Evolution of Professional Certification Markets: Evidence from Field Experiments

Ginger Jin, Andrew Kato, and John A List*
University of Maryland and NBER

10 March 2004

Abstract

Market economies devote substantial resources to certify product quality. While the theoretical literature provides a rich assortment of equilibrium predictions on the informational role of certifiers, empirical investigation remains scant. This study uses field experiments to investigate issues related to an evolving market of professional certification. Via implementation of two field experiments we obtain several unique insights. First, casual observation suggests that the evolution of our chosen certification market—the sportscard grading industry—is consistent with theoretical predictions: the first entrant adopted a coarse grading scheme and subsequent entrants adopted finer grading systems. Second, even under the coarse grading system, the monopolist certifier added valuable information to the marketplace, a result that is inconsistent with theory. Yet it is important to note that the monopoly certification intermediary reveals no information to experienced market participants. Third, the second and third entrants in the industry sharpen grading precision and adopt finer grading cutoffs in an attempt to differentiate from the incumbent, a result consonant with theory. Finally, we find a consistent mapping between prevailing market prices and our empirically estimated grading cutoffs and signal precision, a result suggestive of a high degree of informational efficiency in this particular market.

JEL: D8 (Information and Uncertainty), C9 (Experiments)

Key words: decision-making under uncertainty, field experiment

Contact Information: Ginger Z. Jin, the University of Maryland, Department of Economics, College Park, MD 20742; telephone: (301) 405-3484; email: jin@econ.umd.edu; website: http://www.glue.umd.edu/~ginger/. John A. List, the University of Maryland and NBER, AREC, 2200 Symons Hall, College Park, MD 20742, telephone: (301) 405-1288, fax: (301) 314-9091; email: Jlist@arec.umd.edu; website: http://www.arec.umd.edu/jlist/.

---

*Thanks to the University of Maryland for providing funds to support this research. We are grateful to three sportscard dealers who kindly participated in the second field experiment and provided us insights regarding the sportscard industry. Rachel Croson, Liesl Koch, Marc Nerlove, Tigran Melkonyan, Michael Riordan, Kyle Bagwell, Christopher Mayer, Raymond Fisman, and Raphael Thomadson provided useful remarks and discussion on an earlier version of this paper. Seminar participants at the University of Maryland, Columbia University, and the ASSA meetings in San Diego also provided comments that helped shape the study. Any errors remain our own.

I. Introduction

When buyers lack information on product quality, independent certification is often proposed as a solution (Akerlof 1970). Along these lines, we observe Educational Testing Services (ETS) offering SAT tests for college applicants, U.S. News & World Report ranking universities, Underwriters Laboratories certifying consumer and industrial products, Moody's reporting corporate bond ratings, and accounting companies auditing financial reports for public corporations. While certification of product quality is ubiquitous, many important questions, both positive and normative, remain. For example, what are the appropriate incentives for private, for-profit certifiers to provide truthful and complete information? How well does the market for professional certification function and what principles govern its evolution? What role does competition play in the revelation of information? These questions have attracted theoretical attention, but empirical tests are rare.[1] We fill this void by using two field experiments to investigate an evolving market of professional certification.

In theory, an independent, for-profit certifier may not have sufficient incentive to reveal full information. For example, a monopoly certifier who commits to a uniform service fee may certify all applicants to maximize its grading revenue (Lizzeri 1999); an investment bank may release noisy stock evaluation in order to boost its own mutual funds (Admati and Pfleiderer 1990); and a university may adopt coarse and uninformative

[1] Numerous empirical studies have examined consumer response to information provided by government agencies (such as nutrition labeling), sellers (such as advertising), media (such as airline safety) or rating agencies (such as bond ratings). Other empirical studies have investigated the incentives sellers face in disclosure of information and quality enhancement when they are allowed (or mandated) to provide product information via government agency or independent certifiers. A theoretical literature in finance examines the certification role of investment banks, venture capitalists, and newsletter producers. In correspondence, a number of empirical studies link underwriter fee and stock prices with underwriter reputation or the presence of venture capital. To our best knowledge, no empirical study has examined the evolution of independent certification markets.

grades to market its mediocre students (Ostrovsky and Schwarz 2003). These equilibria are often contrasted with full information revelation, which many theorists argue should exist if the market for certification becomes sufficiently competitive.[2]

These theories, reasonable and intuitive in their own right, pose challenges for empirical tests. Not only do they differ a great deal in specific settings, but none of them specifies the evolutionary path from partial to full information revelation. Aside from the theoretical ambiguity, researchers face several empirical challenges. For instance, researchers are less informed than certifiers, the number of certifiers is often small and rarely changes, and most importantly, competing certifiers adopt different grading criteria and therefore generate sorting in the quality of products seeking certification. This implies that observational data alone might confound criteria differences and sorting effects.

To overcome these difficulties, we conduct controlled field experiments in the certification market for sportscards and memorabilia. Several features of this market make it particularly attractive for an empirical study. First, the nature of the commodity renders it difficult to detect authenticity and quality, thus experts represent an important component of the revelation process. In this sense, experts may not only provide a necessary signal for sellers (dealers) to overcome the asymmetric information problem, but also provide an informative signal to sellers. Second, there is a generally agreed upon set of traits for grading sportscards, and quality is a major determinant of price. Third, the industry is relatively young, and thus far has been unregulated. Thus, we are provided with a unique opportunity to examine the early evolution of an unfettered certification industry: the first grading service, PSA (Professional Sports Authenticators), began

[2] See Section III for a more detailed theoretical overview.

operations in 1987 and is now part of a publicly traded company. Subsequent major competitors entered the market in early 1999 (Sportscard Guaranty LLC (SGC)) and late 1999 (Beckett Grading Services (BGS)). All three services continue operating today.[3]

We make use of the market evolution in two ways: first, we conducted a field experiment in 1997 to examine the information content of PSA grading when PSA acted as a monopolist. We then carried out a second field experiment, in 2002, comparing the information content of PSA grades to that of the subsequent entrants SGC and BGS. The two experiments, and subsequent results, are directly comparable because PSA has not changed its grading criteria between 1997 and 2003 (indeed, its grading system has not changed since its inception).[4]

Specifically, the first field experiment took place at a sportscard show in 1997 where we (i) auctioned off 4 ungraded sportscards, (ii) purchased them back from the auction winners, (iii) had them graded by PSA, and (iii) auctioned them again as graded cards (on the same day). By comparing bidding distributions for the identical card—graded and ungraded—we can estimate whether PSA had any "information content" as the monopolist grader. The second field experiment was carried out in 2002. We had 216 sportscards graded by *all three* major graders—PSA, SGC, and BGS—as well as by three professional dealers. By making use of a random "round-robin" design, we ensure proper inference about the relative information content across all graders. Data gathered in the second field experiment allows us to estimate specific grading cutoffs and signal

[3] As discussed below, at least 14 other "fringe" grading companies have joined the market since 1999. Section II offers an explanation as to why the monopoly broke down in 1999.

[4] PSA never indicates when the certification was issued, and thousands of previously and newly graded copies are traded daily in the same market, forcing PSA to commit to one grading standard over time. This is not surprising considering the fact that PSA still holds the largest market share today and is often viewed as the industry standard. This evidence suggests that PSA has learned an important lesson from the coin market—one major coin certifier increased its grading upper bound from 60 to 64 in the 1970s, which generated a major market upset and was believed to contribute to the decline of coin trading afterwards.

precision, thus allowing a particularly stringent test of market efficiency.[5]

Several interesting insights emerge. First, casual observation suggests that the evolution of the sportscard grading industry is consistent with theoretical expectations, as PSA adopted a coarse grading scheme, while SGC and BGS utilized finer grading systems. Second, even under the coarse grading system, PSA—as the monopolist certifier—adds valuable information to the marketplace. However, the monopoly certifier reveals no information to experienced market participants. Third, SGC and BGS sharpen grading precision and adopt finer grading cutoffs in an attempt to differentiate from PSA. This change in grading adds important informational content to *all* market participants, which is beyond theoretical predictions as most theories assume perfect information on the sellers' side. Finally, we find a consistent mapping between market prices and our empirically estimated grading cutoffs and signal precision, which provides a robustness check of our methods and suggests that the market is "efficient" in the sense that it accurately values signals of the multiple certifiers.

The remainder of our study proceeds as follows. Section II provides a brief description of the sportscard certification market. Section III reviews the literature. Section IV discusses our experimental design and empirical results. Section V concludes.

II. Sportscard Grading

Each year, card companies design and print sets of cards depicting players and events from the previous season. Once the print run of a particular set has been completed, the supply of each distinct card in the set is fixed.[6] The value of a particular

[5] Our field experimental approach highlights the usefulness of controlled field experiments more generally. By combining the control afforded by an experiment with the realism of the field, we are able to overcome aforementioned difficulties associated with observational data while observing behavior in naturally occurring settings where the key theoretical factors are identifiable and arise endogenously.

[6] The exact number of copies printed for a specific card is regarded as an industry secret.

card depends on its scarcity, the player depicted, and the physical condition of the card—i.e., condition of the edges, corners, surface, and centering of the printing. To track card condition, people often use a 10-point scale. For example, a card with flawless characteristics under microscopic inspection would rate a perfect "10" while obvious defects to the naked eye, including minor wear on the corners, would decrease the card's grade to a "7". The card's overall grade is computed via the aggregation of the various characteristics, and post-1980 sportscards that merit a grade below "7" are rarely traded among serious collectors.[7]

PSA began offering grading services in 1987 and its parent company became publicly traded in 1999 (Collectors Universe, under Nasdaq ticker symbol CLCT). SGC entered the professional grading market in 1999, soon followed by BGS. As of 2002, PSA, BGS, and SGC remained the largest and most respected of the existing 10-15 grading services. We believe the breakdown of the PSA monopoly in 1999 is due partly to the onset of the Internet, as detailed in Jin and Kato (2003). In 1998, eBay, the most popular auction site for sportscard transactions, went public. The Internet not only substantially reduces transaction cost, but also intensifies the information asymmetry between buyers and sellers. To overcome the information problem, the demand for professional grading services considerably increased after 1998. The demand shock plus PSA's commitment to its initial grading criterion opened profitable opportunities for potential entrants.

Professional grading is voluntary and costs $6-$20 per card, depending on package size and requested turnaround time; further, the fee is independent of the actual

[7] Indeed, because grading is voluntary and costly, better quality cards are more likely to be graded. This is why very few post-1980 graded cards are ever observed in the 1 to 6 range, even though such grades exist and are given out when warranted. In practice, graded cards are usually "8" or above (Jin and Kato 2003a).

grade received. Graded cards are encased in plastic and sealed with a sonic procedure that makes it virtually impossible to open and reseal the case without evidence of tampering. The casing indicates the grading service, grade received, and a bar code with serial number that identifies the particular copy of the card. Anyone with Internet access can visit the grader's web site and verify the card's grade by serial number. Figure 1 provides an example of a PSA-graded 1985 Topps #401 Mark McGwire (*rookie*), an example of a BGS-graded 1993 Topps Traded #1T Barry Bonds, and an example of an SGC-graded 1991 Topps Tiffany #352 Ken Griffey Jr. *All Stars*.

PSA adopted integer grades from 1 to 10, whereas BGS adopted a slightly finer grading scheme, which included half grades from 1 to 10: 7.5, 8, 8.5, etc. SGC initially used a 100-point grading scale—e.g. 88, 92, 96—but soon provided equivalent conversion to a half-grade system similar to BGS, where 88 means 8, 92 means 8.5, 96 means 9 and 98 means 10. Interestingly, because SGC used only a limited number of grades in the original 100-point grading scale, the converted grades do not exhaust all possible half grades between 1 and 10. One curious omission is 9.5 – the converted SGC system has 7, 7.5, 8, 8.5, 9, and 10, but no 9.5. In comparison, the BGS scale includes all possible half grades, although BGS rarely gives a perfect grade of 10. Among the three certifiers, BGS is also the only one that offers sub-grades for centering, corner, edge and surface, in additional to the overall grade.

A casual comparison of grading scales suggests that the evolution of the sportscard grading industry is consistent with theoretical expectations: the first entrant, PSA, adopted a coarse grading scheme, the second entrant, SGC, adopted a finer scheme, and the third entrant, BGS, adopted an even finer grading scheme.

Another attractive feature of using the sportscard grading industry in our case study is that, whether buying or selling, all trading parties refer to a standard price guide for sportscards—*Beckett Baseball Cards Monthly* for baseball cards, *Beckett Football Cards Monthly* for football cards, etc. For each single type of ungraded card, Beckett collects pricing information from about 110 card dealers throughout the country and publishes a "high" and "low" price reflecting current selling ranges for Near Mint-Mint (8) copies. The high price represents the highest reported selling price and the low price represents the lowest price one could expect to find with extensive shopping. For graded cards, Beckett follows the same practice but lists price ranges for each grade level (usually 7 to 10) of frequently graded cards. When trading volume is high, Beckett reports separate prices for PSA, BGS, and SGC, and pools all other companies as "Others". Jin and Kato (2004) report that market-clearing prices of graded cards closely track the "low" price listed in the Beckett price guide. This particular market feature allows us to treat Beckett "low" prices as a proxy of market-clearing prices and to map them with our empirically estimated grading cutoffs.

III. Literature Review

The theoretical literature relevant to our study derives from two branches. Starting with Grossman (1981) and Milgrom (1981), the first branch examines how intermediaries induce the market to reach a state of full information. In a general setting of "middlemen," Biglaiser (1993) presents some guidelines on which markets benefit from expert intermediaries. Although the model does not exactly match the structure of the sportscard grading industry (because sportscard graders do not act as final retail sellers), the sportscard market matches the guidelines very well: a large proportion of low quality (ungraded) cards, a large price and quality difference between high and low quality ungraded cards, and quality that is difficult to evaluate and involves significant experience/human capital. Under such conditions, Biglaiser (1993) finds that middlemen who can evaluate quality better than the average buyer can be welfare improving.

The second branch examines the strategic incentives of informed intermediaries, who make decisions on the level of information revelation. In Admati and Pfleiderer (1990), an investment bank is informed of the value of risky investments and may sell the information directly via its financial analysis service, or indirectly via a bundle of investments assembled based on its private information. Though the bundling reduces the information's value by adding noise, the investment bank benefits by controlling the manner in which buyers use the information. Lizzeri (1999) examines a situation closely related to our field experiments: the informed intermediary is an independent certifier and is not involved in trade of the certified commodity. Under certain conditions, a monopoly for-profit certifier will choose an information disclosure rule that provides no information but maximizes the demand for certification.[8] In another example, educational institutes have incentives to adopt coarse grading (or inflate grading), if the gain for mediocre students outweighs the harm to good students (Ostrovsky and Schwarz 2003, Chan et al. 2003).

Intuitively, the equilibria of incomplete information revelation may evolve to full information if the intermediary market becomes sufficiently competitive. Following this conjecture, several models predict that competition could support the existence of full information equilibrium, but none of them specifies the evolutionary path from incomplete to complete information revelation. Biglaiser's (1993) analysis depends on fundamental assumptions about competition, namely that middlemen must have sufficient

[8] In some other conditions (e.g., the number of sellers is greater than the number of buyers), the monopoly certifier may design multiple certificates to extract maximal rents from sellers (Lizzeri 1999). In the sportscard industry, printers intentionally short print, so it is reasonable to assume the number of buyers is greater than the number of sellers. In that case, the one-certification equilibrium applies.

competition to induce honest quality reporting and that the larger the size of the market, the greater the benefits of having middlemen. Lizzeri (1999) shows that under carefully chosen buyer beliefs, full information revelation is an equilibrium when there are at least two competing information providers. Okuno-Fujiwara et al. (1992) present a situation in which insufficient variety in credible signals leads to incomplete information revelation, implying that more competition would result in fuller information revelation.

Overall, this rich assortment of studies provides two key predictions: first, in the absence of competition, a monopoly certifier may not reveal full information; second, competition in the certification industry should improve the information content of certificates.[9] In doing so, we naturally examine whether the certification industry reveals information to all market participants or just the uninformed agents. Moreover, we are able to measure the effects of the monopoly certification industry evolving to an oligopoly, and how the level of information varies with the nature of competition among certifiers. To our best knowledge, no previous study conducts these empirical tests.

IV. Experimental Design and Results

A. Field Experiment I

The goal of the first field experiment is to infer the information content of PSA grades when PSA acted as a monopolist certifier. Unlike the bond market, which has a plethora of information sources on most companies, a certifier in the sportscard industry may actually add information to the market. Alternatively, following an equilibrium

[9] Parallel to the theoretical literature, a substantial empirical literature examines whether bond ratings (such as Moody's) provide new information to the financial market. The evidence is inconclusive. Katz (1974), Grier and Katz (1976), and Hettenhouse and Sartoris (1976) found evidence that bond rating increases provided unanticipated information, but decreases did not. Hand et al. (1992) and Ederington and Goh (1998) and others have found the opposite: that bond rating decreases provided new information but increases did not. Finally, Pinches and Singleton (1978), Wakeman (1981), and Weinstein (1977) found no evidence that bond rating changes provided new information in either direction.

reported in Lizzeri (1999), PSA may provide buyers with minimal information on product quality.

To distinguish between these two possibilities, we carried out an experiment on the floor of a sportscard show located in a major Southern city in 1997 using four steps: (1) we auctioned 4 ungraded sportscards and determined the winner, (2) we purchased the cards back from the auction winners,[10] (3) we immediately had PSA grade the cards via their 1-hour, $50 per card, on-site grading system, and (4) we auctioned the same card as a graded variant. The entire procedure took place at the same card show in the morning or afternoon, allowing us to match the cards identically across the ungraded/graded treatment, and to control whatever factors might affect the demand for sportscards over time or across locations.[11]

Each participant's auction experience typically followed three steps: (1) inspecting the good, (2) learning the rules, and (3) concluding the transaction. In Step 1, a potential subject approached the experimenter's table and inquired about the sale of the sportscard displayed on the table. The experimenter then invited the potential subject to take about five minutes to participate in an auction for the sportscard displayed on the table. In Step 2, the subject learned the allocation rules. To perform the simplest possible test of the effect of information on bids, we chose an allocation mechanism−William Vickrey's (1961) second-price auction−which has proven straightforward in other field experiments (List 2001). And, to ensure that the graded and ungraded auctions could be run in the same few hours, we limited the number of participants to 30 in each auction.

[10] We were able to re-purchase all four of the ungraded cards from the auction winners at, or just above, the winner's bid.

[11] We also considered reversing the order (i.e., auctioning off graded cards, buying them back, cracking the seal, auctioning off the identical ungraded cards), but we wished to avoid inadvertently damaging the cards when cracking the seals, which would lead to incorrectly rejecting the null of a treatment effect because the ungraded card would not be the "identical" card of the graded card.

Finally, in Step 3 the subject filled out a survey (see the Appendix for the graded card survey), while the experimenter explained that the subject should return at the top of the hour to find out the results of the auction (in some cases the auction did not "clear" until the top of the next hour). If a subject did not return for the specified transaction time, she would be contacted and would receive her cards in the mail (postage paid by the experimenter) within three days of receipt of her payment. For each ungraded auction, we also asked the participating subject what PSA grade she thought the auctioned card would receive if it were graded.

Panel A of Table 1 summarizes our 4X2 field experimental design. Panel A can be read as follows: *n=30* at the intersection of row 1 column 1, denotes that we included 30 bidders in the Ripken 1982 *Topps* ungraded sportscard auction. For comparison, note that 30 different bidders competed in an auction for the *identical* Ripken 1982 *Topps* sportscard after it had been graded by PSA to be an "8".

We followed several steps to maintain experimental control. First, no subjects participated in more than one treatment. Second, if the individual agreed to participate, she could pick up and visually examine each card (in sealed cardholders, with the graded card condition clearly marked if they were participating in the graded auction). The experimenter worked one-on-one with the participant, and imposed no time limit on her inspection of the cards. Third, treatment type was changed at the top of each hour, so subjects' treatment type was determined based on the time they visited the table at the card show. To further control for temporal selection effects, the ungraded/graded auctions were paired so the bidding in any ungraded/graded pair took place in either the morning or the afternoon. Further, our dealer table was situated at the front of the card show and thus consumers entering the market were the auction participants.

Fourth, the sportscard market naturally includes subjects of varying experience. Thus, we can capture the distinction between those consumers that have intense market experience (dealers) and those that have less market experience (nondealers). Fifth, since our main interest revolves around examining individual willingness to pay, rather than testing the efficiency of the allocation mechanism, the experimenter informed the subjects of the optimal strategy (bidding true value) via several examples. Finally, in the ungraded card auction treatments, we had each subject submit an estimate of the card's grade on the survey (if it were to be PSA graded).

Results: Experiment I

We are particularly interested in whether PSA, acting as a monopoly certifier, provided any information in addition to what market participants already knew about card quality. If the answer is in the negative, the bidding distributions for the same card should be similar before and after PSA grading. If PSA did provide information, we expect the bidding distribution to change in at least two ways. First, if the PSA grade provides a more precise estimate of true card quality, the bidding distributions for graded cards should have lower variances. Second, if PSA grades the card higher (lower) than the market expects, the average bid should adjust upwards (downwards). Moreover, if market experience and evaluation skills are correlated, the PSA grade should provide more information to relatively inexperienced participants than to experienced ones, implying that the above two predictions should be more prominent for the less experienced consumers.

Panel A of Table 1 contains a summary of the auction data for field experiment I. In total, we observed bids from 240 subjects: 120 for ungraded cards and 120 for graded cards. Table 1 can be read as follows: row 1, column 1 shows that 30 bidders placed bids for the ungraded Ripken Jr. 1982 *Topps* card—consistent with the auctions for the other cards, we included 15 non-dealers and 15 dealers. The median bidder believed the card would grade at "PSA 7" if it were graded (s.d. = 2.5). The mean bid was $34.7 with a standard deviation of $32.2. Non-dealers bid on average $27.9 (s.d. = $40.9) and the median non-dealer believed the card would grade at PSA 7 if it were graded (s.d. = 3.3). Dealers bid on average $41.0 (s.d. = $20.6) and the median dealer believed the card would grade at PSA 8 if it were graded (s.d. = 0.6).

An interesting data pattern readily emerges: ubiquitously, the variances of bids in the ungraded auctions are much larger than the bid variances observed in the graded card auctions for the same sportscard. For example, for the Ripken Jr. card, the bid variance is $32.2 in the ungraded card auction and $17.2 in the graded card auction, a difference that is statistically significant at the $p < .05$ level using an F-test for homogeneity of variances. The three other bid variances are also significantly different at the $p < .05$ level.

We also find that the first moments tend to be influenced by the informational content of PSA grades. In this case, however, the bid differences in the graded and ungraded auctions depend on the beliefs about what grade the ungraded card would receive if it were graded. For the Ripken, Thomas, and Griffey cards, the graded card garnered considerably higher bids than the ungraded card (Ripken: $48.0 versus $34.7; Thomas: $90.0 versus $70.8; Griffey: $56.3 versus $41.0). All three differences are statistically significant at the $p < .05$ level using either a standard t-test of means or a Mann-Whitney rank-sum test of treatment differences.[12]

For the Sanders card, its PSA grade (7) was the same as the median dealer's

[12] The rank-sum test has a null hypothesis of no treatment effect, or that the two samples are derived from identical populations.

prediction but was lower than what the median non-dealer expected. Overall, the mean bid drops from \$34.3 to \$30.7, which is consistent with our conjecture that bids will adjust downwards if PSA grading indicates consumer overestimation of card quality. However, since the degree of consumer overestimation is not sufficiently large, the drop in the mean bid is not statistically significant at conventional levels.

To push the data a bit harder, we examine whether the information content influences market participants in a heterogeneous fashion. Panel A of Table 1 contains a summary of the bidding patterns split by dealers and nondealers. An interesting insight emerges: while the mean and variance of nondealers' bids are considerably influenced by the PSA information in every case, dealers are largely unaffected. For nondealers, both parametric and non-parametric Mann-Whitney tests suggest that the bid distributions observed across the graded and ungraded auctions are statistically different at the $p < .05$ level for the Ripken, Thomas, and Griffey card (no statistical significance is achieved for the Sanders card). Furthermore, the bid variances in all four of the graded auctions are significantly less than the bid variances in each of the ungraded auctions at the $p < .05$ level. Alternatively, neither the bid mean nor variance is significantly different across the graded and ungraded cards in the dealer data at conventional levels. Yet, it is important to note that the bid variance decreases for three of the four cards.

Our findings reveal the importance of subject experience levels in testing the theory: consistent with theory, the information does not influence experienced market players. Yet, the monopolist certifier provided information that is unknown to an average market participant.[13] Moreover, the results are in line with the literature that finds that

[13] If we split the non-dealers into "experienced" and "inexperienced" groups based on their number of years of market experience, we find tendencies in the predicted direction: experienced non-dealers are not influenced by the grade while the inexperienced group is influenced.

corporate bond prices are influenced by rating changes (e.g., Katz 1974; Grier and Katz 1976; Hettenhouse and Sartoris 1976). Given that the grading in the sportscard market revolves around detecting authenticity and quality, experts represent an important component of the information revelation process.

B. Field Experiment II

Our second field experiment complements field experiment I by exploring the actual *grading* patterns of both experienced sportscard dealers and professional certifiers. Examining the grading patterns of both experts in the field and professional grading services allows a strict test of whether, and to what extent, the various certifiers reveal information above and beyond what experienced dealers know. As aforementioned, because there is a generally agreed upon set of characteristics that determine the grade of a sportscard—printing defects, centering of the card's picture, corner wear, gloss of the card, color, picture focus, creases, gum and wax stains—and card quality is difficult to evaluate, sportscards present a naturally occurring market in which to explore differences in grading practices across expert sellers and independent certifiers.

We began field experiment II by equally distributing 216 sportscards into 9 groups in February 2002. Upon performing this grouping, we randomly allocated the cards first to the three sportscard dealers (Kevin, Rick, and Rodney) and then to the three certifiers (PSA, SGC, and BGS). Specifically, Kevin received groups A, B, C; Rick received groups D, E, F; and Rodney received groups G, H, K. Once all three dealers finished grading, we mailed groups A, D, G to PSA; B, E, H to BGS, and C, F, K to SGC for official grading. All certifiers returned the cards by April 29, 2002, which marked the end of Round 1. In the next two rounds, we rotated the cards to be graded by one of the other graders until all 6 graders had graded *each* of the 216 cards. Panel B of Table 1 presents the rotation details: each row represents a card group and each column represents one of the six graders.

The round-robin aspect of the experimental design is especially important for two reasons. First, each of the three professional certifiers places the graded card into a sonically sealed plastic casing upon certification and grading. To avoid confounding influences, when we received the graded cards from the certifiers, we recorded the card's grade and carefully chiseled off the plastic casing before re-sending the card to be graded by the other graders. Because the case is designed to prevent tampering, we may have inadvertently damaged the card. The round-robin rotation prevents one certifier from receiving systematically worse cards than another certifier. Indeed, we damaged 4 of the cards accidentally during the process; hence, our final data analysis uses 212 cards.

Second, for the three dealers who do not seal cards in plastic cases, grading entails physical handling. Although they are all experienced dealers and promised to handle the cards with care, there exists a chance that the grading process generated some minor damage to the cards. Such damage would upset future grades, but would not be easily detectable by even the trained eye. This fact represents the impetus for rotating the cards among dealers in such a way that even if the handling differed by dealer, each certifier on average faced the same distribution of card quality. Also note that in each round, dealer grading took place before certifier grading. In case dealers introduced an additional noise in card quality, we would capture it as part of a certifier's signal noise, thus *understating* the signal precision difference between certifiers and dealers. Since in the data we find that all certifiers are at least as precise as dealers, our conclusion is potentially strengthened.

Prior to moving to our empirical results, we should mention a few interesting aspects of our field design. First, none of the professional certifiers knew that we were running an experiment on the certification market and so they graded the cards under the assumption that they had been mailed to their company as “normal” cards to be graded. This was not a difficult task, as these three companies grade, on average, at least 10,000 cards per year. Nevertheless, when mailing the cards to each of the certifiers we took special precautions not to tip them off by using different consumer names and addresses in each round. Second, to ensure that this was a naturally occurring transaction, we paid the typical grading fee for PSA ($8), SGC ($6.5), and BGS ($9) to grade the cards, and we paid a flat-fee ($108) to our three dealers (whose requested fees were lower because they could grade the cards during slow times of the day at their retail shops). We were careful to choose professionals that had been shop owners in the sportscard market for several years and who had heterogeneous experience levels (Kevin: 8 years; Rick and Rodney: 14 years) to provide a demanding test of the professional certifiers.

Results: Field Experiment II

To complement the insights gained from field experiment I, our goals in examining the data from field experiment II are twofold: to estimate the grading cut-offs of each grader, and to identify the amount of noise in each grader's grading system. These empirical estimates allow us to compare directly grading criteria within certifiers and to detect any grading differences between certifiers and dealers.

Before discussing results from our full estimation model, we present summary statistics. Note that different graders may adopt different grading cutoffs, so the grades are ordinal and the raw grades are not readily comparable across graders (e.g., PSA 10 may not be equivalent to SGC 10). Moreover, because most grades are 8 or above and each grader has at most 5 possible grading categories at 8 or above (i.e., 8, 8.5, 9, 9.5, 10), a number of cards obtain identical grades from the same grader, thus creating ties. Inevitably, each grader has a lumpy distribution (see Table 2). Depending on how we order ties, the rank correlation of any two graders could be as low as 0.4 or as high as 0.9. For this reason, it is difficult to make sharp inferences from raw rank correlations.

To deal with these difficulties, we adopt an alternative approach. For any two cards randomly selected from the pool of 212 cards (call them A and B), we examine whether grader j and grader j' agree on their relative quality. If both j and j' agree that the quality of card A is superior to the quality of card B (i.e., $q_A > q_B$), we define the two graders as *strongly consistent* for this card pair. If grader j rated $q_A > q_B$ but grader j' rated $q_A < q_B$, they are *strongly inconsistent*. If one grader rated $q_A > q_B$ but the other rated $q_A = q_B$, they are *weakly inconsistent*. After finishing this comparison for all possible card pairs (22,366 in total), we computed the percentages in which grader j and grader j' are strongly consistent, strongly inconsistent, or weakly inconsistent. By repeating this exercise for every possible grader pair, we obtain three matrices in Table 3: panel A for strong consistency, panel B for strong inconsistency, and panel C for weak inconsistency. The three percentages, by definition, must sum to one in every cell.

Of particular interest is Panel B. The degree of strong inconsistency among professional certifiers is roughly 5%-7%, much lower than that among dealers (10%-13%), or that between professional certifiers and dealers (7%-13%). This suggests that professional certifiers, as a whole, are more compatible and more precise than dealers. Should all professional certifiers systematically miss some important component of card quality, the inconsistency between certifiers and dealers would have been much higher than that among dealers. In the last row, we compute the average strong inconsistency for each grader as compared to the other five. Among professional certifiers, it is clear that

BGS, the last entrant, achieves the highest level of consistency with the other certifiers, and that PSA, the first entrant, is the least in accord. All dealers are similar with, or noisier than, PSA. Again, this is roughly consistent with theoretical predictions: latter entrants should have the most informative grading schemes. Panel A in Table 3 offers similar insights: professional certifiers are more likely to be strongly consistent with each other than are certifiers with dealers, or dealers with dealers. Again, in terms of consistency, BGS is the sharpest and PSA is the least in accord.

While these summary statistics are suggestive, they do not provide explicit estimates of grading cutoffs or grading precision, and therefore do not offer a strict comparison across all graders. We attempt to overcome these shortcomings by examining the data in a full structural model. Suppose card i has an unknown quality q_i, $\forall i = 1,...212$. Grader j observes an unbiased noisy signal of q_i. The signal $s_{ij} = q_i + \varepsilon_{ij}$, where the iid noise $\varepsilon_{ij} \sim N(0, \sigma_j)$ and σ_j denotes the degree of noise in grader j's grading system. Internally, grader j has a set of cutoffs, such as J_8, J_9, J_{10}, etc. Once grader j observes signal s_{ij}, she fits the signal within those cutoffs and assigns corresponding grade g_{ij}. For example, if $J_8 \le s_{ij} < J_{8.5}$, then $g_{ij} = 8$.

Of course, we observe only the final grade $\{g_{ij}\}$. According to the raw grade distribution in Table 3, g_{ij} could be one of (7, 8, 9, 10) if grader j is PSA, (7.5, 8, 8.5, 9) if j is BGS, (7.5, 8, 8.5, 9, 10) if j is SGC, (7.5, 8, 8.5, 9, 9.5) if j is Kevin or Rodney, or (6, 7, 7.5, 8, 8.5, 9, 9.5) if j is Rick. Note that we do not observe any card receiving a BGS 9.5 or BGS 10, implying that the cutoffs for BGS 9.5 and BGS 10 are higher than any cutoff we can estimate from our data.

The unknown parameters are, therefore, true card qualities $\{q_i\}$, grading cutoffs

$\{J_g\}$, and signal precision $\{\sigma_j\}$. Defining a binary variable $1_{i,j,g}$ equal to 1 if grader j gave card i a grade of g, we have the overall likelihood function

$$\ln L = \sum_{i=1}^{212}\sum_{j=1}^{6}\sum_{g}\left\{1_{i,j,g}\cdot \ln\left[\Phi\left(\frac{J_{g+}-q_i}{\sigma_j}\right)-\Phi\left(\frac{J_g-q_i}{\sigma_j}\right)\right]\right\},$$

where Φ and J_{g+} denote the cdf of a standard normal and the cutoff directly above grade g within grader j's grading system. Our model is estimated via maximum likelihood. Note that only Rick and Rodney grade two cards under 7. To facilitate discussion, we treat these below-7 grades as 7 and focus on grading cutoffs of 7.5 or above. Results are robust without this simplification.

Identification Essentially, the true card quality $\{q_i\}$ is identified from within-card, cross-grader variation, similar to a "fixed effects" model, because every grader faces the same "truth." The signal noise $\{\sigma_j\}$ is identified from within-grader cross-card variation similar to a grader "fixed effect," because noise realizations are iid within each grader. Grading cutoffs $\{J_g\}$ are identified from a mapping between card quality and discrete grade, taking into account the fact that signals are noisy.

Although the idea of maximum likelihood is straightforward, we face three identification challenges. First, because every component of our likelihood function takes the form $\frac{J_x - q_i}{\sigma_j}$, the likelihood does not change if all three items are multiplied by the same constant. In response, we normalize $\sigma_{PSA} = 1$. For the same reason, nor does the likelihood change if we add a constant to both the cutoff and the true card quality. In response, we normalize $q_1 = 0$ (the order of the 212 cards is random).

The third identification problem lies in the ordinal nature of the grading data.

Since we do not know the true card quality, maximizing the overall likelihood function will always terminate with an order of quality that is perfectly consistent with the most agreeable grader. Regardless of the initial values, the maximization procedure always finds BGS to be this grader, implying that we cannot estimate σ_{BGS} and that BGS cutoffs are not uniquely identified. To see the latter, suppose the maximum quality of all BGS 8.5 cards is x, and the minimum quality of all BGS 9 cards is y; then the BGS 9 cutoff can be anywhere between x and y. To deal with this problem, we choose $(x+y)/2$ as the cutoff. Of course this choice is arbitrary, but it does not affect our qualitative comparison of graders in grading cutoffs and signal precisions.

Due to the eventual perfect fit with BGS, we maximize the overall likelihood sequentially. To be specific, we initiate q_i as the raw grade average across all six graders. Given $\{q_i\}$, we then search for the best grading cutoffs and signal noise $\{J_g, \sigma_j\}$ to maximize the likelihood function. Because the log likelihood function is additive and separable across graders, this amounts to a standard ordered probit for each grader. In our model, $s_{ij} = q_i + \varepsilon_{ij}$ imposes a coefficient restriction on q_i, which allows us to identify the magnitude of σ_j.[14] Once we obtain estimates on $\{J_g, \sigma_j\}$, we search for the optimal $\{q_i\}$ to maximize the overall likelihood. The iteration continues until both $\{q_i\}$ and $\{J_g, \sigma_j\}$ converge.

The sequential estimation does not provide correct standard errors, because step 2 does not take into account the standard errors of parameters passed along from step 1. To

[14] Specifically, for grader j, we estimate the ordered probit with index function $g_{ij}* = \beta_j \cdot q_i + e_{ij}$ where $e_{ij} \sim N(0,1)$. Converting it to $s_{ij} = q_i + \varepsilon_{ij}$ means that $\sigma_j = 1/\beta_j$. We obtain σ_j's standard error by the delta method, and all cutoffs are deflated by β_j.

solve this problem, we bootstrap the sample 100 times. Each bootstrap sample contains 2,120 cards randomly drawn from the original sample with replacement. The large sample size is chosen to ensure that all parameters remain valid in each random sample.

Panel A in Table 4 presents our estimation results of grading cutoffs and signal precision $\{J_g, \sigma_j\}$, along with bootstrapped standard errors. Panel B conducts statistical tests for comparable coefficients of BGS vs. PSA, SGC vs. PSA, and BGS vs. SGC, accounting for the bootstrapped variances and covariances. Note that true card qualities are incidental parameters and may not be consistently estimated due to the short length of our panel data. The structural parameters, $\{J_g, \sigma_j\}$, however, should be consistently estimated (Neyman and Scott 1948; Hsiao 1986; Hsiao 1991).

Consistent with theoretical models of certification evolution, the latest entrant – BGS – is the most consistent with the other graders, as the final results converge to a perfect fit with BGS grading. For the other two certifiers, the second entrant, SGC, is less noisier than the first entrant PSA ($\sigma_{SGC} < \sigma_{PSA}$), but the difference is not significant at conventional levels. The degree of noise is very close between PSA and the most experienced dealers (Rick and Rodney), while the least experienced dealer (Kevin) is much noisier than each of the other five graders. All of these findings are consonant with theoretical predictions.

Our results also identify grading cutoffs, on which theory remains silent. Comparing the first two certifiers, PSA and SGC, we find that their adopted cutoffs are quite similar on each of their common grade categories: SGC 10 is only marginally significant from PSA 10, SGC 9 is non-distinguishable from PSA 9, and SGC 7.5 is very close to PSA 8. The finer categories that SGC tends to add – SGC 8 and SGC 8.5 – are between PSA 8 and PSA 9. In contrast to these results, the third entrant, BGS, adopted a rather different strategy: it defined finer categories on the high end, as BGS 9 is between PSA 9 and PSA 10, but not close to either end, while BGS 9.5 and BGS 10 are certainly above PSA 10.

It is worth mentioning that, although SGC and BGS use finer scales than PSA, the whole system encompassing all three certifiers is much finer than any of the three alone. This suggests that, although new entrants may attract some business away from the incumbent, they do not replace the existing grading system. Rather, they add value to the whole industry. In response, facing multiple (noisy) certification systems, a seller can strategically maximize the grade of a specific card quite easily. For example, he could send the card first to BGS, crack it open and resend it to PSA if the BGS grade is lower than 9.5, crack open the PSA case if the PSA grade is less than 10, and try it again with SGC. Of course, this practice will stop at some point when the cost of repeated grading is too high. The result of such practice is not uncommon in other industries, however: at least 15 MBA programs claim to be in the top 10, and multiple producers within the same industry claim to have the single best quality.

Mapping with price data One interesting detail thus far not discussed is whether the market is efficient in parsing the signals from the professional certifiers, especially considering that each individual seller has *less* precise information than any certifier on any single good. In the long run, accurate learning is possible if there is a significant group of dealers who make use of card grading frequently. Even if on each single item a dealer's signal is less precise than that of certifiers, each certifier uses the same criteria on different cards and, therefore, the dealer can learn about the grading criteria via the law of large numbers. If this logic is correct, then we should observe a consistent mapping between the ordering of various grading cutoffs and corresponding market prices.

We empirically examine this relationship as follows. We take the Beckett "low" book price as a proxy of market-clearing price, because as aforementioned Jin and Kato (2004) have shown a close relationship between market transaction price and the Beckett "low" price for various types of baseball cards. Our price sample consists of 32 baseball cards that were similar to our experimental cards (i.e., identical technologies), and have detailed book prices by grade and certifier.[15] We use Beckett guides dated February 2002–October 2003 to maximize sample size. Defining the unit of observation as card-certifier-grade, we have 2,022 observations in total, and all available grades are 8 or above. To deal with demand changes across cards and over time, we deflate each price by the PSA 8 price of the same card in the same month. So a deflated price of 2 should be interpreted as 200 percent of its benchmark price. We then compute the average of deflated prices by grade and certifier.[16]

Figure 2 plots grading cutoffs in the upper panel and contrasts them with the average deflated prices in the lower panel. In the upper panel, the horizontal axis is the grading cutoffs estimated in the full model, and the vertical axis is the grading scale ranging from 7 to 10. Each vertical line in the graph denotes the grading cutoff for a specific grade and a specific certifier. To distinguish among certifiers, we use blue lines

[15] The card identities are 1989 Upper Deck #1 Ken Griffey Jr., 1989 Upper Deck #25 Randy Johnson, 1990 Leaf #220 Sammy Sosa, 1990 Leaf #300 Frank Thomas, 1990 Upper Deck #17 Sammy Sosa, 1991 Bowman #569 Chipper, 1991 Upper Deck Final Edition 2F Pedro Martinez, 1992 Bowman #82 Pedro Martinez, 1992 Bowman #461 Mike Piazza, 1992 Bowman #532 M. Ramirez, 1993 Bowman #511 Derek Jeter, 1994 Upper Deck #24 Alex Rodriguez, 1995 Bowman's Best #B2 Vlad Guerrero, 1995 Bowman's Best #B7 A. Jones, 1998 Fleer Tradition Update #U87 T. Glaus, 1998 Fleer Tradition Update #U100 Drew, 1999 Bowman #350 A. Soriano, 1999 Fleer Tradition Update U5 A. Soriano, 1999 Topps Traded T65 A. Soriano, 1991 Upper Deck Final #17F Thome, 1999 Upper Deck Ultimate Victory #136 A. Soriano, 2001 SP Authentic #211 Prior, 2001 SP Authentic #212 Teixeira, 2001 SP Authentic #91 Ichiro Isuzu, 2001 SP Authentic #126 Pujols, 2001 Upper Deck Victory #564 Ichiro, 2001 Bowman #254 Pujols, 2001 SPx #206 Pujols, 2001 Upper Deck #295 Pujols, 2001 Upper Deck Sw Spt #121 Pujols, and 2001 Upper Deck Sw Spt #139 Prior.

[16] Regression analysis controlling for card type and time trend yields the same rank of prices; hence our discussion focuses on the raw averages rather than on regression coefficients.

for PSA, black lines for SGC, and pink lines for BGS. In the lower panel, the horizontal axis is the deflated prices (interpreted as multiples of PSA 8 price) and the vertical axis is the grading scale from 7 to 10. The observed price schedule is a convex, increasing function of grade within each certifier – BGS 9.5 is priced as high as 12.26 times the benchmark price, while that number drops to 2.79 for BGS 9, 1.336 for BGS 8.5, and 1.022 for BGS 8. This confirms the industry understanding that the main action in card grading is to seek a grade at the very high end.

Focusing on ranks, we find that the ordering of grading cutoffs is consistent with the price order. Comparing PSA versus BGS, we find that both cutoffs and prices have $BGS9.5 > PSA10 > BGS9 > PSA9 > BGS8.5 > BGS8 > PSA8$. The relative position of SGC grades at the high end is also consistent: the cutoff (and price) of SGC 10 is less than PSA 10. The only inconsistency between the two panels is that SGC is usually priced significantly lower than PSA at the same grade, but their cutoffs are not statistically different. This result could be due to our small sample sizes, or due to the first mover advantage that PSA has over latter entrants. BGS is better able to overcome this disadvantage, either due to its superior name recognition from sharing the price guide's name, or due to its strategy of differentiating grading categories on the high end.

V. Concluding Comments

The evolution of certification markets merits serious consideration. Theoretically, the role and evolution of the professional certification market represents a rich assortment of equilibrium predictions. Yet formal empirical testing of the theory is rare. We fill this gap by making use of two distinct field experiments. Several insights are obtained. First, the evolution of our chosen certification market—the sportscard grading industry—is consistent with theoretical predictions: the first entrant adopted a coarse grading scheme and subsequent entrants adopted finer grading systems. Second, even under the coarse grading system, the monopolist certifier added valuable information to the marketplace, a result that is inconsistent with theory. Yet it is important to note that the monopoly certification intermediary reveals little information to dealers. Third, the second and third entrants in the industry sharpen grading precision and adopt finer grading cutoffs in an attempt to differentiate from the incumbent. Finally, we find a consistent matching between market prices and our empirically estimated grading cutoffs and signal precision, suggesting the efficiency of the sportscard market.

Besides providing a test of the important theoretical conjectures, these findings raise several additional questions/insights that are not explored in the extant theoretical literature. First, most theories assume that sellers possess perfect information about their own products and that therefore the only role of professional certification is to bridge the informational gap between buyers and sellers. Our results suggest that some certifiers provide better information to *all* trading participants, including sellers. This raises the question of why certifiers have incentives to do so and whether it is necessary to certify the certifiers. We believe high volume dealers could police independent certifiers by learning about the grading criteria via the law of large numbers. These high volume dealers, however, may not act as certifiers themselves because of the possibility that their impartiality could be compromised by their motives to make sales.[17]

These findings also have an important relationship to the literature in finance. Since the seminal work of Leland and Pyle (1977), financial institutions such as investment banks are believed to play a role in "certifying"[18] the quality of firms going

[17] This represents a typical problem in an expert market (see Biglaiser 1993; Biglaiser and Friedman 1994, Albano and Lizzeri 2001, and Pesendorfer and Wolinsky 2003 for detailed theoretical arguments.

[18] Strictly speaking, financial institutions do not fall into the exact definition of independent certifiers because the role of intermediary coexists with other incentives. For example, investment banks charge fees

public. Both theoretical and empirical literatures show that investment banks have incentives to establish a good reputation and that reputable banks are rewarded by higher underwriting fees and higher IPO stock prices (Beatty and Ritter 1986, Carter and Manaster 1990, Chemmanur and Fulghieri 1994, Johnson and Miller 1988, Carter et al. 1998). These intermediaries face a natural police: the true quality of certified projects will be revealed to the public via future stock prices. Unfortunately, such a natural policing mechanism is not always present in markets where professional certification is important and necessary. Our results are comforting in the sense that, in the absence of an obvious policing mechanism, the market of professional certification functions well.

An important normative consideration is that new entrants in a professional certification market provide both benefits and costs, and therefore may not unequivocally be welfare-improving. The benefits arise from better information content embedded in the entrants' grading scales that are often finer and differentiated. Given that there is a fair amount of noise in the new and old grading systems, however, the increased competition in the certification industry might generate incentives for repeated grading, which possibly results in duplicate and excessive certification. Another cost lies in learning the market positioning of the new grader—for every new certifier, the market not only needs to learn its grading criteria, but also must determine the relative position of the newcomer's grading scale to that of all existing certifiers. Since each individual often has less information than any one certifier, this learning process could be long and costly. On this front, any normative model would require more formal theoretical structure.

---

for marketing equities, and the amount of such fees may depend on equity sales. Venture capitalists are equity holders and play an active role in monitoring or managing the project. These settings generate moral hazard or agency incentives that do not exist in a market with pure certifiers as set up in this paper.

References

Admati, Anat and Paul Pfleiderer (1990):"Direct and Indirect Sale of Information," *Econometrica*, 58(4): 901-928.

Akerlof, George (1970): "The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism", *Quarterly Journal of Economics* 84: 488-500.

Albano, Gianluigi and Alessandro Lizzeri (2001): "Strategic Certification and the Provision of Quality," *International Economic Review*, February 2001.

Beatty, R. and Ritter, J. "Investment Banking, Reputation and Underpricing of Initial Public Offerings." *Journal of Financial Economics*, 1986, 15, pp. 213-232.

Biglaiser, Gary (1993): "Middlemen as Experts," *The RAND Journal of Economics*, 24, 212-223.

Biglaiser, Gary and James W. Friedman (1994): "Middlemen as Guarantors of Quality" (1994), *The International Journal of Industrial Organization*, 12, 509-531.

Carter, R.B. and Manaster, S. (1990): "Initial Public Offerings and Underwriter Reputation." *The Journal of Finance*, September 1990, 45 (4), pp. 1045-1068.

Carter, R.B.; Dark, F.H. and Singh, A.K. "Underwriter Reputation, Initial Returns and Long Run Performance of IPO Stocks." *The Journal of Finance*, March 1998, 53(1), pp. 285-311.

Chan, William; Li, Hao and Suen, Wing (2003): "Why Are Grades So High?" University of Toronto Working Paper.

Chemmanur, Thomas J. and P. Fulghieri (1994) "Investment Bank Reputation, Information Production, and Financial Intermediation," *The Journal of Finance*, March 1994.

Ederington, Louis H. and Jeremy C. Goh (1998) : "Bond Rating Agencies and Stock Analysts: Who Knows What?," *The Journal of Financial and Quantitative Analysis*," 33(4): 569-585.

Grier, Paul and Steven Katz (1976): "The Differential Effects of Bond Rating Changes Among Industrial Public Utility Bonds by Maturity," *Journal of Business* 49: 226-239.

Grossman, Sanford (1981): "The informational Role of Warranties and Private Disclosure about Product Quality." *Journal of Law and Economics*, 24: 461-489.

Hand, J., R. Holthausen, and R. Leftwich (1992): "The Effect of Bond Rating Agency Announcements on Bond and Stock Prices," *The Journal of Finance* 57: 733-752.

Hettenhouse, George W. and William L. Sartoris (1976): "An Analysis of the Informational Value of Bond-Rating Changes," *Quarterly Review of Economics & Business*, Summer 1976, Vol. 16: 65-78.

Hsiao, Cheng (1991) "Identification and Estimation of Dichotomous Latent Variables Models Using Panel Data" *The Review of Economic Studies*, 58(4): 717-731.

Hsiao, Cheng (1986) *Analysis of Panel Data*, Cambridge University Press, 1986.

Jin, Ginger Z. and Andrew Kato (2004): "Consumer Frauds and the Uninformed: Evidence From an Online Experiment," *working paper.*

Jin, Ginger Z. and Andrew Kato (2003): "Dividing Online and Offline: A Case Study," *working paper*.

Johnson, J. and Miller, R. "Investment Banker Prestige and the Underpricing of Initial Public Offerings." *Financial Management,* 1988, 17(2), pp. 19-29.

Katz, Steven (1974): "The Price and Adjustment of Bonds to Rating Reclassifications: A Test of Bond Market Efficiency," *The Journal of Finance*, 29(2): 551-559.

Leland, Hayne and David Pyle (1977). ”Informational Asymmetries, Financial Structure and Financial Intermediation”, *The Journal of Finance*, 32, 371-387.

List, John A. (2001): “Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards,” *American Economic Review* (2001), 91[5], pp. 1498-1507.

Lizzeri, Alessandro (1999): “Information Revelation and Certification Intermediaries,” *The RAND Journal of Economics*, Summer 1999.

Milgrom, Paul R. (1981): “Good News and Bad News: Representation Theorems and Applications”, *The Bell Journal of Economics*, 12: 380-391.

Neyman, J. and Elizabeth L. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations" *Econometrica,* 16(1):1-32.

Okuno-Fujiwara, Masahiro, Andrew Postlewaite, and Kotaro Suzumura (1992): “Strategic Information Revelation,” *The Review of Economic Studies*, 57(1): 25-47.

Ostrovsky, Michael and Michael Schwarz (2003): “Equilibrium Information Disclosure: Grade Inflation and Unraveling,” Harvard Business School Working Paper.

Pesendorfer, Wolfgang and Asher Wolinsky (2003): “Second Opinions and Price Competition: Inefficiency in the Market for Expert Advice," *Review of Economic Studies*, April 2003.

Pinches, George E. and J. Clay Singleton (1978): “The Adjustment of Stock Prices to Bond Rating Changes,” *The Journal of Finance*, 33(1): 29-44.

Vickrey, William (1961): “Counterspeculation, Auctions, and Competitive Sealed Tenders.” *Journal of Finance*, 1961, 16(1), pp. 8-37.

Weinstein, Mark (1977): “The Effect of A Rating Change Announcement on Bond Price” *Journal of Financial Economics* 5: 29-44.

Appendix: Subject Instructions for Graded Card Vickrey Auction

Welcome to Lister's Auctions. You have the opportunity to bid in an auction for the baseball card on the table.

The card up for auction is the 1982 Cal Ripken Jr. *Topps Traded PSA 8* on the table.

Auction Rules:

A sealed bid second-price auction will be used to determine the winner of each item. Thus, if your bid of $X is the highest bid and the next highest bid is $X-5, you win the card but only pay $X-5. Under this bidding mechanism it is best for you to bid your true value because overbidding may cause you to pay too much and underbidding decreases your odds of winning the item.

I will accept 30 bids in this auction and therefore will clear the auction at the top of the next hour. I will order the bids from highest to lowest in order to determine the winner of the card.

For example, if the top four bids are ranked highest to lowest as follows:

$A
$B
$C
$D

The bidder who bid **$A** wins the card and pays **$B**.

Given that the winner of the card will pay a price equal to the amount of the second-highest bid, please place your bid below:

Cal Ripken Jr. 1982 Topps "Traded" NRMT PSA 8 $_____________

After the winner pays me cash or check for the card, the card will be awarded to the winner (we pay postage). Please sign the line below to verify your bids. Also, please provide your name, telephone number and mailing address below:

Signature________________________________

Name___________________________________

Address____

Phone#__________________________________

We now want to ask you a few more questions. These questions will be used for statistical purposes only. THIS INFORMATION WILL BE KEPT STRICTLY CONFIDENTIAL AND WILL BE DESTROYED UPON COMPLETION OF THE STUDY.

1. How long have you been dealing with sportscards and memorabilia? _______yrs

2. If you are a dealer, how long have you been an active dealer? _______yrs

3. Gender: 1) Male 2) Female

4. Age ______ Date of Birth ____________

5. What is the highest grade of education that you have completed. (Circle one)

   1) Eighth grade 3) 2-Year College 5) 4-Year College
   2) High School 4) Other Post-High School 6) Graduate School Education

6. What is your approximate yearly income from all sources, before taxes?

   1) Less than $10,000 5) $40,000 to $49,999
   2) $10,000 to $19,999 6) $50,000 to $74,999
   3) $20,000 to $29,999 7) $75,000 to $99,999
   4) $30,000 to $39,999 8) $100,000 or over

Table 1. Experimental Design

Panel A: Field Experiment I – ungraded/graded auction

Card Type	Ungraded	Graded
Ripken Jr. 1982 *Topps*	n=30 (PSA 7; 2.5) Bid = $34.7 (32.2) Non-dealer bid = $27.9 (40.9) (PSA 7; 3.3) Dealer bid = $41.0 (20.6) (PSA 8; 0.6)	n=30 (PSA 8) Bid= $48.0 (17.2) Non-dealer bid = $51.7 (13.0) Dealer bid = $44.3 (20.3)
Sanders 1989 *Score*	n=30 (PSA 7; 2.2) Bid = $34.3 (32.3) Non-dealer bid = $44.3 (40.8) (PSA 8; 3.0) Dealer bid = $22.0 (15.2) (PSA 7; 1.1)	n=30 (PSA 7) Bid= $30.7 (22.5) Non-dealer bid = $40.2 (24.5) Dealer bid = $21.1 (15.9)
Thomas 1990 *Leaf*	n=30 (PSA 8; 2.3) Bid = $70.8 (43.4) Non-dealer bid = $66.3 (53.5) (PSA 7; 3.2) Dealer bid = $75.3 (31.4) (PSA 8; 0.8)	n=30 (PSA 9) Bid= $90.0 (22.3) Non-dealer bid = $96.9 (21.4) Dealer bid = $83.0 (21.7)
Griffey Jr. 1989 *Upper Deck*	n=30 (PSA 7.5; 2.8) Bid = $41.0 (35.9) Non-dealer bid = $36.7 (47.8) (PSA 5.5; 3.5) Dealer bid = $45.3 (18.7) (PSA 8; 0.8)	n=30 (PSA 8) Bid= $56.3 (22.3) Non-dealer bid = $65.0 (24.6) Dealer bid = $47.6 (16.2)

Notes: Row 1, column 1 shows that 30 bidders placed bids for the ungraded Ripken Jr. 1982 *Topps* card. The median bidder believed the card would grade at PSA 7 if it was graded (s.d. = 2.5). Mean bid was $34.7 (s.d. = 32.2). Non-dealers bid on average $27.9 (s.d. = $40.9) and the median non-dealer believed the card would grade at PSA 7 if it was graded (s.d. = 3.3). Dealers bid on average $41.0 (s.d. = $20.6) and the median dealer believed the card would grade at PSA 8 if it was graded (s.d. = 0.6). Each auction had 15 non-dealers and 15 dealers.

Panel B: Field Experiment II – the round-robin design

Total 216 Cards	PSA	SGC	BGS	Kevin	Rick	Rodney
Card Group A	Round 1 Step 2	Round 2 Step 2	Round 3 Step 2	Round 1 Step 1	Round 3 Step 1	Round 2 Step 1
Card Group B	Round 2 Step 2	Round 3 Step 2	Round 1 Step 2	Round 1 Step 1	Round 3 Step 1	Round 2 Step 1
Card Group C	Round 3 Step 2	Round 1 Step 2	Round 2 Step 2	Round 1 Step 1	Round 3 Step 1	Round 2 Step 1
Card Group D	Round 1 Step 2	Round 2 Step 2	Round 3 Step 2	Round 2 Step 1	Round 1 Step 1	Round 3 Step 1
Card Group E	Round 2 Step 2	Round 3 Step 2	Round 1 Step 2	Round 2 Step 1	Round 1 Step 1	Round 3 Step 1
Card Group F	Round 3 Step 2	Round 1 Step 2	Round 2 Step 2	Round 2 Step 1	Round 1 Step 1	Round 3 Step 1
Card Group G	Round 1 Step 2	Round 2 Step 2	Round 3 Step 2	Round 3 Step 1	Round 2 Step 1	Round 1 Step 1
Card Group H	Round 2 Step 2	Round 3 Step 2	Round 1 Step 2	Round 3 Step 1	Round 2 Step 1	Round 1 Step 1
Card Group K	Round 3 Step 2	Round 1 Step 2	Round 2 Step 2	Round 3 Step 1	Round 2 Step 1	Round 1 Step 1

Notes: Round 1 in blue, Round 2 in black, and Round 3 in pink. The total number of cards in use is 216. Four of them were damaged, so the final sample size is 212.

Table 2. Field Experiment II: Grade Distribution by Grader

	PSA	BGS	SGC	KEVIN	RICK	RODNEY
4	0	0	0	0	1	0
4.5		0		0	0	0
5	0	0	0	0	0	0
5.5		0	0	0	0	0
6	0	0	0	0	1	2
6.5		0		0	0	0
7	1	2	2	1	2	0
7.5		3	3	4	3	2
8	66	43	11	37	45	25
8.5		124	49	129	92	62
9	134	40	134	40	57	120
9.5		0		1	11	1
10	11	0	13	0	0	0
Total	212	212	212	212	212	212

Notes: Each cell represents frequency. Blank means the grade is not applicable to the grader.

Table 3. Summary Statistics by Degree of Consistency

Panel A: % strongly consistent (both graders said A>B, A=B or A<B)

	psa	bgs	sgc	kevin	rick	rodney
PSA	1.000					
BGS	0.491	1.000				
SGC	0.537	0.465	1.000			
Kevin	0.409	0.399	0.418	1.000		
Rick	0.377	0.492	0.414	0.402	1.000	
Rodney	0.408	0.492	0.475	0.428	0.429	1.000
sum (except self)	2.223	2.339	2.308	2.057	2.114	2.232
average (except self)	0.445	0.468	0.462	0.411	0.423	0.446
Ranks by average	4	1	2	6	5	3

Panel B: % strongly inconsistent (one grader said A>B, and the other said A<B)

	psa	bgs	sgc	kevin	rick	rodney
PSA	0.000					
BGS	0.059	0.000				
SGC	0.053	0.070	0.000			
Kevin	0.111	0.109	0.100	0.000		
Rick	0.130	0.089	0.109	0.131	0.000	
Rodney	0.111	0.069	0.091	0.103	0.118	0.000
sum (except self)	0.463	0.396	0.423	0.554	0.577	0.492
average (except self)	0.093	0.079	0.085	0.111	0.115	0.098
Ranks by average	3	1	2	5	6	4

Panel C: % weakly inconsistent (one grader said A=B and the other said A>B or A<B)

	psa	bgs	sgc	kevin	rick	rodney
PSA	0.000					
BGS	0.450	0.000				
SGC	0.411	0.465	0.000			
Kevin	0.480	0.492	0.482	0.000		
Rick	0.493	0.419	0.478	0.467	0.000	
Rodney	0.481	0.438	0.435	0.469	0.453	0.000
sum (except self)	2.314	2.265	2.269	2.389	2.309	2.276
average (except self)	0.463	0.453	0.454	0.478	0.462	0.455
Ranks by average	5	1	2	6	4	3

Table 4. Field Experiment II: Full Model Estimation

Panel A: Estimates

	PSA		BGS		SGC		KEVIN		RICK		RODNEY	
	coeff.	std.err.	coeff.	std.err.	coeff.	std.err.	coeff.	std.err.	coeff.	std.err.	coeff.	std.err.
σ	1	0	0	0.092	0.855	0.141	1.843	0.222	0.952	0.089	0.977	0.089
cutoff 7.5			-2.274	0.624	-3.453	0.536	-5.727	0.922	-3.101	0.558	-3.602	0.564
cutoff 8	-3.990	0.559	-2.058	0.589	-2.923	0.513	-4.519	0.784	-2.788	0.549	-3.201	0.551
cutoff 8.5			-1.061	0.456	-2.194	0.494	-2.197	0.580	-1.343	0.489	-1.890	0.502
cutoff 9	-1.092	0.482	0.229	0.500	-1.072	0.484	1.252	0.466	0.101	0.484	-0.679	0.485
cutoff 9.5							4.897	0.668	1.544	0.510	2.688	0.563
cutoff 10	1.605	0.491			1.313	0.533						

Note: (1) Only Rick and Rodney give two grades below 7. To facilitate comparison, we group them as 7 so we only identify cutoffs for 7.5 or above. This simplification does not change any conclusion.
(2) The grading noise of PSA (σ_{PSA}) is normalized as 1 and the true quality of the first card (q_1) is normalized as 0.
(3) Coefficient estimates are based on the original sample. Blank cells indicate non-applicable. Standard errors are based on 100 bootstrapped samples, each containing 2,120 cards randomly drawn from the original sample with replacement.
(4) In the original sample, estimation converges to a perfect fit with BGS, and therefore $\sigma_{BGS} = 0$. In bootstrapping samples, 99 out of 100 converge to a perfect fit with BGS as well. The only sample that does not converge to a perfect fit with BGS converges to point estimates $\sigma_{BGS} = 0.134$, $\sigma_{SGC} = 0.95$, and $\sigma_{PSA} = 1$. This non-zero σ_{BGS} contributes to the standard error of σ_{BGS}.

Table 4. Panel B: Tests for significant difference across coefficients

BGS vs. PSA

	σ of PSA	PSA 8 cutoff	PSA 9 cutoff	PSA 10 cutoff
σ of BGS	-1.000 *** (0.092)			
BGS 7.5 cutoff		1.716 *** (0.248)	-1.182 *** (0.237)	-3.878 *** (0.267)
BGS 8 cutoff		1.932 *** (0.226)	-0.966 *** (0.199)	-3.663 *** (0.237)
BGS 8.5 cutoff		2.929 *** (0.226)	0.031 (0.101)	-2.666 *** (0.140)
BGS 9 cutoff		4.219 *** (0.230)	1.321 *** (0.102)	-1.376 *** (0.099)

SGC vs. PSA

	σ of PSA	PSA 8 cutoff	PSA 9 cutoff	PSA 10 cutoff
σ of SGC	-0.145 (0.141)			
SGC 7.5 cutoff		0.537 (0.325)	-2.361 *** (0.299)	-5.508 *** (0.344)
SGC 8 cutoff		1.067 *** (0.283)	-1.832 *** (0.191)	-4.528 *** (0.228)
SGC 8.5 cutoff		1.796 *** (0.241)	-1.102 *** (0.174)	-3.799 *** (0.223)
SGC 9 cutoff		2.918 *** (0.220)	0.020 (0.092)	-2.677 *** (0.133)
SGC 10 cutoff		5.303 *** (0.298)	2.405 *** (0.191)	-0.293 * (0.158)

SGC vs. BGS

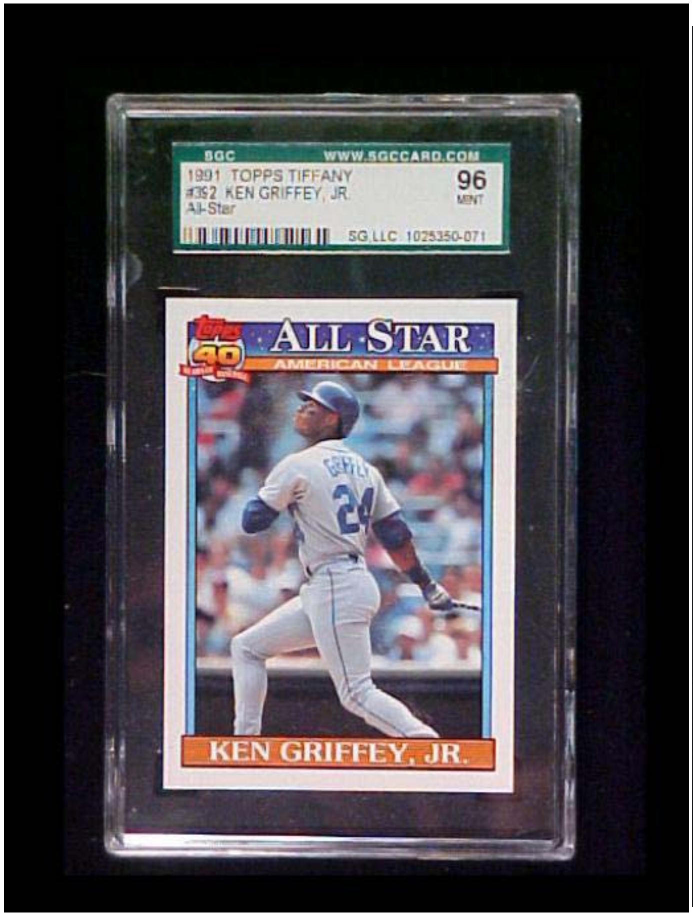
	σ of BGS	BGS 7.5 cutoff	BGS 8 cutoff	BGS 8.5 cutoff	BGS 9 cutoff
σ of SGC	0.855 *** (0.141)				
SGC 7.5 cutoff		-1.179 *** (0.335)	-1.394 *** (0.304)	-2.392 *** (0.310)	-3.682 *** (0.321)
SGC 8 cutoff		-0.649 * (0.363)	-0.864 *** (0.336)	-1.862 *** (0.333)	-3.152 *** (0.342)
SGC 8.5 cutoff		0.080 (0.260)	-0.136 (0.219)	-1.133 *** (0.187)	-2.243 *** (0.198)
SGC 9 cutoff		1.202 *** (0.231)	0.986 *** (0.190)	-0.011 (0.112)	-1.301 *** (0.116)
SGC 10 cutoff		3.587 *** (0.309)	3.370 *** (0.289)	2.347 *** (0.202)	1.084 *** (0.182)

Note: All comparisons take into account bootstrapping variances and covariances.
***p<0.01, ** p<0.05, * p<0.1.

Figure 1. Examples of Graded Cards

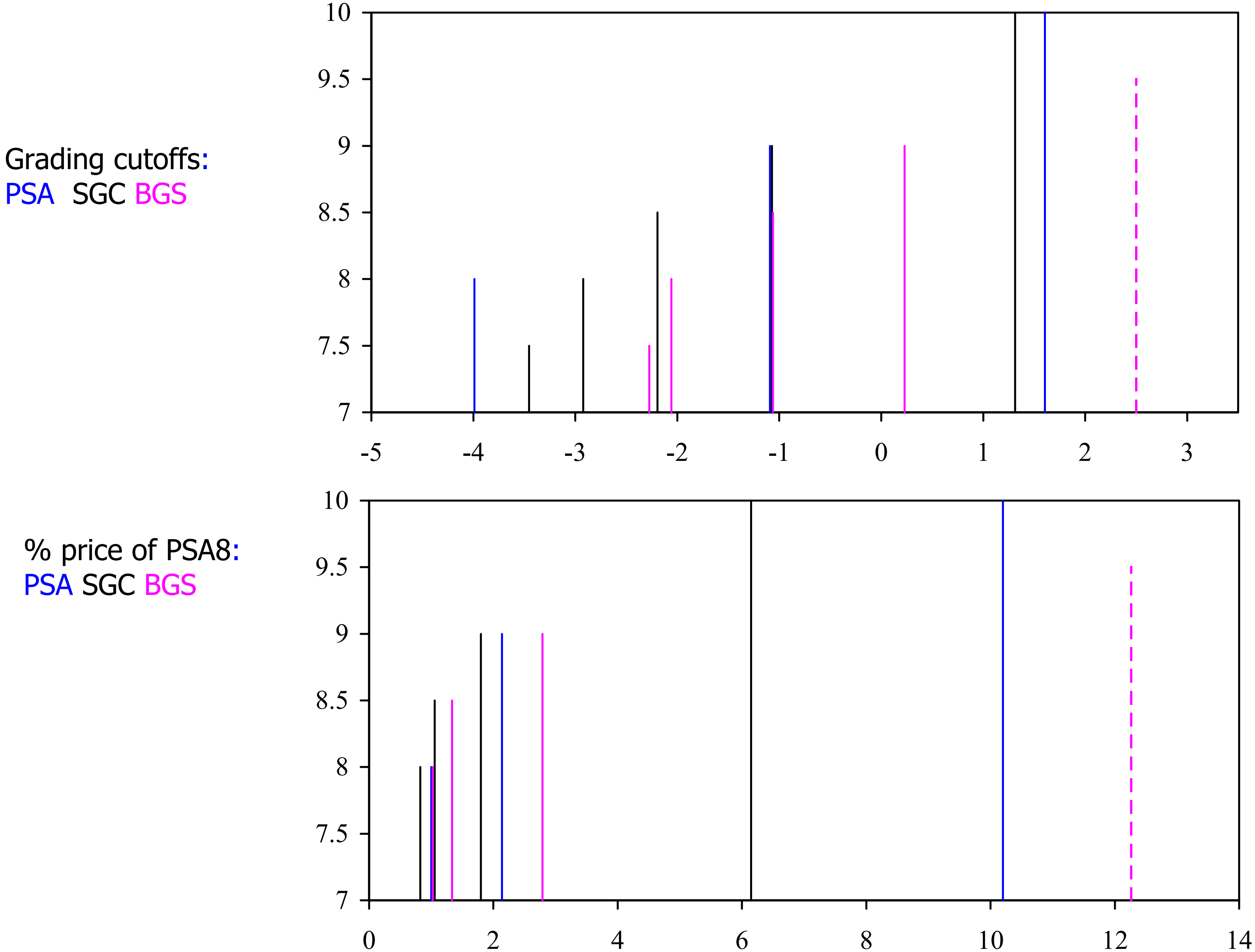
BGS (serial number at the back)

SGC (96 is equivalent to 9 in a 1-10 scale)



PSA

Figure 2. Contrast of grading cutoffs and deflated price by grade and grader



Note: The magnitude of BGS9.5 cutoff is hypothetical, because we do not observe any card graded BGS9.5 in our experiment. However, the deflated price of BGS9.5 is precisely estimated based on Beckett low book price.