specifically, we use a data set of MLB players to test our theory by examining the impact of certifications on voting for the Hall. In this setting there is a perceived tight coupling between measured performance and capabilities, minimizing the uncertainty regarding the level of an actor's capabilities and thus isolating the impact of certifications on the assessment of whether the perceived capabilities meet the standard of desirability.

Theory and Hypothesis Development

Reputation and Certifications

In contexts where assessing the overall quality or capabilities of an actor is difficult, research has suggested that one important third-party quality signal that may influence an actor's reputation is the outcome of certification contests (e.g., Rao 1994, Rindova et al. 2005, Wade et al. 2006). We define a certification contest as a "competition in which actors in a given domain are ranked based on performance criteria that are accepted by key stakeholders as being credible and legitimate" (Wade et al. 2006, p. 644). Certification contests allow observers to distill myriad data points into one ranking, and make clear and comparable attributions of an actor's relative worth or standing (Elsbach and Kramer 1996, Fombrun and Shanley 1990). Thus, the resulting certifications signal the relative quality of an actor within a given domain (e.g., industry, organizational field, or time period) and, in turn, create a hierarchical ordering of the actors who were evaluated (Elsbach and Kramer 1996, Podolny 2005, Wade et al. 2006).

In this paper, we distinguish between two different types of evaluations that often have been clumped generically under the label of certification contests, but in fact represent different types of evaluation and outcome (Table 1). These fundamental types of evaluation are distinguished by whether the evaluation of an actor is done relative to other actors (relative evaluations) or against a fixed standard (yardstick evaluations). Within these two broad categories, there are three specific types of evaluations: certification contests (a relative evaluation), accreditations (a yardstick evaluation), and ratings (a yardstick evaluation). Relative evaluations, such as our conceptualization of certification contests, compare actors to one another to determine a relative rank-ordering of actors where some actors are necessarily certified and others are not. The resulting rank-ordering may simply be between winners and nonwinners (e.g., CEO of the Year Contest; Graffin et al. 2008, Wade et al. 2006) or may provide a more comprehensive ranking with multiple levels of ranking (e.g., *Fortune* magazine's 100 Best Companies to Work For; Fulmer et al. 2003). As such, in relative evaluations what is being assessed is an actor's performance vis-à-vis other actors, not their absolute level of performance. In other words, evaluators cannot determine a priori what level of performance an actor needs to win a given certification contest. For example, in any given year, there may be many films deemed "Oscar-worthy," yet only one film can win the Academy Award for Best Motion Picture. Similarly, there may be a year when no film stands out as great, yet one film *must* be certified as winner of the Academy Award for Best Motion Picture.

In contrast, yardstick evaluations, such as accreditations or ratings, compare actors against a categorical

Table 1 Typology of Social Evaluations

Type of contest	Type of evaluation	Type of outcome	Examples	Measures (studies)
Certification	Actors measured against one another.	Relative rank ordering of actors against one another.	—*Business Week* rankings of business schools —Academy Awards —All-Star and MVP awards in sports —CEO of the year contests	—CEO of the year contest (Graffin et al. 2008, Malmendier and Tate 2005, Wade et al. 2006) —Speed and reliability contests in the early automobile industry held by magazines and newspapers (Rao 1994) —*Fortune* magazine's survey of corporate reputations (Fombrun and Shanley 1990) —*Fortune* magazine's 100 Best Companies to Work For (Fulmer et al. 2003) —*Business Week's* ranking of business schools (D'Aveni 1996, Elsbach and Kramer 1996, Martins 1998, Rindova et al. 2005, Segev et al. 1999) —*Business Week's* ratings of boards of directors of U.S. corporations. (Johnson et al. 2005)
Rating	Actors measured against categorically defined standards without regard to the performance of other actors.	A clumped rank ordering of actors that meet the standards of different categories.	—*Michelin Guide* to restaurants —*Consumer Reports* ratings of consumer products	—Ratings of wine quality (Benjamin and Podolny 1999)
Accreditation	Actors measured against an absolute standard (regardless of whether that standard is known or defined with certainty).	A dichotomous outcome such that the actor meets the standard or does not meet the standard.	—AACSB accreditation —ISO 9000 accreditation —MLB Hall of Fame	—ISO 9000 or 14001 (Beck and Walgenbach 2005, Boiral 2003, Cascio et al. 1996, Darnall 2006, Guler et al. 2002, Jiang and Bansal 2003, Terlaak and King 2006, Uzumeri 1997) —Third-party accreditation of colleges (Durand and McGuire 2005, Wiley and Zald 1968) —Financial records accredited by auditors (Wilson 1985) —Issuance of a charitable registration number by the government (Singh et al. 1986) —A commission certification indicating compliance with government regulations (Sine et al. 2007)

Note. Association to Advance Collegiate Schools of Business (AACSB); International Organization for Standardization (ISO).

standard, resulting in a clumped rank ordering of actors that meet or do not meet the standards of different categories. The difference between accreditations and ratings is that in an accreditation, the outcome is binary—the actor is accredited or not, e.g., the Association to Advance Collegiate Schools of Business (AACSB) accreditation for business schools or ISO9000 accreditation for businesses. As such, accreditations are non-rival assessments where actors are compared against a fixed standard of performance (Deephouse and Suchman 2008), whether that standard is known or defined with certainty. In accreditations, therefore, any number of actors or no actors at all could be considered to have met or exceeded the standard.[1] In rating contests, on the other hand, the outcome for an actor falls in one of any number of categories. For example, when the *Michelin Guide* rates a restaurant, the restaurant will be awarded a number of stars based on the particular standards of the guide. Any number of restaurants can warrant a rating in any of the categories, i.e., restaurants are not compared directly against one another, but are compared against the category standards established by the guide. The result, then, is a rank ordering of actors (restaurants with three stars versus two stars, etc.), but without distinction between actors within a category (e.g., between restaurants with three stars). One other distinction between types of evaluations is that the outcomes of certification contests can accumulate, whereas accreditations and ratings do not. For example, an actor may be referred to as a "three-time Academy Award winner" or a baseball player as a "seven-time All-Star," whereas a firm is or is not ISO9000-accredited or a business school is or is not

AACSB-accredited, regardless of how many times this accreditation is renewed.

While the above discussion recognizes that there are different types of third-party quality signals, the current study focuses on how one type of quality signal, certifications, influences the overall assessment of actors. Prior research suggests that certifications influence the reputation-building of both individuals and organizations by providing a clear signal regarding the relative capabilities of actors, based on expert evaluation (e.g., Rao 1994, Rindova et al. 2005, Wade et al. 2006). Given the consistent support at both levels of analysis, Podolny (2005) contends that the influence of quality signals on reputation operates similarly at both the individual and organization level. For instance, at the individual level, Wade et al. (2006) found that winners of a "CEO of the Year" certification contest were granted greater increases in compensation than their noncertified counterparts. Wade et al. (2006, p. 643) took this result to mean that: "... under conditions of evaluative uncertainty, one mechanism by which the capabilities of social actors are assessed is through certification contests." At the organization level, Rao (1994) found that certified firms were more likely to survive in the early automobile industry. Rao (1994, p. 32) also contended that "(certification) contests structure search in crowded and confused markets and circumvent the issue of measuring capabilities." In noting this prevailing view, Rindova et al. (2005, p. 1045) challenged this notion by issuing a call to examine reputation-building in a context "where rankings are common and pervasive, but product quality is not so difficult for stakeholders to evaluate." In answering this call, we contend and demonstrate that certifications still influence the assessment of actors when technical uncertainty is minimal.

Types of Uncertainty

Despite our contention that reputation is based on two types of assessment, previous studies of certifications specifically and reputations generally have asserted or implied that it is technical uncertainty that makes an actor's reputation meaningful. In this vein, Rindova et al. (2005, p. 1034) wrote, "... reputation is valuable because it reduces the uncertainty stakeholders face in evaluating firms as potential suppliers of needed products and services." Similarly, Washington and Zajac (2005, p. 284) comment that reputations are "... a summary categorization of real or perceived historical differences in product or service quality among organizations, given imperfect information." Thus, within the broader reputation literature, it is thought that reputation is valuable mainly to the extent to which it reduces the technical uncertainty surrounding the actor's capabilities or quality (Fombrun 1996, Rindova et al. 2005),[2] while the question of resolving performance standard uncertainty has received little attention.

We conceptualize performance standard uncertainty as existing when there is uncertainty as to the yardstick, or standard, against which an actor's capabilities should be evaluated. Thompson (1967, p. 84) referred to this kind of uncertainty as dealing with the "standards of desirability." Thus, even if technical performance is perfectly observable and is perceived to correspond perfectly to an actor's capabilities, performance standard uncertainty exists if it is not clear which standard is appropriate. Making an assessment in light of performance standard uncertainty thus represents an equivocal situation (Weick 1979). According to Weick (1995, p. 27), the problem in equivocal situations is not a lack of information, i.e., technical uncertainty, but rather, "the problem is that there are too many meanings, not too few ... The problem is confusion, not ignorance." In Weick's (1995) conception, this confusion may not necessarily be reduced by simply gathering more information about the quality of the actor's capabilities or more details about an actor's technical performance, as that is not where the uncertainty lies. In other words, it is the standard itself that is uncertain, not the assessed capabilities of the actor.

Performance standard uncertainty may arise from at least two different sources. The first potential source stems from the context in which the evaluation takes place. If multiple performance metrics are available, observers may have conflicting interpretations of what metric or combination of metrics constitutes the appropriate standard of technical performance that is desired. Indeed, even if all observers have the same information about the actor's performance, they may disagree over what combination of performance metrics should be considered to meet or exceed the standard. Thus, one source of performance standard uncertainty is the observers' disagreement regarding the weighting and level of different performance metrics that are considered acceptable.

A second potential source of performance standard uncertainty is the deliberate infusion of ambiguity by those who are assessing actors relative to the standard. Observers may actually prefer ambiguous standards, because this ambiguity enables them to maintain control over the process. If there was an unambiguous standard, then the assessment of an actor no longer requires expert observers, but instead is merely a simple comparison of an actor against a checklist of criteria. In other words, an explicit a priori standard shifts the power of judgment from expert observers to the standard itself. Additionally, as discussed above, it may be difficult or impossible to a priori define every possible criteria against which the actors should be assessed. However, even if this were possible, observers may still wish to be able to include or exclude actors based on criteria other than direct measures of performance, some of which may be difficult or impossible to measure (e.g., leadership of a CEO, or the social responsibility of a corporation). Finally, observers may also wish to recalibrate the acceptable standard over time. That is, even if an exhaustive list of criteria were known and measurable, what level of performance observers deem appropriate may shift, incrementally or dramatically, over time. Despite Thompson's (1967) recognition of this type of uncertainty some four decades ago, this idea has received little discussion in the reputation literature in general or the certification literature in particular. The few studies that have explicitly mentioned this sort of uncertainty have been conducted in contexts where there are high levels of both technical and performance standard uncertainty (e.g., Pfeffer et al. 1976), which makes disentangling the influence of certifications on these different types of uncertainty difficult.

In sum, the current study conceptualizes that certifications may address two distinct types of uncertainty regarding the assessment of actors, technical uncertainty and performance standard uncertainty, and either or both may be present in a given context. The next section examines how certifications impact the long-term reputation of actors, even when technical uncertainty is minimal because of the presence of performance standard uncertainty.

How Certifications Influence Reputations by Reducing Performance Standard Uncertainty

Prior studies have posited that certifications are an important means by which the uncertainty regarding an actor's capabilities is reduced (Rao 1994, Wade et al. 2006). This idea contains an implicit assumption that for third-party quality signals to matter, there must be technical uncertainty present (Festinger 1954, Pfeffer et al. 1976, Podolny 2005). However, we propose that even when there is little technical uncertainty, certifications may still influence the assessment of actors because of the presence of performance standard uncertainty. This stems from the fact that certifications may not only provide new information about how the actor performed, but also about how others have interpreted the actor's earlier performance relative to his or her peers. Thus, certifications may act as a filter through which observers interpret the appropriate standard against which the actor's overall capabilities are assessed. Cumulative certifications speak to the fact that the actor, over multiple periods, was thought highly capable relative to his or her peers and thus his or her capabilities should meet or exceed any reasonable standard. As the standard against which actors are being measured may be uncertain, the perceived capabilities of certified actors may inform what level of capabilities observers believe is needed to meet or exceed the standard of desirability. We thus expect that, even for actors with similar levels of technical performance along well-documented dimensions, those who have won and/or accumulated certifications are more likely to be perceived as having exceeded the standard of desirability. Therefore, we predict that, regardless of actors' level of technical performance, certifications will positively impact the long-term reputation of all actors. Accordingly, we hypothesize:

Hypothesis 1. *Certifications will positively influence the long-term reputation of actors in situations that involve minimal amounts of technical uncertainty.*

Interaction of Performance Standard Uncertainty and Technical Performance Level

Our first hypothesis focuses on the impact of certifications on the long-term reputation of an actor, which speaks to the final overall collective judgment as to whether an actor has exceeded the uncertain standard of desirability. This hypothesis thus provides a baseline test of the importance of certifications for all actors because of the presence of performance standard uncertainty. Our second hypothesis examines the iterative process by which observers arrive at this assessment of long-term reputation. Recall that reputation is defined as the collective judgment of observers regarding the quality or capabilities of a focal actor within a specific domain that is earned over time (Fombrun 1996, Podolny 2005, Rindova et al. 2005, Washington and Zajac 2005). This definition implies that reputations are gained through a process of iterative assessments. The current hypothesis suggests that the importance of certifications on iterative assessments will increase as an actor's technical performance approaches the uncertain standard of desirability. In other words, contrary to other studies, which implied that the influence certifications have on the assessments of actors is homogeneous (e.g., Rao 1994, Wade et al. 2006), we propose that this influence will vary depending on their level of technical performance.

For actors whose performance is low, there is little uncertainty that the actor fails to meet the standards of desirability even though this standard remains uncertain. Similarly, for actors whose performance is high, there is also little uncertainty that their performance exceeds any reasonable standard of desirability. For such high- and low-performing actors, there is little uncertainty as to whether they have met any reasonable standard of desirability, and observers quickly arrive at their final overall assessment. Therefore, technical performance dominates the assessment for such actors as their level of technical performance clearly puts them above or below any reasonable, meaningful standard. Consequently, for such actors, the influence of certifications on their long-term reputation is likely to be low. However, for an actor whose technical performance approaches the reasonable range within which the uncertain standard resides, the uncertainty in the minds of observers as to whether the actor has met the standard increases. Thus the uncertainty surrounding the evaluation of such actors shifts from uncertainty surrounding the actor's capabilities to the uncertainty surrounding the standard. Accordingly, for an actor whose perceived capabilities approach the standard, certifications increasingly dominate the assessment, because it is increasingly difficult to distinguish whether the actor met the standard. As such, it may take iterative assessments for observers to coalesce around a final assessment of the actor's overall capabilities.

Recall that one reason performance standard uncertainty exists is that multiple performance metrics may make it difficult for observers to agree on exactly what combination of technical performance metrics constitutes the standard. However, even if observers could agree on what level of performance is required along performance metrics, they have a vested interest in ensuring that performance standard uncertainty persists so that they retain their ability to control such assessments. Therefore, observers are only likely to seek to minimally reduce performance standard uncertainty. That is, they seek to ensure that performance standard uncertainty persists because the persistence of this uncertainty is what allows them to retain their judgment role and not forfeit such judgments to an explicit yardstick, where meeting or exceeding the standard becomes a simple formulaic procedure that no longer requires their expertise or opinion.

To avoid diminishing their role in this manner, certifications allow observers to make judgments about actors while allowing performance standard uncertainty to persist through the following process. In evaluating the capabilities of actors across the full range of technical performance, as described above, observers are likely able to quickly make final judgments about those actors who clearly fall short of or exceed the standard of desirability despite the uncertainty that surrounds the standard. In evaluating the remaining actors, technical performance will be reasonably similar, or at least similar enough that observers maintain doubt as to whether the technical performance meets or exceeds the standard of desirability. In what they term the isolation effect, Kahneman and Tversky (1979) propose that when people evaluate choices, information that is similar between choices is set aside, isolating the areas of difference between choices. Accordingly, when technical performance is similar between actors close to the uncertain standard of desirability, observers may put this information aside and instead turn to third-party quality signals, such as certifications, to guide their assessments of these actors. In this regard, certifications allow observers to make judgments about whether an actor does or does not meet the standard of desirability without ever fully defining what that standard actually is in terms of specific technical performance.

At the population level, performance standard uncertainty is reduced as the range of technical performance for actors remaining under consideration narrows. However, for individual actors who remain under consideration, the uncertainty surrounding assessment of their capabilities increases as their level of technical performance approaches the uncertain standard, because it is increasingly uncertain whether they have exceeded the standard. The use of certifications allows judgments to take place regarding individual actors, even when performance standard uncertainty remains, through the use of prior expert evaluations to replace, or at least supplement, technical performance information. Indeed, using certifications to make judgments while retaining performance standard uncertainty may be appealing to observers. It is defensible because it is "objective" in that it represents a summary judgment of expert observers at an earlier time (Wade et al. 2006), and yet allows observers to retain their judgment prerogative. Accordingly, we expect that certifications will not benefit all actors equally, but rather the influence of certifications will be amplified when the technical performance of actors approaches the uncertain standard, and so we hypothesize:

HYPOTHESIS 2. *Increasing performance standard uncertainty amplifies the relationship between certifications and the assessment of actors. That is, across levels of technical performance there will be an inverted U-shaped relationship between certifications and the assessments of actors in situations that involve minimal amounts of technical uncertainty.*

We test the relationship proposed in Hypothesis 2 on not only the initial assessment, but also on the iterative assessments of actors. This will allow us to examine the degree to which performance uncertainty persists. It is consistent with our theoretical framework that performance standard uncertainty would continue to persist given the vested interest observers have in this occurring, and we thus expect that the curvilinear relationship proposed in Hypothesis 2 will persist as long as performance standard uncertainty persists. Indeed, because such comparisons will become more fine-grained along technical performance dimensions, as actors whose performances are judged to exceed or fall short of the standard are no longer under consideration. We expect that certifications may become the primary means of distinguishing between which actors have or have not met the uncertain standard. Thompson (1967, pp. 86–87) recognized this problem in writing: "when standards of desirability are ambiguous, the assessor must find other (nontechnical) means of resolving his dilemma."

Method

Empirical Context

We studied the relationship between certifications and reputation in the voting for MLB's Hall of Fame by the Baseball Writers Association of America (BBWAA). This setting provides a desirable context in which to test our theory for a number of reasons. First, this is a setting with low technical uncertainty. In baseball, nearly every action that a player takes is measurable and a statistic exists to document how well a given player is performing on that dimension. As such, baseball has been described as "the world's best documented sport" (Frick 1973). Thus, there is a tight coupling between an actor's technical performance over his career and his assessed capabilities. Second, the five-year gap between the end of a player's career and his first eligibility for the Hall means there is ample time for voters to evaluate a player's career in terms of performance statistics and awards accumulated. Third, the annual voting allows us to examine how actors are evaluated who continue to warrant further evaluation by not being inducted into the Hall, but receiving enough votes to continue consideration.

Voting for the Baseball Hall of Fame. The BBWAA has held the voting responsibility for induction into the Hall since the inception of the process in 1936. Former players who have played in a minimum of 10 seasons and have been retired for 5 years are eligible. The BBWAA Screening Committee was introduced in 1968, and its function is to eliminate players from consideration whose careers clearly do not warrant consideration for induction into the Hall (i.e., players whose technical performance is well below the uncertain standard of desirability).[3]

A writer must have covered Major League Baseball for at least 10 years to qualify as a BBWAA voter. Nominees are inducted into the Hall if they receive 75% or more of the ballots cast and can remain on the ballot for up to 15 years, but to remain on the ballot from one year to the next, they must receive at least 5% of the ballots. According to the Hall's website, "Voting shall be based on the player's record, playing ability, integrity, sportsmanship, character, and contributions to the team(s) on which the player played. No automatic elections based on performances such as a batting average of 0.400 or more for one (1) year, pitching a perfect game or similar outstanding achievement shall be permitted." Despite this guidance, statistics are most often cited as the reason voters side for or against a player's candidacy. In this regard, Skipper (2000, p. xiv) wrote, "more than anything else, they (statistics) are the documentation—the overriding reasons for election of players to the Hall of Fame." Furthermore, Tracy Ringolsby, who has been a member of BBWAA since the 1970s and was its president in 1986, stated, "Stats are the tool I can use to feel I have a handle on a player (in voting for the Hall). I do not pretend to be able to visually break down a player like a scout" (Lederer 2004).

Sample

Our sample includes all MLB-eligible position players, i.e., nonpitchers[4] who retired between 1931 and 1990. We elect to begin our sample in 1931, as this date coincides with the advent of the modern awards in baseball as well as being five years before the Hall opened. We cut off our sample in 1990, so all players in our sample have had a full opportunity to be inducted into the Hall.[5] This resulted in an overall sample of 1,042 players. In testing Hypothesis 2, our sample is iteratively reduced as some players are inducted into the Hall and others drop out of consideration in each round. See Appendix 2 for detailed information regarding the number of players under consideration each round.

Dependent Variables

Long-Term Reputation. Our variable of interest in Hypothesis 1 is whether a player was voted into the Hall. We thus constructed a dummy variable that equals one if the player is inducted into the Hall as a result of voting by BBWAA and zero if a player is not. While we recognize that induction into the Hall is the result of an accreditation, this still represents an omnibus assessment of the quality of a player's career, and as such is consistent with our definition of reputation.

Assessment of Actors. To test iterative assessments of actors for Hypothesis 2, we used the percentage of votes a player received in each round. In rounds subsequent to round one, we also include the voting percentage received in the previous round of voting as a control variable.

Independent Variables

Certifications. For MLB players, certifications exist in the form of awards players win during the course of their careers as a result of assessments by observers. We include the major certifications accumulated by MLB players: Most Valuable Player (MVP), World Series MVP, Rookie of the Year, All-Star, and Gold Glove. These certifications were consolidated into one variable, which we call certifications. To construct this variable, we standardized the number of each type of award won and then summed these standardized scores into one variable for each player. We did so because some awards, such as All-Star and MVP, differ significantly in that only two players (one in each league) win the MVP award annually, while over 50 players are named to the All-Star team each year.[6] Thus, standardizing the awards allowed each award to have a similar impact on the overall variable.

The MVP, in its current form, has been awarded since 1931 and is presented annually by the BBWAA. The voting is done by two members of the BBWAA from each major league city. The World Series MVP Award was originally given by the editors of *Sport Magazine* and started in 1955. The award is now voted on during the final game of the World Series by a committee of reporters and officials in attendance (http://www.baseball-almanac.com). The Rookie of the Year Award, now known as the Jackie Robinson Award, is given to the individual player from each league (National and American) who was deemed to have the best rookie season during his first year of eligibility. This award began in 1947, and in 1947 and 1948, only one winner was selected from MLB. Since 1949, two players have been selected each year, one from each league. Since 1980, this award has been granted through voting by the BBWAA.

All-Star teams were originally selected by the managers and the fans for the 1933 and 1934 games. From 1935 through 1946, managers selected the entire team for each league. From 1947 to 1957, fans chose the team's starters. From 1958 through 1969, managers, players, and coaches made the All-Star Team selections. Since 1970, the fans' votes determine the starters, while back-up players are selected by the manager of the All-Star team. The Gold Glove was first presented by the Rawlings Corporation in 1957 to honor the 18 best fielders at their positions. Rawlings used a combination of defensive statistics along with visual effect to determine the winner and honored the recipient with a large gold glove. Voting is currently done by the managers and coaches from each team who are not able to select their own players. One player from each position, in each league, receives a Gold Glove annually.

While the process of determining the winner of some of the awards varied during our sample, what the award represented remained substantially the same. For instance, selection of the All-Star team has been determined at times by managers and at other times by fans, and, since 1970, fans have determined the starters and managers have determined the nonstarters. However, during the entire time period, the point of the All-Star team has remained to select the best players by position. It is worth noting, however, that our results are robust to samples that exclude the awards that start later in our sample (Gold Glove and World Series MVP) as well to different sample periods (see Endnote 3).

Level of Technical Performance. To test Hypothesis 2, we calculated the career standards statistic (James 1994) to assess the overall level of a player's technical performance. This statistic was developed by Bill James, who has written more than 45 books that focus on MLB statistics and has been referred to as the "Mozart of baseball statisticians" (*Chicago Sun-Times*). (See Appendix 1 for a detailed explanation of the calculation of this score.) According to James (1994, p. 175), based on this statistic, "If there was a perfect player, he'd score 100." James considers those who score 60 or higher as exceptional and thus likely to be considered by BBWAA observers as clearly exceeding the standard of desirability, while players become "a viable candidate" for the Hall at 35 (James 1994, p. 181). Thus, players who score 35–59 are thought to have a level of technical performance that is near the uncertain standard for induction into the Hall. Using this guidance, we expect that the influence of certifications will increase as actors' technical performance approaches this range of technical performance. Thus, while we do not directly measure performance standard uncertainty in our study, we believe that at the population level, it may be inferred at any point by the range of technical performance attained by players who remain under consideration. (See Appendix 2.) However, for an individual actor under consideration, as the actor's overall technical performance approaches the uncertain standard, uncertainty increases as to whether this actor met the standard of desirability.

Consistent with the recommendations of Cohen et al. (2003, pp. 292–295), we generated three additional variables using the career standards statistic to test the curvilinear (inverted-U) relationship proposed in Hypothesis 2. Prior to the generation of these additional variables, the career standards statistic was centered to reduce collinearity between the squared and nonsquared terms (Cohen et al. 2003). First, we interact the career standards statistics with certifications a player has won, which tests for the presence of a linear interaction between certifications and technical performance. Second, we generate the squared value of the career standards statistic and then we interact this squared value with certifications, which tests for the nonlinear portion of the curvilinear relationship. Support for Hypothesis 2 is found when the linear interaction is positive and the squared interaction term is negative.

Control Variables

Player Offensive Variables. To control for specific aspects of the actual performance of players in our sample, we included the most common career offensive statistics. We controlled for runs batted in, stolen bases, batting average, slugging percentage, and on-base percentage. We multiplied all variables expressed as percentage (batting average, slugging percentage, and on-base percentage) by 1,000 to ease their interpretation. Initially, we also controlled for the number of home runs hit by a player in his career, but we dropped this variable because of its collinearity ($r = 0.87$) with runs batted in.[7] We also include a variable that counts the number of widely recognized statistical milestones a player achieved over the course of his career: 500 home runs, 3,000 hits, or a 0.300 batting average. We dummy-coded the achievement of each of three milestones: 1 = met milestone; 0 = did not meet milestone. Thus, this variable can take on a value from 0 to 3.[8] MLB players who achieve 500 career home runs or 3,000 hits are often referred to as having earned admission to a "club." Further evidence of the importance of 500 career home runs or 3,000 career hits is the fact that every player who has achieved either milestone ultimately has been inducted into the Hall. Last, we controlled for whether a player's career batting average was over 0.300, since this level of performance is considered the demarcation point between great hitters and all other hitters. Regarding the importance of hitting 0.300, McConnell (2003) wrote that, "The magic number for a batting average is 0.300." Furthermore, books that track statistical achievements of players often list the names of players with career 0.300 batting averages (e.g., LaMar 1993).

We also created a variable that sums the number of times a player led his respective league (American or National) in each "Triple Crown" category: home runs, runs batted in, or batting average. The winner of each category in each league is monitored closely, because the player who leads the league in any of the categories is referred to as that year's "batting champion" or "home run champion." This variable essentially creates a scale counting the number of times a player was the best in his respective league along a known performance dimension, as leading the league in one of these categories may be an important means by which a players' technical performance is assessed.

Player Defensive Variables. First, we capture the efficiency with which the player plays defense, calculated by adding together the number of successful defensive plays a player makes (in baseball terms, putouts + assists) and dividing that total by the total number of defense plays a player attempted (putouts + assists + errors). Second, we include a dummy variable controlling for whether a player played a middle infield position. This variable equals one if the player played catcher, second base, or shortstop as each of these positions is thought to be more defensively demanding. Thus, if a player played one of these positions, his offensive performance may be interpreted differently by voters (Findlay and Reid 2002).

Other Control Variables. We also control for the total number of years in each player's career, as many baseball statistics are cumulative over the course of a player's career (runs batted in, batting championship). Controlling for the duration of a player's career helps control for players whose relatively long or short career may affect the meaning of their statistics. Since the performance of a player's team may influence perception of his career, we also control for the winning percentage of the team on which he played, the number of World Series championships won, and the average number of fans attending the games in which he played. The first two variables control for how well a player's team performed, while the third controls for the relative visibility of the player. Last, we add a dummy variable for each decade in our sample that takes on the value of one in the decade in which the player retired and zero for all other decades, with the 1930s as the omitted category to control for possible heterogeneity in the selection process.

Analysis

To test Hypothesis 1, which uses induction into the Hall as the dependent variable, we used a logistic model, as this is a dichotomous outcome. As such, the dependent variable has an S-shaped association with its predictors (Liao 1994), which violates the assumption of linearity in many regression models. To test Hypothesis 2, which uses the percentage of votes a player received in each round as the dependent variable, we used ordinary least squares regression with robust standard errors to correct for heteroskedasticity in the error terms.

Results

Table 2 presents descriptive statistics for each of the variables in our study as well as the bivariate correlations. Table 3 presents the logit models that test Hypotheses 1.[9] Table 3 tests Hypothesis 1, which asserts that the certifications won by an actor are positively associated with his long-term reputation (entry into the Hall). Model 1 presents the effects of the control variables, and Model 2 adds the certification variable (Certifications). As certifications have a positive and significant effect ($p < 0.01$) on induction into the Hall, Hypothesis 1 receives support.

Table 4 presents the models that test Hypothesis 2; Models 1 and 2 present the control models and Models 3–7 test Hypothesis 2 across rounds 1–5 of voting for the Hall. Support for Hypothesis 2 is indicated when the linear interaction of certifications and the career standard statistic is positive and the interaction of certifications and the squared career standards statistic is negative.[10] We found that the linear interaction of certifications and the career standard statistic was positive and significant in Rounds 1 and 2, and the interaction of certifications and the squared career standards statistic was negative and significant in Rounds 1 and 2 as well, supporting Hypothesis 2. Furthermore, the linear interaction is positive and significant ($p < 0.05$) and the squared interaction term is negative and moderately significant ($p < 0.10$) in Round 3, suggesting that performance standard uncertainty persists through multiple iterations of evaluation even when controlling for the votes in the prior round. However, by Round 4 only the linear interaction is significant ($p < 0.05$), suggesting that as observers continue to confront iterative evaluations of a diminishing group of actors, they coalesce around their final evaluation of each actor. The results of the curvilinear (inverted-U) interaction for voting in the first round are represented graphically in Figure 1.

Discussion

This study examined one of the implicit assumptions of the reputation literature, that technical uncertainty is necessary for third-party quality signals, such as certifications, to influence the assessment of an actor (e.g., Rao 1994, Wade et al. 2006). The reputation literature has asserted that the information provided by certifications helps to fill in incomplete information regarding an actor's capabilities. Contrary to this implicit assumption, we found that certifications influence an actor's reputation, even in a context of low technical uncertainty. Indeed, in a context where an actor's performance was easily observable and considered tightly coupled with the actor's capabilities, we found that the existence of performance standard uncertainty, which we

Table 2 Descriptive Statistics

Variable	Var	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Inducted into hall (1 = Yes)	1	0.05	0.22	1.00																						
Voting percentage in the 1st round	2	3.07	13.84	0.77	1.00																					
Awards (standardized)	3	0.00	0.89	0.65	0.76	1.00																				
Career standard statistic	4	0.00	11.80	0.66	0.59	0.63	1.00																			
Career standard statistic squared	5	139.08	366.74	0.74	0.68	0.58	0.71	1.00																		
Years in career	6	14.01	3.40	0.33	0.35	0.39	0.48	0.35	1.00																	
Average annual attendance (millions)	7	1.13	0.42	−0.03	0.07	0.14	−0.13	−0.06	0.04	1.00																
Career runs batted in	8	526.30	368.82	0.56	0.54	0.65	0.76	0.60	0.67	−0.03	1.00															
Career stolen bases	9	61.81	85.22	0.23	0.26	0.24	0.30	0.22	0.33	0.11	0.32	1.00														
Career batting average	10	264.99	25.39	0.33	0.26	0.33	0.55	0.41	0.34	−0.24	0.61	0.27	1.00													
League leader	11	0.38	1.43	0.60	0.63	0.57	0.61	0.71	0.31	−0.01	0.60	0.10	0.37	1.00												
World Series championships	12	0.80	1.28	0.22	0.22	0.29	0.24	0.22	0.17	0.14	0.25	0.11	0.16	0.18	1.00											
Career team winning percentage	13	505.23	39.77	0.20	0.20	0.29	0.28	0.18	0.12	0.27	0.25	0.14	0.22	0.14	0.57	1.00										
Career fielding percentage	14	975.30	12.13	0.03	0.06	0.13	−0.05	0.03	0.11	0.20	0.07	−0.08	0.01	0.06	0.05	0.06	1.00									
Career milestones achieved	15	0.11	0.35	0.54	0.52	0.42	0.66	0.68	0.28	−0.18	0.50	0.20	0.57	0.58	0.11	0.17	−0.02	1.00								
Middle infield position (dummy)	16	0.33	0.48	−0.01	−0.01	−0.01	0.09	−0.07	−0.05	−0.03	−0.20	−0.09	−0.28	−0.09	0.01	0.02	−0.27	−0.08	1.00							
Career slugging percentage	17	389.16	60.73	0.42	0.38	0.46	0.59	0.48	0.35	−0.05	0.77	0.12	0.70	0.54	0.20	0.23	0.14	0.47	−0.41	1.00						
Career on-base percentage	18	333.41	32.52	0.35	0.30	0.35	0.62	0.42	0.32	−0.20	0.59	0.21	0.75	0.41	0.18	0.24	0.07	0.47	−0.26	0.70	1.00					
40s decade dummy	19	0.12	0.33	0.05	−0.05	−0.01	0.14	0.08	−0.01	−0.44	0.10	−0.02	0.22	0.03	0.00	0.04	−0.14	0.14	0.00	0.08	0.18	1.00				
50s decade dummy	20	0.13	0.33	−0.01	−0.03	0.00	0.01	−0.05	0.01	−0.10	−0.05	−0.11	0.04	0.01	0.04	0.03	−0.03	−0.04	−0.01	0.02	0.13	−0.14	1.00			
60s decade dummy	21	0.16	0.37	−0.03	0.00	0.04	−0.03	−0.02	−0.11	−0.01	−0.01	−0.13	−0.04	−0.01	0.06	0.02	0.06	−0.06	0.02	0.08	0.01	−0.16	−0.17	1.00		
70s decade dummy	22	0.21	0.40	0.01	0.05	0.04	−0.12	−0.01	0.00	0.10	−0.09	−0.01	−0.28	−0.03	−0.06	−0.04	0.08	−0.05	0.02	−0.16	−0.27	−0.19	−0.20	−0.22	1.00	
80s decade dummy	23	0.28	0.45	−0.05	0.01	0.00	−0.11	−0.06	0.10	0.62	0.00	0.16	−0.12	−0.04	−0.07	−0.04	0.12	−0.13	−0.05	−0.06	−0.13	−0.24	−0.24	−0.27	−0.32	1.00

Table 3 Induction Into Hall of Fame (Yes = 1; No = 0)

Variable	Model 1	Model 2
Years in career	0.138 (0.161)	0.187 (0.176)
Average annual attendance (millions)	−1.364 (1.768)	−1.182 (1.889)
Runs batted in	0.001 (0.002)	0.002 (0.03)
Stolen bases	0.007*** (0.003)	0.008*** (0.003)
League leader	0.363* (0.208)	0.169* (0.220)
Career batting average	−0.036 (0.026)	−0.048 (0.030)
On-base percentage	−0.017 (0.023)	0.000 (0.027)
Slugging percentage	0.012 (0.014)	0.010 (0.015)
Team World Series championships	−0.216 (0.222)	−0.245 (0.252)
Career team winning percentage	0.023** (0.011)	0.009 (0.012)
Career milestones	0.372 (0.928)	0.123 (1.087)
Career fielding percentage	0.058 (0.034)	0.042 (0.037)
Middle infield position (dummy)	1.842** (0.826)	1.156 (0.925)
40s decade dummy	0.112 (1.081)	−1.263 (1.199)
50s decade dummy	1.217* (1.313)	−0.764 (1.501)
60s decade dummy	0.665 (1.937)	−2.097 (2.159)
70s decade dummy	0.770 (1.805)	−3.359 (2.337)
80s decade dummy	−0.507 (2.228)	−3.870 (2.757)
Career standards statistic	0.226*** (0.061)	0.195*** (0.072)
Certifications (standardized)		1.357*** (0.450)
Constant	−73.713** (32.074)	−52.799 (34.712)
Pseudo R^2	0.798	0.828
Observations	1,042	1,042
Chi^2 (degrees of freedom)	324.91 (19)	337.17 (20)
Log likelihood	−41.149	−35.109

* = $p < 0.10$; ** = $p < 0.05$; *** = $p < 0.01$.

define as the uncertainty that exists when the standards or yardstick against which the actor's capabilities are to be judged in order for them to be considered acceptable are uncertain (Thompson 1967), allowed certifications to influence an actor's reputation. These results support our theory that two types of uncertainty may influence an actor's reputation, technical uncertainty and performance standard uncertainty, and that even when technical uncertainty is minimized, there may remain performance standard uncertainty regarding how an actor should be judged. Thus, certifications may influence an actor's reputation by reducing technical uncertainty and also by reducing performance standard uncertainty regarding whether an actor's capabilities meet the standard of desirability. By disentangling the two separate types of uncertainty and establishing that performance standard uncertainty can exist independently of the presence of technical uncertainty, we hope to open the doors to future research to understand the interplay between the two. We see this as an important broadening of the scope of the influence of third-party quality signals.

Second, we found that certifications became more influential when an actor's level of technical performance was near the uncertain standard of desirability. This finding suggests that the influence of certifications may vary from actor to actor, and as an actor's technical performance approaches the uncertain standard of desirability, the importance of these third-party quality signals rises as observers have increasing difficulty assessing an actor's capabilities relative to the uncertain standard. An implication of this finding is that, contrary to previous studies that suggested that certifications have homogeneous effects on actors within a given domain, the technical performance of actors and the influence of certifications on their reputations are inextricably linked in that the level of an actor's technical performance influences the degree of performance standard uncertainty associated with the assessment of that actor. Indeed, as actors' performance approaches the uncertain standard, certifications will influence the long-term reputation of these actors by providing a ready-made assessment by previous observers and may speak to whether the actor has exceeded the standard of desirability.

Given our initial encouraging results, future research could test whether the influence of other types of third party assessments varies across levels of technical performance in other contexts. For example, a recently developing literature examines how reports by the financial press may impact organizational outcomes (e.g., Deephouse 2000, Pollock and Rindova 2003). These studies have shown that the amount and type of coverage provided by the financial press may affect firm performance (Deephouse 2000) and initial public offering valuation (Pollock and Rindova 2003). Both studies posited that technical uncertainty allowed the financial press to influence the opinion of external observers. However, our results suggest that press coverage may have greatest impact on firms with the greatest amount of performance standard uncertainty. Future research may wish to examine whether the reports offered in the financial press are more or less influential depending on how well a given organization has performed.

We also theorized that the persistence of performance standard uncertainty may stem from the presence of multiple performance metrics or that the persistence may be at least partially attributed to a deliberate act on the part of the observers themselves. Indeed, if performance

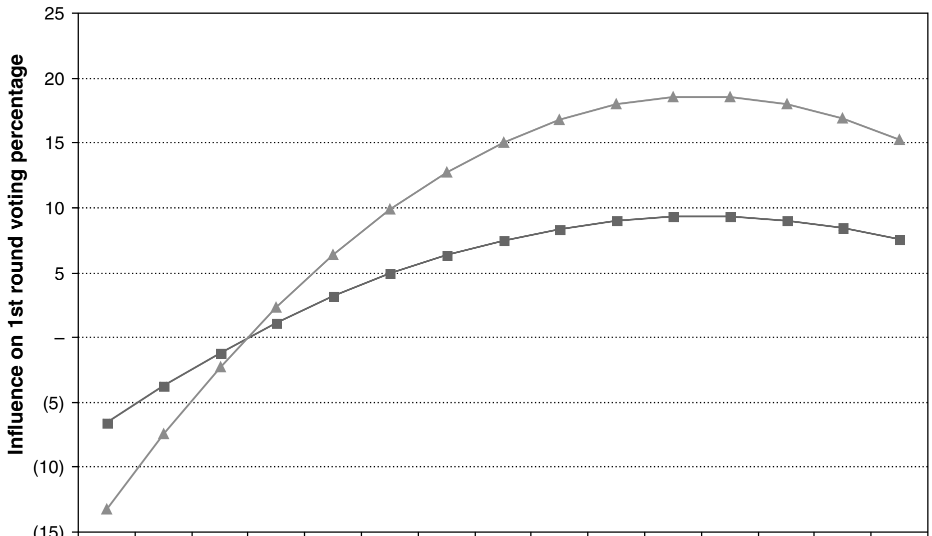
Table 4 Voting Percentage by Round

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Variable	Round 1	Round 1	Round 1	Round 2	Round 3	Round 4	Round 5
Years in career	−0.108	0.249**	0.164*	−0.164	−0.291	−0.123	−0.164
	(0.119)	(0.099)	(0.098)	(0.119)	(0.313)	(0.388)	(0.580)
Avg. annual attendance (millions)	−1.118	−1.480	−1.475	0.024	1.389	−3.567	−2.844
	(1.561)	(1.230)	(1.095)	(1.306)	(4.196)	(5.452)	(7.058)
Runs batted in	0.012***	−0.004	0.001	0.005*	0.007	0.005	0.010
	(0.003)	(0.002)	(0.002)	(0.003)	(0.005)	(0.006)	(0.008)
Stolen bases	0.021**	0.013	0.013	0.004	0.002	0.005	0.005
	(0.010)	(0.009)	(0.009)	(0.003)	(0.009)	(0.008)	(0.014)
League leader	3.557***	1.322**	1.183	0.276	−0.486	0.086	−0.077
	(0.636)	(0.650)	(0.747)	(0.549)	(0.526)	(0.703)	(1.007)
Career batting average	−0.090***	−0.083	−0.056***	0.001	−0.056	−0.029	−0.131
	(0.025)	(0.021)	(0.020)	(0.023)	(0.059)	(0.064)	(0.108)
On-base percentage	0.013	0.007	0.010	−0.006	−0.025	−0.033	−0.010
	(0.014)	(0.013)	(0.012)	(0.016)	(0.043)	(0.044)	(0.060)
Slugging percentage	−0.021*	0.005	−0.006	−0.012	−0.001	−0.005	0.006
	(0.012)	(0.009)	(0.008)	(0.011)	(0.028)	(0.030)	(0.039)
Team World Series championships	0.755	−0.012	−0.009	0.711*	0.957*	1.242*	0.258
	(0.515)	(0.403)	(0.423)	(0.385)	(0.551)	(0.697)	(0.860)
Career team winning percentage	0.010	0.000	0.008	−0.006	−0.007	0.007	−0.005
	(0.013)	(0.009)	(0.008)	(0.010)	(0.024)	(0.038)	(0.048)
Career milestones	10.956***	5.413***	5.045***	1.263	2.880	4.103	4.078
	(2.290)	(2.095)	(1.850)	(1.606)	(2.287)	(3.263)	(3.819)
Career fielding percentage	0.035	−0.026	0.002	−0.004	0.123*	0.037	0.058
	(0.025)	(0.019)	(0.018)	(0.027)	(0.067)	(0.072)	(0.106)
Middle infield position (dummy)	1.454	−0.378	0.252	−0.276	−0.387	−1.240	0.743
	(0.898)	(0.894)	(0.804)	(0.827)	(1.844)	(2.280)	(2.933)
40s decade dummy	−0.361	−3.577***	−3.233***	1.610	2.967	2.625	4.258
	(1.233)	(1.051)	(1.027)	(1.136)	(2.128)	(2.323)	(3.044)
50s decade dummy	3.278**	−1.568	−1.119	2.927*	3.811	10.510**	13.503***
	(1.563)	(1.164)	(1.144)	(1.612)	(3.196)	(4.046)	(4.406)
60s decade dummy	4.433***	−1.349	−0.200	2.2022	0.836	3.852	4.711
	(1.525)	(1.019)	(0.935)	(1.517)	(3.721)	(4.434)	(4.870)
70s decade dummy	4.830	−1.105	−0.553	0.269	−1.688	5.267	9.226
	(1.864)	(1.163)	(1.043)	(1.526)	(4.394)	(5.542)	(6.477)
80s decade dummy	4.383**	−0.352	−0.341	−0.697	−1.958	2.908	−1.182
	(2.049)	(1.441)	(1.286)	(1.775)	(6.048)	(8.788)	(8.879)
Career standards statistic		−0.201	0.032	−0.016	−0.024	−0.148	−0.204
		(0.057)	(0.054)	(0.073)	(0.215)	(0.267)	(0.364)
Career standards statistic squared		0.009***	0.002	−0.001	0.008*	0.010	0.014
		(0.003)	(0.003)	(0.002)	(0.005)	(0.007)	(0.009)
Certifications (standardized)		9.019***	2.510***	−0.359	−1.402	−1.237	−0.160
		(1.083)	(0.790)	(0.656)	(1.654)	(1.629)	(2.352)
Certifications ∗ Career standard statistic			0.522***	0.253***	0.371**	0.376**	0.114
			(0.085)	(0.093)	(0.147)	(0.172)	(0.232)
Certifications ∗ Career standard statistic squared			−0.006***	−0.005**	−0.005*	−0.004	0.001
			(0.002)	(0.003)	(0.004)	(0.004)	(0.005)
Voting percent in prior round				0.936***	0.863***	0.999***	0.965***
				(0.081)	(0.079)	(0.099)	(0.174)
Constant	−20.492	44.698**	5.869	10.956	−94.991	−20.672	−21.629
	(25.935)	(20.904)	(19.597)	(27.089)	(63.467)	(67.478)	(98.026)
R^2	0.529	0.706	0.755	0.832	0.763	0.829	0.722
Adjusted R^2	0.520	0.700	0.750	0.817	0.726	0.792	0.655
Observations	1,042	1,042	1,042	295	174	137	125

$^{*} = p < 0.10$; $^{**} = p < 0.05$; $^{***} = p < 0.01$.

standard uncertainty did not persist, assessing whether actors surpassed this standard would be a simple matter of checking a given actor against a predetermined list of objective criteria. In other words, the complete elimination of performance standard uncertainty would disempower expert observers, as specialized knowledge would no longer be required to make such assessments. Indeed, we suggest that it may be the case that it is actually in the best interest of expert observers for performance standard uncertainty to persist. Performance standard uncertainty deliberately infused by observers would suggest that the standard of desirability against which actors are measured is itself socially constructed in the sense that what is acceptable may vary between actors

Figure 1 Influence of Certifications on First Round Voting for Hall Across Levels of Technical Performance



	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
Certifications +1 sd	(6.6)	(3.7)	(1.1)	1.2	3.2	4.9	6.4	7.5	8.4	9.0	9.3	9.3	9.0	8.5	7.6
Certifications +2 sd	(13.2)	(7.5)	(2.3)	2.3	6.4	9.8	12.7	15.0	16.8	18.0	18.6	18.6	18.0	16.9	15.2

Performance points (Bill James statistic)

or organizations. Thus, while the capabilities of certain actors are perceived as clearly above or clearly below this uncertain standard, for actors whose technical performance approaches the uncertain standard, the persistence of performance standard uncertainty is what allows observers discretion to make judgments that include or exclude actors who may be very similar along technical performance dimensions. Specifically, we found that over multiple rounds of voting for induction into the Hall, certifications continued to influence the percentage of votes players received. As Appendix 2 demonstrates, even after 15 rounds of voting, the career standard statistic of players remaining under consideration ranged from 19 to 50. Thus, despite a minimal level of technical uncertainty, performance standard uncertainty persisted throughout multiple rounds of voting for the Hall, which allowed certifications to continue to be influential even after the influence of technical performance measures became insignificant. It is also the case that in inducting players for the Hall, some players are inducted with slightly worse technical statistics than other comparable players who are not inducted (James 1994), suggesting that, as we contend, observers may deliberately retain performance standard uncertainty to maintain their ability to judge actors with similar technical performance differently, effectively introducing social desirability in assessing whether a particular actor met the standard of desirability.

That performance standard uncertainty may persist because of active agency on the part of expert observers has implications for groups or organizations that set standards. Our theory suggests that the more uncertain the standards, the more powerful the assessing agency becomes. Consider environmental groups that may attempt to influence the conduct of large corporations such as Walmart or Microsoft to become more environmentally or socially responsible. If the standard for being considered "green" or "socially responsible" were clear and quantifiable, such interest groups would quickly lose influence, as meeting a particular standard would simply amount to checking off "green" practices. As such, it may behoove such organizations to make sure that such standards remain uncertain. Along these same lines, it will be telling to examine the relative influence of corporate governance watchdog groups in the United States prior to and following the passage of the Sarbanes-Oxley Act in 2002. The outcome of this particular legislation is that corporations' governance practices can now be assessed against a very specific compliance standard that lists in detail what it means for directors to be considered independent or how the audit committee should be structured for a corporation to be in compliance. Our theoretical framework suggests that, going forward, the power of governance watchdog groups should decrease, as Sarbanes-Oxley removed a great deal of performance standard uncertainty regarding the governance of firms in the United States. Alternatively, to remain powerful and relevant, such groups would need to shift the criteria by which organizations are evaluated. Indeed, further exploration of the complex web of interdependencies between performance standard uncertainty and various interest groups represents fertile ground for future research.

Last, we also theoretically distinguished between different types of evaluations which had previously been clumped together under the broad label of "certifications." Within our conceptualization, certifications involve evaluations where actors' relative quality is assessed and some actors must win certifications as a result of such evaluations. On the other hand, accreditations and ratings are evaluations where actors are assessed relative to a standard without regard to the relative quality of the actors under consideration. This distinction helps clarify the types of information provided by different types of assessments and may also have implications for the types of uncertainty that are reduced as a result of these different types of evaluations. Future research can examine how these different types of social evaluations may inform each other. Since certifications contests typically focus on ranking actors of high quality within a given domain (e.g., CEO of the Year or *Fortune's* Most Admired Companies) and accreditations typically focus on whether actors have surpassed some objective standard of quality (e.g., ISO 9000 or AACSB), we expect that the information provided by these different types of quality signals is complementary. Our results suggest that performance standard uncertainty is what allows certifications to influence the outcome of an accreditation, i.e., induction into the Hall. We also expect that in a context with high technical uncertainty, accreditations may provide information that will signal to observers that actors within this context have met a specific standard of technical performance, which could influence subsequent certification contests. Indeed, future research may examine how these different types of third party quality signals may interact and influence the reputation of those under consideration under varying conditions of uncertainty.

Conclusion

When observers make a collective judgment about an actor's overall reputation, there are two different assessments that need to be undertaken: an assessment of the technical capabilities of the actor, and how those capabilities measure up against the standards of desirability (Thompson 1967). Uncertainty can exist around both of these assessments. While much of the current literature on reputations has implied that third-party quality signals, such as certifications, are important to the assessment of reputation because of the uncertainties surrounding the assessment of the technical capabilities of the actor, we demonstrate that performance standard uncertainty can exist even when technical uncertainty is minimal, and that third-party quality signals also can independently reduce uncertainty surrounding the standards of desirability. We also found that the influence of certifications varies depending on the level of an actor's technical performance, i.e., as an actor's performance approaches the uncertain standard of desirability, the influence of certifications increases.

Acknowledgments

The authors thank the Editor, Tina Dacin, and two anonymous *Organization Science* reviewers for their insight and support; Bill James for the statistical advances he contributed to the baseball analysis, which was leveraged in this paper, and Sean Lahman and his creation, the Baseball Archive. The authors also gratefully acknowledge support provided through a University of Georgia Terry School of Business Terry-Sanford research grant.

Appendix 1. Calculation of Hall of Fame Career Standards Statistic

The Hall of Fame career standard statistic is calculated by awarding points in the following manner (James 1994 pp. 174–175):

Awarded for batting statistics:

- One point for each 150 hits above 1,500, limit 10.
- One point for each 0.005 of batting average above 0.275, limit 9.
- One point for batting over 0.300.
- One point for each 100 runs over 900, limit 8.
- One point for scoring more than 0.500 runs per game.
- One point for scoring more than 0.644 runs per game.
- One point for each 100 runs batted in over 800, limit 8.
- One point for driving in more than 0.500 runs per game.
- One point for driving in more than 0.600 runs per game.
- One point for each 0.025 of slugging percentage above 0.300, limit 10.
- One point for each 0.010 of on-base percentage above 0.300, limit 10.
- One point for each 200 home runs.
- One point if home runs are more than 10% of hits.
- One point if home runs are more than 20% of hits.
- One point for each 200 extra base hits over 300, limit 5.
- One point for each 200 walks over 300, limit 5.
- One point for each 100 stolen bases, limit 5.

In addition, award points for a player's primary position:

- 20 points for catcher, 16 for shortstop, 14 for second baseman, 13 for third baseman, 12 for center fielder, 6 for right fielder, 3 for left fielder, 1 for first baseman, 0 for designated hitter.

Endnotes

[1]Within accreditations, there is also a distinction to be drawn between those that require renewal, and those that do not. Examples of organization level accreditations that require renewal include ISO9000 and AACSB accreditations. An example of an accreditation that does not require renewal is the Federal Energy Regulatory Commission accreditation of start-ups in the independent power sector (Sine et al. 2007). At the individual level, CPAs and MDs require ongoing education and periodic reaccreditation. Examples of accreditations that do not require renewal include the British honor system, whereby honors are granted for life by the monarch, and the Academy of Motion Picture Arts and Sciences which grants lifetime achievement awards. The MLB Hall of Fame used in this paper is also an example of an individual accreditation that does not require renewal.

[2]We focus on the discussion of reputation, although some other constructs, particularly status, may seem conceptually similar to reputation; however, the earned quality of reputation sets it apart from status. Status is conceived as "a fundamentally relational concept" (Washington and Zajac 2005, p. 286) that is not simply an atomistic attribute of an isolated actor but rather

Appendix 2. Players Considered for Induction Into Hall of Fame, by Round

	Round of voting														
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th
Number of players															
Considered each round	1,042	295	174	137	125	106	84	76	63	56	48	39	31	19	15
Inducted each round	22	1	1	4	4	4	2	4	4	2	1	1	1	0	0
Dropping out each round	725	120	36	8	15	18	6	9	3	6	8	7	11	4	15
Players to be considered in next round	295	174	137	125	106	84	76	63	56	48	39	31	19	15	0
Range of career standard statistic															
Minimum number of players considered	1	8	8	15	15	15	15	15	15	19	19	19	19	19	19
Maximum number of players considered	82	79	79	79	79	66	66	64	64	64	54	54	54	54	50
Descriptive statistics															
Total players considered	1,042														
Total players not inducted into the Hall	990														
Total players inducted into the Hall	52														
	Mean	Std. dev.	Min	Max											
Players elected in the hall annually	0.87	0.81	0	4											

Note. Players are inducted when they receive 75% or more of the ballots cast by BBWAA in a given round and they drop out of consideration if they receive less than 5% of the ballots cast in a given round.

is "directly tied to the pattern of relations and affiliations in which the actor does and does not choose to engage" (Podolny 2005, p. 13). Thus, status is gained through association with other high-status actors, and, as Washington and Zajac (2005, p. 282) note, is "not based on traditional performance considerations" and can be influential "irrespective of performance." On the other hand, reputation does not focus on affiliations, but rather on the evaluation of the quality of an actor. As our measure of actor reputation, voting for the Hall, represents an omnibus assessment of a player's quality or capabilities over time, it is consistent with our definition of reputation.

[3]Any player on MLB's "ineligible list" is not an eligible candidate. During the time period of our study, only one player, Pete Rose, was on the ineligible list and he was excluded from our sample.

[4]A player's position was determined by the position at which he played the most games during his career.

[5]The authors verified that no players in our sample were still under consideration for induction into the Hall. As an aside, we reran all models using samples using cut-off dates of 1985, 1980, and 1975, and our results were substantively unchanged in all cases.

[6]To examine the robustness of our results, we also constructed separate measures for awards where only one or two are given each year (MVP, World Series MVP, Rookie of the Year) and awards where there are multiple winners for each position (Gold Glove and All-Star); our results and conclusions are substantively unchanged.

[7]Our results are substantively unchanged when we include home runs and exclude runs batted in.

[8]If each milestone dummy variable is entered separately, our results and conclusions are substantively unchanged.

[9]We used the Collin procedure in Stata to test whether multicollinearity is a concern. Based on the variance inflation factors generated, all of which fell well below 10.0, multicollinearity does not appear to be an issue with the variables used (Belsey et al. 1980).

[10]We also performed alternative analyses where we used a spline function to generate three certification variables across different levels of players' technical performance. Using James' guidance, the first variable took on the value of certifications for players whose performance was below 35 using James' career standards statistic and zero otherwise; the second took on the value of certifications for players whose performance was between 35 and 59 and zero otherwise; and the final variables took on the value of certifications for players whose performance was 60 or higher and zero otherwise. The only variable that was positive and significant was the certification variable for players whose career performance statistic was between 35 and 59, which is consistent with our suggestion that certifications are most meaningful for players whose performance approaches the uncertain technical standard. While the fit between the models listed in Table 4 and the spline models did not differ significantly, we used the curvilinear interaction models as the continuous nature of this interaction is more consistent with our theory.

References

Beck, N., P. Walgenbach. 2005. Technical efficiency or adaptation to institutionalized expectations? The adoption of ISO 9000 standards in the German mechanical engineering industry. *Organ. Stud.* **26** 841–66.

Belsey, D. A., E. Kuh, R. E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* John Wiley, New York.

Benjamin, B. A., J. M. Podolny. 1999. Status, quality, and social order in the California wine industry. *Admin. Sci. Quart.* **3** 563–589.

Boiral, O. 2003. ISO 9000: Outside the iron cage. *Organ. Sci.* **14** 720–737.

Carter, S. M., T. W. Ruefli. 2006. Intra-industry reputation dynamics under a resource based framework: Assessing the durability factor. *Corporate Reputation Rev.* **9** 3–25.

Cascio, J., G. Woodside, P. Mitchell. 1996. *ISO 14000 Guide*: *The New International Environmental Management Standards*. McGraw Hill, New York.

Cohen, J., P. Cohen, S. G. West, L. S. Aiken. 2003. *Applied Multiple Regression/Correlation Analysis for the Behavior Sciences*. 3rd ed. Lawrence Erlbaum Associates, Mahwah, NJ.

D'Aveni, R.A. 1996. A multiple-constituency, status-based approach to interorganizational mobility of faculty and input-output competition among top business schools. *Organ. Sci.* **7** 166–189.

Darnall, N. 2006. Why firms mandate ISO 14001 certification. *Bus. Soc.* **45** 354–381.

Deephouse, D. L. 2000. Media reputation as a strategic resource: An integration of mass communication and resource-based theories. *J. Management* **6** 1091–1112.

Deephouse, D. L., M. C. Suchman. 2008. Legitimacy in organizational institutionalism. R. Greenwood, C. Oliver, R. Suddaby, K. Sahlin, eds. *The SAGE Handbook of Organizational Institutionalism*. Sage, Thousand Oaks, CA, 49–77.

Durand, R., J. McGuire. 2005. Legitimating agencies in the face of selection: The case of AACSB. *Organ. Stud.* **26** 165–196.

Elsbach, K. D., R. M. Kramer. 1996. Members' responses to organizational identity threats: Encountering & countering the Business Week rankings. *Admin. Sci. Quart.* **41** 442–476.

Festinger, L. 1954. A theory of social comparison processes. *Human Relations* **7** 117–140.

Findlay, D. W., C. E. Reid. 2002. A comparison of two voting models to forecast election into the National Baseball Hall of Fame. *Managerial and Decision Econom.* **23** 99–113.

Fombrun, C. J. 1996. *Reputation*: *Realizing Value from the Corporate Image.* Harvard Business School Press, Boston.

Fombrun, C. J., M. Shanley. 1990. What's in a name: Reputation building and corporate strategy. *Acad. Management J.* **33** 233–258.

Frick, F. C. 1973. *Games, Asterisks, & People*: *Memoirs of a Lucky Fan*. Crown Publishers, New York.

Fulmer, I. S., B. Gerhart, K. Scott. 2003. Are the 100 best better? An empirical investigation of the relationship between being a "great place to work" and firm performance. *Personnel Psych.* **56** 965–993.

Guler, I., M. Guillen, J. Macpherson. 2002. Global competition, institutions and the diffusion of organizational practices: The international spread of ISO 9000 quality certificates. *Admin. Sci. Quart.* **47** 207–233.

Graffin, S., J. Wade, J. Porac, R. McNamee. 2008. The impact of CEO status diffusion on the economic outcomes of other senior managers. *Organ. Sci.* **19** 457–474.

James, B. 1994. *The politics of Glory, How Baseball'S Hall of Fame Really Works*. Macmillan Publishing Company, New York.

Jiang, R. H. J., P. Bansal. 2003. Seeing the need for ISO 14001. *J. Management Stud.* **40** 1047–1067.

Johnson, J. L., A. E. Ellstrand, C. M. Daily, D. R. Dalton. 2005. Board reputation and stockholder wealth: The influence of the financial press. *Strategic Management J.* **26** 461–471.

Kahneman, D., A. Tversky. 1979. Prospect theory: An analysis of decisions under risk. *Econometrica* **47** 313–327.

LaMar, S. 1993. *The Book of Baseball Lists*. McFarland, London.

Lederer, R. 2004. A hall of fame chat with tracy ringolsby. *Baseball Analysts* (December 31). http://baseballanalysts.com/archives/2004/12/a_hall_of_fame_1.php.

Liao, T. F. 1994. *Interpreting Probability Models*: *Logit, Probit, and Other Generalized Linear Models*. Sage, Thousand Oaks, CA.

Malmendier, U., G. A. Tate. 2005. *Superstar CEOs.* 7th Annual Texas Finance Festival Paper. http://ssrn.com/abstract=709861.

Martins, L. 1998. The very visible hand of reputational rankings in US business schools. *Corporate Reputation Rev.* **1** 293–298.

McConnell, B. 2003. Career 300 batting averages. *Baseball Res. J.* (January 1).

Pfeffer, J., G. R. Salancik, H. Leblebici. 1976. The effect of uncertainty on the use of social influence in organizational decision making. *Admin. Sci. Quart.* **21** 227–245.

Podolny, J. M. 2005. *Status Signals*. Princeton University Press, Princeton, NJ.

Pollock, T. G., V. P. Rindova. 2003. Media legitimation effects in the market for initial public offerings. *Acad. Management J.* **46** 631–642.

Rao, H. 1994. The social construction of reputation: Certification contests, legitimation, and the survival of organizations in the American auto industry, 1895–1912. *Strategic Management J.* **15** 29–44.

Rindova, V. P., I. O Williamson, A. P. Petkova, J. M. Sever. 2005. Being good or being known: An empirical examination of the dimension, antecedents, and consequences of organizational reputation. *Acad. Management J.* **48** 1033–1050.

Scott, W. R. 1994. Institutional analysis: Variance and process theory approaches. W. R. Scott, J. R. Meyer and Associates, eds. *Institutional Environments and Organizations*: *Structural Complexity and Individualism.* Sage, San Francisco, 81–99.

Segev, E., A. Raveh, M. Farjoun. 1999. Conceptual maps of the leading MBA programs in the United States: Core courses, concentration areas and the ranking of the school, *Strategic Management J.* **20** 549–565.

Sine, W. D., R. J. David, H. Mitsuhashi. 2007. From plan to plant: Effects of certification on operational start-up in the emergent independent power sector. *Organ. Sci.* **18** 578–594.

Singh, J., D. Tucker, R. House. 1986. Organizational legitimacy and the liability of newness. *Admin. Sci. Quart.* **31** 171–193.

Skipper, J. C. 2000. *A biographical Dictionary of the Baseball Hall of Fame.* McFarland, London.

Terlaak, A., A. A. King. 2006. The effect of certification with the ISO 9000 quality management standard: A signaling approach. *J. Econom. Behav. Organ.* **60** 579–602.

Thompson, J. D. 1967. *Organizations in Action*. McGraw-Hill, New York.

Uzumeri, M. 1997. ISO 9000 and other metastandards: Principles for management practice? *Acad. Management Executive* **11** 21–36.

Wade, J., J. Porac, T. Pollock, S. Graffin. 2006. The burden of celebrity: The impact of CEO certification contests on CEO pay and performance. *Acad. Management J.* **49** 643–660.

Washington, M., E. J. Zajac. 2005. Status evolution and competition: Theory and evidence. *Acad. Management J.* **48** 281–296.

Weick, K. 1979. *The Social Psychology of Organizing*, 2nd ed. Addison-Wesley, Reading, MA.

Weick, K. 1995. *Sensemaking in Organizations*. Sage, Thousand Oaks, CA.

Wiley, M., M. Zald. 1968 The growth and transformation of educational accrediting agencies: An exploratory study of the social control of institutions. *Sociol. Ed.* **41** 36–56.

Wilson, R. 1985. Reputations in games and markets. A. E. Roth, ed. *Game-Theoretic Models of Bargaining*. Cambridge University Press, London, 27–62.

Zuckerman, E. W. 1999. The categorical imperative: Securities analysts and the illegitimacy discount. *Amer. J. Sociol.* **104** 1398–1438.