Machine Assisted Grading of Rare Collectibles through the Computer-based Objective Interactive Numismatic System framework

By

Richard A. Bassett

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Professional Studies
in Computing

at

School of Computer Science and Information Systems

Pace University

August 26, 2003

Dissertation Signature (Approval) Page

We hereby certify that this dissertation, submitted by Richard A. Bassett, satisfies the dissertation requirements for the degree of Doctor of Professional Studies and has been approved.

_____-_____
Dr. Charles Tappert Date
Chairperson of Dissertation Committee

_____-_____
Dr. Fred Grossman Date
Dissertation Committee Member

_____-_____
Dr. Allen Stix Date
Dissertation Committee Member

School of Computer Science and Information Systems
Pace University 2003

Abstract

Machine Assisted Grading of Rare Collectibles through the Computer-based Objective Interactive Numismatic System framework

By

Richard A. Bassett

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Professional Studies
in Computing

August 2003

Three overreaching findings emerged from the details of this research investigation:

1. The Internet makes a feasible alternative to the palpable examination of rare coins for the purpose of grading: human numismatists accepted the validly of determinations they reached on the basis of images delivered on Web pages, and an automated grading system successfully operated with images of the same resolution as the delivered images.

2. The aggregated interagreement on coin quality among human graders working individually over the Internet was substantially increased when they were furnished with a machine-generated reference point.

3. The development of automated systems for grading coins that work with effectiveness of human experts appears to be a technical possibility based on the successful prototype described herein.

Taken collectively, these findings suggest that coin grading may be done more economically and in a timelier manner than it is done now through the mailing of coins to professional services.

There are two empirical thrusts. One examines the use of machine grades supplied as a grading aid to experienced human graders working over the Internet. The goal was to gain insight into the process of remote coin grading over the Internet. The presence of accurate computer-generated grades improved the human performance. The validity of

the grading process was substantiated by significant test-retest reliability on repeated gradings of the same coins.

The other thrust explores the effectiveness of an automated system yielding expert grades on the 70-point Sheldon scale for Lincoln cents. The system matches a coin's histogram of red, green, and blue sub pixel frequencies against histograms in a database of coins rated by human experts. While there is little construct validity for this technique, data confirms that it works well.

Acknowledgments

The process of this dissertation has taken many twists and turns along the way. I evaluated five ideas in some considerable depth before deciding on this one. Even within this idea my initial thought was that this study was about developing an automated coin grader, when in fact the automated coin grader turned out to be a minor component of the overall dissertation.

There are many people to thank who either contributed to the dissertation directly or supported me along the way. My biggest supporter during this journey was my wife, Jennifer Bassett, as she provided love, moral encouragement and made many sacrifices during these past three years. The rest of my family also endured missing my company on many occasions. Accommodating my work on this dissertation has become such a large component of our family life that my 9 year old son now thinks that working on the dissertation is something that I do for fun.

In 2002 several students from Dr.Tappert's CS631Q class framed out most of the programming for the technical grading model. They are: Ping Gallivan Eric Heinen Xiang Gao & Akarsh Sakalaspur. Dr.Cha from Pace contributed his expertise to our group on the Histogram Distance algorithm and Eric Heinen continued his assistance with the research framework into CS615. Thank you Ping, Eric, Xiang and Akarsh.

The weekly Sunday night 8pm phone calls with team members Kevin Burns, Eddie Davis, Fred Dreyfus and Paul Meisinger was a tremendous help in getting through the DPS program. The moral, educational and collaborative support derived from these meetings was invaluable. I especially appreciated that everyone was willing to work around my lifetime television schedule of watching 60 Minutes at 7pm.

The students in my Fall 2002 MIS405 class at Western CT State University, along with programmer Jed Wood, were a great help in the design and development of the second generation of the online visual grading websites. This exercise proved to be invaluable in getting to the finished methodology even though this application wasn't used in the final analysis.

One of the best things about the DPS program at Pace University is all of the great people that you meet. I have had the pleasure of meeting such a diverse collection of good people in this program including classmates, professors, administrators, librarians and graduate assistants. I am especially thankful for the tireless efforts of Christine Longo and for the many talks, emails and IM sessions with my new friend and classmate Carlos Buskey.

Lastly I wish to thank my committee advisors Dr.Charles Tappert, Dr.Fred Grossman and Dr.Allen Stix for hanging in there with me. Double thanks go to Chuck Tappert for spending so much time with me and providing so much guidance in the role of chairperson.

Table of Contents

List of Tables

List of Figures

Chapter 1 - Introduction

1.1 Statement of problem

Accurately identifying, grading and then determining the authenticity of rare collectible items such as coins, stamps, cards, comic books and artwork is a subjective non-automated process conducted by appraisers or graders. Appraisers and graders are usually experts in their respective fields who draw on large established pools of domain knowledge, opinions of other experts in their field and make comprehensive comparisons to other 'works' in the field to assist them in arriving at their findings. As with many professions the credentials for a person establishing himself or herself as 'an expert' can range from being non-existent to possessing a long list of impressive industry certifications.

There are two empirical thrusts of this research:

One examines the use of machine grades supplied as a grading aid to experienced human graders working over the Internet. The goal was to gain insight into the process of remote coin grading over the Internet. The presence of accurate computer-generated grades improved the human performance. The validity of the grading process was substantiated by significant test-retest reliability on repeated gradings of the same coins.

The other thrust explores the effectiveness of an automated system yielding expert grades on the 70-point Sheldon scale for Lincoln cents. The system matches a coin's histogram of red, green, and blue sub pixel frequencies against histograms in a database of coins rated by human experts. While there is little construct validity for this technique, data confirms that it works well.

1.2 Significance of the study

The Internet has provided a major boon to the rare collectibles marketplace as dealers and auction houses are now able to reach vast numbers of collectors, investors and other potential buyers of their offerings. Collectors accumulate everything from rare coins and stamps to baseball cards, autographs, antiques, posters, comic books, beer cans and artwork. Investors seek to purchase rare collectibles that are appealing to their tastes and will appreciate over time.

The selling of collectible items isn't limited exclusively to dealers and auction houses as collectors are also able sell their duplicates or extra collectibles items by reaching large numbers of other collectors through the Internet. Collectors with wares to sell can access other collectors through auction sites such as eBay.com and Half.com as well newsgroups, chat rooms, email and their own personal websites.

Ultimately collectors, appraisers, dealers, auction houses, markets and insurance companies are concerned with the value, or worth, of a rare collectible as a basis to determining the proper wholesale and retail pricing.

The substantial increase in buyers, sellers and trading activity that the Internet has generated has magnified the problem of over-graded collectibles and thus has led many to purchase items at improperly inflated valuations.

1.3 Coin Grading Overview

The act of determining the condition of a collectible is known as grading. The assessment of the amount of circulation wear or the level of preservation that a particular coin exhibits, or in the case of uncirculated coins by how closely it resembles its "perfect" condition when they were newly struck is then referred to as the grade. The grade of a collectible item usually plays a significant role in relationship to the value of the item.

Two types of grading exist within the coin-collecting domain; they are technical and market grading.

- Technical grading implies the strict adherence of certain objective and published grading rules, such as the those by the American Numismatic Association [4] without the inclusion of subjective qualities. Thus arriving at a technical grade involves only evaluating the merits of the coin by assessing the wear and defects that occur after it was struck. Technical grading is based primarily on fact and can be thought of as science in that it represents the undistorted views of a grader, which are free of emotion or personal bias that are based on observable phenomena, appraisal and evidence without the distortion of personal feelings or interpretation.

- Market grading is a grade that takes into account some or all of the technical grading features plus certain subjective features such as strike, luster, die state, and overall eye appeal of a collectible item. Market grading is the judgment exercised by a grader that may be based on that individual's personal impressions, visual observations, and cognitive abilities, past experiences, feelings and opinions rather than strict objective external facts. Market grading is more of an art form as it is largely concerned with the aesthetic appeal and how the market will accept a certain collectible.

Collectors and dealers are often at odds with each other on the issue of whether grading is a science or an art form [33]. This research will attempt to demonstrate that effective grading of collectibles is both science and an art as effectively identifying the technical rules based features in conjunction with determining where and how a collectible item will be received by the marketplace are both of critical importance. A large part of the confusion is that there is great inconsistency in how grading is done by various graders. Dealers, collectors, casual observers or third-party grading services can all perform the grading of collectibles, dealers and third-party grading services are thought to have the status of expert, or professional, graders. Expert graders are known to use either or both technical or market grading in assessing collectibles. The method of grading, technical or market is usually not obvious nor is it always identified with the grade that is assigned. The values for grades are also not published based on technical or market grades, which also contributes to the subjectivity in pricing.

In order to eliminate the confusion between technical and market grades within this research, grades will be discussed in terms of expert grades and not as technical or market grades. The expert grades are the baseline grades, which represent the output results from the machine-based system of the digitized coin images, which were previously graded by the third party grading services.

1.4 Research Focus Areas

Three overreaching findings emerged from the details of this research investigation:

1. The Internet makes a feasible alternative to the palpable examination of rare coins for the purpose of grading: human numismatists accepted the validly of determinations they reached on the basis of images delivered on Web pages, and an automated grading system successfully operated with images of the same resolution as the delivered images.

2. The aggregated interagreement on coin quality among human graders working individually over the Internet was substantially increased when they were furnished with a machine-generated reference point.

3. The development of automated systems for grading coins that work with effectiveness of human experts appears to be a technical possibility based on the successful prototype described herein.

Taken collectively, these findings suggest that coin grading may be done more economically and in a timelier manner than it is done now through the mailing of coins to professional services.

1.5 The COINS (Computer-based Objective Interactive Numismatic System) Model Defined

This research centers on the grading of collectible business strike coins. Business strike coins are the coins that are given only one blow or strike from the dies when they are minted and are intended for normal circulation and use in commerce. Coins that are introduced into commerce, or spent, are identified as circulated while coins, which are withheld from circulation and maintain much of their original luster are identified as mint-state. Circulating coins are graded on the Sheldon 70-point scale [64] with grades ranging 1 to 59 while mint-state coins are graded in the ranges of 60 – 70.

Recognizing the need for objective and subjective assessment of collectibles, a model known as COINS has been developed for this research. COINS is an acronym for Computer-based Objective Interactive Numismatic System. COINS is intended to enhance the ability of graders.

The COINS model is a dual component framework of which the first component is a machine-based system that performs grading through the use of a template-matching algorithm known as the Histogram Distance Algorithm that is described in detail within

Chapter 3 and yields a expert grade as the resultant output. The machine-based component of COINS is a software system that processes scanned images against a stored database of pregraded images and determines the grade of a coin by pattern matching. The second component of the COINS model utilizes the output from the first component as the baseline grade for subjective human/machine evaluation process. This human/machine component utilizes the Internet as a transport vehicle for facilitating on the grading of collectibles by allowing expert grading consultants to performing grading on the digital images of collectibles.

The purpose of COINS is to provide grading experts with a baseline machine generated grade thus allowing the grading experts with a starting point to obtain grades in a narrower grade range than in the absence of guidance. By narrowing the grading ranges, COINS can assist with consensus grading.

The value that COINS would have in a commercial setting is the ability to grade coins quickly, reliably, and inexpensively. The obvious limitation of COINS is that it would have little appeal in the grading of common collectibles or super rare collectibles. Common collectibles normally command low or minimal value and as such, condition doesn't often affect the pricing. Super rare collectibles, which are on the opposite side of the scale, are normally graded in a mode of consensus grading, which is the process of determining the condition of a coin by using multiple graders. Major auction houses attempt to overcome the problem of expert's varied interpretation in grading by securing the opinion of multiple graders when the financial stakes are significant, such as in the case of the 1933 $20 Gold Double Eagle auctioned off by Sotheby's [5]. Bringing in multiple veteran graders to arrive at a consensus for an item that sells for several million dollars has obvious return on investment (ROI) benefits, as the stakes are large. The majority of transactions involving rare collectibles cannot muster the justification for multiple graders mulling over the salient points of a collectible for hours on end, as the fees to experts are substantial while the value of the average collectible is considerably smaller. Some third party grading services such as Numistrust [52] offer consensus grading as the cornerstone to their service offering. Third party grading services are discussed in more detail within Chapter 2 of this study.

1.6 Research Approach

The initial development work by a student team from Pace University from January 2002 to May 2002 on an automated grading system offered encouraging evidence that a machine-based system to grade collectibles can be built [8].

Similar endeavors in to the realm of machine-based grading systems were previously attempted by PCGS [33] and CompuGrade [36] in the 1990s but were quickly abandoned due to their lack of commercial success and user acceptance. Both PCGS and CompuGrade attempted to build systems that they anticipated would become commercially viable and profitable. They soon discovered that the development of software could be a long and expensive process, which at times never seems to have a conclusion. In the wake of rising development costs, missed deadlines, ever increasingly complex rule sets the hope of all

profitability diminished, therefore, both companies quietly withdrew their systems from the market.

The Pace University effort was unconstrained by the financial obstacles that plagued the earlier systems by PCGS and CompuGrade, as there was no profit incentive and no pressure to get a system to the marketplace. The machine-based system was built to grade collectibles as an academic research exercise. Thus the development team was free of the typical commercial risks that many software development projects undergo and was able to focus in on just moving the technology forward.

The machine-based component of COINS was able to take a scanned image of a collectible and pattern match the grade of the item under consideration. This was accomplished through a series of Java programs that performed pattern recognition through template matching [60] using the Histogram Distance Algorithm on the images and thus produced baseline grades for the items under scrutiny. While only one algorithm was employed in this research future researchers and developers may wish to consider the employment of additional algorithms. Chapter 3 covers the merits of why the Histogram Distance Algorithm was chosen and which other algorithms were considered.

An initial problem that the machine-based component of COINS ran into was in the area of user acceptance. This problem also overshadowed the efforts of PCGS and CompuGrade. The COINS machine-based system is capable of producing an expert grade. Many domain experts argue that a machine based system is not capable of addressing at least six critical subjective areas which must considered when market grading a coin [9, 40, 42]. Included in these six subjective areas are: color, toning, planchet quality, strike quality and aesthetic appeal. The difference between what is an acceptable grade, technical and market has been a point of contention in the collecting domain for a long period. Technical grades are compared to science, and market grades are compared to being an art form.

An initial focus of this study was to build a system capable of producing acceptable grades on a consistent basis. While challenging, producing expert grades with a machine-based system was clearly obtainable as demonstrated by the results of the machine-based experiments within Chapter 4 of this research.

From the onset of this research, it was hypothesized that the machine-based grading system would yield consistent and repeatable results while producing grades on an expert grading level. To verify the reliability of the machine-based system the results were measured against the baseline grades of the coins, which were obtained from the third-party grading services. The reliability of the machine-based system in evaluating the images of coins that were graded by the third-party grading services is documented within Test E of the machine-based reliability tests of Chapter 4.

It is further hypothesized that the larger the database that the machine was working from the better it would perform. This hypothesis is worthy in as much as anomalies could result from the histogram distance measurement algorithm, and such anomalies would detract substantially from the utility of the machine-based system. In the absence of sufficient

database representation of the different grade bands, (Good, Very Good, Fine…etc) the machine-based system couldn't possibly yield proper results as the nearest-neighbor pattern match might come from an improper grade band. For instance matching an image with a grade of Good-4 as a grade of Good-6 isn't as much of an anomaly as matching the same image as a grade of Fine-15 as Good-4 and Good-6 are in the same grade band but Good-4 and Fine-15 are not. To ensure closeness in matching, the database had to be populated with images that were representative of all of the major grade bands.

1.7 Scope and limitation of the study

The collectibles market is both dynamic and vast. Dynamic as new issues of collectibles are constantly being produced while the population of previously issued collectibles decreases in size with attrition over time. Historical significance and market demand change over time, as do the value of items. Most highly collectible type items have weekly wholesale and retail price sheets that are published. The market is vast in that thousands and even millions of current collectibles are turned out each year, some with slight series differences, which may make a single series in a single year in strong demand. Contributing to the vastness is the increased number of collectors and collectable series in the past several years. As previously indicated, the Internet has contributed to the increased number of collectors and has opened the door for other types of previously obscure collectibles items, such as Paint by Numbers sets.

The continuously changing nature and the vastness of the collectibles marketplace makes it extremely difficult for collectors to specialize in more than one collectible as a considerable amount of knowledge is required to efficiently and effectively accumulate and maintain a quality collection. For this reason many collectors tend to specialize in a single collectible type.

This research concentrates on the grading of one area of collectibles, United States business strike coinage. For testing purposes, the focused attention will be placed on the limited series of Lincoln Cents in United States coinage. The technology employed within the machine-based system of COINS was developed in such a way that it could be adaptable with minimal modifications to work with other coin series and other collectibles, although this research did not branch out and test these other areas. The trained database of the COINS framework is made up on 160x160 dpi GIF images of Lincoln Cents, these images could be substituted in future research by images of other series or other collectible items. The resolution of 160x160 dpi was chosen as testing deemed it a high enough resolution to provide the desired of detail required on the images of the collectible being evaluated while being simultaneously small enough to ensure quick loading while using dial-up connections. Nothing was specifically hard coded into COINS that would force it to work with just this one series. COINS draws on the same stored images for the machine-based pattern matching techniques and the machine-assisted Internet evaluations.

1.8 Rarities and Common Collectibles

Providing discriminating scrutiny to certain collectibles over others is justified in that a limited supply of collectibles are considered rarities and can potentially command considerable values. However, the vast majority of collectibles are common and usually command very little value or attention from collectors. Professional grading and appraisal of collectibles comes with long lead times and has a tendency to be expensive. Thus, it is important to screen out the less sought after collectible items from the rarities when undergoing expensive professional grading.

Numerous factors make a collectible item rare, even though a common misconception is that age alone is the most decisive parameter. Some collectible items such as raw precious metals, i.e. gold and silver, have only to contend with supply and demand when determining value. However, the value of non-commodity type rare collectibles is more difficult to calculate as it is dictated by numerous complex factors such as condition, authenticity, age, number originally produced, estimated surviving population, historical significance, as well as market demand.

Factors that contribute to the value of a rare collectible

- Condition
- Authenticity
- Age
- Number originally produced
- Estimated surviving population
- Historical significance
- Market demand

Figure 1.8.1 – Rarity Contributing Factors

Condition, which is often referred to, as the grade is significant, as the most discriminating collectors prefer their collectible holdings to be in the best condition/grade possible. Prices for perfect and near perfect condition, also known as mint state condition, collectibles tend to be considerably higher than for those in similar worn collectibles.

Determining *authenticity* of a potentially rare collectible is a major concern due to the increased financial incentive to produce counterfeit collectibles and the emergence of the technological capabilities to do so. Counterfeiting rare collectibles dates back to the days that coins were first produced. [69] Counterfeiting can be achieved by producing a near duplicate item from scratch or by altering an existing item in series so that it conforms to the attributes of a rarer and more highly sought after item.

With some collectibles, *age* can be a significant driving factor in determining the value, such as in the case of a 1795 Draped Bust Silver Dollar. The value of collectible antiques is often thought to increase markedly with age. However, with many collectibles, age in itself is

insignificant in determining value. For instance, a 1995-W One Dollar Silver Eagle made at West Point is considered a great rarity with a present day value of just over $2,000 while an uncirculated Silver Dollar made in 1900, which is over 100 years old, is worth just $50. Another example of age having little influence on the value of a coin is the Ancient Greek coin identified as Tiberius 14-37 AD. In very good condition, this 1900+-year-old coin can be currently obtained for approximately $100.

The number or amount of a collectible that was *originally produced* is a value-contributing factor as it helps to give understanding to the largest potential size of the population. The production records for mass-produced items such as stamps, coins, comic books, prints and cards in modern times are quite accurate. The records for custom made items that haven't been mass-produced such as artwork and antiques are not as reliable. Still the population records for both custom and mass-produced items in the non-modern periods tend to be non-available and generally unreliable.

In general terms, the *surviving population* of a collectible item is thought to diminish with time. This surviving population can also be thought of as the supply. As in the case of silver coinage, much of the silver coinage was melted in 1965 and 1980 respectively when the US Government switched over to the clad copper coins for circulation and the Hunt Brothers attempted to corner the world silver market. Many coins are damaged and lost. When the surviving population of a particular appealing collectible is thought to be low, the value or price of that collectible would usually be on the high side. Conversely, when an ample amount of a collectible exists, the value is thought be low. If the documented surviving population exceeds the number production population, it is quite possible that a sufficient number of counterfeits may exist or that improper production record keeping occurred contributing to the excessive numbers. [78]

A collectible may be an object of historical significance that is tied closely to a period, a person or a special event in history. Often collectible items with historical significance may start as ordinary products. For instance, certain antiques from the Victorian Period command hefty premiums and are quite sought after as some collectors romanticize and feel nostalgic about this timeframe in history. Barry Bonds 600th homerun ball is an example of an ordinary item. In this case, a mass-produced baseball gained sufficient value by being part of a historical event. Items with historical significance do not always hold their values of time. For instance, a special event in history may shine the light of historical significance on to a collectible series temporarily causing the value to rise while the demand is strong. However, once the collecting public losses interest or the supply catches up to the demand levels, the collectible will drop in value and no longer be considered rare. A classic example of this was Beanie Baby mania that spurred the value of many of these stuffed animals to astonishing high temporarily levels only to have the prices crash back to earth after the market was flooded by these mass-produced items. Short-term demand spikes can occur that temporarily drive up the price of a collectible.

Market demand contributes to value fluxuations as collectible items fall in and out of favor with collectors over time. New hoards can be discovered which increase the supply in the market and thus drive down the demand. Newly revealed low supply numbers can suddenly

cause a collectible to enjoy strong demand. Publicity of an item, the artist or production facility can create sudden demand for items. Unexpectedly high prices realized at auctions for similar items may cause a sudden surge in demand for a collectible. The passing of an artist or designer can create the realization that the supply is suddenly finite and thus drive up demand. When demand for a collectible gets stronger, the value of that collectible may rise at exponential rates. By example: In 1995, 4.6 million $1 Silver Eagles were minted in Philadelphia and 395,000 in San Francisco as normal. [37] These coins were not considered rarities of any great value when they were acquired. However, in the same year just 30,000 $1 Silver Eagles were minted at the West Point Mint. These coins were made to celebrate the 10th Anniversary of the Silver Eagle and were essentially given away free with the purchase of the Gold Eagle 4 piece set. The Gold Eagle set had a price of $995, which was a steep enough price deterrent to cause many would be collectors to pass on purchasing the Gold set just to obtain the $1 Silver Eagle. The final mintage numbers of the $1995-W $1 Silver Eagle were relieved in early 1996, just after the sale of 1995 Gold sets were completed. Demand for the 1995-W $1 Silver Eagle surged after the mintage numbers were released, as did the value of the coin. At present day, the 1995-W $1 Silver Eagle has a value in excess of $2,200 thus demonstrating that a particular series of a collectible can become significantly more valuable than other series that were produced in the same year.

Collectibles that maintain their value and appreciate in the long term usually have several or more of these prime factors in common and are often the most sought after items in their respective categories or series. These items are referred to as rare collectibles. Only the minority of items in a particular collectible category is usually deemed rare and have high values attached to them. Thus, the vast majority of collectibles in a category or series are more readily available and easier to obtain at lower costs; these items are usually referred to as being common.

Condition (grade) and authenticity are the two major factors that sellers can fake, with condition (grade) being by in large the most widely abused factor. The value of a collectible item can be substantially more or less than its true value if the condition or grade of a collectible is improperly represented.

Two examples:

> A potential buyer of a rare collectible may attempt to purchase a collectible item for a lower value than fair market by claiming that the offered collectible is of a lesser grade than it actually is. A uniformed seller may accept the buyer's opinion of the grade and the monetary offer made, as they lack the ability to obtain the knowledge on true grading. Undergrading is a common problem that people face when dissolving estates and hoards.
>
> A potential seller often represents an item at a higher grade than it actually is, perhaps only by a single grade. Because small differences in grading rare collectibles can often result in large differences in the value, a potential buyer stands to loose a great deal of money by purchasing an over-graded collectible. Overgrading is a common practice by many dealers and collectors that sell uncertified collectibles.

1.9 The objective and subjective components in Grading

In the context of purchasing decisions, buyers arrive at their potential purchasing decisions by undergoing objective and/or subjective assessments of the products under their consideration. [55] Objective purchasing decisions are based on facts and are uninfluenced by emotions or personal prejudices [72] while subjective decisions are greatly influenced by the buyers judgment and based on personal impressions, past experiences, sensory perceptions, feelings and opinions rather than objective hard facts. [74]

When a buyer is considering only facts in the purchasing process they are thought to be exhibiting objectivity. Objectivity represents the undistorted views of a buyer, which are free of emotion or personal bias that are based on observable phenomena, appraisal and evidence without the distortion of personal feelings or interpretation.

"Economists generally assume that people know their preferences with certainty and that associated choices are based on observable, well-understood and objective measures of the goods and services. Economic models typically contain the assumption that quantities and qualities of goods and services are measured without error." [55] While this is likely true for similar commodities with easily identifiable characteristics it may not be the situation when evaluating unique items where some component characteristics are not easily observed or identified.

In 1987 Puto developed a buying decision model, which suggests that purchasing decisions of dissimilar products tend to be based upon the buyer's subjective expectations of the product [57]. Puto's model assumes that a buyers approach to a purchasing decision comes with a set of subjective expectations about the item being considered as well as a set of specific buying objectives. In 1982, Payne also suggested that buying decisions are dependent upon the buyer's perceptual or subjective factors. [53] Thus, it is important to consider the subjectivity in a buyer's product assessment process as it relates to purchasing decisions.

Subjectivity is the judgment exercised by a buyer that is based on that individual's personal impressions, visual observations, feelings and opinions rather than objective external facts. Subjectivity can be largely dependent on the mental state or reactions of the person making the statement. "Subjectivity of sensory qualities", such as visualization, is the acceptance that the qualities experienced by the senses are not something belonging to the physical beings, but are subject to interpretation. [74]

According to Loki Jörgenson [38] visualization is not strictly repeatable as one user may perceive a product one way while another user may something entirely different. Visualization relies on subjective interpretation through contextual knowledge, which is usually derived from an experience. Jörgenson also claims that visualization is resistant to the systematic evaluation and objective assessment procedures. Thus, one could conclude that visualization is a subjective assessment. The following studies of subjective evaluation of X-rays in CT head images and wool lot sales are offered as examples to support Jörgenson's

position that one user or buyer user may perceive a product, or environment one way while another user may something entirely different.

A subjective evaluation study which examined X-ray CT head images was conducted by Woobin Kee at the University of Washing in 1995 [39]. In the study three patients were selected as the test set and multiple X-ray readers were given nine films and asked to evaluate, for each film, the quality of 10 images compared to the reference images located at the upper-left and lower-right corners. Each reader was asked to assign one of the following four grades to each of the images in the film:

1. Better than the reference image
2. Same as the reference image
3. Poorer than the reference image, but acceptable
4. Unacceptable

The quality of each image was given a single grade based on overall evaluation of clarity, sharpness, freedom from artifacts, and amount of noise. There was a large variation among readers in judging the image quality. Some radiologists were quite liberal in evaluating the quality of the compressed images, while others clearly distinguished the difference in quality of the compressed images from that of the originals, with a strong correlation depending upon compression method and ratio.

Another study which was conducted in west Texas during 1990 and 1991 to quantify the effects of presale objective data and the economic impact of making the objective data available to bidders prior to wool lot sales. Wool lots were tested ahead of time for yield. All lots were available for presale visual subjective evaluation by potential buyers. The objective yield measurement information was provided on only half of the lots to buyers prior to bidding. All lots contained other data collected, which included date of sale, lot size, and price received. Small insignificant quality differences existed between objectively and subjectively sold wool lots in every characteristic. However, wool buyers paid more, sometime significantly more, for wools that had objective measurements made available to the buyers compared to the subjectively evaluated wools. Thus, providing objective measurements to the buyers had a positive impact on prices paid for original bag wool in west Texas in 1990 and 1991. A major implication of this study is that providing buyers with objectively measured wool data prior to the sale can increase prices paid for wool while providing only subjective evaluation yields lower prices. [43]

Differences in the perceptions of objectivity and subjectivity are not limited to strictly to visualization alone as demonstrated by an ongoing problem of subjectivity between expert wine tasters. Currently wine tasting is largely a subjective activity conducted through the sensory preceptors of taste and smell. Visualization is less important in wine tasting as color is easily altered with dies. Taste and smell help wine experts to determine the quality and the appeal of the wine under consideration. Considerable debate is has been underway for some time among many wine-tasting experts on the adoption of a standardized scoring system. Some experts feel that a standardized system for scoring wine would bring about a much-needed objective level of measurement to wine tasting and thus eliminate much of the

subjectivity. Objective wine measurements would make it easier to compare wines and rank them, than it is to taste one wine and attach a score to it without reference to its peers. Some experts advocate that objective wine tasting be done in peer-group, single blind conditions, with ratings that reflect an independent, critical look at the wines without taking into account price or the reputation of the producer/grower. [2]

As in other domains, the rare collectibles marketplace endures the inherent problems between objective and subjective assessments or grades. Objective grades are thought to be technical grades and subjective grades are thought to be market grades. A technical grade includes only facts and not opinions while a market grade can include the personal impressions, past experiences, sensory perceptions, feelings and opinions of the grader.

In order to arrive at a technical grade a considerable amount of comparison, evaluation and analysis must be performed by using the many volumes of published domain data that exists. Arriving at an objective grade can be an arduous task without the benefit of a return on investment of time. Generally great efforts are expended on a collectible item when the financial stakes are significant, such as in the case of the 1933 $20 Gold Double Eagle auctioned off by Sotheby's. [5]

Properly configured machine-based systems are best suited for determining the expert grades of collectibles, especially when the values of the collectible under consideration is not worthy of extreme human scrutiny. Machine-based systems will probably not be able to produce market grades, as they will always have to be provided by humans due to subjective interpretation associated with them. Therefore, it is understandable that previous attempts at commercialized machine-based systems that attempted to produce market grades were unsuccessful. [36, 56]

1.10 Similarity in Grading Concerns between rare collectibles

Numerous similarities exist between the different types of collectibles (stamps, coins, comic books, and cards). For instance, each of these collectibles has a defined set of grading criteria, i.e. each has a large base of domain knowledge, many experts exist is each area, high priced rarities exist in each of these collectible markets, third-party grading companies exist that offer grading services and there is great incentive to get authentication and grading right.

Numerous similarities exist between the different types of collectibles

- Coins
- Stamps
- Comic Books
- Cards

Figure 1.10.1 Grading Similarities in Collectibles

A brief overview of several major US collectible categories that could ultimately benefit from the model developed in this study:

Coins – US coins have been officially minted since 1793 at various mints. Current denominations for business strike coins are .01, .05, .10, .25, .50 and 1.00. Each of these denominations is represented by a current series. For instance, the current .01 series is the Lincoln Cent and the current .05 series is the Jefferson Nickel. Over time, the series in the respective denominations are replaced. The transition of series from 1793 to present includes: Large Cents, Flying Eagle Cents, Indian Head Cents and the Lincoln Cents. Other denominations have undergone similar series changes. Coins are graded on the 1 – 70 Sheldon Rarity Scale [64] where 1 is the lowest or poorest grade and 70 is the best grade that a coin can obtain.

Stamps – US Postal Stamps have their origins dating back to 1775 when Benjamin Franklin became the first Postmaster General. The Post Office has released thousands of series of stamps since 1775 in many different denominations. The denomination / series rate change for postal stamps is considerably quicker than that of coin mintage. A postal series may be around for a year or less while a coin series is often minted for a minimum of 25 years [79] and usually much longer. Condition is a major factor in the collecting of stamps as stamps are collected in the conditions of Superb, Extremely Fine, Extra-Fine, Very Fine, Fine/Very Fine, Fine, Good or Average and Poor [49] in unused (mint condition) and used.

Cards - Most sports cards were originally promotional items given out by tobacco companies to promote their products. In the 1930s, the tobacco was replaced by gum, the cards became more of the focus, as companies such as Goudey, and Play Ball produced cards. It wasn't until after the Second World War that cards began to be produced by companies on a regular basis, first with Bowman in 1948, then with Topps in 1951. Since the late 1980s, there has been an explosion of card sets, with each of the four card companies producing dozens of sets in each sport under a variety of labels and set names. Scarcity and condition are key things to consider when determining prices but it is ultimately the player on the card that determines the demand. Ultimately, everyone wants the most attractive card. A clean card with good centering and color, sharp corners and edges, and a focused picture, is the goal of almost every collector. Since cards are so relatively plentiful, condition is the big purchase decision on a particular card. As one would expect, the better the condition, the higher the price (sometimes exponentially so.) Cards are graded on a scale from Poor (the worst) to Mint (the best) [19].

Comic Books – "The Yellow Kid created by Richard Felton Outcault in 1895 is recognized as the first comic book. Outcault was the first person to use the balloon, a space where what the characters said was written" [10]. In the past 20 years, comic book collecting has become big business with certain series and issues becoming quite valuable. Condition is a major factor when determining the value of rare comic books.

Recognized companies in the market that once dealt in a single domain area of rarity collecting have now moved into multiple market segments. Two such prominent examples include:

- The Certified Collectibles Group [1] which is an umbrella organization consisting of Numismatic Guaranty Corporation of America (NGC) a grading service in rare coins, Sportscard Guaranty, LLC (SGC) a grading service in sports cards and CGC a third party grading service in comics.

- Industry leader Heritage Rare Coins and Currency of Dallas, TX [35] has recently branched into the sphere of Comic Books [34]. Heritage recognized that due to the many similarities in rare collectibles that they could leverage their knowledge in the foundations of buying, selling and grading rare collectibles into their Internet based auction technologies.

For complete details on how grades/condition matters for different types of collectibles: see Appendix C2 – Standard Grades of Collectibles

1.11 Summary

A major problem with human grading is that it lacks accuracy and consistency [48]. Machines might become good at grading but will clearly fail to take into account the subjective features that many grading professionals feel are important [33]. A hybrid human-machine grading process extends the grade into an expert grade by allowing expert graders to include certain subjective features that they feel are worthy of consideration. Among the most common subjective features are color, toning, planchet quality, strike quality and aesthetic appeal. These systems are currently not available within the coin-collecting domain. Third-party grading companies have yet to see the value of such an approach, although it as has been discussed extensively in the major coin collecting newsgroups such as rec.coin.collecting and PCGS Forums.

This research includes the study of the feasibility of grading coins over the Internet by viewing coin images and examining the differences between human grading, machine grading and hybrid machine assisted human grading. Showing that such a hybrid human-machine effort has merit, could be the catalyst for change within the third party grading services and for other web-based grading initiatives that involve others in the grading process.

Chapter 2 – Relevance of Research

2.1 Overview of Research

This chapter explores numerous issues associated with human, machine-based and human-machine based grading in the context of previous research in these areas. This research explores the grading results and methodology of an objective machine-based system and how the resultant output potentially changes when the grading model is altered to incorporate subjective human-machine interaction. The grading of rare collectibles, in particular coins is the application and the example of this research study, although the framework of the COINS model can be extended to include other dominations and series of coins as well as other collectible domains. Considerable attention is focused on the variances that are known to occur within grading [40, 46, 48] and the reasoning for those differences [33, 52, 69] as well as how the grading process can be improved by using computer based technology.

As previously stated in Chapter 1 the major goals of this research are:

1. The examination of the use of machine grades supplied as a grading aid to experienced human graders working over the Internet. The goal was to gain insight into the process of remote coin grading over the Internet. The presence of accurate computer-generated grades improved the human performance. The validity of the grading process was substantiated by significant test-retest reliability on repeated gradings of the same coins.

2. The exploration of the effectiveness of an automated system yielding expert grades on the 70-point Sheldon scale for Lincoln cents. The system matches a coin's histogram of red, green, and blue sub pixel frequencies against histograms in a database of coins rated by human experts. While there is little construct validity for this technique, data confirms that it works well.

2.2 The issues associated with Human Grading

When humans grade items, such as coins, it is assumed that they carefully examine and visually recognize all of the objective and subjective detailed features on the item under consideration before arriving at a grade.

Depending upon the series, the number of detailed features to consider on a coin may range from 10 – 20. These 10 - 20 features then need to be translated to an industry-accepted ANA grade, which is on a 1 – 70 scale.

The cognitive ability of the grader also comes into play with grading. [11, 45, 47] Grading is a visual recognition process, as it requires humans to identify visualized items and match them against the stored knowledge in their long-term memory. If the grader's long-term memory doesn't contain the building blocks for the recognition of a particular collectible item then they are apt to misidentify or improperly grade it. A person with great skills in the

Large Cent series (1793 – 1857) may possess considerable knowledge on the nuances of the series but lack any depth of knowledge in the Lincoln Cent series. When presented with a Lincoln Cent to grade the Large Cent expert would draw on their abilities in grading Large Cents and attempt to apply them to the Lincoln Cent. As so many features are different between the series, it is unlikely that the Large Cent expert would be correct in their assessment. However, due to humans ability to do complex visual pattern matching it seems they may come close and in extreme cases such as a coin is a cull or a nice specimen their odds of getting the grade correct increases [77]. There is diversity of the interpretations of human experts in the grading of rare collectibles. In order for humans to evaluate items for grading, they must draw on their natural cognitive ability for visual recognition.

The natural ability for humans to recognize patterns is something that is often taken for granted as visual pattern recognition comes to us so easily. Humans are able to recognize a large number of items quickly with little or no cognitive effort. The recognition process usually occurs so rapidly and automatically that we often do not even appreciate that any type of internal processing has taken place.

To humans, pattern recognition is a seemingly effortless endeavor that is employed everyday, as we perceive items that we come in contact with and with in the world around us.
The process of visual pattern recognition is a complex problem, so much so that the visual areas that are responsible for this process occupy up to one-half of our cortex. [68]

The memory area for visual stimuli is referred to as iconic memory. In the case of visualization, iconic memory is the sensory memory for visual, which acts as a buffer for incoming stimuli received through the senses. Iconic memory can be defined as brief sensory memory for some visual stimuli that occur in the form of mental pictures. It is important to note that iconic memory is temporary and fades quickly. Information is then passed from sensory memory into short-term memory by attention; thereby filtering the stimuli to only those items, which are of interest at a given time, see Figure 2.2.1. This passing takes place since iconic memory cannot hold much information for long (less than 0.5 sec for iconic memory [67]). The transferring process is filtered, by only allowing stimuli of current interest to pass. Without filtering, our short-term memory will be overloaded easily. We tend to focus our attention to stimuli that makes us excited or captures our interest.

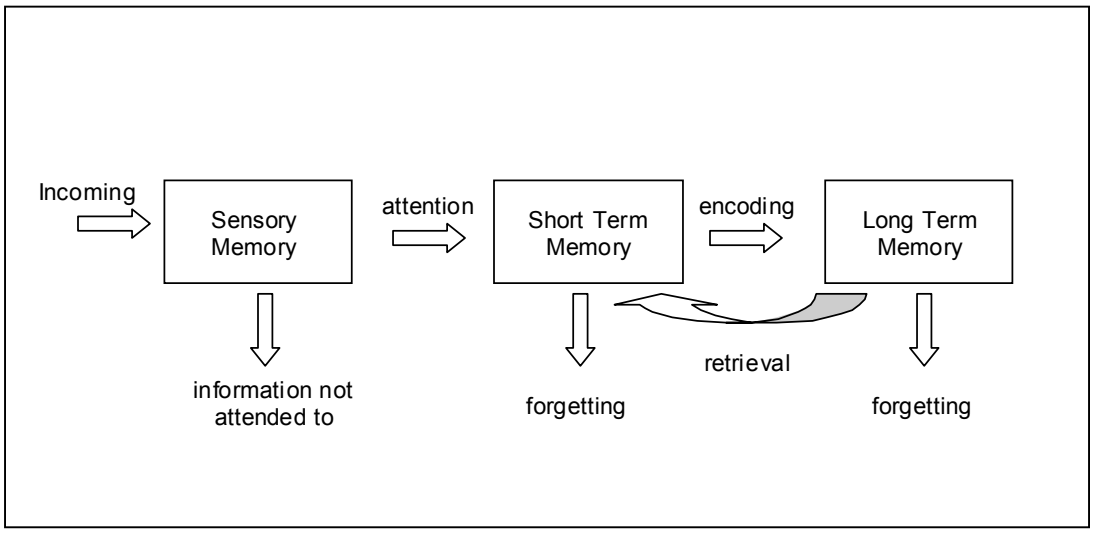


Figure 2.2.1 – Human Visualization Model

Short-term memory or working memory acts as a "scratch-pad" [24], much like computer random access memory, for the temporary recall of information that is being processed. Short-term memory can be accessed rapidly in around 70ms, but will be swapped out of memory in 200ms as it decays rapidly leading to forgetting. It also has limited capacity [59]. The capacity of short-term memory appears to be rather limited. Humans can hold only about 7 "chunks" of information in short-term memory at a time; the size of a chunk is relative, not absolute [47].

2.3 Machine Grading Issues

Machines can be configured and trained to perform an evaluation of all of the grading aspects that humans are supposed to consider in the grading process but rarely do [9, 33]. Such a system should be able to consistently produce grades at a higher accuracy and consistency level than humans. In order for a machine to be able to carry out many of the same tasks as the expert human graders the machine would have to be trained through the process of machine learning on the domain knowledge approaches and features that the experts draw upon [26]. Then the machine would therefore have to acquire information, organize it and then be able to make use of this knowledge [6].

Machines typically learn in the following three phases [26]:

1) Training – machines are trained via predefined rule sets, which include examples of correct behavior of the freshly stored knowledge.

2) Validation – in the validation phase a human tester determines if the machine requires additional training or additional rules sets, which were stored during the training phase. Alternatively, the human tester may validate the rules using other knowledge bases and automated tools. In the validation phase, the expert, acting as the tester, can then critically evaluate the inferred rules and verify their

reasonableness. If the rules are incorrect, the expert may be able to suggest alternative examples that can guide further learning. [30]

3) Application - the rules are used in responding to some new situations.

A primary advantage in using a machine-based system in an application with a large amount of data stored in its database is that the system should always produce the same consistent outputs given the same inputs and operating conditions. [16] If the system gets something right once it should consistently get it right, conversely if the system makes a mistake once it will consistently make the same mistake until programmed, taught or trained otherwise. Humans are able to learn from their mistakes and attempt to work out why things went wrong and are able to try alternative solutions. Through visual pattern recognition [6, 11, 12, 23, 45, 68] humans are intuitively able to notice similarities between things, and therefore can generate new thoughts about their environment and approaches to problem solving.

Machines are built to emulate limited components of the human cognitive system and can be configured to specific perform tasks, which produce consistent repeatable results. A human cognitive system is capable of reasoning in a variety of ways, using considerable amounts of stored knowledge, can learn from its prior experiences thus improving its performance as new knowledge accumulates, can be conscious of its own behavior and contemplate on its own capabilities; and finally can take action in a vigorous manner when surprises occur. [13]

In the case of robotics being deployed in a manufacturing setting a robot may be used to weld the same joint in the same location in the same model of car on an assembly line 24 hours a day. With proper initial training and maintenance, the robot wouldn't tire and should perform its appointed, but limited, task as long as required. The value that machines serve is that they produce consistent results and they tend not to be influenced by some of the outside factors that humans are such as conflicting financial incentives and physical impairments. However machines lack the complete cognitive and deductive qualities that humans possess and must be trained, or programmed, with the details of all tasks they are to perform. If a machine has only the Lincoln Cent series in its database and a Large Cent is presented for the first time the machine would not be able to identify the item at even the most basic levels. A human in the same situation would recognize the Large Cent as US Coinage with a denomination of 1 cent and may be able to come close, within several grades, to guessing the grade.

2.4 Human/Machine Collaborative Grading

Humans have tremendous cognitive abilities for visual pattern recognition and don't specifically need to be trained for every combination and permutation of items in order to visually recognize them. The cognitive recognition process allows a human to identify an item based on similar items that they are familiar with even if they haven't seen the exact item before [45].

Dr Michael Tarr states, "that a single recognition system can not support categorization at many different levels". [68] Tarr further goes on to say that the template and feature matching models, which are described with Chapter 3, do not have an identifiable method for

segmenting complex input into a series of objects that need to be recognized; instead, the matching models appear to rely on being 'fed' items to recognize

Human/Machine collaborative grading provides the output results of the machine-based system as input to the expert grader thus allowing him/him the ability to apply subjective opinions within the grading process.

2.5 The Evolution of Grading & Authentication

Assigning a grade to coins, or other rare collectibles, helps to establish the condition and the state of preservation of the collectible. Accurately determining the condition of a coin is significant as it is a large contributing factor to determining the value of a collectible in the marketplace.

History of Grading:

In 1948, a well-known numismatist by the name of Dr. William Sheldon attempted to standardize coin grading by proposing what is now known as the Sheldon Scale and his work was first published in 1949 under the title Early American Cents. In 1958, the work was updated with Walter Breen and Dorothy Paschal under the new title Penny Whimsy [64]. Dr Sheldon's scale, which runs from one to 70, was originally devised specifically for large cents, but it is now applied to all series. This scale uses the first 59 numbers to deal with circulated coins and the last 11 numbers with uncirculated coins. Not every number is used. The Sheldon Scale was a vast improvement over grades such as Good and Fine, but there was still substantial room for disagreement between two parties based on subjective opinion.

In 1958, Martin R. Brown and John W. Dunn published A Guide to the Grading of United States Coins, which included illustrations with sketches. Before this book, there was no reference devoted to the subject of coin grading [54]. Prior to the 1970's, the individual Coin Dealer usually did the grading of coins. However, this system proved inefficient because opinions about grading would differ from one dealer to another. In the mid 1970's, there were three uncirculated grades. The grades were uncirculated, brilliant uncirculated and gem brilliant uncirculated.

On June 12, 1972, ANACS began accepting coins for authentication, certifying them as genuine and producing photo-certificates of authenticity. Official grading at ANACS would not begin until 1978 [63]. In the late 1970's the ANA established more precise grading standards than those previously published by both Brown and Dunn and James Ruddy [22]. The ANA used the 70-point scale conceived by Dr. Sheldon in his book Penny Whimsy. [64] In the ANA grading standards the uncirculated grades composed of values from 60 to 70 from the then current three-grade system. Coins that were previously sold as uncirculated became MS-60, coins previously gem BU became MS-63 and gem BU coins were sold as MS-65.

The Birth of Third-Party Grading Services

In the mid-1970's the American Numismatic Association (ANA) the nation's largest coin club, recognized the need for uniform grading standards in the coin market. ANA appointed a panel to evaluate possible grading standards. While considered numerous options, including a system with grades from 1 to 100 it was the Sheldon [64] 70 Point Grading Scale that was selected in the final analysis. The ANA apparently was influenced by the fact that the 70-point scale was a known commodity in at least one series. [62] The Sheldon scale was adopted and applied to the grading of all US Coins series.

Dr Sheldon's passion in coin collecting and research was limited to the Large Cent series and his 1 – 70 scale was developed specifically with this coin series in mind. This scale has shortcomings as its most obvious deficiency is that of the 70 numbers available and only 11 of those from 60 to 70 are set aside for Mint State (MS) coins and unimpaired Proofs. The other 59 grades (1 – 59) are reserved for circulated coinage. Given the sphere of interest in which Dr Sheldon was operating in, the large cent series, it is understandable that a disproportional larger set of numbers would be allocated to circulated specimens as higher quality mint state examples of large cents, which were minted from 1793 to 1857, were considerably scarcer than circulated examples.

Since Dr Sheldon's work, mint state coins have assumed far greater importance than circulated coins, and it would be desirable to have more space for them on the grading spectrum to reflect their degree of difference more precisely. Nonetheless, PCGS and other third party grading services have used the 1-70 scale since opening for business in 1986. [62]

In 1986, the Professional Coin Grading Service (PCGS) began operations followed in 1987 by the Numismatic Guaranty Corporation (NGC). These grading services differed from anything that had come before as these companies sealed coins in tamper-evident plastic holders, known throughout the industry as slabs [46], see Figure 2.5.1. A certified coin, or "slab", is a coin that has been authenticated, graded and encased in a hard plastic holder by a professional certification service. The main premise behind "slabbing" as coin is that once a coin has been graded and labeled no user should be able to swap the coin with another so it is almost impossible to open a slab without destroying it. This allows dealers and collectors to place "sight-unseen" bids for PCGS and NGC certified coins as there is supposed to be consistency in the grades being assigned.

Figure 2.5.1 – A Morgan Silver Dollar Certified as MS65 by PCGS
Source: *Wexford Capital Management - Stephens City, Virginia*

Many dealers initially resisted change and questioned the viability of the standardized grading systems being touted by the third-party grading companies. But more grading services quickly appeared in the market filling a previously unknown and unanticipated market need.

During the late 1980's, the major grading services PCGS, NGC and ANACS had defined grading to levels that are more detailed. The higher quality uncirculated coins, previously known as Brilliant Uncirculated (BU) were now expanded into a range of MS60 to MS70. MS-60, MS-63, MS-64, MS-65, MS-66 and MS-67 are the most common uncirculated grades issued by the grading services. The odd grades of business strikes of MS-6l, MS-62, MS-68 and MS-69 are relatively scarce as collector demand isn't often strong for these grades. [22]

Until 1986, there were no consistent interpretations of grading standards for coins as "value based" grading standard used through 1985. Literature and guides existed but they were of poor quality and subject to wide interpretation. Author Ray Wyman notes "The mention of a few minor marks said nothing about the size and location of the blemish on the coin". [78] As of June 2002 there are 72 documented 3rd Party Grading companies [63]. Many of these companies are no longer in business today.

Some of the current grading companies, which certify and slab coins, are listed in Table 2.5.2.

Company Acronym	Company	Address Information
ANACS	Amos Press	P.O. Box 182141 Columbus, OH 43218-2141 800-888-1861
ICG	Independent Coin Grading Company	7901 E. Belleview Ave. Suite 50 Englewood, CO 80111 877-221-4ICG or 303-221-4424

ICCS	International Coin Certification Service Inc.	Toronto, Ontario, Canada 416-488-8620
NCG	Numismatic Guaranty Corporation	P.O. Box 4776 Sarasota, FL 34230 800-NGC-COIN or 941-360-3990
PCI	PCI Coin Grading Service	P.O. Box 8609 Chattanooga, TN 37414 800-277-2646 or 423-485-0997
PCGS	Professional Coin Grading Service	P.O. Box 9458 Newport Beach, CA 92658 800-447-8848 or 949-833-0600
SEGS	Sovereign Entities Grading Service	401 Chestnut St. Suite 103 Chattanooga, TN 37402 888-768-7261

Table 2.5.2 – Current third party grading services
See Appendix C3 for a complete list of all known grading companies

Problems and Challenges of Third-Party Grading Services

The grades issued as certifications by the grading companies are opinions. The same coin may receive widely diverse grades if submitted to different services and may also receive a different grade if it is cracked out of the holder and resubmitted to the same service.

Several years ago, Kevin Foley, the editor of The Centinel, official journal of the Central States Numismatic Society, conducted a study on third-party Grading Services. He sent 10 different coins to four different professional grading services [54]. The four services failed to agree on the grade on any of the submitted coins. On one of the 10 coins submitted, a 1919 Standing Liberty quarter, the professional opinions ranged all the way from Almost Uncirculated (AU) 55 to Mint State (MS) 65. Norman Stack, a late well-known dealer, had a Liberty Head $20 piece which he sent four times to grading services and had it returned in slabs marked with grades of MS-61, MS-62, MS-63, and MS-64. Harvey Stack, CEO of Stack's, reports sending a gold dollar to a grading service, having it certified as AU-50, then sending it back to the same service, after which resubmission it "improved" to MS-60. [54]

Grading standards for some, uncirculated coins have changed, some would say evolved, since slabs were first produced in 1986. Therefore, a coin in an early slab may potentially receive a different more liberal grade if resubmitted now.

Some grading services have better reputations for more accurate grading than other services. Thus, the certification from some grading services may not be regarded as being reliable by experienced numismatists. In such cases the coins encased in holders by questionable grading services tend to sell and trade at lower values than coins encased by those grading services perceived in the market to be deemed more reliable. Experienced numismatists understand this but novice and uninformed collectors do not as they will often be duped into paying regular market price for coins encased by grading services with questionable skills.

Certification and slabbing prices range from $15.00 to $175.00 per coin depending on a number of factors. On a high-end rare coin, the charge is insignificant relative to the coins' market value. However, on a common coin, such as an uncirculated 1921 Morgan Silver Dollar, the cost of certification can far exceed the value of the coin.

2.6 Summary of what is known and unknown about the research questions

Human coin grading

Humans have been grading coins since long before the third-party grading services came into existence. Prior to the adoption of the Sheldon 70 point scale by the ANA in the 1970's dealers and collectors were assigning grades to coins as the basis of valuation on a less defined scale than Sheldon's [64]. The 70-point scale offered more grading choices for graders and expanded the possibility for more diverse human grading.

Stu Miller's grading challenge Website [48] is a good demonstration of the differences in opinions that graders have when grading the same item.

STUJOE- Grading Challenge Results (11/18/02)

Series	Date / MM	Low	High	Midpoint	Number	Count	Total Score	Weighted Ave
Indian .01	1905	12	58	35	208	8	8288	39.85
Lincoln .01	1913D	6	58	32	184	11	6446	35.03
Liberty .05	1902	6	58	32	138	9	2870	20.80
W Libert .50	1944	40	66	53	145	12	8795	60.66
Wash .25	1962D	40	67	53.5	125	13	7699	61.59
Buffalo .05	1936S	8	55	31.5	137	12	4815	35.15
Merc .10	1937	6	50	28	163	11	4215	25.86
Franklin .50	1954	45	69	57	117	12	7118	60.84
Liberty .25	1928	6	40	23	115	8	1281	11.14
Barber .10	1903	8	58	33	124	12	2558	20.63
Kennedy .50	2002D	45	69	57	100	12	6389	63.89
Jefferson .05	1938	1	58	29.5	105	14	2589	24.66
Morgan 1.00	1989O	58	68	63	146	9	9354	64.07
Shield .05	1866	12	67	39.5	91	19	4329	47.57
Sac 1.00	2002D	1	69	35	100	19	5913	59.13
Roos .10	1963	8	66	37	71	15	4135	58.24
	Averages	18.88	61.00	39.94	129.31	12.25	5424.63	43.07

Table 2.6.1 – Results of Stujoe Grading Challenge

Table 2.6.1 represents 16 different grading tests that Stu Miller ran on his Website in which an average of 129 graders participated. Looking at these tests as an aggregate reveals some interesting observations on the differences of opinions that humans have in grading as

graders have picked grade ranges from 18.88 (Almost Fine) to 61 (MS 61) with an average of 43.07 (Extra Fine).

As long as grading has existed so have differences of opinions in the grades being assigned [3, 40, 52, 54, 69]. Numerous reasons typically are given for the variations found in human grading: financial incentive, lack of domain knowledge and experience, gradeflation, vision impairments, lighting and the cognitive abilities of the grader (see figure 2.6.2).

Reasons why Humans grade collectibles incorrectly:

- financial incentive
- lack of domain knowledge and experience
- gradeflation
- vision impairments
- lighting
- cognitive abilities of the grader

Figure 2.6.2 – Grading Challenges for Humans

Financial incentive may motivate a prospective buyer to deliberately grade lower than the actual grade in order to purchase a specimen for less than its true market value. Conversely, a seller may assign a higher grade to an item hoping to sell it for more than it is worth. Unfortunately, these unethical practices occur at an alarming rate within the collectibles-trading domain.

Lack of domain knowledge and depth of experience affects novice collectors, collectors with scattered and diverse interests, estate recipients whom may happen to inherit collections and even seasoned experts that aren't staying up on their trade. Persons that lack proper depth of knowledge in a series, which they may be trading in, such as Lincoln Cents, not only, have to content with the basic 70 point grading scale but they must understand the particular nuisances of the series when grading or evaluating items. For Example: In the Lincoln Cent series the 1909-S VDB coin is quite valuable, especially in higher grades, a single grade difference can change the value by hundreds of dollars. If a person who has inherited a collection possesses a 1909-S coin and doesn't recognize the small VDB letters on the reverse of the coin then they could potentially sell the coin for considerably less than it is truly worth. If unethical financial opportunism is used by a potential buyer, they may still attempt to undergrade the misidentified coin in an attempt to purchase for even less than it is worth. This is just one simple example within each series of collectibles there are at least 3 or more such examples.

Gradeflation has to do with a fundamental shift in general grading opinions within the marketplace. Within the coin collecting arena grading became more liberal after 1986, meaning what was a AU-55 prior to 1986 may have be accepted with a market grade of AU-58 or even MS-60 by today's standards. This doesn't mean that the coins somehow improved in quality over time, instead experts in the field became for liberal with their

assignment of grades. This shift occurred around the same timeframe that 3rd part grading services started gain a toehold in the market.

Physical impairments such as poor vision and poor lighting contribute to the problem that humans have with grading. As coins are small, it becomes difficult for aging experienced graders to properly see all of the required detailed features. Deteriorating eyesight, poor lighting or lack of available magnification make it difficult to properly examine a small item such as a coin without missing critical features. To overcome this problem experienced collectors will often examine coins under magnification with a Jewelers Loop.

The cognitive abilities of the grader also come into play with grading [11, 45, 47]. Grading is a visual pattern recognition process, as it requires humans to identify visualized items and match them against the stored knowledge in their long-term memory. If the graders' long-term memory doesn't contain the building blocks for the recognition of a particular collectible item then they are apt to misidentify or improperly grade it. A person with great skills in the Large Cent series (1793 – 1857) may possess considerable knowledge on the nuances of the series but lack any depth of knowledge in the Lincoln Cent series. When presented with a Lincoln Cent to grade the Large Cent expert would draw on their abilities in grading Large Cents and attempt to apply them to the Lincoln Cent. As so many features are different between the series, it is unlikely that the Large Cent expert would be correct in their assessment. However due to humans ability to do complex visual pattern matching they may come close and in extreme cases (such as a coin is a cull or a nice specimen) their odds of getting the grade correct increase dramatically.

Previous attempts at machine-based coin grading

There have been a several previous documented attempts at machine-based coin grading. Two commercial companies, PCGS [33] and CompuGrade [36] systems in the market in the 1990's but quickly withdrew them due to their lack of commercial success. In 1990, PCGS announced a computerized system for grading coins. The system, which was known as the PCGS Expert, utilized robotics, image enhancement, image processing and an online image database for its integrated computer system [33].

Professor Daniel Power [56]: "PCGS claimed that their "Expert System" combined state-of-the-art computer technology with leading edge software and peripheral hardware. The system employed mechanical positioning, expert systems, real time video, image enhancement, image processing and an on line image data base. The PCGS Expert included rules to evaluate many grading characteristics including strike and overall eye appeal. The system was built during a two-year development period. The tool that PCGS developed was both a hardware/software system. Coins were placed on a carousel, designed to hold more than 40 coins. The carousel rotated and an individual coin was selected for grading and then by means of robotic adjustments the coin was moved to a precise fixed position for grading by the computer. In roughly three minutes, the computer assembled relevant grading information and made its final determination of a grade or quality score."

Both PCGS and CompuGrade attempted to build systems that they anticipated would become commercially viable and profitable. They soon discovered that the development of software could be a long and expensive process, which at times never seems to have a conclusion. In the wake of rising development costs, missed deadlines, ever increasingly complex rule sets the hope of all profitability diminished and both companies quietly withdrew their systems from the market.

Rick Montgomery, President PCGS, when asked about computer grading at a Long Beach show by collector Bill Chin [18]. Montgomery said" that 80% of the time the computer program did pretty well, but when the program missed, it missed by a mile. Also the computer program was only set up for the Morgan dollar series."

CompuGrade, a company run by Jim Diefenthal, had a system, which could consistently grade Morgan dollars to a standard of a tenth of a point. One aspect of the CompuGrade system was that the system could be used as a grading teaching tool as a method for learning grading [29]. No formalized research could be located that described the approach of the CompuGrade system. However, coin expert John Baumgart attended a presentation of CompuGrade at the 1991 Chicago ANA show and reported his observations of the CompuGrade system [9].

According to Baumgart, the CompuGrade system graded coins based on digital images taken in a multistep proprietary technology process within a controlled environment. First, a defect map was generated by subtracting an ideal coin from the coin that was being graded and thus producing the defect map. This map contained all defects discovered such as bag marks and scratches, which appeared on the coin. Then an algorithm rated the marks contained in the defect map based on location and severity to arrive at part of the grade. The next step required several more images of the coin to estimate the eye-appeal from the coins luster. Lastly, a human evaluator made sure there weren't any catastrophic errors.

Some of the problems reported [9] with the CompuGrade system included:

- Toning fooled the system
- Abnormal die strikes created significant challenges
- One algorithm for grading didn't work for all coins in the Morgan Series
- The system didn't assure authenticity
- The eye-appeal algorithm didn't provide the entire picture of eye-appeal, which is arguably subjective from person to person

Avid collector and dealer Byron Reed pointed out that the CompuGrade system was "great from the standpoint of determining detail, but was really poor when determining pretty verses ugly" [58].

An important aspect of the system was the maintenance of a database, which included the previously graded coins. This database was potentially useful for identifying specific individual coins by contact marks, wear and for the purpose of insurance, theft protection, rarity evaluation, and provenance [9, 58].

2.7 The contribution this study makes

Interest in the accumulation and trading of rare collectibles has enjoying unprecedented growth and attention in recent times. The Internet and the global economy have greatly contributed to the explosion in terms of the number of dealers and collectors that have become actively involved in the collectibles domains. The increased population of dealers and collectors has led to a substantial increase in transactional activity with many more collectible items being bought, sold and traded than ever before. The problem of undergrading and overgrading has been on the rise in proportion to the increase in transactional activity. Thus, getting the grade as close to a valid match as possible has become more important than ever as in order to properly assess the value of a rare collectible item, as it is a significant contributing valuation factor. Despite the existence of standard grades in the respective collectible domains, differences of opinion on the grades assigned to collectibles are common [48].

The rampant problem of undergrading and overgrading exposes the collectible domains to possible financial implosion, similar to what the stock market has gone through in the 2000 – 2003 timeframe. As more and more collectors discover that they have become duped by improper grading they will become disenfranchised and will withdraw their money and interest from the respective collectibles areas. Collector interest is the demand component in valuation. With diminished demand, the value of all collectibles will drop with the rarer collectible items suffering the most pressure.

Getting accurate grading is an important problem with potentially serious long-term financial consequences for many participants (dealers and collectors) in the collectibles domains, especially since values are so tightly related to the condition of the collectibles. Still, a small percentage of participants are profiting nicely by the present subjective nature of grading, these people do not want the system to change. The collectibles area has grown too large to allow the current subjective methods of grading to continue. The time has arrived for a standard and consistent method of grading to be established, a method that has universal acceptance among the majority of interested parties.

Chapter 3 – Machine and Online Experiment Methodology

3.1 Background:

Visual Pattern Recognition

The three most popular theories on how human achieve visual pattern recognition are known as template matching, feature analysis and object recognition [11, 23, 45, 60]. This research utilizes the template-matching model but future researchers may wish to employ feature analysis and object recognition.

The template-matching model is the task of pattern recognition using the template matching theory is to find the memory template (or pattern) that best matches the information, or input, that is stored in iconic memory. This involves the act of retrieving stored templates from our long-term memory for each visual pattern that we recognize. Each stored template is in the form of an intensity array and the values of in that array are the ones that are produced when the corresponding pattern is actually seen. Representing the input and the memory templates in the same format simplifies the template matching process and shows the relationship between the mental representations and the processes that operate on them. [60] When a good match is found between the patterns, in iconic memory against the patterns stored in long-term memory then the stimulus is identified based on that match. This is a "bottom-up" model in the sense that the visual "data" from the environment is processed completely and passed on for additional processing level in memory.

While simplicity is the largest benefit afforded by the template matching model several significant drawbacks are known to exist. Among these potential drawbacks are:

- Great difficulty with pattern variability - This is due to the many possible variations in form, size, shape and style of the patterns to be recognized. Take the case of a single letter such the letter A, it can be represented in many fonts, font sizes, point sizes, upper case, lower case just to name a few variations. When a person reads an A into there iconic memory a pattern must be matched against long-term memory before the stimuli is notified that the image is an A. To overcome the problem of variation a virtually unlimited number of templates must be stored in long-term memory to achieve a high level of accuracy of pattern matching. Many theorists believe that storing such a large number of templates is unreasonable. [23] Recognition speed is another byproduct problem of variation as the number of templates stored in long-term memory the longer that it takes the pattern match to occur.

  Template models are rigid in their ability to cope with variation. When a template search and match algorithm within long term memory is taking pace the match algorithm doesn't give an all-or-none output to the stimuli, but instead reports on the degree of match between the input and a template. [60] The search and match algorithm compares the degree of match to a viewed object and produces producing the yes-or-no decision about the match. Thus, template models can tolerate some degree of variability, if the degree of match doesn't get too small.

- Context effects and novel patterns also create a problem for the template matching model as this model of recognition fails to account for the recognition of any novel pattern [23], even if it is similar in form to the original. For instance, it is difficult to recognize an odd shaped chair, such as a beanbag chair, when you've never seen such an object represented as a chair before. In addition, how do you recognize an acquaintance that has just dyed his hair blue if an exact template of this representation has not already been stored in long-term memory?

The template-matching model works reasonably well for situations in which there are a limited number of shapes to be recognized, but it is too limited to be a holistic model of pattern recognition [60].

Previous attempts at machine-based grading

History has demonstrated that rudimentary applications can be developed which are capable of determining grades for collectibles. PCGS, CompuGrade and Pace University have all developed partial systems that accomplished a good portion of the automated grading task. The grading systems that were developed by PCGS and CompuGrade were both commercial ventures, while the system developed by the team at Pace University was for the purpose of research.

In 1990 PCGS, a now prominent third-party grading service, had developed a system called The Expert that was capable of determining the technical grade of Morgan Dollars [33]. From all indications [9] PCGS discovered that it was more difficult and more costly to develop a fully working system than they had anticipated. They also discovered that human intervention could not be avoided in the automated grading process.

In 1991, another grading company known as CompuGrade brought another automated grading system to the market. According to a 1991 press release on CompuGrade [44]: "The system scanned coins using a high definition television camera. The location and severity of each detracting mark, such as scratches, the luster, brightness and sharpness of detail were all evaluated. A perfect coin received a grade of MS-70, with MS-60 being the lowest grade." At the time of the press release, CompuGrade was only grading Mint State Morgan dollars. This short-lived CompuGrade system was unique in that it attempted to provide more precise technical grades within the Sheldon scale [64] as their system was able to extend grades to five places to the right of the decimal place [36] although only one place to the right of the decimal was actually published. The Sheldon scale provides for whole number grades of 1 to 70, with 1 being the lowest grade and 70 being the highest grade possible. Prior to the introduction of this system, no known commercial venture had attempted to provide decimal place precision to the Sheldon scale. Through their approach of decimal place grading the CompuGrade, system had attempted to address the ongoing problem with grades that fell between two whole numbers. The CompuGrade system focused primarily on the Morgan Dollar series and the system was withdrawn from the marketplace shortly after introduction and never completed.

A student team at Pace University in the Spring of 2001 [8] was able to develop a system that determined the grade of a whole coin. This was accomplished by having the computer application compare the vectors of a scanned coin image against a database of pregraded and prescanned coin vectors. The Pace system focused on the Lincoln Cent series but was designed to handle any other series with the inclusion of the appropriate trained databases by working with standard 160x160 GIF images. It should be noted that this system has not tested with other series and that could be an opportunity for future research. This system was intended as an academic research exercise and not a commercial venture.

The limited domain training that they were provided also hampered the previously developed systems. Each system focused on a specific series within a collectible domain. The PCGS and CompuGrade systems primarily graded Morgan Silver Dollars while the Pace system focused on Lincoln Cents.

According to the experts in the coin-collecting field some of the shortcomings of a fully automated system for grading collectibles include:

- Lack of a true picture of the eye-appeal [9] as the eye-appeal algorithm didn't provide the entire picture of eye-appeal, which is arguably subjective from person to person.

- Toning can have a drastic affect on the image of a coin [40]. Toning could not be measured accurately by the system, as measuring toning in with an automated system is a complex task. Toning is a natural discoloration of a coin's surface by the atmosphere over a long period. Many collectors often consider toning attractive and desirable and they tend to prefer coins with natural toning. Toning is a subjective quality that is an important factor in determining the value of a coin. Toning colors of major concern are white, copper, nickel, and gold depending upon the metal composition of the coin. However these major colors can include: red, red-brown, brown, white, full white, original color, dark color, light tone, pleasing tone, rainbow tone, unusual tone, dark fields and light fields.

- The grading of collectibles is arguably a subjective process.

- The measurement of weak strike versus wear for a given issue [42]

- The difficulties in differentiating strike through, die chips, filled dies from other characteristics and contact marks [42]. Die strikes are created during the minting process of impressing the design from a die into a planchet to make a coin. They are important because they indicate the completeness of detail (as in weak strike, full strike, etc.).

3.2 Research Overview:

In order to address the shortcomings of the earlier fully automated market grading systems of PCGS and CompuGrade this research develops a grading model with a framework for interactive human grading on subjective qualities during the final grading process. As such,

the development of the COINS model utilized in this study has been be divided into three phases:

1. Measure the grading abilities of domain experts
2. Develop the systematic framework for an expert grading model.
3. Design an interactive systematic framework for a human grading model across the Internet.

Computer programs in Java and ASP were developed as tools to test the underlying hypotheses being examined. Extensive testing was conducted on these software tools and the results obtained were carefully scrutinized. The skills and programming efforts of two Graduate teams at Pace University in White Plains, NY were utilized for the development of the software tools during the January 2002 – March 2003 timeframe [8, 28].

Outside domain experts were consulted, via email, telephone and in personal meetings during this study for their expertise on the framework of the approach, their opinions on the pregraded samples that were used to train the system, on the results; and to utilize the interactive hybrid system. The experts that participated in these various capacities include: Chris Schopfer, Jim Halperin, Sam Turgeon, David Zaniewski and Andrew Schreck.

3.3 Summarized Research Steps Employed

Phase 1: Measure the grading abilities of domain experts

An area of concern in this study was to test how well humans do at assessing the condition or grade of rare collectibles. While most experts and dealers feel that they have good grading skills the feeling of this author is that, there are often varied interpretations by the domain experts when it comes to grading.

To test this research question two Web based tools were utilized. One tool known as the Automated Rare Coin Grader was developed by a graduate student team at Pace University [28] and the other is an independently developed website known as The Grading Challenge by Stuart Miller [48]. The goal of both sites is to have a pool of graders look at coin images and to identify what they think the grades should be. These human grading models are described in considerable detail later in this body of work. This study takes the results of the data obtained from these sites and measures how well graders do in the grading process of pregraded coins.

Phase 2: Develop the systematic framework for a grading model.

The only known previous attempts at computerized grading are the disbanded commercial ventures by PCGS and CompuGrade, which date back over a decade. As both of these ventures were proprietary endeavors by the respective companies, little documentation on these systems is available and no scientific research papers were produced. The scant documentation that does exist includes some generic descriptions on inputs and outputs. No information is available about the actual processing methodology or the algorithms and

assumptions used. Interested collectors and hobbyists have contributed much of the information that is available.

Before assembling a team to develop the software tools, this researcher reviewed what had been done previously in this area. The motivation for this preliminary review was first determine if a system could in fact be developed and second to reuse whatever known best practices happened to be available from previous attempts to avoid reinventing the wheel.

The concept of the summarized grading model essentially involved doing an entire unit comparison to determine the grade. This phase yielded: a grading quiz website and a comparison program that utilizes two comparison methods: Histogram Distance and Edge Detection.

Phase 3: Design an interactive systematic framework for a human grading model that assists and enhances grade testing across the Internet.

Assembling a virtual pool of expert grading consultants was made possible by transmitting digital images of collectibles across the Internet. The web based software tools had to be constructed to address the inherent problems of browser compatibility, limited connection speeds and being non-intrusive. These problems translated into constraints, which limited the possible development platform options. The question as to which browser was likely to be used by the expert grading consultants was a concern that was addressed by taking into account that Microsoft Internet Explorer is the dominant browser used by over 80% of the users on the Internet followed by Netscape Navigator [21]. The issue of potential connection speed was another factor, as it could not be assumed that the expert graders would have fast broadband or better connectivity. Since many Internet users still use dial-up connections the online experiments had to work and be responsive in connection speeds as slow as 24K bps. Lastly, the experiments had to be designed to work with commonly available PC technology. Thus, all online software tools had to be delivered as standard web pages that did not require users to download anything special such as Flash or Java and the tools had to work nicely within the confines of the standard configurations of browsers, anti-virus software and firewall protection.

As such, the online experiments were constructed to work with both Internet Explorer and Netscape Navigator browsers and were delivered as dynamic Active Server Pages (ASP) with no JavaScript being used and no cookies being required. The benefit of using Active Server Pages is that they are constructed dynamically on the web server in visual basic scripting based on the request made via the browser by the expert grader. All of the technical assembly is done at the web server and the recipient gets a simple HTML type page delivered to their machine. Sending a smaller JPG image of the coin with each test page and then giving the grader the option of downloading a larger image if required addressed the potential problem of connection speed.

The browser-based tools drew on a database that contained stored images, machine-based grades and the test sequence for each new participant. A web based Access 2000 database was used to store all of this information. The same database was also used to record each

test interaction of the expert grader consultants as they participated in the grading experiments.

3.4 Experimental Overview & Controls – Machine Based Grading Experiments

Overview of Machine-based Experiments:

The machine-based experiments have been constructed to measure a machine's ability to perform recognition on collectible images by using an automated template-matching model. The mechanics of this test were designed to measure the template-matching model in machines visual pattern recognition capabilities [60] while capturing the overall grade assigned by the expert system of the presented scanned images. The captured grading results for each image later are compared with the results of the human graders to measure variances between the average human grades for each item against the machine grade.

Testing Methodology:

In the machine experiments, the machine-based system is presented with the scanned obverse (front) images of 20 different sample Lincoln Cents against a stored database of 105 pregraded images. The pregraded images were assigned grades by Richard Bassett and Sam Turgeon. To insure that the machine-based system was trained with reliable data the average grade between the two grades was used when the variance was less than or equal to 3.5 grade points, or 5% difference. Images with a variance larger than 3.5 grade points were rejected. The scanned images presented to the expert system represented by the majority of the spectrum of possible business strike grades in the Sheldon [64] 1 – 70 range. As each coin image was submitted to the system it returned a screen as shown in Figure 3.4.1 showing a numeric grade and a corresponding descriptive text grade. Additionally the system provided the ability to scroll through the comparative results of the analysis so that the user could determine what the closest image match was.

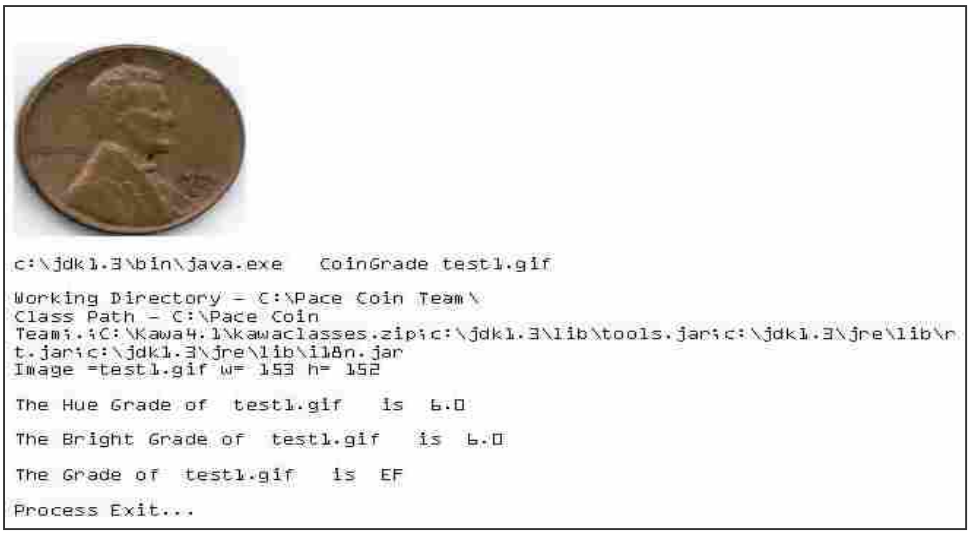


Figure 3.4.1 – Machine Grading Screen

Machine Model Components:

The grading model has three major components: Input, Process and Output (Figure 3.4.2). The component makeup of the grading model is largely transparent to the end-user unless it is malfunctioning. The proper functioning of each component of the model is critical to the successes of the experiment as it each assumes an appropriate role in handling the expert grading of the presented collectible image.

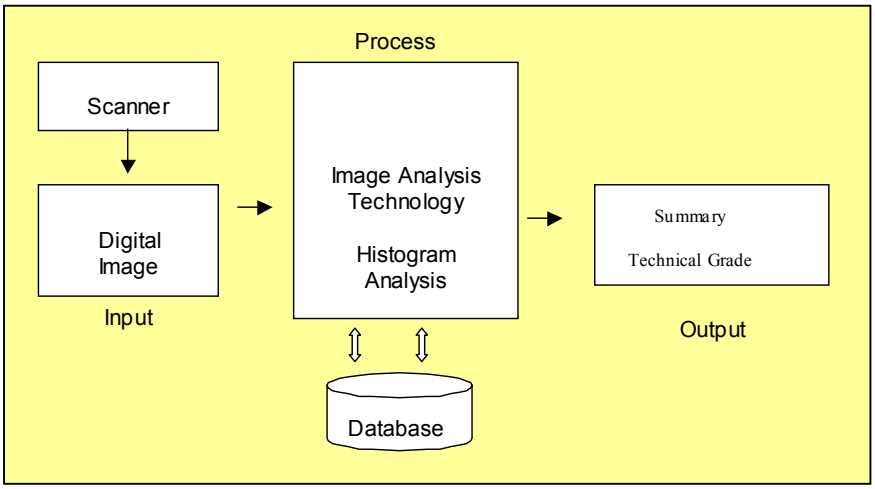


Figure 3.4.2 – Grading Model

The Input Component of the grading model provides users with the flexibility of either digitally scanning their collectible item via a GUI interface or to submitting a captured digitized image in a 256 color, 96 pixels/inch GIF format. The color and pixel requirements of the machine-based system were selected they were the formats that were most readily available of most scanners and digital cameras. The multimode nature of the input process is designed for user flexibility as some users may lack access to a scanner or lack the knowledge of how to operate a scanner.

When using the system locally, that is the software components are installed on their local hard drive, a user can scan a collectible or capture an image with a digital camera and perform the entire grading process on their own computer. The results of the grading can be privately maintained in a locally stored database.

In remote mode a user would acquire a digitized image and then submit it to a browser based processing system on the Internet. This allows users the safety of keeping their items in a private and secure setting and not bringing their rare collectible items out into a public domain. The Internet facilitates user flexibility in these areas while having the ability of obtaining appraisals or engaging in commerce by submitting their images remotely. This option is a significant convenience factor in the design consideration of this framework.

The Process Component has two distinct approaches to processing the grading request of the user: summary grading and detailed grading. Summary grading involves the examination of

the collectible as a whole unit rather than using a detailed feature set. The summarized approach, which was used in this research, is the quicker of the two but it is also the least reliable because it is subject to false readings in certain conditions. In this method the entire digitized image of the collectible item is examined by using Histogram Distance Detection comparison algorithms. Histograms are simply arrays of bins, where each bin is associated with a range of data values and contains a frequency count that is the number of occurrences of that particular data value in an input dataset. This measurement creates three internal arrays, representing hue, saturation, and brightness information of the input image.

Procedures Employed in the development and construction of this machine-based expert system:

The development of this machine-based system consists of seven distinct phases: Determine Operating Platform, Select Appropriate Algorithms, Consult Expert Graders, Training, Data/image acquisition, Testing and Automated grading.

- Determine the Operating Platform: The determination of what type of technology to be employed would be utilized was given careful consideration in the construction of the framework of this system. The system was developed in Java so that it could ultimately be transported to the largest number of operating platforms possible. For purposes of this study the machine-based system would live exclusively on a local machine, future post dissertation work in this area could extend the operating environment to the Internet in a browser-based environment.

- Selection of the Appropriate Algorithms: A number of different algorithms, including The Roberts Operator, The Sobol Operator, The Robinson Compass Masks were evaluated to match the functional requirements of the machine-based system to perform template matching [60]. These algorithms were given consideration due to the extensive amount of documentation available on them and encouraging previous research results in the area of computer vision and image processing [70]. Ultimately these algorithms, which are Edge detectors, were not selected, as they would be better suited for feature detection [70] rather than template matching. The algorithm that was selected was the Histogram Distance Algorithm based on its ability to perform template matching [17] on an entire image rather than detailed feature segmentation. The Histogram Distance Algorithm is a near-neighbor pattern-matching algorithm that converts the image being evaluated into histograms. The histograms are then viewed as vectors of the digitally scanned coin color images and are represented in three distinct areas: hue, saturation and brightness. The histogram area is measured in terms of nominal, ordinal and modulo vector values and combined into a single numeric value.

- Consult Experts Graders: Two expert graders, Richard Bassett and Sam Turgeon, graded the 105 coins, which were used to train the database. Each grader graded the training coins independent of the other graders. The average grade of all both graders was used as the grade, which was attached to the image. Coin images were rejected when grader consensus was more than 5% off from the average grade, or +/- 3.5

points on the 70-point Sheldon scale. Although this disqualification criterion may have introduced some bias, it expedited the development of the prototype of the machine-based system.

- Data/image acquisition: The accepted coins were scanned as .gif files and each image was coded with an incrementing filename in the order received. For instance, the first coin was saved as COIN1.GIF and the thirty-sixth coin was saved as COIN36.GIF.

- Train the system. The coin images to be tested by the machine-based system were matched against the average grades assigned and stored into the Java based system. Researchers doing future work in this area might want to extend the system to include an SQL database to accommodate a much larger training database with thousands of samples and to decrease the speed of evaluation. Coding the training data within the Java system was a sufficient method [8] in getting the system completed in a timely fashion, if not the most efficient approach, in executing the requirements of the machine-based tool in this research study.

- Testing: In order to ensure that the machine-based system was evaluating the images properly a series of five reliability tests was constructed. The machine-based system was tested with various combinations of images in the trained database to determine if reliability increased of decreased substantially as the number of trained images was significantly changed. After testing the machine-based system with a trained database size of 36, 55, 65 and 85 coins it was determined that ample representation of the various grade bands (Good, Fine…etc) were under represent thus the database size was increased to 105 images include better representation in the each of the major grade bands.

    a) Run 50 of the 105 pregraded training coin images through the system (and removing these 50 images from the training database leaving a net of 55 trained images) to determine the level of accuracy of the machine-system.

    b) Run 40 of the 105 pregraded training coin images through the system (and removing these 40 images from the training database leaving a net of 65 trained images) to determine the level of accuracy of the machine-system.

    c) Run 20 of the 105 pregraded training coin images through the system (and removing these 20 images from the training database leaving a net of 85 trained images) to determine the level of accuracy of the machine-system.

    d) Run each pregraded training coin image through the system (and remove this image from the training database leaving a net of 104 trained images) to determine the level of accuracy of the machine-system. This test was done to determine if there were any weaknesses in the training data by ensuring that each grade band was properly represented. For instance if only one coin images with the grade of F12 were in the trained database and then removed while being tested, the test results should not be able to return F12 as a grade

for the image being tested, thus altering the researcher to the possibility of under representation of the grade F12 within the trained database.

e) Run the 20 coins that were evaluated by the third-party grading services against the machine-based system with a training database size of 105 that had the level of accuracy (as determined by tests a, b, c & d above). This test was done to see how closely the machine-based system was able to return grades to those assigned by the third-party grading services.

Note: The images that were being tested were removed from the trained database so that exact image to images matches would not occur.

- Automated grading of coins: The machine-based system was setup to compare the extracted histogram values against a stored database of trained histogram value measurements. The image being evaluated was converted to Hue, Saturation and Brightness vectors (HSB vectors) and compared against the HSB vectors of the images stored in the training database. The template matching logic was programmed to detect the closest distance measurement between the image being evaluated and the images in the database.

  Each image in the database is represented using three primaries of the color space chosen. The most common color space used is RGB. Each color channel is quantized into m intervals. So the total number of discrete color combinations (called bins) n is equal to m^3. For example, each channel is commonly quantized into 16 intervals. Therefore, we have 4096 bins in total. A color histogram H(M) is a vector (h1, h2,..., hn), where each element hj represents the number of pixels falling in bin j in image M. These histograms are the feature vectors (indexes) to be stored as the index of the image database.

  The 24-bit RGB color for each pixel will be counted in eight intensity bins in each dimension. Each color has eight bits (255 color intensity), which are sampled. This gives a three-dimensional color vector for each pixel, with three bits representing each color dimension. Then the RGB vectors can be transformed to HSB color vectors. Histograms will then be performed by calculating the L2 norm distance between the histograms. Similarity results from HSB and RGB histogram searches will be compared, and weighting the HSB, components for better comparison are compared.

  During image retrieval, a histogram is found for the image. A calculation is then made to measure the distance between the histograms of the image, which is being evaluated, and the stored images in the database. (If images are of different size, their histograms can be normalized.) Evaluated images with the smallest distance are retrieved from the database and noted as the closest match, or nearest neighbor. [41]

Number and characteristics of Samples

Twenty sample coin images of Lincoln Cents were assembled as the experimental control group. The control group of samples was selected because of sending the twenty physical coins out to third-party grading services to be professionally graded. Once the twenty coins came back from the third-party grading services they were cracked out of their encapsulated cases and scanned. See figure 3.4.3 for what an encapsulated Lincoln Cents looks like. The images of the coins were then tested in machine reliability test e test to help determine the effectiveness of the machine-based system. The reliability test was constructed so that the machine would have the opportunity to grade all 20 unique samples and compare them to the trained database. The test was also setup to capture the amount of time that each grader spent grading a particular coin in an effort to determine if ample consideration was being given in the evaluation process.

Figure 3.4.3 – Lincoln Cent Varieties

Each coin sample used in this study is part of the Lincoln Cent family, which has been minted from 1909 to present. Only the obverse (front) of the Lincoln Cent was evaluated in all of the experiments of this study. The obverse was selected rather than the entire coin as there was a major design change that occurred in 1959 when the reverse of the coin was changed over from the "Wheat Ears" to the "Lincoln Memorial". Figure 3.4.3 provides illustrative samples of the changes in the reverse designs. Each sample was unique in that each coin was represented by a different date & mintmark combination (see Appendix B1 for a list of the samples).

Resources Used

The resources used for this study included hardware, software, graduate study teams and experts with domain knowledge.

- Hardware: With image management being central to this project, it was critical to employ the proper hardware that afforded the ability to capture high-resolution images in a clear and consistent manner. For this several Hewlett Packard scanners were used, as was the Kodak DC260 digital camera with the zoom lens option. These devices provided for 2400 x 2400 dpi and with 48-bit bit depth capacity.

- Software: Scanning software, Photoshop editor, HTML, Image Processing Tool, ASP, Java Programming Resource kit

- Graduate Study Teams: Two teams of graduate students worked under the strict direction of this author to develop specific pieces of the coin grading system and to the test results of the grading system. The graduate student teams were part of Dr. Tappert's Pervasive Computing (CS 631Q) and Software Engineering (CS 615) classes, and each team was required to work with a mentor on a significant technology project in order to fulfill the requirements of the course.

Projected Outcomes:

One research focus was on the development of a creditable outmoded coin grader. The question was whether outmoded grades tended to conform to grades provided by standard third-party grading services. In order to gain acceptance of the experts it was thought that the machine-based system should be able to grade as well as if not better than, that of the standard third-party grading services.

The overall expectation was that the machine-based grading system would yield consistent and repeatable results while producing grades on an expert grading level. The accuracy of the machine-based system was measured against the baseline grades of 20 sample coins, which was obtained from the third-party grading services.

Reliability and Validity:

Producing consistent grading results that were objective was an important consideration of this experiment. The experimental controls were put in place to reduce evaluation variables, to increase validity and to ensure that the results were reliable. The controls implemented during this experiment included:

- The machine-based system graded the same images as the expert grading consultants would in the online grading tests.

.

3.5 Machine Research Area #01: Machine Reliability

Research Questions Explored:

- Does the size of the database in the expert machine-based system have an impact on the overall grading results? Our hypothesis is that improvement results as the database increases in size.

- How well does machine-based grades match the grading of the third-party services? The null hypothesis is no correlation.

Overview of Machine Reliability Issues:

An area of interest of this research was to determine why all known attempts at commercialized automated grading systems had failed [9, 36, 44, 58]. From all sparse documentation obtainable, it appears that the failures were likely attributed to trying to get the machine-based systems to be complete graders, which were capable of returning market grades.

We soon realized the enormity of trying to develop a system that was capable of producing expert grades once the development and testing of our machine was in progress. Producing market grades required first obtaining the grade, then having the machine-based system perform subjective analysis on the coins based on the collective opinions of many human graders and then obtaining expert grader acceptance of the finished result. Given the widespread diversity of opinions [48] that humans have in grading it seemed rather unlikely that the development a machine-based system that was capable of producing widely accepted market grades was obtainable.

The issue of the machine generated grades being accepted by the expert grading consultants is viewed as a crucial point given that the lack of expert acceptance contributed to the failure of previous commercial systems. Consequently, in order to gain expert user acceptance of the machine-based grades a series of reliability tests were constructed to determine the effectiveness of the machine-based system in determining grades.

Testing Methodology:

Five separate tests were constructed (tests a – e) in order to measure the grading reliability of the machine-based system.

a) Fifty of the 105 pregraded training coin images were run through the system (while removing these 50 images from the training database leaving a net of 55 trained images) to determine the level of accuracy of the machine-system. This test used the smallest trained database size in the machine-based system with limited grade band representation.

b) Forty of the 105 pregraded training coin images were run through the system (while removing these 40 images from the training database leaving a net of 65 trained images) to determine the level of accuracy of the machine-system. This test used the second smallest trained database size with a slight increase in the grade band representation from test a. The purpose of increasing the database size and grade representation was to see if the level of accuracy increased as the database population increased.

c) Twenty of the 105 pregraded training coin images were run through the system (while removing these 20 images from the training database leaving a net of 85 trained images) to determine the level of accuracy of the machine-system. This test used the third smallest trained database size with a slight increase in the grade band representation from test c. The purpose of increasing the database size and grade representation was to see if the level of accuracy increased as the database population increased.

d) Each pregraded baseline coin image was run through the system (while removing the image under evaluation) leaving a net of 104 trained images. This test had the largest database represented of all of the tests.

e) The 20 coins that were evaluated by the third-party grading services were run against the machine-based system with a training database size of 85 that had the level of accuracy (as determined by tests a, b, c & d above). This test was conducted to determine how close the machine-based system came to grading compared to the third-part services.

Note: The images that were being tested were removed from the trained database so that exact image to images matches would not occur.

Projected Outcomes:

The preliminary version of the Pace machine-based system contained a database of just 36 coin images. The initial tests with this limited database yielded encouraging results by demonstrating that the system was capable of pattern matching on like grades [8, 28]. However, there were a considerable number of times when pattern matches were not made as close as they could be due to under-representation in the database. Machine-based systems usually perform better at their tasks with larger training databases so it was expected that the larger the database that the machine was working from the better it would perform [14, 15]. As only a single recognition algorithm was being used, the histogram distance algorithm, anomalies in the machine-based grading was expected but the exact extent of the variation was unknown at the onset.

In order for the machine-based results to be accepted by the domain community a baseline target for the level of desired accuracy being sought was set at 95% with an error rate of no more than 5%. This is a tight tolerance in terms of coin grading as a 5% error rate on the 70-

point Sheldon [64] grading scale translates to just 3.5 grading points which is generally within the same or nearest grading band.

Outcome Measurements:

The outcome measurements in the 5 reliability tests included the following data items:

Number of Coins Tested
Number of Coins in Trained Database
Average Coin Grade
Average Machine Grade
Average Grade Variance (in Points)
Sheldon Grading Scale (in Points)
Error Rate (% of inaccurate grades)
Level of Accuracy

3.6 Experimental Overview and Controls – Online Grading Experiments

The online grading experiments is a model that utilizes the results obtained in the machine-based grading model as input accompanied by interactive input supplied by the user to arrive at a consensus grade. (See Figure 3.6.1)

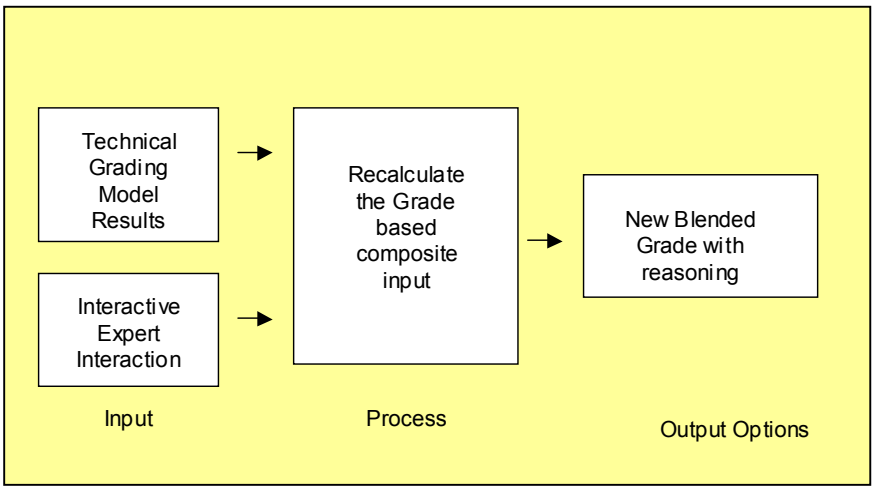


Figure 3.6.1 – Interactive Online Grading Model

The mechanics of this test were designed to combine the attributes of human visual pattern recognition [60] with the ability of the computer to produce an expert grade.

Testing Methodology:

The approach of this experiment was to allow expert graders the ability to utilize a dynamic web-based application in a browser across the Internet. This tool, was developed by this author using Microsoft Active Server Pages (ASP), is capable of presenting the human grader with a scanned image of a coin, a machine grade, a dropdown list for grade selection, subjective quality selections, comment recording (Figure 3.6.2) and the ability to store and record the selected grade into a database.

Each expert grader was to evaluate a series of 20 coin images presented to them (one at a time). The images presented were the same coin images that were evaluated by the machine-based system and that were graded by the third-party grading services. Upon completing the 20 online grading tests each expert grader was prompted for some closing data about the online grading experience and their level of expertise. (Figure 3.6.4)

In order to assign a grade in the detail feature recognition model graders must perform detailed feature matching by examining up to 20 points on the item. Graders are allowed to weight the features, which their experience tells them, are more important and they are able to exclude, or assign a weight of zero, to features that they feel are insignificant.

Humans would normally just look at several features, or chunks [47] when grading a coin. By using a graphical user interface (GUI) on the computer to do this we are able to present the graders with all possible features on every coin that they should consider.

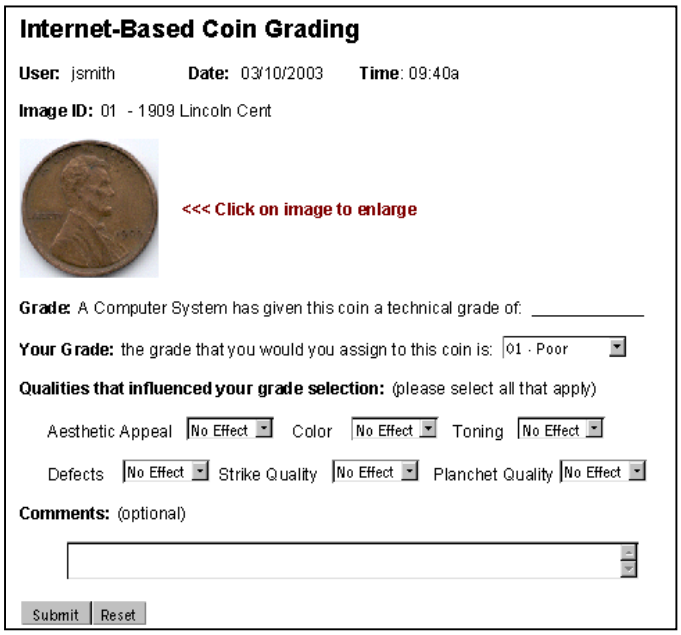


Figure 3.6.2 – Screen Shot of Online Grading Test #1 of 20

Upon completion of grade, selection the user clicks on the Next Coin Button (not shown in figure 3.6.2 above) and the grade selected is recorded into a data file. After the last coin is graded, the user gets a message indicating that the session is over at that point the system automatically emails a text file to the control email account with the all of the grading results for the entire session.

Figure 3.6.4 – Screen Shot of Online Welcome Screen

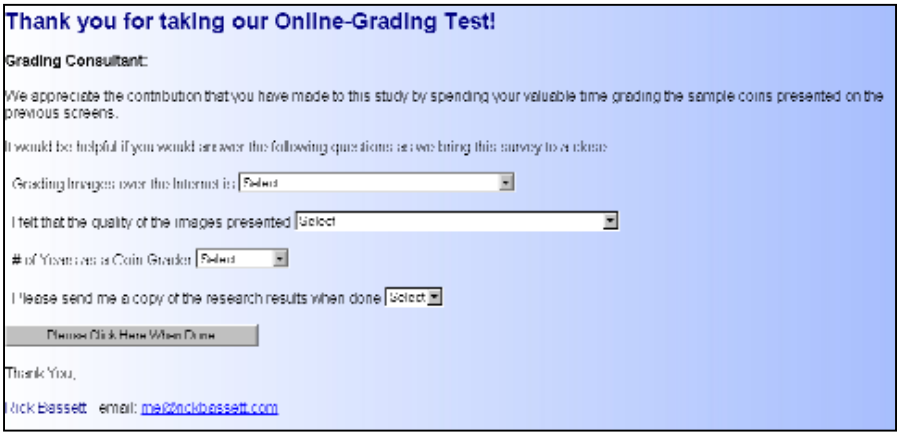


Figure 3.6.5 – Screen Shot of Online Ending Screen

Logistical Considerations:

The method of testing across the Internet was devised to decrease the amount of time required to complete testing to span the geographic distance between the expert graders and to maintain experimental consistency. The expert graders that participated in the study were located all over the country and the logistics of getting the expert graders into a single physical location for testing was not possible. Getting the physical coin samples to the expert graders was also a logistical challenge, which presented too many experimental risks.

An earlier approach that was considered, and quickly rejected, was to send the actual coins out to each of the graders. This approach was rejected based on the time requirements and the potential loss/damage risk to the sample coins. By sending the physical coins out to the graders, testing could only be done sequentially, one grader at a time, as the coins could only be in one location at any point in time. The time that it would take to perform the test on the grading would be burdened by the extra layers of overhead which included the speed of the US Postal Service (to and from the grader) and the amount of time that the grader maintained the test coins in their possession. If a grader was in California, the mailing process alone could potentially add 10 days to the amount of time before another grader could test the coins. The projected extra time from shipping could easily have added 50 days to the study as an aim of the study was to have at least 5 expert graders participate in the experiment.

Still a larger risk facing the experiment was the potential loss or damage of the physical samples. In order to ensure consistency for later result measurement and data analysis it was important that all graders graded the same coins and the all coins remained in the same condition throughout the entire experiment. By sending the samples out to the graders there was always the possibility that the samples could be lost in shipping, damaged anywhere along the way or kept by the grader. The probability of loss or damage increased substantially with each additional grader that was added to the test. The loss or alteration of the samples at any point would essentially mean that the experiment would have to stop with the results thus far received or that the entire process would have to be restarted with a new sample set.

The methodology of putting the experiment on the Internet overcame the issues of turnaround time and risk to the samples. The browser-based testing tool was capable of working interactively with numerous expert graders simultaneously while preserving the integrity of the samples.

Participants:

The 40 participants of this online grading experiment included expert coin collectors and coin dealers with considerable domain knowledge. The collective grading knowledge base of the participants represented over 200 years with an average of just over 5 years per participant. The total population of expert graders and dealers worldwide is estimated to be 5,000 to 10,000 according to expert Jim Halperin [33].

Participants for this study were located by placing postings (see appendix A1 for a copy of the postings) in the online newsgroups rec.collecting.coins, uk.rec.collecting.coins and the PCGS Forums. These newsgroups are for the discussion of all things related to coin collecting. Typical on topic subjects minimally include: coins, tokens, currency, patterns, medal collecting, news relevant to coin collecting, coin grading, cleaning, upcoming coin shows and coin history. These groups tend to attract a large diversity of coin collectors, dealers, buyers, sellers and other coins enthusiasts that have skill levels that range from novice to seasoned expert. The rec.collecting.coins has been in existence since 1994 and has a large following within the collecting community with thousands of new posts appearing each month.

Potential participants filled out a screening form (see appendix A1) for selection consideration. Screening was done in an attempt to get the most experienced graders possible for the experiment. As such the screening form requested key experience information such as the number of years as a collector, the number of years as a dealer and a URL of any coin related website that they may maintain. The request for information in the screening form was balanced against the privacy needs of collectors. In order to minimize the potential of theft most collectors discretely will not provide detailed contact information [20] which can then link their physical address to the location of their collections.

Number and characteristics of Samples

The number of samples that each participant would need grade was determined to be 20. This number was arrived at after considerable dialog with various graders over a several month period. The initial thought was to have each grader review several hundred samples each. After presenting this proposition to potential graders, the overwhelming feedback was that the grading test should take a grader no more than 30 to 60 minutes and that graders usually take 2 – 3 minutes per coin when they are carefully examining a coin. To arrive at the desired grading session window the sample set size of test 20 coins was selected with the assumption that an average time of 2 – 3 minutes would be spent grading each coin. A two hundred-coin sample could potentially take a grader 8 – 12 hours, which is considerably more time than any of the graders, said that they would be willing to contribute to this research.

Each coin sample used in this research is part of the Lincoln Cent family, which has been minted from 1909 to present. Only the obverse (front) of the Lincoln Cent was evaluated in all of the experiments of this study. The obverse was selected rather than the entire coin, as there was a major design change that occurred in 1959 when the reverse of the coin was changed over from the "Wheat Ears" to the "Lincoln Memorial". Figure 3.4.3 provides illustrative samples of the changes in the reverse designs. Each sample was unique in that each coin was represented by a different date & mintmark combination (see Appendix B1 for a list of the samples). The test was constructed so that each of the participants would get the opportunity to grade all 20 unique samples in the same order. The test is also setup to capture the amount of time that each grader spends grading a particular coin.

Projected Outcomes:

An area of concern with the online experiments was to determine how close to consensus expert graders would come to each other when grading the presented coin images with the inclusion of subjective features. The expectation was that this research in this area would yield more variation in terms of grading results than the machine-based experiments as subjective qualities were being introduced.

Outcome Measurements:

The outcome measurements were to vary depending upon the actual experiment that was being conducted. The information that was captured for use in all of the online grading experiments was:

- Coin ID
- Year & Mint Mark
- Machine Grade
- Grader Grade
- Grade Presented to Grader
- Average Grade
- Grade Difference from Machine Grade
- Standard Deviation
- IP Address of Grader
- Screen Name of Grader (see Figure 3.6.3)
- Email Address of Grader (see Figure 3.6.3)
- Subjective Qualities 1 – 6 (see Figure 3.6.2)
- Comments (see Figure 3.6.2)
- Closing Questions 1 – 4 (Figure 3.6.4)

Reliability and Validity:

Producing consistent grades was an important consideration of this experiment. Tight experimental controls were put in place to reduce evaluation variables, to increase validity and to ensure the results yielded the best possible in terms of reliability. The controls implemented during this experiment included:

- A minimum level of 3 years grading experience with Lincoln Cents was required for all graders. A person with three or more years of grading experience is usually considered an expert.

- The same coin images were used for the online grading experiment as the machine-based grading experiment in order to measure variances in the experiments.

- All graders graded the same 20 coin images.

- The same evaluation procedure for the grading test was utilized for all graders.

- There was no financial incentive for overgrading & undergrading.

- The testing interface allowed the graders to enlarge the images in double size 320x320 of the original image size of 160x160 for better viewing.

- Varied and measured machine grades were provided to the graders in advance.

- Twenty sample coins were pregraded by independent third party certification services in order to arrive at a baseline grade for measurement in the experiments.

- The methodology of putting the experiment on the Internet overcame the issues of turnaround time and risk to the samples. The browser-based testing tool is capable of working with numerous expert graders simultaneously while preserving the integrity of the samples.

Since security and theft is a significant concern for many collectors and dealers, expert graders were allowed to provide a screen name in lieu of the real name. Valid email addresses were required so that confirmations could be sent at the completion of the test. There were a number of special circumstances that would get the entire results of an online grading expert disqualified, which include:

- When graders failed to provide a valid email address or failed to respond to the confirmation email they had their results disqualified. The results of nine candidates were disqualified in this fashion.

- Graders that selected all of one grade (the same grade) for all 20-coin tests had their results disqualified. The results of six candidates were disqualified in this fashion.

- In order to eliminate outliers, graders that appeared to have less experience in grading than they represented were disqualified. This was measured by 50% or more of a graders grades being off by more than two standard deviations from the machine-grade. The results of three candidates were disqualified in this fashion.

- The graders that took less than 5 minutes or 15 seconds per coin, for the entire 20-coin test had their results disqualified as these efforts were deemed to be insufficient in time to provide a valid and through evaluation. The results of four candidates were disqualified in this fashion.

- Graders that took the test more than once, based on duplications in email address and/or IP addresses, had all results after their first series of tests disqualified. Only their first test was included. The results of six duplicate tests were disqualified in this fashion.

- Graders that graded less than five coins out of 20 had their results disqualified. The results of four candidates were disqualified in this fashion.

The overall disqualification of results for unique expert graders numbered 15 as the grading results for many of the ‘problem graders’ were disqualified on several levels while a net of 37 expert graders results were used for data analysis.

3.7 Online Research Area #01: Experts Acceptance of Advance Grade Knowledge

Research Questions Explored

- Does knowing a grading opinion in advance influence the grade given by graders?
- Does providing the correct grade yield a more consistent result than providing perturbed grades?
- Do graders accept the expert grade provided by the machine as being accurate or does their consideration of subjective factors considerably bias the results?

Hypothesis: The following hypothesis derive from these general research questions:

1) More convergence will be evident in instances when participants receive accurate guidance than when they receive no guidance.

   The test of this hypothesis rests upon the rudimentary application of the F-Distribution for inferring that the standard deviations of two populations are significantly different. To illustrate, consider coin 1, a 1919D Lincoln penny with the grade of eight (i.e. "very good" - design and legend clear but worn flat; other details worn away). Twelve graders graded this coin in the presence of this knowledge (apparently derived by the computer), while 15 others graded this coin without a hint.

   The F-test offers ground for concluding that the presence of a valid suggestion causes a convergence among the determinations made by a set of graders working independently. That the observed convergence is about the suggested grade offers further evidence that an operative process lies behind the data, not sampling artifact. This process involves the proffered suggestion being internalized to structure the perception of an ambiguous stimulus.

2) More convergence will be evident in instances when participants received perturbed guidance than when they received no guidance.

3) More convergence will be evident in instances when participants received accurate guidance than when they received perturbed guidance.

Online grading experts who believe that they are provided with a reliable machine-based grade will do better at grading that those graders that are not provided an accurate grade.

Overview of Online Experiment Methodology

All coins that are being presented to the grading consultants have been graded by the machine-based coin grading system, which was developed exclusively for this research. A common problem with earlier attempts of machine-based grading systems was the lack of acceptance of the results by the experts within the collectibles domain. When machine-based systems are used to grade coins they are best at determining the grade of a collectible as the machine can be programmed and trained on the necessary logic to evaluate the salient

features at predefined locations based on specific rules. The area that machine-based systems are not particularly good at is the evaluation of subjective qualities such as color, aesthetic appeal, strike, defects, planchet quality and toning. Subjective qualities are opinions of the non-technical areas of the collectible item that are extremely difficult to quantify as opinions on these subjective qualities vary greatly from expert to expert and are influenced by many factors such as years of experience, personal preference, and expert knowledge in a particular series, physical abilities and market demand.

The grading results of machine-based system were presented to the expert grading consultants with machine-based grades, no grades and deliberately perturbed information to determine if knowing the a grading opinion in advance influenced the opinion of the expert.

Thirty-two of the participants went through the experiment in its entirety, but an additional five participants did not complete the full set of grading. Thus, somewhat more data is available for analysis than indicated above.

Testing Methodology

For experimental purposes, the grades given to the expert grading consultants were rotated each time a new grader arrived at the web site. The first grader that signed into the system got machine test 1, the 2nd got known test 2 and the third got known test 3 and the 4th gets cycled back to known test 1 again. This rotation is done so that each user would evaluate coin images with actual machine grades, with unknown grades and with perturbed grades. The following Table 3.7.1 shows how this rotation occurred. Thus, the experiment was constructed so that each grader was evaluating coins with varied information provided.

Grading Consultant Sign-in Order	Machine or Service Grade Provided	No Machine Grade Provided	Perturbed Machine Grade Provided	Duplicates
1	Coins 1 – 6	Coins 7 – 12	Coins 13 - 18	2 (5,7)
2	Coins 7 – 12	Coins 13 – 18	Coins 1 – 6	2 (5,7)
3	Coins 13 – 18	Coins 1 –6	Coins 7 -12	2 (5,7)

Table 3.7.1 – Rotation Order of Grades Presented

This experiment was designed to present the expert graders with an image of the coin, a machine grade, choices for subjective qualities and the ability to assign a grade of their own from a series of drop down choices. (See Figure 3.6.2) Expert graders were all under the impression that the grade, which they were being presented, was the machine-grade.

Detail of Grades Presented

Coin ID	Image	Yearmm	Machine Expert Grade	Perturbed Grade	Test 1	Test 2	Test 3
01	COIN11	1919D	8-VG	4-G	MG	PG	NG
02	COIN03	1911	6-G	12-F	MG	PG	NG
03	COIN19	1941	12-F	8-VG	MG	PG	NG
04	COIN21	1946	53-AU	60-MS	MG	PG	NG
05	COIN60	1951D	15-F	10-VG	MG	PG	NG
06	COIN61	1952	45-EF	55-AU	MG	PG	NG
07	COIN39	1944	53-AU	45-EF	NG	MG	PG
08	COIN46	1946S	60-MS	63-MS	NG	MG	PG
09	COIN32	1954	12-F	6-G	NG	MG	PG
10	COIN82	1968D	63-MS	65-MS	NG	MG	PG
11	COIN31	1953S	6-G	10-VG	NG	MG	PG
12	COIN23	1947S	12-F	20-VF	NG	MG	PG
13	COIN74	1959	45-EF	30-VF	PG	NG	MG
14	COIN16	1935S	25-VF	30-VF	PG	NG	MG
15	COIN40	1944D	55-AU	50-AU	PG	NG	MG
16	COIN25	1949	40-EF	50-AU	PG	NG	MG
17	COIN17	1940	12-F	10-VG	PG	NG	MG
18	COIN22	1946S	53-AU	60-MS	PG	NG	MG
19 *	COIN60	1951D (05)	15-F	10-VG	NG	MG	HL
20 *	COIN39	1944 (07)	53-AU	45-EF	MG	HL	NG

Where MG = Machine Grade provided
PG = Perturbed (or misleading) Grade provided
NG = No Grade provided
* = Duplicates

Table 3.7.2 – Detail of Grades Presented

Table 3.7.2 identifies the pattern detailed pattern rotation in which the expert graders were presented the grades in which to evaluate.

Expert graders would login into the initial introduction page (see figure 3.6.3) and provide their screen name and email address, the system would also capture their IP address. The IP address was used as a secondary method of validation to determine if the same person was taking the tests multiple times with different screen names. After completing the initial introduction page, they would be directed to the first coin test page (see figure 3.6.2) where they would select a grade and provide feedback on the coin that they were evaluating based on the information provided. Upon the completion of the first coin, they would be directed to the second coin page to repeat the process until they completed 20 coins grading tests. After the twentieth coin, the grader was directed to the completion page (see figure 3.6.4) where they answered four quick wrap up questions.

Projected Outcomes:

1) Expert graders often disagree with the grades provided by third-party grading services [65]. It was suspected that being provided a grade in advance would bias or influence an expert graders opinion of the grade that they would assign to a coin. The system of providing some accurate grades, some inaccurate grades and no grades was devised in order to eliminate potential bias and to really draw the expert out in terms of their own grading abilities.

2) Since the machine-based system was designed to produce an expert grade, it was expected that expert grading consultants would take exception with the opinion of the machine-based grading result and apply their own subjective evaluation to the grade and thus changing the grade of the coin. What was unknown was the extent to which

expert graders would change the grades and if the changes would be more liberal or more conservative than those of the machine-based grade.

Outcome Measurements:

As each expert grader evaluated the coin images, the following data was being captured during this experiment:

- Coin ID
- Year & Mint Mark
- Machine Grade
- Grader Grade
- Grade Presented to Grader
- IP Address of Grader
- Screen Name of Grader (see Figure 3.7.3)
- Email Address of Grader (see Figure 3.7.3)
- Comments (see Figure 3.7.2)
- Closing Questions 1 – 4 (Figure 3.7.4)

The expert grade, which was presented to the grader in this experiment, was key as the experiment was designed to measure mean differences between the expert grade and the grade, which they assigned.

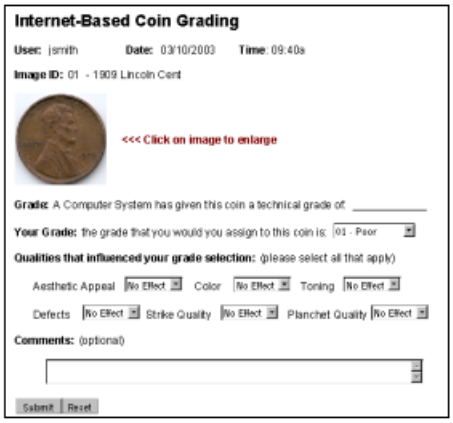


Figure 3.7.3 - Screen Shot of Online Grading Test #1 of 20

Figure 3.7.4 - Screen Shot of Online Welcome Screen

3.8 Online Research Area #02: Perception of Internet based Grading

Research Question:

- Was the Internet an effective vehicle for grading?

Hypothesis: Expert graders would see the value in grading images over the Internet although they would rather grade from the actual coins. The graders that participated in this test are predisposed to the use of technology as they are on the Internet.

Overview of Internet based Grading:

The methodology of putting the experiment on the Internet overcame the issues of turnaround time and risk to the samples. The browser-based testing tool is capable of working with numerous expert graders simultaneously while preserving the integrity of the samples.

An earlier approach that was considered, and quickly rejected, was to send the actual coins out to each of the graders. This approach was rejected based on the time requirements and the potential loss/damage risk to the sample coins. By sending the physical coins out to the graders, testing could only be done sequentially, one grader at a time, as the coins could only be in one location at any point in time. The time that it would take to perform the test on the grading would be burdened by the extra layers of overhead which included the speed of the US Postal Service (to and from the grader) and the amount of time that the grader maintained the test coins in their possession. If a grader was in California, the mailing process alone could potentially add 10 days to the amount of time before another grader could test the coins. The projected extra time from shipping could easily have added 50 days to the study as an aim of the study was to have at least 5 expert graders participate in the experiment. Still a larger risk facing the experiment was the potential loss or damage of the physical samples. In order to ensure consistency for later result measurement and data analysis it was important that all graders graded the same coins and the all coins remained in the same condition throughout the entire experiment. By sending the samples out to the graders there was always the possibility that the samples could be lost in shipping, damaged anywhere along the way or kept by the grader. The probability of loss or damage increased substantially with each additional grader that was added to the test. The loss or alteration of the samples at any point would essentially mean that the experiment would have to stop with the results thus far received or that the entire process would have to be restarted with a new sample set.

Testing Methodology

Upon completion of the 20 coin tests, each grader was asked a few summary questions of their Internet based grading experience (see figure 3.6.4). These questions were focused first on the graders comparative perception of grading coins over the Internet verses grading the physical coins and second on the quality of the digital images.

An additional test was constructed which involved having a small population of the expert graders grade the physical coins in order to further test the hypothesis. The grading experts would see the value in grading images over the Internet as opposed the actual coins. An area concern of this test was to determine if there were substantial differences in the online grading results compared to the results of grading the physical coins in person. This test required face-to-face meetings with five of the expert graders that participated in the online tests. They were shown the same 20 coins in the same order with the same grades disclosed as they were shown previously. A proper sample size of five experts graders was chosen based on the population size of thirty-seven expert graders, which participated in the online research. There was nothing remarkable about the qualifications of the five graders selected other than their geographic location of being on in the Northeast United States which made visiting them more practical for the research.

Projected Outcomes:

Many graders often complain of how difficult, or inconvenient, it is for them to grade slabbed, or encapsulated, coins and many threads in the PCGS Form and the rec.Coin.Collecting newsgroups often echo the same some complaints about grading coin images in online auctions. It is commonplace to see comments such as "it is extremely difficult to grade anything other than the physical raw unslabbed coin" in these online community forums. With this advance knowledge, this researcher didn't expect to change a rather popular and often voiced opinion of the collectors and dealers.

The expectation was that a large number of expert graders would say that they would rather grade the physical coin instead of a coin image. The purpose of this experiment was to gain a quantifiable reference point to this sentiment.

Outcome Measurements:

As each expert grader evaluated the coin images, the following data was being captured during this experiment:

- IP Address of Grader
- Screen Name of Grader (see Figure 3.7.3)
- Email Address of Grader (see Figure 3.7.3)
- Closing Questions 1 – 4 (Figure 3.7.4)
- Comparative results for the five graders, which did online grading and physical coin grading.

Closing questions #1 and #2 were the primary basis for the analysis of this experiment. The first question asked the experts how grading images on the Internet is compared to grading physical coins. The second question asked the experts about the quality of the images presented.

3.9 Online Research Area #03: Grader Consensus

Research Question Explored:

- To what extent is there a convergence of opinion among human graders that use the online grading system?

Hypothesis: The expert grading consultants will have widely diverse grading opinions on the images that they are grading. Informal experience that humans grade inconsistently is supported by the Stujoe grading tests [48] and the Kevin Foley tests [33] on human grading.

Overview

An interesting area of examination in this study was comparing how the grading experts did against each other when assigning grades. The Stujoe grading challenge [48] was the basis for suspecting that graders of all experience levels have diverse opinions. Commercial transactions of collectibles also indicate that buyers and sellers have different interpretations of grades when financial incentives are involved.

Expert graders are known to use either or both technical or market grading when assessing collectibles. The method of grading, technical or market is usually not obvious nor is it always identified with the grade that is assigned. The values for grades are also not published based on technical or market grades, which also contributes to the subjectivity in pricing.

Testing Methodology

This experiment was designed to measure the consensus amongst graders during their participation in the online experiments Within this experiment, we captured the years of experience data and the identity of the graders to ensure that we knew something about the experts participating in the study. Part of the problem with the Stujoe grading challenge [48] is the lack of experimental controls in place allowing anyone to take the challenge as often as they like with no way to determine the skill level, or the level of seriousness, of the grader. In this study, inexperienced graders and thumpers, those that selected all of one grade for all of the coins that they graded, were disqualified and their input was not included in the final tabulations. The banded results of all the qualified expert graders were compositely compared to determine the extent of which agreement amongst the expert graders exists.

Known Issues with Grader Consensus:

A person only has to read a handful of posts on the PCGS Forum [54] or on the newsgroup rec.collecting.coins to get a sense of the wide diversity in grading opinions that graders have. There are frequent posts in both of these popular venues for coin collectors and dealers, which take to task the grades of coins, which are offered for sale on the Internet. A large number of these heated threads center on offerings that are made on eBay.com. This experience combined with the documented diversity in grading shown in Table 2.7.1, the

results of Stujoe Grading Challenge [48] have led this researcher to expect that there would be considerable diversity amongst the expert graders in this experiment despite their expert status level. This experiment documented the range of grades that experts assign to a collectible and to show how wide the range of interpretation between all of the experts is.

Outcome Measurements:

As each expert grader evaluated the coin images the following data was being captured during this experiment:

- Coin ID
- Year & Mint Mark
- IP Address of Grader
- Machine Grade
- Grader Grade
- Screen Name of Grader (see Figure 3.7.3)
- Email Address of Grader (see Figure 3.7.3)

The frequency range, the smallest grade and the largest grade, were obtained from these data points and compared to the machine-based grade to determine how wide the extreme grading points were. During the data analysis component of this research the range was used in the computation of the variance and the standard deviation and analysis of variances using the F-Test were performed.

3.10 Online Research Area #04: Duplicate Grades

Research Question Explored:

- How did grading experts do with the duplicates?

Hypothesis: It is observed that humans provide widely divergent grading opinions [33, 48] therefore it is expected that graders will offer differing opinions, but not significant, when grading the coin images which they have previously graded.

Overview:

This experiment was designed to determine if there would be any variation in grading if duplicate coins were presented to the expert graders without the identification of duplicates to the expert graders.

Testing Methodology

Coins #19 & #20 were duplicates of coins #05 & #07. The only change on the screens for these grading tests was the machine-grades that the graders saw. The online tester scrambled the grades that the expert graders saw on tests #19 and #20 based on what they saw in tests #05 & #07. For instance if the grader was originally presented with a reliable machine for coin tests #05 or #07 then they were presented with either a misleading grade or a no-grade for tests #19 or #20.

In addition to the 18 coin viewings and grading discussed above, there was a 19th and a 20th. For these, participants were shown a coin they had dealt with earlier. The purpose was to investigate the stability of individuals' grading standards.

Projected Outcomes:

This experiment was established to measure both the expert graders consistency in grading the same coins as done previously and to see how varying the machine-grade presented would affect the grade that an expert would assign to a previously graded coin. The expectation was that there would be a certain amount of change that occurred in the grading of the duplicates but the hope was that the change, or variance, would not be significant. This test was also used to determine if the expert was really an expert as the credentials of the expert was more carefully scrutinized if there was a significant amount of variance in the grading of the duplicate coin images.

Outcome Measurements:

As each expert grader evaluated the coin images the following data was being captured during this experiment:

- Coin ID
- Year & Mint Mark

- Machine Grade
- Grader Grade
- Grade Presented to Grader
- Average Grade
- Grade Difference from Machine Grade
- Standard Deviation
- IP Address of Grader
- Screen Name of Grader (see Figure 3.7.3)
- Email Address of Grader (see Figure 3.7.3)

The two measurements that were most important in the data analysis of this experiment were the difference in the average expert grade of the grader from the machine grade and the differences in standard deviation as they were used to determine how the population of expert graders behaved in their evaluations in terms of data clustering.

Chapter 4 – Results

4.1 Machine-based Grading Experiments

This research initially had a very early target of producing a fully automated coin grader. However an extensive review of the literature revealed that even it were possible to produce such as system it would probably not gain the acceptance of the collectibles' community. The ever-raging debate by experts and novices over whether coin grading is an art or a science seems to overshadow the potential success of a fully automated grader.

The area that most experts agree on is that there is great inconsistency in grading rare collectibles, including those that are graded by third-party grading services [25, 33, 65]. Some grading services use the ANA standards as a baseline while other services use their own standards. Some services are considerably more liberal in their grading than others. Consequently the ability to trade rare encapsulated coins on a sight unseen basis is risky. The grading of raw unencapsulated coins experiences the same problems of inconsistent grading. Still expert dealers and collectors would rather purchase an encapsulated coin than a raw coin if the slab has come from one of the top grading third-party grading services. A popular saying in the collecting community is "buy the coin not the slab"[31]. What this statement means is: look at the coin and make your own grade determination, don't accept the grade on the slab.

The differences in interpretations are demonstrated in section 4.6 of this study where graders often apply their own subjective measurements within the grading process. The subjective qualities that matter to each grader vary with their experience, expertise of a series and personal preferences.

The long-term aspiration of a machine-based grading system is to return a consistent expert grade that has a high level of reliability and ultimately can become a baseline grade for human grading. The results of this research demonstrates that the machine-based system, developed for this study, has a reliability rating in excess of 95% when it is trained with a sufficient representation of coins in the 1 – 70 Sheldon scale. While these results are encouraging, still better results are likely if the number of evaluation algorithms and the number of sample images in the trained database [15, 30] were increased.

The results of the online Research Area #2 (Advance Grade Knowledge – Section 4.4) demonstrate that graders that have advance of a grade do better at grading than those graders without advance knowledge. With this result, one could hypothesize that an expert grader could do even better at grading if they had a high degree of confidence in the baseline grade that they are starting. As experts often doubt that the actual grade of a collectible in a slab is the real grade, they usually do feel that it is close if the slab comes from one of the more reliable third-party grading services. This gives experts a more comfortable starting position in grading. Relying on a select few grading services would be fine if they also weren't subject to inconsistent grading over time but they are. Over the past decade and a half, grading services have become more liberal with their grades and coins that were graded by

the major grading services often receive better grades now if they are cracked out of their slabs and resubmitted.

All of the coin samples used in the online grading experiments of this study was first graded by the machine-based system. The machine-grade became the baseline grade. These samples were also sent out to several third-party grading services for a market opinion on their grades. This provided afforded the online experiments a reasonable baseline grade to start with. It also provided a benchmark for statistical measurement of the graders accuracy. Operators of a future version of the machine-based system probably wouldn't want to incur the expense of sending their samples out for a second opinion to grading services should this system ever evolve into a production system.

Machine Research Area #01 (Machine Reliability – Section 4.2) was designed to test the reliability of the machine. It is actually a series of five small tests combined into a larger experiment. The tests varied the number of trained items in the database and the number of samples in order to determine the strengths and weaknesses of the machine-based system. The reliability rating of over 95% in test e (Machine Reliability – Section 4.2) was particularly significant since these were the coins used as the sample set in the online experiments.

4.2 Machine Reliability Testing

Questions Addressed:

How reliable is the machine-based grader developed for this study?
This can be divided into two hypotheses.

- Does the size of the database in the machine-based system have an impact on the overall grading results? Our hypothesis is that improvement results as the database increases in size.
- How well does machine-based grades match the grading of the third-party services. The null hypothesis is no correlation.

Projected Outcomes:

A machine-based system usually performs better at the tasks with a larger training database. [14, 15] Thus, it is hypothesized that the larger the trained database that the machine was working from the better it would perform. This hypothesis is worthy in as much as anomalies could result from the histogram distance measurement algorithm, and such anomalies would detract substantially from the utility of the machine-based system. As only a single recognition algorithm was being used, histogram distance measurements, anomalies in the machine-based grading were expected but the exact extent of the variation was unknown at the onset.

In order for the machine-based results to be statistically significant, the desired level of accuracy being sought in the reliability tests needed to be at least 95% with error rate of no more than 5%. This is a tight tolerance in terms of coin grading as a 5% error rate on the 70-point Sheldon [64] grading scale translates to just 3.5 points.

Results:

Machine reliability tests A – C involved increasing the size of the trained database incrementally as a method of measuring if increasing the number of items in the trained database had a positive or negative effect on the overall reliability of the machine-based system.

Test A Results:

Fifty coin images were randomly pulled out of the training database leaving 55 images remaining in the machine-based system. These 50 images were then run through the machine-based system one at a time, with the results of 20 coin images (40%) coming back with grade matches while 30 images (60%) were not matched exactly. The average grade variance in this test was 5.70 points, which was calculated by taking the total of all the differences between the expected grade and the grade the machine returned (171) and then dividing that by the number of incorrect grades (30), thus 171 / 30 = 5.70 points. The correlation between the machine grading results and results of the

expert third party grading services results were .967 with an explained variance, or r**2, of .935 while the error rate was .065.

Test B Results:

Forty coin images were randomly pulled out of the training database leaving 65 images remaining in the machine-based system. These 40 images were then run through the machine-based system one at a time, with the results of 16 coin images (40%) coming back with grade matches while 24 images (60%) were not matched exactly. The average grade variance in this test was 4.71 points, which was calculated by taking the total of all the differences between the expected grade and the grade the machine returned (113) and then dividing that by the number of incorrect grades (24), thus 113 / 24 = 4.71 points. The correlation between the machine grading results and results of the expert third party grading services results were .979 with an explained variance, or r**2, of .958 while the error rate was .042.

Test C Results:

Twenty coin images were randomly pulled out of the training database leaving 85 images remaining in the machine-based system. These 20 images were then run through the machine-based system one at a time, with the results of 9 coin images (45%) coming back with grade matches while 11 images (55%) were not matched exactly. The average grade variance in this test was 4.09 points, which was calculated by taking the total of all the differences between the expected grade and the grade the machine returned (45) and then dividing that by the number of incorrect grades (11), thus 45 / 11 = 4.09 points. The correlation between the machine grading results and results of the expert third party grading services results were .987 with an explained variance, or r**2, of .974 while the error rate was .026.

Test A-C Results:

Table 4.21 summarizes the results of reliability tests A – C and demonstrates the larger the trained database that the machine was working from the better the machine would perform in terms of an increasing explained variance and a decreasing error rate.

Machine-Based Grading Tests - Machine Reliability

	Test A	Test B	Test C
# Coins Tested	50	40	20
# Coins in Trained Database	55	65	85
Average Expert Coin Grade	36.6	37.0	34.3
Average Machine Coin Grade	37.0	36.5	35.5
Expert to Machine Mean Difference	3.4	2.8	2.3
Correlation between machine & expert grades (r)	0.967	0.979	0.987
Explained Variance (r**2)	0.935	0.958	0.974
Error Rate (unexplained variance, 1-r**2)	0.065	0.042	0.026

Table 4.2.1 – Machine-Based Reliability Grading Results Tests A-C

Figure 4.2.2 supports the hypothesis that grading results improve with an increased trained database size. The explained variance (level of accuracy) increases and the error rate decrease as the number of items in the trained database is increased.

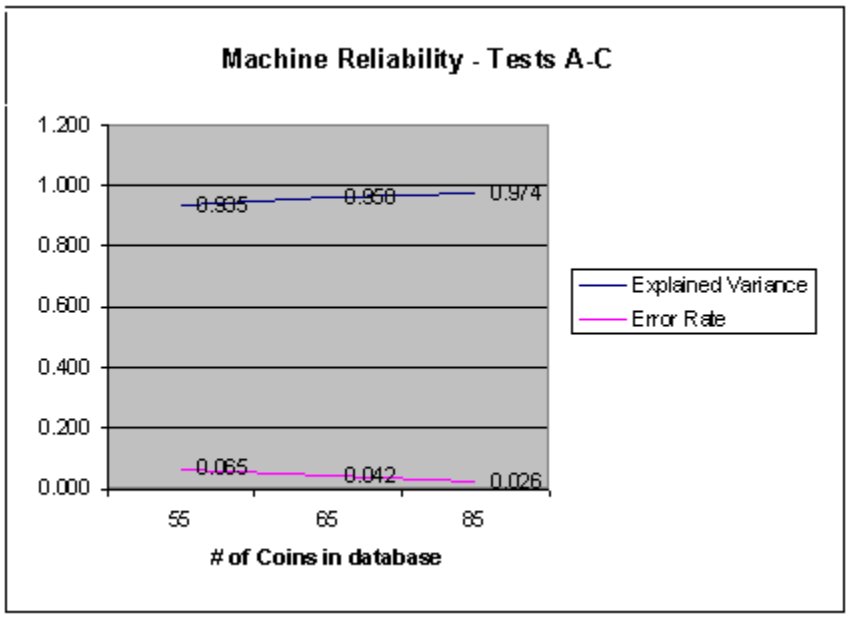


Figure 4.2.2 – Machine Reliability Tests A-C

Test D Results:

Test D was the most exhaustive and the most rigorous test as every coin image was pulled out of the training database one at a time leaving 104 images remaining in the machine-based system thus exploring every possible existence of coins across the Sheldon 70-point spectrum. These 105 images were then run through the machine-based system one at a time, with the results of 54 coin images (51.43%) coming back with grade matches while 51 images (48.57%) were not matched exactly. The average

grade variance in this test was 3.08 points, which was calculated by taking the total of all the differences between the expected grade and the grade the machine returned (157) and then dividing that by the number of incorrect grades (51), thus 157 / 51 = 3.08 points. The correlation between the machine grading results and results of the expert third party grading services results were .991 with an explained variance, or r**2, of .982 while the error rate was .018.

Table 4.2.3 summarizes the results of reliability tests D which continues to demonstrate the larger the trained database that the machine was working from the better the machine would perform in terms of an increasing explained variance and a decreasing error rate.

Machine-Based Grading Tests - Machine Reliability

	Test D
# Coins Tested	105
# Coins in Trained Database	104
Average Expert Coin Grade	34.9
Average Machine Coin Grade	35.0
Expert to Machine Mean Difference	1.5
Correlation between machine & expert grades (r)	0.991
Explained Variance (r**2)	0.982
Error Rate (unexplained variance, 1-r**2)	0.018

Table 4.2.3 – Machine-Based Reliability Grading Results Tests D

Test E Results:

The twenty coins that came back from the third-party grading services were run against the database of 85 images in the machine-based system. These images were then run through the machine-based system one at a time, with the results of 11 coin images (55.00%) coming back with exact grade matches while 9 images (45.00%) were not matched exactly. The average grade variance in this test was 3.11 points, which was calculated by taking the total of all the differences between the expected grade and the grade the machine returned (28) and then dividing that by the number of incorrect grades (9), thus 28 / 9 = 3.11 points. The correlation between the machine grading results and results of the expert third party grading services results were .995 with an explained variance, or r**2, of .990 while the error rate was .010.

Table 4.2.4 summarizes the results of reliability tests E which continues to demonstrate the larger the trained database that the machine was working from the better the machine would perform in terms of an increasing explained variance and a decreasing error rate.

Machine-Based Grading Tests - Machine Reliability

	Test E
# Coins Tested	20
# Coins in Trained Database	85
Average Expert Coin Grade	31.3
Average Machine Coin Grade	32.2
Expert to Machine Mean Difference	1.4
Correlation between machine & expert grades (r)	0.995
Explained Variance (r**2)	0.990
Error Rate (unexplained variance, 1-r**2)	0.010

Table 4.2.4 – Machine-Based Reliability Grading Results Tests E

Test A-E Results:

Table 4.2.5 summarizes the results of reliability tests A – E and demonstrates the larger the trained database that the machine was working from the better the machine would perform in terms of an increasing explained variance and a decreasing error rate.

The predictive ability of the machine-based system is particularly impressive when compared to the spread in coin grades likely to emerge among different experts or services. The Stujoe site, for instance, documents a normal distribution of grades with tails that span 70% of the Sheldon continuum. While a particular expert may disagree with the grade assigned by the machine-based system, our system remains consistent relative to the expertise it encapsulates.

The machine-based system returned a level of accuracy of at least 93.5% on all reliability tests, which were conducted when there were 50 or more trained images in the database. Table 4.2.5 shows the summarized results of all tests that were conducted; the detail of all tests is available in Appendix F.

Machine-Based Grading Tests - Machine Reliability

	Test A	Test B	Test C	Test D	Test E
# Coins Tested	50	40	20	105	20
# Coins in Trained Database	55	65	85	104	85
Average Expert Coin Grade	36.6	37.0	34.3	34.9	31.3
Average Machine Coin Grade	37.0	36.5	35.5	35.0	32.2
Expert to Machine Mean Difference	3.4	2.8	2.3	1.5	1.4
Correlation between machine & expert grades (r)	0.967	0.979	0.987	0.991	0.995
Explained Variance (r**2)	0.935	0.958	0.974	0.982	0.990
Error Rate (unexplained variance, 1-r**2)	0.065	0.042	0.026	0.018	0.010

Table 4.2.5 – Machine-Based Reliability Grading Tests A - E

Figure 4.2.6 supports the hypothesis that grading results improve with an increased trained database size. The explained variance (level of accuracy) increases and the error rate decrease as the number of items in the trained database is increased.

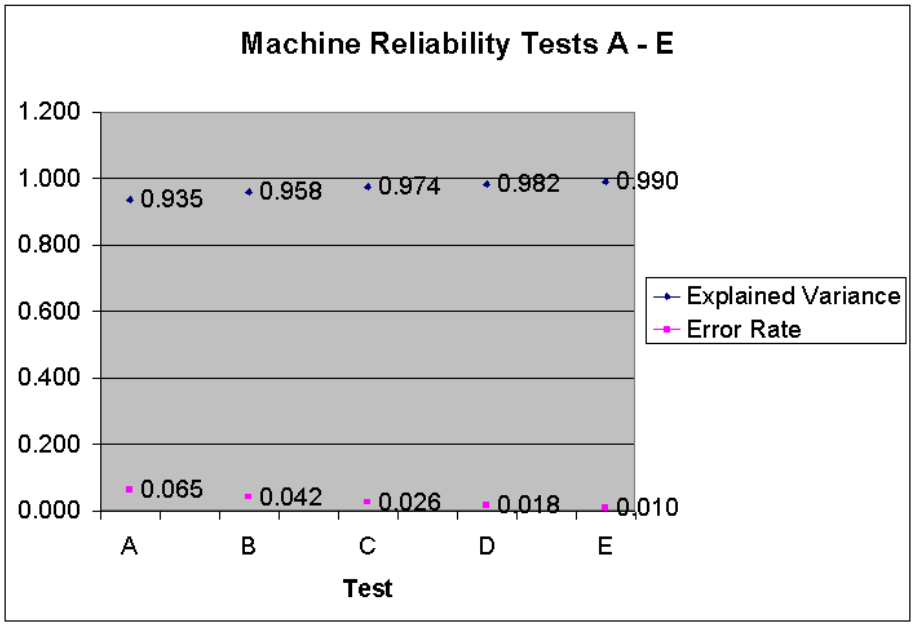


Figure 4.2.6 – Machine Reliability Tests A-E

Recommendations:

Extend the number of trained images in the database. Adding several hundred more properly graded images would probably increase the percentage of accuracy a few percentage points but the real gains would probably come in the area of the percentage of coins that were exact matches.

Employ multiple grading algorithms. Presently just the histogram distance measure algorithm was used for this research with impressive accuracy results. Perhaps by adding several additional matching algorithms the results could be increased by a using an interpolation of the results. Stand-alone machine-based systems that used a single new algorithm would need to be constructed and tested first to measure reliability.

4.3 Online Grading Experiments

The online grading experiments were designed to give expert grading consultants the ability to look at digitized images of collectible items over the Internet and assign grading opinions to them. The format of these experiments has undergone two previous generations of development and redesign before arriving at the current format. The input of coin collectors, dealers, grading consultants and casual observers was sought in order to construct the best experiments possible with each generation of online grading web sites.

The online grading experiment investigates the persuasive impact of computer-generated Sheldon scores on experienced human graders. The impact of mechanized grades is an issue of interest inasmuch as these may not always be accurate. For instance, a pattern matching program, or an expert system, may be applied beyond its appropriate domain. For instance, a system capable of providing accurate grades for Lincoln Cents would not do well with Morgan Dollars. The program could be deliberately skewed in favor of buyers or sellers (e.g. to report scores somewhat exceeding what might be expert consensus).

- How susceptible to the bias of an expert grade are experienced graders likely to be?

- How susceptible are experienced graders likely to be to a perturbed suggestion?

As previously established, ambiguity is inherent to the human hand-and-eye process of rating coins; and where objective anchoring is lacking, even arbitrary cues carry sway.

Each of 32 experienced graders was asked to rate twenty coin images on the Sheldon Scale, the continuum running from 1 to 70 for ranking the quality of coins as collectables. The images were displayed on a computer monitor via the Internet. In a randomized fashion, ratings were made in three different conditions:

- A third of the time an expert grade was furnished along with the coin's image

- A third of the time a perturbed expert grade was furnished along with the coin's image; for instance, on such a trial a coin with a true grade of 8 might be accompanied with an alleged grade of 4

- A third of the time the coin's image was displayed with no grading guidance, neither accurate nor perturbed

The overarching hypothesis is that the disparity among the ratings of an ad hoc coin declines in the presence of suggestion. The convergence of ratings about a mean is tighter when the suggestion is accurate than when error is introduced to cast doubt upon its credibility.

Online Grading Website:

The third generation online grading experiment was published on the Internet on March 30, 2003 at which time postings were made in the rec.coin.collecting newsgroup and the PCGS forum, which invited expert graders to participate research experiments. (See appendix A1 for the posting).

Interested expert graders were directed to a Web site where the expert graders provided their screen name and email address before being presented with a series of 20 grading tests in which they were required to provide their opinion of what they though the grade should be. Figure 4.3.1 shows coin screen 1 of 20. The graders were prompted for some closing questions at the completion of the 20-coin test. The online grading experiments were active for 10 days in which time 51 graders participated. It was determined that approximately 14 of the 51 participants were not experts and their results were discarded and disqualified. This still left a pool of 37 expert grading consultants, which were 25 more than originally planned. The criteria for disqualification documented in Section 3.5 of this study.

The design of this web based software tool was constructed to address the inherent problems of browser compatibility, limited connection speeds and being non-intrusive. As such, it was designed to work with both Internet Explorer and Netscape Navigator browsers and was delivered as dynamic Active Server Pages (ASP) with no JavaScript being used and no cookies being required. The benefit of using Active Server Pages is that they are constructed dynamically on the web server in visual basic scripting based on the request made via the browser by the expert grader. All of the technical assembly is done at the web server and the recipient gets a simple HTML type page delivered to their machine. Sending a smaller JPG image of the coin with each test page, and then giving the grader the option of downloading a larger image if required, addressed the potential problem of connection speed.

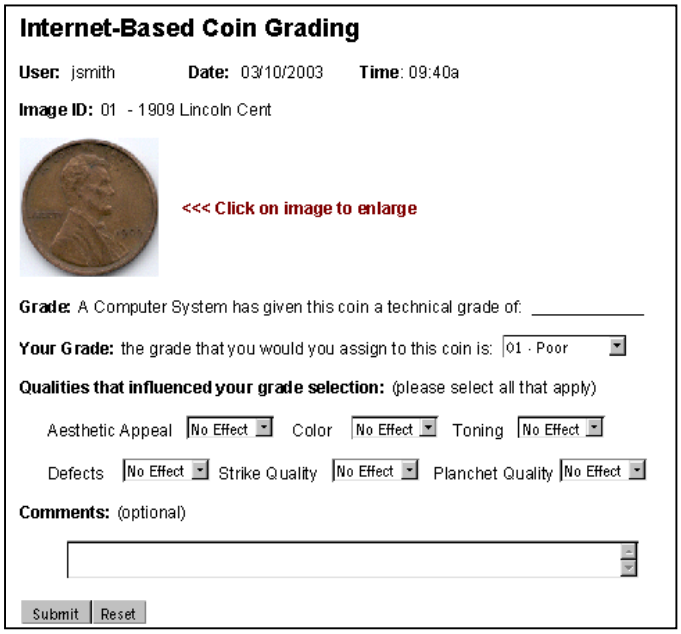


Figure 4.3.1 –Current Generation Online Grader

The results of all of the graders were captured as each coin image was graded and written into an Access 2000 database and data mined in Visual FoxPro then summarized and statically tabulated in Excel.

Rationale for the approach to the Data Analysis

The F-test does not constitute unequivocal support for pronouncing the observed difference in judgment clustering between the two conditions to be significant. Yet, the large value of the ratio, in light of test's robustness, definitely conveys the idea that the presence of the expert's grade affects the grader's opinion. This explains why we have taken mainly a descriptive approach to the data analysis in this group of experiments. The F-test, which is the foundational tool for showing that the presence of a computer-generated coin rating exerts a persuasive impact upon experienced graders working alone (i.e. apart from the forces of group dynamics), can only be used as a sign; it cannot be used as an inferential litmus test. Other methodological complications also point us in the direction of a hypothesis generating as opposed to a hypothesis-testing approach. One is that the Sheldon scale, although easily treated as interval, is actually ordinal. This excludes all the most common, powerful parametric statistics from use. For instance, very fundamentally, rigor precludes testing a set of Sheldon scores for normality because normality depends upon equal conceptual increments in equally sized spans of the spectrum. The t-test, the F-test, and the analysis of variance are, according to strict standards, inappropriate.

Another complication resides in the experimental design. Participants were asked to perform twenty coin ratings consecutively. One presentation schedule began with the expert suggestion accompanying. Another schedule began without an accompanying suggestion. (And a third schedule began with an accompanying suggestion that was allegedly "expert" but in deliberately distorted.) Work in experimental psychology shows that initial experiences in an experiment are biasing. Participants beginning without an accompanying suggestion may remain more self-reliant. Participants beginning with distorted ratings may ignore, or actively reject, the genuine expert ratings. When these possible "sequence effects" are controlled for, n becomes too small for even meaningful descriptive statistics. For coin 1, above, instead of having 12 observations on the simple condition of "expert grade suggestion accompanying," there would be set of compound conditions; for instance "expert grade suggestion accompanying when the expert grades are seen first followed by no grade suggestion accompanying."

If this is not enough, "order effects" may also be present. These relate to changes in response to experimental conditions because of time. For instance, as the experiment goes on participants get tired. Their last five gradings may be made with less attention to detail than their first five. Alternatively, they may become more intent on finishing-up than on accuracy. In fact, of the 37 participants who started the experiment, only 32 completed it. That five dropped-out en route is a good indication that participants' state changes as time elapses.

With an ordinal-level dependent variable, factors that may be clouding the effect of treatment conditions, and the small n; the prudent course is sift through the data in search of suggested

findings. The data is trustworthy and settling for suggested findings allows the necessary statistical liberties. Thus, while no hypothesis is tested in a strict sense, sufficient support is garnered to make them worthy serious consideration and good candidates for further research.

One very basic finding stands out from all the others, and it is of greatest numismatic significance: none of the participants displayed any doubts about the feasibility of grading a coin from the displayed image. Exactly how the Internet will affect procedures for coin valuations remains to be seen, but the potential is there for it to become prominent.

4.4 Online Research Area #01: Experts Acceptance of Advance Grade Knowledge

Questions Addressed

- Does knowing a grading opinion in advance influence the grade given by graders?
- Does providing the correct grade yield a more consistent result than providing perturbed grades?
- Do graders accept the expert grade provided by the machine as being accurate or does their consideration of subjective factors considerably bias the results?

Hypotheses:

1. More convergence was evident in instances when graders received accurate guidance than when they received no guidance.

   The test of this hypothesis rests upon the rudimentary application of the F-Distribution for inferring that the standard deviations of two populations are significantly different. To illustrate, consider coin 1, a 1919D Lincoln penny with the grade of eight (i.e. "very good" - design and legend clear but worn flat; other details worn away). Twelve graders graded this coin in the presence of this knowledge (apparently derived by the computer), while 15 others graded this coin without a hint.

   The F-test offers ground for concluding that the presence of a valid suggestion causes a convergence among the determinations made by a set of graders working independently. That the observed convergence is about the suggested grade offers further evidence that an operative process lies behind the data, not sampling artifact. This process involves the proffered suggestion being internalized to structure the perception of an ambiguous stimulus.

2. The convergence evident in the presence of accurate guidance, relative to no guidance, was more pronounced for coins whose rating were more open to subjective variation.

3. Experienced graders are less influenced by perturbed coin ratings, alleged to be expert, than by accurate expert ratings.

We have previously established the machine-based system's success at predicting coin grades from third party graders. Now, the question is on its more general success with a pool of experts using it over the Internet. This time the human grading will explicitly invite qualification based on subjective factors beyond the ware and defects of concern in formulating a grade. In other words, we will be correlating the system's grade with the average expert grade prescribed by our participating experts.

Our hypothesis is that the correlation will be lower than seen between the machine grades and those of the services on the twenty selected coins.

Projected Outcomes:

Since the machine was developed to produce an expert grade, it was expected that expert grading consultants would take exception with the opinion of the machine-based grading result and apply their own subjective evaluation to the grade and thus changing the grade of the coin. What was unknown was the extent to which expert graders would change the grades and if the changes would be more liberal or more conservative than those of the machine-based grade would.

Expert graders often disagree with the grades provided by third-party grading services [65]. It was suspected that being provided a grade in advance would bias or influence an expert graders opinion of the grade that they would assign to a coin. The system of providing some accurate grades, some inaccurate grades and no grades was devised in order to eliminate potential bias and to really draw the expert out in terms of their own grading abilities.

The effect of advance grade knowledge on expert graders
(prompting condition: 1 = machine grade, 2 = no grade, 3 = misleading grade)

Coin	Year	Machine Grade	Mean Grader Grade	Mean Condition1 Grade	St Dev. Condition1 Grade	Mean Condition2 Grade	St Dev. Condition2 Grade	Mean Condition3 Grade	St Dev. Condition3 Grade	Smallest Mean	Smallest ST Dev
01	1919D	8	9.5	7.5	1.4	12.1	4.7	7.9	1.4	Machine	Machine
02	1911	6	7.4	4.7	1.2	7.9	1.6	9.8	3.6	Machine	Machine
03	1941	12	23.6	15.8	6.1	29.8	10.2	24.4	10.4	Machine	Machine
04	1946	53	60.7	57.7	3.5	62.6	2.8	61.7	2.4	Machine	Misleading
05	1951d	15	24.2	14.8	4.9	36.4	9.6	18.4	7.4	Machine	Machine
06	1952	45	49.9	45.4	4.6	52.3	4.9	52.0	3.7	Machine	Misleading
07	1944	53	52.0	52.9	2.6	52.4	4.9	50.9	5.6	Misleading	Machine
08	1946s	60	61.1	63.0	1.9	58.6	4.0	61.6	1.4	No Grade	Misleading
09	1954	12	33.7	28.5	10.1	43.2	5.8	30.1	12.7	Machine	No Grade
10	1968d	63	61.2	62.5	1.2	58.1	3.9	62.6	1.6	Misleading	Machine
11	1953s	6	18.9	16.9	9.7	24.7	11.4	15.9	8.3	Misleading	Misleading
12	1947s	12	24.4	19.2	7.1	34.0	13.9	20.4	8.0	Machine	Machine
13	1959	45	32.3	38.1	7.7	24.4	15.5	32.7	11.3	Machine	Machine
14	1935s	25	18.8	17.5	6.2	16.1	6.0	22.9	3.9	No Grade	Misleading
15	1944d	55	55.0	55.2	2.4	58.0	2.4	52.0	3.8	Misleading	Machine
16	1949	40	48.5	47.2	5.8	49.1	9.3	49.6	5.0	Machine	Misleading
17	1940	12	21.2	17.3	5.9	25.9	8.1	21.5	10.5	Machine	Machine
18	1946s	53	62.3	61.9	2.9	63.7	1.8	61.3	2.9	Misleading	Misleading
19	1951d	15	19.5	19.7	5.7	17.3	4.6	21.3	6.4	No Grade	No Grade
20	1944	53	51.0	50.6	3.5	52.9	6.0	49.2	10.3	No Grade	Machine
	Averages	32.2	36.8	34.8	4.7	39.0	6.6	36.3	6.0		
			# Smallest	11	11	4	2	5	7		
			% Smalles	55.00%	55.00%	20.00%	10.00%	25.00%	35.00%		
			Color Codes:		Red = Smallest Mean			Blue = Smallest Standard Deviation			

Table 4.4.1 - The effect of advance grade knowledge on expert graders

Results

Expert graders were presented with 20 screens similar to those in Figure 3.6.2 giving them 20 images (18 of which are unique and 2 of which were duplicate) to evaluate. The results in Table 4.4.2 document that in 14 out of 20 grading tests (70% of the time) expert graders assigned grades higher than those of the machine-based system while in 6 out of 20 tests (30% of the time) expert graders assigned lower average grades than the machine-based system.

The explained variance when subjective factors are incorporated into the grading mix is lower in comparison to that of the machine-based experiments (See Figure 4.4.2). The error rate also is greater than double the worst machine-based test A with the lowest number of trained images. The substantive reason is the subjective factors now brought into active consideration. The statistical reason is that with a greater number of human grades, there is bound to be those with idiosyncratic standards.

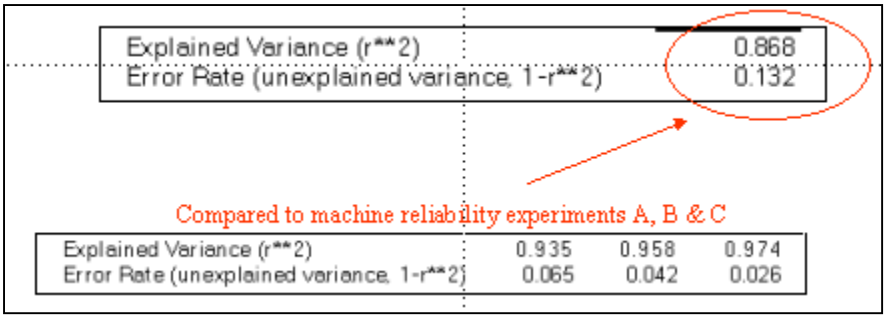


Figure 4.4.2 – Explained Variance & Error Rate – Experts Acceptance Testing

Coin	Year	Machine Grade	Average Grade	Average Difference	Graders view of machine grade
01	1919D	8	9.486	1.486	Graders more liberal than machine
02	1911	6	7.361	1.361	Graders more liberal than machine
03	1941	12	23.611	11.611	Graders more liberal than machine
04	1946	53	60.722	7.722	Graders more liberal than machine
05	1951d	15	24.194	9.194	Graders more liberal than machine
06	1952	45	49.917	4.917	Graders more liberal than machine
07	1944	53	51.972	-1.028	Graders more conservative than machine
08	1946s	60	61.057	1.057	Graders more liberal than machine
09	1954	12	33.743	21.743	Graders more liberal than machine
10	1968d	63	61.171	-1.829	Graders more conservative than machine
11	1953s	6	18.943	12.943	Graders more liberal than machine
12	1947s	12	24.441	12.441	Graders more liberal than machine
13	1959	45	32.324	-12.676	Graders more conservative than machine
14	1935s	25	18.824	-6.176	Graders more conservative than machine
15	1944d	55	54.971	-0.029	Graders more conservative than machine
16	1949	40	48.529	8.529	Graders more liberal than machine
17	1940	12	21.206	9.206	Graders more liberal than machine
18	1946s	53	62.281	9.281	Graders more liberal than machine
19	1951d *	15	19.531	4.531	Graders more liberal than machine
20	1944 *	53	51.031	-1.969	Graders more conservative than machine

Table 4.4.3 Experts Acceptance of Machine Grade (* *denotes duplicates*)

Still if one were to interpret these results with a 5% chance of error than any average grade difference of +/- 3.5 points would be deemed to be acceptable based on the 70-point Sheldon grading scale. The acceptance results of the machine-based grade would then appear to be more favorable.

Experimentation showed that knowing an expert grade in advance did overly influence the results of the grading consultants in terms of how close the experts came to the machine-based grade. The examination of the size of the standard deviation measurement was studied to determine how close expert graders came to the machine-based grade. In table 4.5.2 it is shown that in 11 of 20 experiments, or in 55% of the cases, knowing the machine-grade in advance yielded the tightest standard deviation but in 9 out of 20 experiments, 45% of the time, knowing the machine-grade in advance did not produce a result closest to the machine-based grade.

Table 4.4.4 shows that the error rate was the lowest and the explained variance (level of accuracy) was the highest when expert graders were provided a more accurate baseline grade, in this case the machine grade. Also shown is that even when expert graders are provided with a misleading grade that the error rate is considerably lower that when they are not provided any grade at all. Figure 4.4.5 graphically illustrates the changes in explained variance and error rates as the conditions change between Machine Grade, No Grade and Misleading Grade.

Advance Grade Knowledge

	Machine Grade	No Grade	Misleading Grade
Correlation between machine grades & expert graders	0.962	0.825	0.945
Explained Variance (r**2)	0.925	0.681	0.893
Error Rate (unexplained variance, 1-r**2)	0.075	0.319	0.107

Table 4.4.4 - Advance grade knowledge on expert graders

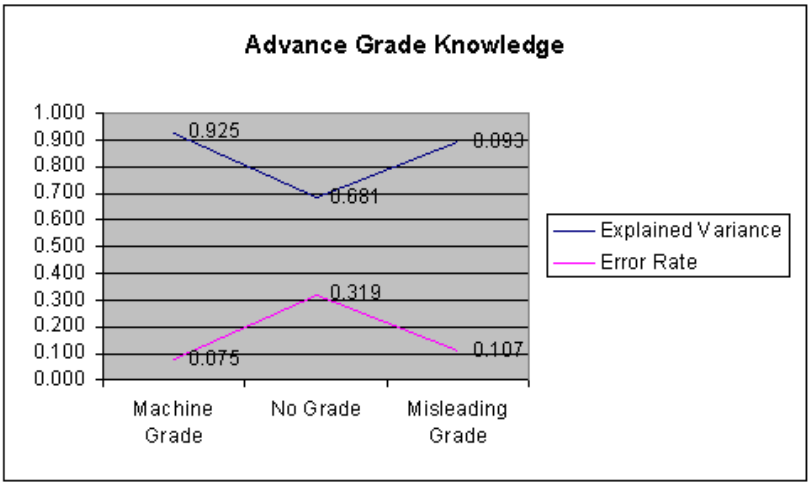


Figure 4.4.5 – The correlations between the different conditions

When analyzing the data with F-Test variance comparisons with 5% F-Test significant level, as noted in Table 4.4.6 there is a tendency for greatest convergence when an expert grade or a perturbed grade is given relative to no grade, or no suggestion. The F-Test also shows that there is the least convergence when no grade is provided.

Variance F-Test Comparision - All Coins

Dispersion vs Clustering

	Less Convergence (F < -3)	Same Convergence (-3> F <3)	More Convergence (F >=3)
Expert Grade Given *- relative to No Suggestion*	1	11	8
Perturbed Grade Given *- relative to No Suggestion*	3	13	4

Table 4.4.6 – F-Test Variance Analysis

Recommendations:

All online grading should include accurate / true grades.

4.5 Online Research Area #02: Perception of Internet based Grading

Question Addressed:

- Was the Internet an effective vehicle for grading?

Hypothesis: Expert graders would see the value in grading images over the Internet although they would rather grade from the actual coins. The graders that participated in this test are predisposed to the use of technology as they are on the Internet.

Projected Outcomes:

Many graders often complain of how difficult or inconvenient it is for them to grade slabbed or encapsulated coins and many threads in the PCGS Form and the rec.Coin.Collecting newsgroups often echo the same complaints about grading coin images in online auctions. It is commonplace to see comments such as "it is extremely difficult to grade anything other than the physical raw unslabbed coin" in these online community forums. With this advance knowledge, this researcher didn't expect to change a rather popular and often voiced opinion of the collectors and dealers.

The expectation was that a large number of expert graders would say that they would rather grade the physical coin instead of a coin image. The purpose of this experiment was to gain a quantifiable reference point to this sentiment.

Results:

The first question asked the experts how grading images on the Internet is compared to grading physical coins

- 78.37% said that it is more difficult than grading coin images over the Internet than the physical coins
- 20.03% said that it is the same or better than grading the physical coins
- 1.6% said that the practice of grading over the Internet should be banned

The second question asked the experts about the quality of the images presented

- 64.87% said that the images were good enough to properly grade the coins
- 35.13% said that the images were of poor quality and made it hard to grade properly

Internet verses Physical Coin Grading Test

An additional test was constructed which involved having a small population of the expert graders grade the physical coins in order to further test the hypothesis that grading experts would see the value in grading images over the Internet as opposed the actual coins. This test was to determine if there were substantial differences in the online grading results compared to the results of grading the physical coins in person. This test required face-to-face meetings

with five of the expert graders that participated in the online tests. They were shown the same 20 coins in the same order with the same grades disclosed as they were shown previously.

The results of the face-to-face grading sessions are contained within Table 4.5.1. Overall, the aggregate difference in the grading results of the 100 coins graded was 5.10%. The smallest variance in grading was a difference of 1.71% shared by Coins 01, 10 and 14 while the largest difference was 14.29% for Coin 09.

Physical verses Online Grading Results

Coin Id	Year	Machine Grade	Grade Difference	Percentage Difference
01	1919D	08-Very Good	1.20	1.71%
02	1911	06-Good	1.80	2.57%
03	1941	12-Fine	3.00	4.29%
04	1946	53-Almost Uncirculated	3.40	4.86%
05	1951d	15-Fine	2.40	3.43%
06	1952	45-Extra Fine	4.00	5.71%
07	1944	53-Almost Uncirculated	5.20	7.43%
08	1946s	60-Mint State	1.40	2.00%
09	1954	12-Fine	10.00	14.29%
10	1968d	63-Mint State	1.20	1.71%
11	1953s	06-Good	7.00	10.00%
12	1947s	12-Fine	6.00	8.57%
13	1959	45-Extra Fine	4.40	6.29%
14	1935s	25-Very Fine	1.20	1.71%
15	1944d	55-Almost Uncirculated	2.60	3.71%
16	1949	40-Extra Fine	3.60	5.14%
17	1940	12-Fine	5.40	7.71%
18	1946s	53-Almost Uncirculated	2.00	2.86%
19	1951d	15-Fine	2.00	2.86%
20	1944	53-Almost Uncirculated	3.60	5.14%
		Averages	3.57	5.10%

Table 4.5.1 – Physical verses Online Grading Test

The results of this test are encouraging as they demonstrate that the expert graders assigned the same or a like grade to the coin as the image that they evaluated in 94.9% of the evaluations. The expert graders assigned the same or like grade to the physical coins as they did to the coin images presented to them on the Internet. This demonstrates even though expert graders would prefer to grade the physical objects rather than images the medium, in this case the Internet, did not seem to be a limiting factor as an effective grading vehicle.

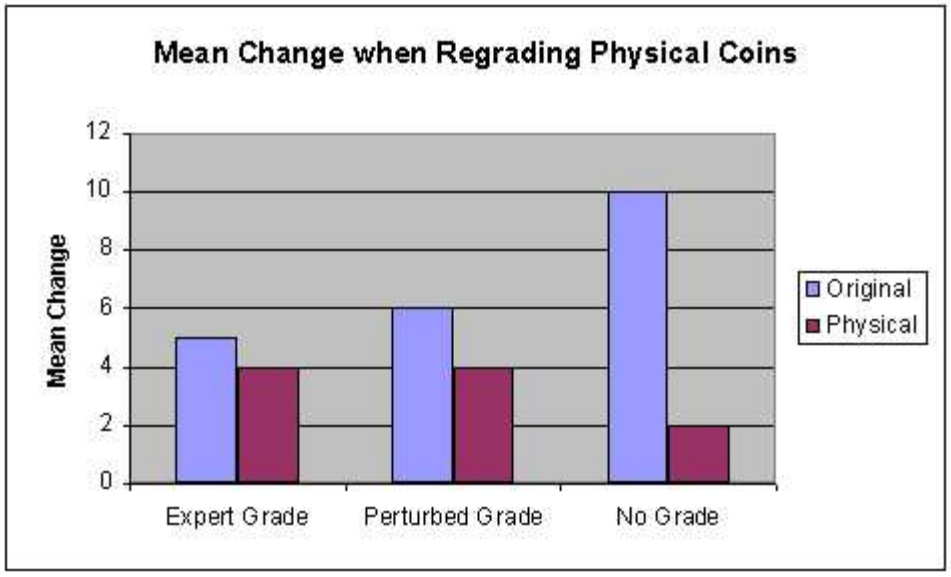


Figure 4.5.2 – Mean Change when regrading Physical Coins

Figure 4.5.2 shows that the mean change when graders are provided with the expert grade is the smallest compared to when graders are provided with a perturbed grade or no grade in the original online tests. However, expert graders did significantly better when grading the physical coins when they were provided with no grade than they did grading the same coins from images.

4.6 Online Research Area #03: Grader Consensus

Questions Addressed:

- To what extent do human graders come to consensus when grading the same coins as other human graders?

Hypothesis: Expert grading consultants will have widely diverse grading opinions on the images that they are grading. Informal experience that humans grade inconsistently is supported by the Stujoe grading tests, Coin World tests, Kevin Foley tests on human grading.

Projected Outcomes:

A person only has to read a handful of posts on the PCGS Forum [54] or on the newsgroup rec.collecting.coins to get a sense of the wide diversity in grading opinions that graders have. There are frequent posts in both of these popular venues for coin collectors and dealers, which take to task the grades of coins, which are offered for sale on the Internet. A large number of these heated threads center on offerings that are made on eBay.com.

Coin	Year	# of Graders	Lowest Grader Grade	Highest Grader Grade	Machine Grade
01	1919D	37	3	25	8
02	1911	36	3	20	6
03	1941	36	3	45	12
04	1946	36	50	65	53
05	1951d	36	1	45	15
06	1952	36	35	62	45
07	1944	36	30	63	53
08	1946s	35	45	65	60
09	1954	35	8	55	12
10	1968d	35	50	65	63
11	1953s	35	6	55	6
12	1947s	34	6	55	12
13	1959	34	6	60	45
14	1935s	34	4	40	25
15	1944d	34	45	63	55
16	1949	34	6	65	40
17	1940	34	4	45	12
18	1946s	32	55	66	53
19	1951d	32	6	40	15
20	1944	32	15	64	53
	Averages	**34.7**	**19.1**	**53.2**	**32.2**

Table 4.6.1 Grader Consensus

This experience combined with the documented diversity in grading is shown in Table 2.7.1. The results of Stujoe Grading Challenge [48] have led this researcher to expect that there would be considerable diversity amongst the expert graders in this experiment despite their expert status level. This experiment was to document the range of grades that experts assign to a collectible and to show how wide the range of interpretation between all of the experts is.

Results

The experts do disagree in their grading opinions. Table 4.6.1 shows the grading results for all twenty-coin images along with the composite averages. The results demonstrate the extent to which experts do disagree with each other. The diversity in the range of grades from lowest to highest is telling in showing how different experts assign grading opinions to each item.

For instance when looking at Coin #01 the grades range from 3 to 25, or 22 grading positions on the Sheldon 70 point scale. The average grade of 9.486 for all of the expert graders is close to that of the machine grade in that it is only off by 2.1% but the range suggests that some graders were off significantly from both the machine-grade and each other. This pattern is repeated many times within these experiments and it didn't really matter if the coin was of a higher grade or a lower grade.

Recommendations

Setup a future experiment that shows the experts how their peers are grading. Increase the number of grading experts in future studies to several hundred to evaluate the diversity in grading of a larger sample.

4.7 Online Research Area #04: Duplicate Grades

Question Addressed:

- How did the grading experts do with the duplicates?

Hypothesis: Given that it is hypothesized that humans provide widely divergent grading opinions it is expected that graders will offer differing grading opinions on coins which they have previously graded.

Projected Outcomes:

This control experiment was established to measure both the expert graders consistency in grading the same coins as done previously and to see how varying the machine-grade presented would affect the grade that an expert would assign to a previously graded coin. The expectation was that there would be a certain amount of change that occurred in the grading of the duplicates but the hope was that the change, or variance, would not be significant. This test was also used to determine if the expert was really an expert as the credentials of the expert were more carefully scrutinized if there was a significant amount of variance in the grading of the duplicate coin images.

Results

The results were interesting in that there was a more significant variation in the grades assigned on the first series of coins #19 and #05 than those of the second series of coins #20 and #07. The average difference in the first series was 4.663 grade points or 6%. While the average difference in the second series was a scant .941 grade points or 1.3%. Table 4.07.1 shows the grading results for the duplicate coins in comparison to their original counterparts.

Coin	Year	Machine Grade	Ave Grader Grade	Difference From Mach	Standard Deviation
19	1951d	15	19.531	4.531	6.187
05	1951d	15	24.194	9.194	11.226
20	1944	53	51.031	1.969	7.210
07	1944	53	51.972	1.028	5.209

Table 4.7.1 – The Results of Graders grading duplicate coins

The expectation was that there would be a certain amount of change that occurred in the grading of duplicates but the hope was the change or variance would not be significant.

The first online experiment focused upon the persuasive impact that computer-generated coin ratings were likely to have on experienced graders working over the Internet. A related issue is the persistence of this impact: is it relatively enduring or ephemeral? Pursuant to an answer, subjects were asked to grade two coins a second time. Table 4.7.2 compares the stability of gradings between subjects for whom an expert grade accompanied the first grading in contrast to those for whom no grade accompanied.

Stability of Repeated Gradings of the Same Coin

	Repeated grading		
	Same as Initial grading	**Closer to Coin's True Grade**	**Farther from Coin's True Grade**
Accurate rating first	8	6	7
No rating first	4	15	3

Chi-square = 8.85
Significant at the .02 level

Table 4.7.2 - Stability of Repeated Gradings of the Same Coin

The data is suggestive but does not show a significant difference: When participants were presented with an expert's grade at the time that they were asked for a determination, 8 out of 21 times (38% of the time) their judgment remained unchanged. In contrast, when no suggesting accompanied initial gradings, judgments remained unchanged only 4 out of 22 times (18% of the time).

One of the participants mentioned noticing that he was asked to grade the same coin twice. We take this to mean that experienced graders, focused on the task, are aware of what they are doing. When confidence in a grade is higher, uncertainty goes down and reliability goes up.

Figure 4.7.3 portrays the entirety of the data on grade and regrading reliability:

The Chi-square computation:

Observed values:

	Repeated grading: Same as Initial grading	Closer to Coin's True Grade	Farther from Coin's True Grade	
Accurate rating first	8	6	7	21
No rating first	4	15	9	28
	12	21	16	49

Expected Values:

	Repeated grading: Same as Initial grading	Closer to Coin's True Grade	Farther from Coin's True Grade	
Accurate rating first	21*12/49 = 5	21*21/49 = 9	21*16 = 7	21
No rating first	28*12/49 = 7	28*21/49 = 12	28*16 = 9	28
	12	21	16	49

O	E	(O-E)*(O-E)/E
8	5	1.8
6	9	1
7	7	0
4	7	1.3
15	12	0.75
3	9	4
		8.85 ← chi-square

degrees of freedom = (rows-1)*(columns-1) = 2

Figure 4.7.3 - Chi-square computation regarding duplicates

The value to equal or exceed:

for the .05 level of significance is 5.99
for the .02 level of significance is 7.82
for the .01 level of significance is 9.21

Chance alone could produce a distribution with the observed degree of aberration on 2% of the time.

Chapter 5 - Conclusions

Two major areas were examined in this study:

1. To determine if an automated machine-based grading system produces expert grading results that are consistent and reliable enough to be accepted by the grading experts in the field.

2. To determine if having expert graders applying their subjective market grading analysis to the digitized images which were pregraded by the machine-based system enhances the collectibles' grading experience above that of human only or machine only.

Human graders grade collectibles inconsistently and have a spotty record when it comes to grading consensus, even within their own grading. A large problem with human grading is that humans lack accuracy and consistency in grading; experiment #04 of this study on duplicate coin grading demonstrates this point. The StuJoe Grading Challenge [48] demonstrates that graders are all over the board when grading. Experiment #03 of this study also shows the diversity in the grading frequency range between experts.

The machine-based experiments demonstrate that it is possible to design machine-based systems that are capable of grading digitized images of collectibles with consistency [8]. Grades derived from machine-based systems are subject to lack of user acceptance in the marketplace much like coins encapsulated [46]. Machine-based grading systems might ultimately evolve into becoming good at grading in all coin series at a production level but they will clearly fail to take into account the varying subjective features that professionals feel are important.

The approach of grading digitized images over the Internet that provides a reliable baseline grade derived from a machine as a starting point has a better statistical chance of being accurate, as many of the technical components that are often bypassed in the human visual pattern recognition process [47] are reigned in. This hybrid human-machine grading process extends the machine grade with the addition of subjective features enabling the expert grader the ability to arrive at an expert grade.

An extended commercial quality machine-based grading system that is hosted on the Internet could become a major threat to the existence of the third-party grading services. The multifaceted appeal of this approach is that collectors and dealers would be able to submit their own coin images to a web based system to get their expert grade in minutes instead of weeks for considerably less cost than what is changed now and without the risk of loss to their valuable collectibles in the transit process. The added advantage of this approach is that the results submitted to the machine-based system would always yield the same results thus eliminating the cottage industry of cracking out coins from the slabs [65]. Dealers wishing to buy sell or trade collectibles across the Internet could then have a fast, reliable and low cost method for obtaining a baseline grade that would become a reasonable starting point for the laying of subjective qualities.

5.1 Conclusions – Machine-based Experiments

The larger the database in the machine-based system the better the system seemed to do at grading, as the level of explained variance increased from 91.86% to 95.60% as the number of trained images went from 55 to 105. Increases in the number of trained images also seemed to contribute to the increase in percentage of coins, which were exactly matched as this percentage increased from 40% to 51.43% as the number of trained images went from 55 to 105.

The standard deviation got noticeably smaller and narrower, starting at 4.126 and dropping to 2.454, as the number of trained samples increased from 55 to 105. This suggests that the measure of distance between the machine generated grade and the expected grades were getting closer as the number of samples got larger and the machine had a larger knowledge based to draw from.

The comparative results between Test D, where every trained image was evaluated, and Test E, where the 20 images of the coins graded by the third-party grading services, was not significantly different. While the standard deviation dropped the overall accuracy of the machine-based system changed little. This would tend to suggest that the machine-based system was functioning consistently with the trained database of 100+ images and that the grades provided by the third-party grading services may have been more accurate in comparison to the stored coin images within the system.

5.2 Conclusions – Online Experiments

Hypotheses:

1. More convergence was evident in instances when graders received accurate guidance than when they received no guidance.

   Data on coin 1 suggests that ratings of coins are affected when performed in the presence of an expert's rating, and the effect is to pull the ratings toward that of the expert's. This is borne out by similar examinations of the data from the other coins. In 10 out of the full set of 20 coins, the data looked very similar to coin 1. For 9 of the coins, the F-test suggested no elevation in clustering with the presence of an expert opinion, although the expert opinion almost always brought about more clustering focused around it. In only one case, that of coin 9 was the ratings more clustered in the absence of an expert opinion.

   With respect to the mean, with expert grades present the average difference in subjects' ratings from the coins' true ratings was 5 points on the Sheldon scale. With expert grades absent, this difference was 10.

   Accordingly, we regard the collected data as tending to confirm hypothesis 1.

2. The convergence evident in the presence of accurate guidance, relative to no guidance, was more pronounced for coins whose rating were more open to subjective variation.

   A plausible explanation for hypothesis 1 is that grading requires an exact assessment of a constellation of amorphous qualities. Under such circumstances, there is a natural tendency to reify authoritative suggestion. A social psychologist might draw an analogy between deciding upon a Sheldon score and determining the distance moved by a pinpoint of light in a pitch-dark environment. With no points of reference, inherent ambiguity is such that even a stationary point of light will be observed to move. Numismatists, however, regard coin grading as neither ambiguous nor arbitrary. Specific grades mean specific things. All the same, some coins may be more difficult to grade than others may. However, which coins?

   Suppose that coins with higher Sheldon ratings were more difficult to grade than coins with lower Sheldon ratings. If this were so, one would expect the disambiguation of an expert opinion to operate more strongly there. Or, if coins with lower Sheldon ratings were more inherently perplexing, then this is where an expert's opinion would be the stronger magnet. Does the data show any systematic variation? No.

   We grouped our coins into three categories: those showing heavy ware (expert ratings of 6, 8, 12, and 15), those showing moderate ware (expert ratings of 25, 40, and 45), and those showing slight ware or no ware (expert ratings of 53, 55, 60, and 63). The degree of convergence precipitated by the expert suggestion was the same across coin quality. The criterion for a coin's being counted as exhibiting "more convergence" when participant graders were given an expert's opinion than when graders were given no suggestion was a significant F-test (worked out as exemplified at the beginning of the discussion of hypothesis 1).

Degree of Convergence by Coin Quality

	Less convergence with presence of expert rating	Same degree of convergence with presence of expert rating	More convergence with presence of expert rating
Heavy ware	1	4	4
Moderate ware	0	1	3
Slight or no ware	0	4	3

Table 5.2.1 – Degree of Convergence by Coin Quality

While table 5.2.1 provides no gist in favor of hypothesis 2, neither is it directly disconfirming. Quality was the only prospective means we had for classifying coins as being more or less evasive when it came to grading. Perhaps, in fact, quality is unrelated to grading difficulty. Nothing in the numismatic culture or literature suggests that such a relationship exists. There may be other factors obscuring the grading process; for instance in arriving at market grades as opposed to technical grades. This hypothesis must be retained as a supposition worthy of examination when the opportunity permits.

3. Experienced graders are less influenced by perturbed coin ratings, alleged to be expert, than by accurate expert ratings.

Degree of influence is evidenced by the convergence effect exerted by the presence an expert's grade, whether accurate or perturbed. In other words, we influence will be viewed indirectly by comparing the number of coins on which there was significant gravitation toward the suggested grade relative to the scatter of grades when no suggestion was present. The finding, which substantiates the hypothesis, is in table 5.2.2 below:

Degree of Convergence by Rating's Accuracy

	Less convergence with presence of a coin rating	Same degree of convergence with presence of a coin rating	More convergence with presence of a coin rating
Accurate rating	1	9	10
Perturbed rating	3	12	5

Table 5.2.2 – Degree of Convergence by Rating's Accuracy

When a rating is inaccurate, its persuasive pull is reduced. Convergence toward a perturbed rating occurred only half as often as convergence toward an accurate rating (five times versus ten). Moreover, perturbed ratings sometimes had the effect of dispersing the subjects' gradings. Unfortunately, subjects were not immune to misleading influence. In the case of five of the coins, opinions solidified around the perturbed rating.

Summarized Conclusions - Online Experiment #01:

These test results indicate that expert graders accept the expert grade provided by the machine-based system about 35% of the time and apply their own opinions as to the grade

65% of the time. Since the test coins were graded by third party grading services, these results suggest that expert graders do not agree with the third party grading service grades in at least 65% of the time.

Expert grading consultants:

- Came closet to the machine-based grade (55% of the time) when they were provided with a reliable machine-based grade in advance.
- Came next closest to the machine-based grade when (35% of the time) when they were provided with a misleading incorrect grade in advance.
- Were the furthest away from the machine-based grade with only 10% accuracy when they were provided with no grading opinion in advance
- Were 90% effective when they were provided with some grading opinion even if the grading opinion was incorrect.

.

Based on these experimental results and those in table 4.4.1 there is no reason to believe that knowing a reliable machine-based grade in advance has any undue influence on the grading accuracy by expert grading consultants. Expert graders still appear to layer their own opinions within the grading process by changing the machine-based grade 65% of the time regardless of the grade that they are provided. Lastly, knowing any grade as a baseline seems to increase the grading accuracy significantly over not being provided with a grade at all.

Summarized Conclusions - Online Experiment #02:

The majority of the graders indicated that is more difficult grading digitized coin images over the Internet to the physical coins but almost an equal majority indicated that the images were good enough to properly grade the coins. These results can be interpreted as the experts preferring to grade the physical coins but acknowledging that the Internet and the digital images are a suitable substitute in the absence of the actual coins.

Summarized Conclusions - Online Experiment #03:

The opinions of experts are diverse as they consider different factors when they grade. Having a large pool of 37 expert graders helped to produce better results which were close to that of the machine-grade overall. The danger of having significantly fewer expert graders was that we may have only captured the extremes in data, the highest or the lowest, and may not be close to the results of the machine-based system.

Summarized Conclusions - Online Experiment #04:

While variance of both of these tests is small, 6% or less, it is still surprising that they did not come out with the same results given that the graders are experts. Previously we have demonstrated that variation exists between graders. This experiment extends that point by demonstrating that grading variation also exists among the same grader.

It is easy to understand that a grader may assign a different grade to a coin on a different day should mitigating circumstances come into play. However, it is difficult to explain why an expert would grade a coin differently less than 45 minutes after grading the previous one. The argument that great timeframes were spanned in the process of these tests by individual graders cannot be made in support of these results as all grading tests were done in the same session with an upper time limit of 45 minutes.

5.3 Future research opportunities in this subject area

A great number of exciting future research opportunities available for anyone that wishes to extend the work done in this study. Many of these opportunities are in the future development and refinement of the machine-based system and how it interfaces to the next generation of the online grader.

Extend the number of algorithms in the machine-based system and then interpolate
Presently just the histogram distance measure algorithm was used for this research with impressive accuracy results. Perhaps by adding several additional matching algorithms the results could be increased by a using an interpolation of the results. Stand-alone machine-based systems that used a single new algorithm would need to be constructed and tested first to measure reliability.

Extend the number of trained images in the database
Adding several hundred more properly graded images would probably increase the percentage of accuracy a few percentage points, but the real gains would probably come in the area of the percentage of coins that were exact matches.

Extend the machine-based framework to include a detailed technical grading model
A detailed technical grading model could be designed to examine the detailed features of the collectible being graded. There are often 10 – 20 detailed features on most rare collectible items that appraisers or graders should examine when determining the condition. In practice most graders do not examine all of the salient features when grading collectibles due in part to the large number (70^{**20}) of permutation possibilities that could unfold in the process. As a rare collectible, such as a Lincoln Cent, may have 20 detailed features to be examined, each of these features can have 70 possible grades. A possible outcome of this research would be to develop a system that can examine details of a scanned image at predefined locations and to determine the grade of each condition.

Database Reconstruction
As this system was designed for research purposes and the number of trained samples was not large, an elaborate relational database was not used. Researchers doing future work in this area might want to extend the system to include a SQL database to accommodate a larger training database and speed considerations.

Operating Platform of Machine-Based System
The machine-based system was developed in Java so that it could ultimately be transported to the largest number of operating platforms possible. For purposes of this study, the machine-

based system would exist exclusively on a local machine. Future post dissertation work in this area could extend the operating environment to the Internet in a browser-based environment.

Next Generation Online Grading Website

The output of the web based machine system could become the input to the next generation of the online web site where buyers and sellers could collaboratively use visual pattern recognition technology similar to that detailed in the Nagy study [51].

Appendix A

Forms & Communication

This appendix contains communication between Richard Bassett and some of the expert graders, which helps to demonstrate some of the interaction between the participants and the researcher.

Appendix A1

Postings seeking Expert Graders

Posting placed in rec.coin.collecting and PCGS Forums 1/12/03

From: "Rick Bassett" <nospam@bassett.ws>
Subject: Need Expert Coin Graders to participate in Online Grading Test
Date: Sunday, January 12, 2003 10:21 PM

I am seeking a limited number of coin collectors or coin dealers that are experienced in grading Lincoln Cents to take part in my online grading experiment. This experiment, which is part of my dissertation research, essentially involves grading the obverse of 50 circulated Lincoln Cents from scanned computer images.

Experiment Background & Purpose: I am working on my doctorial dissertation in visual pattern recognition in the application area of coin grading. This dissertation will attempt to ascertain how well humans do at grading, how well computers do at grading and how successful collaborative human/computer grading may be.

Here is a link to the details of what I am looking for:
http://www.rickbassett.com/pace/dissertation/test_candidates/default.asp

I will select the most qualified experts for this test as I am looking for the best possible grading data and experience is extremely important for this study.

Thank you for the consideration,

Rick Bassett

(My email address is on the web link above if you wish to communicate with me)

Posting placed in rec.coin.collecting and PCGS Forums 3/30/03

From: "Rick Bassett" <nospam@bassett.ws>
Subject: Need Coin Collectors and/or Dealers to participate in Online Grading Test
Date: Sunday, March 30, 2003 11:27 PM

I am seeking coin collectors and/or coin dealers to take part in my online coin grading experiment. This experiment involves grading the obverse of 20 circulated Lincoln Cents from scanned computerized images. This experiment is very quick and easy to take; I modified it greatly based on the feedback received from the original test in January.

I am working on my doctorial dissertation in human and machine-based (computerized) grading of collectibles in the application area of coin grading. This dissertation will attempt to ascertain how well humans do at grading, how well computers do at grading and how successful collaborative human/machine grading may be.

Here is a link to the test site:
http://www.rickbassett.com/pace/dissertation/new_test/coin.asp

Thank you for the consideration,

Rick Bassett

My email address is on the web link above if you wish to communicate with me

Appendix A2

Feedback & Comments from Expert Graders

Email Communication received from expert graders as the online experiments were progressing in the early stages of the study. Note that the names of the expert graders are not disclosed for security purposes per the privacy agreement.

From BG 2/8/03

Rick,

Don't sweat it too much. Even if you mailed them the actual coins some of them would not be satisfied. I think what they are forgetting is that what they are seeing is what the computer program will most likely 'see'. I would suspect that we need to teach it what to look for with the same images it will use. If it is too difficult to scan the image consistently over a large number of conditions then the whole premise I would think, is in vain.
Ooops, sorry to go off on a tangent :-(

Rick Bassett <me@rickbassett.com> wrote:

> You can complete the test with the existing images if you like. Thank you for the feedback, I am going a little crazy on this end trying to please some of the graders.
>
> Rick

From BG 2/7/03

Thank you for the update. Honestly, the photos are adequate. It is much easier to tell the difference between a raw 25 and a raw 30 than between a raw 66 and 68. Sometimes I needed to kick up the brightness of an image, but for the low grade coins at least in the first half of the survey that was a minor adjustment. Also, remember that a grader will usually assign a grade to a coin within the first 20 seconds. Having high detail images for these circulated cents would be a bit of overkill. I think that some people have gotten spoiled nit picking every ding or striation. Good luck!

From CS 1/15/03

> Rick:
> Regarding variation as the enemy, market grading is a source/cause/subset of inconsistency, so it's really inconsistency (variation) as a single main problem. Another source that relates to your study is different grading standards: Published ones like ANA, PCGS, Photograde, et al.; unpublished standards like NGC, ACG, et al.; and worst, unstated standards, such as many internet dealers use--"I grade this as VF, but you decide for yourself".
>
> I have started to make a habit of stating the standard I am using when I grade. I typically use ANA standards, which are higher than most (all?) others. This often causes me to grade on the conservative end of a group. This happened just yesterday in an rcc thread entitled, "Walking Liberty half dollar grading opinions wanted". Your experiment is targeted at another source of variation, namely uneven wear. Still, you may find it useful to ask the graders to state the standard they use.
>
> Some sources of error I have observed include:
> Economic conflicts of interest
> Differing standards (market grading is a subset of this)
> Deficient skills
> Uneven wear/preservation
> Insufficient effort ("I'm busy, but I'd call it"; "I don't have the book handy, but")
> Physical limitations (eyesight)
> Environmental factors (lighting)
> Ambiguity over strike and die state vs. wear
> Deceptive alteration
>
> Among dealers, I believe economic conflicts are the primary reason for variation. Differing standards would be second, and they are frequently caused by economic conflicts of interest.
>
> Regards,

From CS 1/4/03

> Rick:
> I just signed up for your experiment, and looked at the sample. In the "Detailed Feature Grading", did you intend to add the fields (left and right) and rim, or just the features currently shown? I would recommend the former. You may also want to disaggregate some of the profile features, such as ear, hair, cheek, collar, and bow tie, which sometimes differ in grade on the same coin. Also, I was unable to view the larger image.
>
> Great experiment, by the way--I think the future of coin collecting is largely in cyberspace, and you're working one of the key issues in that arena. If you follow rec.collecting.coins closely, you may have noticed that I frequently say, "grading variation is the enemy!" Anything that economically reduces grading variation is good.
>
> Good luck with it!

The following are unedited Comments which were posted in the PCGS Forum by Grading Experts that took the online test

- Is it me or was seeing the computer grade posted distracting to you too. I see how tireing it could be just grading coins all day. Man I bet pro graders take alot of breaks. I wonder how many coins a pro grader does grade before he says `I need a drink or something,be right back`

- computers grade didn't distract me at all, but i think some of those grades were hilariously way off

- I've taken it, but there is the usual caveat that you cannot grade accurately by picture. You need to see it in person to do it correctly.

- I think I would be a lousy grader!

- To me it looked like some of the coins were reused.

- You know, I kinda thought the same thing.....is hard to grade pics though.

- I didn't even know there were computer grades until I got half way though. Most of the computer grades seemed to be way off. Maybe that is ACG's problem? NOT! The color also made it very hard to gradet the MS items - if they were. All of them looked cleaned or bassy.

- I think some coins were reused as part of the test, to see if your grading opinion changed when a grade was given. Some of the grades given were way off, but again I think it was designed to see if you would raise or lower your grade if a grade was given. I will be very interested in the results of this test.

- I'd love to try, but since I'm not very good at grading lincolns, all I'd do is mess up your results.

- I think the last three coins were reused, perhaps this is an attempt to validate the grades initially given to them.

The following are unedited comments, which were recorded by the expert graders that took the online grading tests.

- Planchet quality.That one is overlooked alot I,ll bet.
- First thing I noticed was rim separation
- extensive wear
- Good honest wear
- Nice coin for the grade
- Not quite as nice as the first
- Minor red corrosion does not matter in this grade.
- Solid fine not quite VF
- Actully is VF-35. The coin's a near miss at the 40 grade.
- Defects,Negative I hope means no defects seen.
- Coin marks on collar
- Detect slight wear in hair
- The color looks funky in the picture. The coin could be cleaned, but I gave i
- Coin has been dipped
- EF35?
- You mean VF35 right?
- Appears to have abrasion on the cheekbone, jawbone and chin. Also looks like
- I've net graded this. It looks like Lincoln took a hit on the jaw, which take
- Spots detract
- Detracting corrosion fleck at 8 o'clock, otherwise a decent AU55 specimen.
- Looks like a cleaned and re-toned AU to me. I net graded it.
- I think it would be a 45 but given the very common date I tend to grade tight
- Again, spots
- Detracting flecks (appears to be a fingerprint) and a white spot on and above
- AU-58 from wear, but ugly spots take it to AU-50 for me.
- Ok.Im,ll try not to look at computer grade first. hehe
- Solid 60
- Appears to be slight rub on cheek and jawbone keeping it from being mintstate
- Once more funky color perhaps from the scan make this hard to assign MS grade
- Not red enough
- 06!!!!!!!!!!!??????????? the computer is broken.
- There's some crud inside the letters. For 19th century coins this can be expe
- I think a 63 is being generous too.
- Appears to have no wear, but the luster weak.
- Its hard not to see the computer grade before formulating ones own opinion.No
- Not quite VG
- Classic VF-20, but net great for eye appeal.
- Only slightly better than VG
- Things are a lttle fuzzy in the photo, but the coin is this grade or maybe AU
- Rim dings detract from the overall appearance of this coin.
- The spot at 9 o'clock makes this less desirable.
- Once more the color in the photo makes grading tough.
- Some red left

- not quite XF
- Crud lifted from obverse at 9 o'clock with less than great skill.
- Nice specimen, Few bag marks.
- Would really need to see this in person. Color does not come well here on red
- hmmm.is everyone having alittle variance from the computer grade also?
- Looks to be abrasion on the cheek, jawbone and chin. Looks like a metal dete
- Is this a repeat? Lincoln appears to have taken a hit on the jaw. EF-40 downe
- oh its that one again.Through in a few duplicates huh.
- Discoloration on cheek may be wear or coin friction discoloration meaning MS6
- This is the same coin as image #7
- Looks to be a fingerprint on the surface, a white fleck on and above the ear
- Definitely a repeat. Spots make this coin ugly.

Appendix B

Design Documents

The design documents within this appendix represent the specifications and design plans for all three versions of the online Internet based grader and the automated machine-based grader. These documents are meant to demonstrate the mindset in design at the point in time and were edited after the design was underway.

Appendix B1

Design Document for Online Grading Experiment Software First Generation

Web Site:

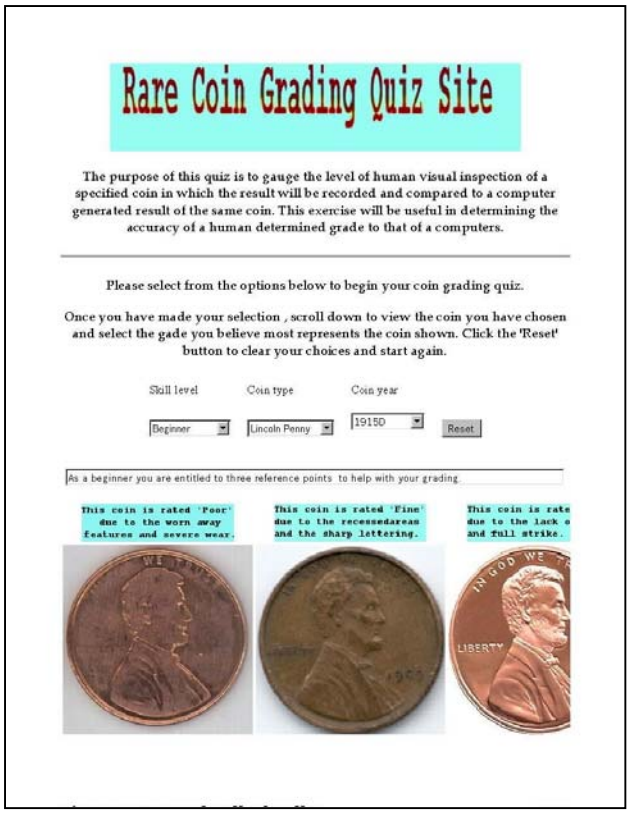
Purpose and Theme
The purpose of the website is to provide an attractive and functional interface in which a user will interact with the coin grading software results. The site will test the ability of the user to determine the actual grade of a coin by visual comparison of several specific features. The theme will be in the style of a quiz. The person will choose a coin from the database, they will answer a few short questions regarding it, and then submit their findings.

Navigation Method
Mouse and keypad will be the tools utilized in the interaction of the user and the website. Each field must be filled in order to continue. A 'GRADE NOW' button will send the entered information to the server to be processed and returned.

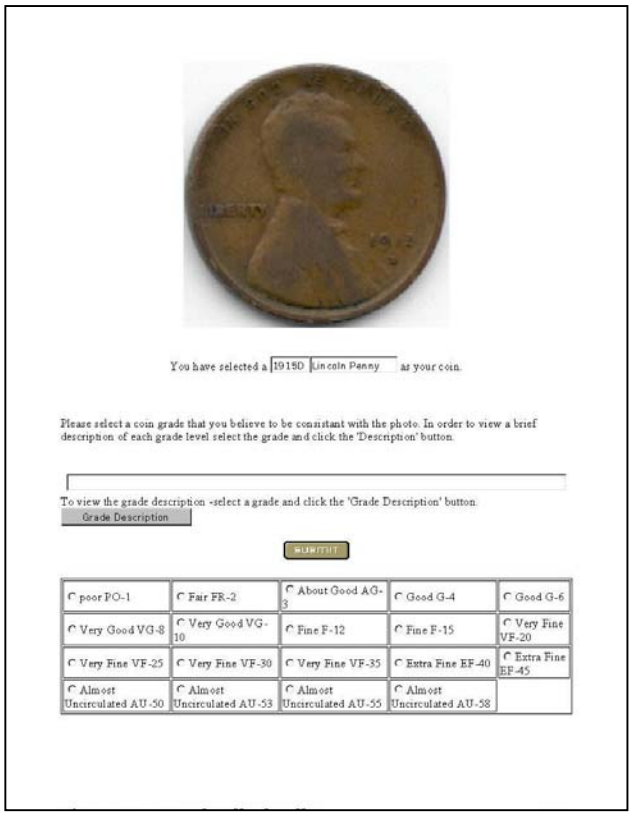
Implementation Tools
The front end will consist of JavaScript, html and java code. The back end will utilize ASP to communicated with the server and the .DBMS files will be storing the entered information.

Appendix B1a

Online Grader Screens - First Generation



Page 1



Page 2

Appendix B2

Design Document for Online Grading Experiment Software Second Generation

Functionality & Features Required

Part 1: Entire Coin Grading (Summarized Grading)

Overview: A summary user coin grading site. This site allows a user to select pregraded coin images from a list (without showing them the grade) and to make a guess at what they think the grade is.

This front end website must perform the following functions:

- Provide a single screen intuitive GUI for users
- Prompt the user for their name
- Allow users to select one of the pregraded & stored coin images and pick the overall grade. The pregraded coin images will be in GIF format. Display the coin image prominently.
- After the user picks the grade on the displayed coin then show them the overall grade that they system thinks the coin is. Prior to showing the grade to the user we need to write the result to a file for statistical measurement. This is important; as we want to be able to see how good humans are at picking the grades of coins. The info that needs to be written to the file should minimally include: Username, Date, Time, Coin Info (date, mintmark or id), The user selected grade, the predetermined computer grade. This output can be written to an Access file that I can download from the web.
- Give the user the ability to do detailed feature grading. So after they get their results prompt them to do detailed feature grading which essentially moves them to **Part 2** with the data on the coin that they are presently working on. *Pass to the following: Username, date, coin id, year, mintmark, the overall technical grade of the coin and the grade of each of the 20 features*

Part 2: Feature Set Coin Grading (Detailed Grading)

Overview: This is the second screen that takes the passed data from Part 1 but expands the grading process into more detail.

A rough prototype of this site can be seen online at **http://matrix.csis.pace.edu/~f02-it608-s11/**

- As pictured on this site we need to identify the 20 key features of a Lincoln Cent on the left. (There are 6 additional features identified below) We would also like to duplicate the slider bar approach so that the grade for the feature changes as the slider bar is moved. The grades range from 1 to 70 but not all grades are used, *the last pages of this document* gives the grade scale ranges.

- On the right side of the prototype we also show a weight note that the weight for any of the 26 features can be 0 to 100% with the total of all features not to exceed 100%.

- We need to accept the data that from Part 1 coming into to this application and use it as the initial values. So if the coin is coming through as VG-8 then the overall technical grade should be held at VG-8. Each of the 20 sliders should be initialized to the feature value that is being received from Part 1. Note: Even though the features will initially be the same grade as the overall grade please use the passed 20 feature grades as down the road the features that we will be sending will be different from the overall technical grade.

- We want to somehow display a graphic of the overall coin and a graphic of each feature to the right of the slide bar and the weight

- The total grade can't exceed 100%

- Can we make the dropdown weight % a bit smaller in size than it is now?

- We need to add 6 subjective categories (with initial weights of 0%) to the end of the list. These categories are: Color, Toning, Defects, Strike Quality, Planchet Quality & Aesthetic Appeal. I think that it would be better if we defined them in a datafile somewhere and then let the user either override the descriptions or select them from a drop down list. The significance of this is that what is important to one user (eye appeal) may not be important at all to another user. Some users may want to include 1 or 3 or all 5 subjective categories in their human/machine grading process and may assign heavy weights while others may assign insignificant weights.

- We will need two output options: Summary Output & Detail Output - Summary will print a summarized record of the grade & Detail will print all of the features and both options will include the technical grade as well as the interactive grade. Perhaps this output can just be framed as a web page in a new window that users can just use the browser's print command to print.

- The evaluation as a % is really the grade. So it should try to give a description of the closest grade. For instance if it comes out at 8% it should say that it is a VG (Very Good)

- We should preserve the original technical grade (the info being passed to this applet) in a sep. field so a comparison to the human / machine interaction can be made later on.

- After the user changes the features (via moving the slide bars and changing the weighting values) they should have to hit a button that does the calculation of the grade instead of showing them the results as they are making changes. This gives us the opportunity to capture the data in this interactive session and write the data to a file for processing later on. It is important to capture the grading results to a file for statistical measurement, as we want to be able to see how rich the human/machine experience is (and the main point of the project for me). The info that needs to be written to the file should minimally include: Username, Date, Time, Coin Info (date, mintmark or id), The user computed grade and the predetermined technical computer grade and each of the 20 features, each of the 6 subjective factors selected and the corresponding weights of each. This output can be written to an Access file.

These are the deliverables that I need at the end:

We (you and I) can test this site in your development area until we agree that it is complete. After it is working I am planning on moving this entire system to my dissertation website by the end of this semester so that I can test it out on experts over the intersession timeframe. I am looking at having this done before 12/15/02 if possible.

- Electronic versions of all programs, websites and graphics
- Instructions on how to install these programs on my dissertation website
- Very brief documentation or a list that describes each file contained on the media and what it does (1 sentence for each would be great)
- I may need help installing getting this system to work on my website and may need to consult with you.

Besides the web sites working in the way that we expect one of the most important things for me is to capture the results so that statistical processing cane be done.

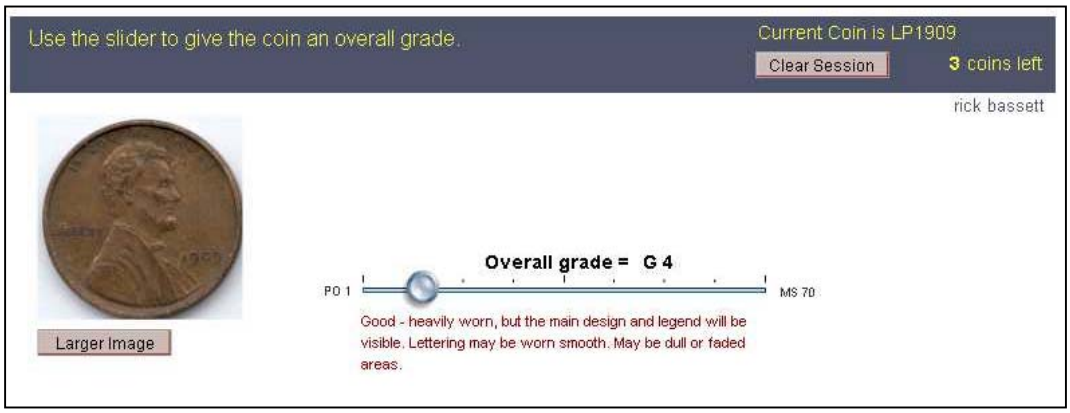
RARE COIN GRADING Scale

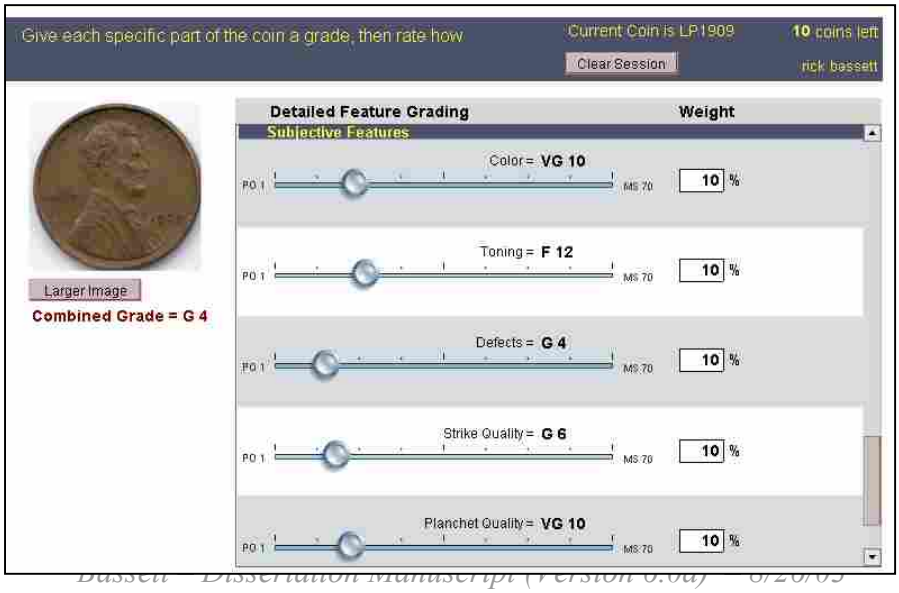
- **(1) Poor**, filler or cull (P) - barely recognizable, may contain holes
- **(2) Fair** (F) - very heavily worn; major portions may be completely smooth
- **(3) About Good** (AG) - heavily worn; date may be barely discernable
- **(4 & 6) Good** (G) - Coin will be heavily worn, but the main design and legend will be visible. Lettering may be worn smooth. May be dull or faded areas.
- **(8 & 10) Very Good (VG)** - Still well worn but more of the rim will be evident. Design and legend will be clear but worn flat. Lacks specific details.
- **(12 & 15) Fine (F)** - Medium to heavy wear but even overall. The design becomes clearer and details begin to appear. Some letters within the design will be apparent.
- **(20 & 25) Very Fine (VF)** - A visibly nicer coin. High spots will show light, even wear. Various major features are visible. Lettering is all readable.
- **(35, 40 & 45) Extra Fine (XF)** - Slight wear will show on the highest points of the main devices. Words are sharp and easily readable. All details are clearly defined.
- **AU 50** - Slight traces of wear on the highest points of the coin; may be dull with some evidence of luster under any toning.

- **AU 53** - Just slightly better than an AU 50 with a little more luster visible. Eye appeal begins to make a difference between the AU grades.
- **AU 55** - An obviously nicer coin than an AU 50 with no major difficulties. More luster shines through the surfaces.
- **AU 58** - This is oftentimes called a slider as it will appear to many observers to be uncirculated. Just the faintest wear on the highest points of the coin. Luster should be quite evident, although some toning can be apparent. Usually coins with poor eye appeal will not make the AU 58 grade.
- **MS 60** - Mint State indicates a coin that has no wear and is uncirculated. It may have numerous bagmarks and/or be toned. MS 60 is the lowest quality of an uncirculated coin.
- **MS 61** - An uncirculated coin that is just slightly better than MS 60. However, no question that it is uncirculated. Whereas, some may debate over the merits of a coin being MS60 because of the excessive bagmarks, the MS61 should be more desirable.
- **MS 62** - This coin should be a much cleaner specimen than an MS 60, yet, just slightly better than an MS 61. There should be fewer bagmarks as the coin takes on more attractive features.
- **MS 63** - This is the grade that many collectors feel is the most collectible in numismatics. Prices are typically reasonable compared to higher grades and the coin should have at least an average strike and eye appeal, with minimal distracting marks.
- **MS 64** - This is the grade where prices in many series begin to increase dramatically. For this reason the coin will begin to show fewer marks and the strike will be the strongest yet. No primary distractions that will draw your eye. A near-gem coin with just a few tiny marks or weakness in strike to keep it from a higher grade.
- **MS 65** - This is the gem category. Coin should be fully struck with eye appeal. Either brilliant or toned but there should not be any unsightly marks or color that negates eye appeal. Any marks should be very minor in appearance. Prices spread out even further.
- **MS 66** - A coin that just jumps out at you as being nicer than an MS 65. The main devices on either side should have no more than very minor ticks and the fields should be cleaner than that of an MS 65.
- **MS 67** - A superior coin that has no major distractions to speak of. The fields should be near flawless with just the slightest contact on the main device. This coin should emit a look of satisfaction from the viewer. Prices increase further especially for coins with short supplies and strong demand.
- **MS 68** - A difficult grade to determine by most experts. When does a coin become MS 68 but is not quite MS69 or 70? A very superior coin with maybe just a minor tick on either side keeping it from perfection.
- **MS 69** - This is a coin that should create a gasp when viewed. There should be no imperfections to the naked eye. With a magnifying glass a minor mark or impediment may be visible.
- **MS 70** - A perfect coin with no imperfections seen with a magnifying glass. There should be no marks whatsoever; the coin must look like it just left the Mint. Very unusual in early coins as the mint did not have the quality they do today. Modern coins have been given this exalted grade although there is debate whether coins can be perfect.

Appendix B2a

Online Grader Screens - Second Generation





Appendix B3

Design Document for Online Grader Third Generation

By: Rick Bassett – 03/10/03 – Revised 03/15/03

I will construct a 2nd experiment that allows for human / machine grading.

This will be a dynamic web-based application that allows users to look at a series of 20 coin images presented to them (one at a time), with some information on the image that they are looking, that extracts feedback from the user about the image.

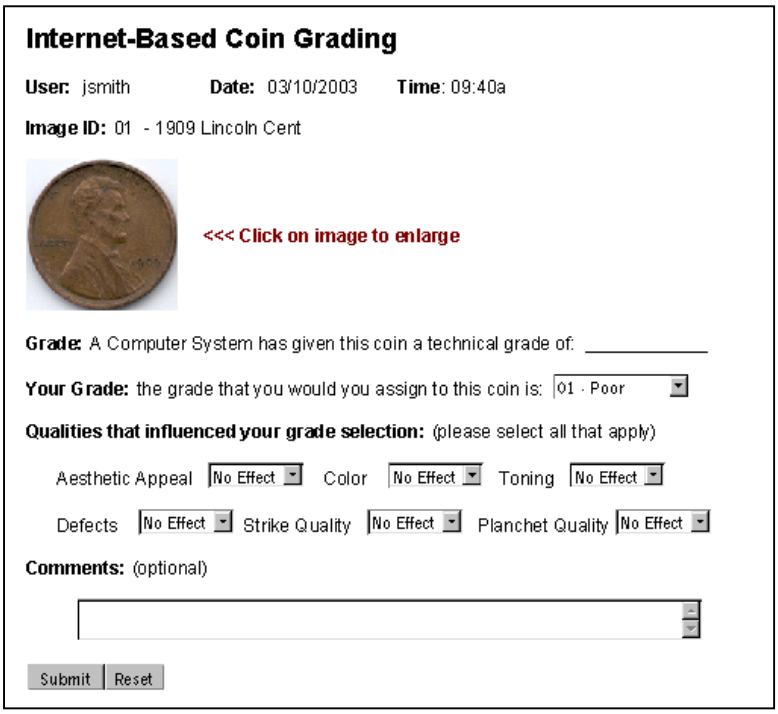
The images presented will be coin images of the coins that were evaluated by the machine-based system and that were graded by the 3rd party grading services.

For experimental purposes we will rotate the machine grade that we display to the user. So the first user that signs into the system gets machine grade 1, the 2nd gets known grade 2 and the third gets known grade 3 and the 4th gets cycled back to known grade 1 again. This rotation is done so that each user will evaluate coin images with actual machine grades, with unknown grades and with misleading grades. The following table shows how this rotation will occur.

Grading Consultant	Machine or Service Grade Provided	No Machine Grade Provided	Misleading Machine Grade Provided	Duplicates w/ Actual & No Info
1	Coins 1 – 6	Coins 7 – 12	Coins 13 - 18	2 (3,9)
2	Coins 7 – 12	Coins 13 – 18	Coins 1 – 6	2 (9,15)
3	Coins 13 – 18	Coins 1 –6	Coins 7 -12	2 (15,3)
4	Coins 1 – 6	Coins 7 – 12	Coins 13 - 18	2 (3,9)
5	Coins 7 – 12	Coins 13 – 18	Coins 1 – 6	2 (9,15)
6	Coins 13 – 18	Coins 1 –6	Coins 7 -12	2 (15,3)
7	Coins 1 – 6	Coins 7 – 12	Coins 13 - 18	2 (3,9)
8	Coins 7 – 12	Coins 13 – 18	Coins 1 – 6	2 (9,15)
9	Coins 13 – 18	Coins 1 –6	Coins 7 -12	2 (15,3)

The information on the images (image id, image description, known grade (1,2 or 3) and URL to the image itself) will be stored in a table within an Access table. User feedback on the image must be written to another table in the Access database that contains the following fields (image id, image description, known grade, date, time, user name, user email, user grade, subjective qualities check box choices 1 – 6 and comments.

There will be at least 9 grading consultants evaluating 20 coins.



Human/Machine Grading Experiment - Prototype

The screen shot above is a prototype of the screen that each grading consultant will see for each image that they must evaluate. Again the grade provided will be an actual, no grade, or misleading grade

Appendix B4

Design Document Automated Machine Grader

1.1 Overview

The grading system will entail three areas, which will be linked via proprietary software to produce an accurate and uniform result. This result will be the product of a comparison of like features. The features will be selected and graded by a predetermined database of graded characteristics consistent with recognized grading values. The information below details the system and its processes.

1. **Database**
    a. Coins
        i. A Lincoln penny was selected due to its low per unit cost and abundance
        ii. 100-200 Lincoln pennies are selected to represent all possible grades throughout its entire run of 1909-present
        iii. Each coin will be scanned and put into a database labeled with a grade (industry recognized value system)
    b. Scan Process
        i. A high resolution scanner/digital camera will be used to capture the image
        ii. The scanner/camera will be manipulated/adjusted in order to ensure environment consistency throughout the image capturing process i.e. uniform background, lighting, shadows etc…
        iii. All images will be saved as .gif files in order to minimize scan time and aid in ease of data processing
    c. Image Gathering
        i. All .gif files will be cataloged and labeled with a recognized industry label ranging from 'about good to uncirculated'
        ii. The standards will be determined according to the A.N.A (American Numismatic Association)
    d. Feature extraction
        i. A select number of coin features will be designated to represent the criteria for the grading scale
        ii. Selected features of each scanned coin will be extracted and coded with its respective grade to provide a baseline for future comparisons
2. **Grade Scan**
    a. A coin will be placed in the secure environment
    b. The coin will be scanned and saved in .gif format
        i. The .gif file will be data based and selected features extracted
    c. The extracted features will then be saved and compared to the preexisting database of grade features related to that particular coin and year

3. **Software**
    a. A custom made software package will compare each individual feature to the existing grade specific identical feature
        i. From this comparison a grade level will be given to each individual characteristic
        ii. The results will be recorded to a separate database
    b. A coding scheme will be implemented to determine the overall grade of the coin (feature #1 = good, feature #2 = fine, feature #3 = very good == overall rating of coin is very good)
4. **Result**
    a. Each result will be printed on white paper
        i. Printed information will include:
            1. Name, date, coin year and name, grade given
               Valuation specific order code for future reference

1.2 Project Scope

The scope of this project for this course is to gather and finalize the design requirements for the new system for processing rare coins and then to expand the usage scope to other valuable collectibles such as postage stamps, cards and possibly antiques. We will analyze the effectiveness of the system designed. To achieve our goals, we will communicate only with the clients presented by Dr. Tappert. By the end of the semester, we will deliver to our clients a working system for grading coins of type 'Lincoln cents'. The final working system will meet all of the requirements stated in the Project Description.

1.3. Risk Analysis

The following includes a list of risks associated with this project

- *Risk 1:* Inconsistent availability of adequate resources.
- *Risk 2:* Negligence of quality assurance due to time constraints.
- *Risk 3:* Performance metrics need to be determined to design a better system. These metrics are difficult to capture.
- *Risk 4:* Changes to requirements have the potential to adversely affect the project schedule and jeopardize the project

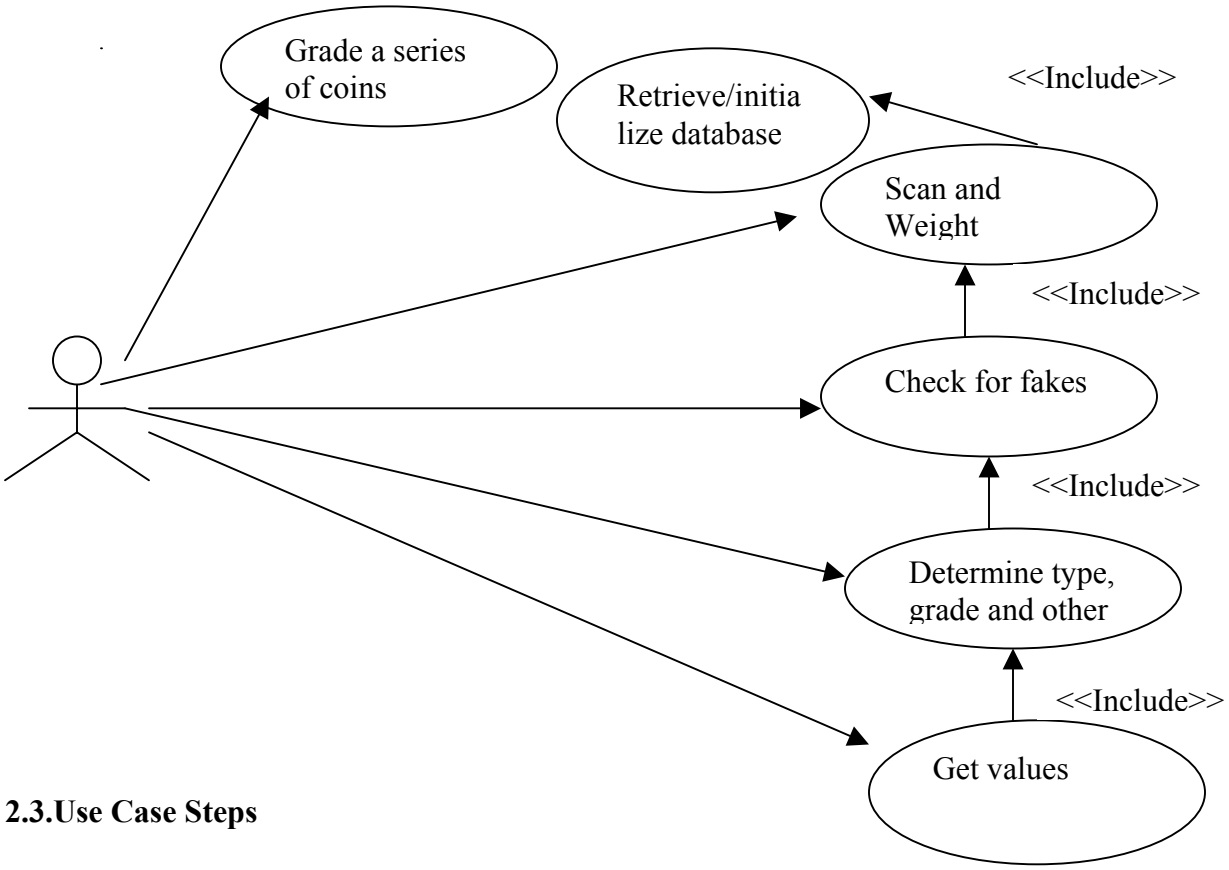
2.1. Use Case Simple Prose

Name: Grade rare coins
Description: A user, most probably a Coin Expert or Dealer, will initialize the system if it has not already been initialized. He/She will then place a coin on a scanner at which time a high-resolution scan of the coin will be made. The weight in grams will also be taken and recorded. After the scanning and weighing are completed the user can choose one of three functions:

- Make a determination to see if the coin is a counterfeit or if the coin has been altered by comparing the image to a database of known fakes.
- Compare the scanned visual image to a large graphical database to determine the type, denomination, date and an accurate grade of the coin.
- Cross-reference the information obtained and compare it against a database of market values to assign an appropriate asset value. Obtaining a proper asset value is extremely important for securing proper insurance as well as knowing what price should be charged in asset disposition

After this the user can choose another function for the same coin or begin with another coin.

2.2. Use Case Diagrams



2.3.Use Case Steps

Name: Grade a series of coins
Description: Grade the Rare Coins
Preconditions: A user is eligible to use this system
Post conditions:

- The coin grader will provide the values of a number of coins to the user

Basic Course of Action

1. The user wants to grade coins
2. The user initializes the system
3. The user scans and weighs a single coin
4. The system will record all information about the coin such as the coin image and weight etc
5. The system will prompt the user to choose to check for a fake, or determine grade, date, etc., or to determine value
6. The system determines the information about the coin using stored images and information in the database.
7. The system reports the information to the user and records the information in the database.

8. The system prompts the user to examine another coin or to check for more information about this coin.
9. If the user chooses to examine another coin, return to step 3, otherwise return to step 5.

Name: Retrieve/Initialize Database
Description: Read the database in from the disk if necessary and flag it as initialized.
Preconditions: A user is eligible to use this system
Post conditions:

- The Database is initialized.

Basic Course of Action

1. The user wants to use the system.
2. The database is checked to see if it has already been initialized. If it has, the use case ends.
3. The database is read in from the disk and flagged as initialized.
4. The use case ends.

Name: Scan and Weigh Coin
Description: Scan and weigh an individual coin
Preconditions: A user is eligible to use this system
Post conditions:

- The Coin has been scanned and weighed.

Basic Course of Action

1. The user wants to use the system.
2. The database is checked to see if it has already been initialized. If it has not, perform the use case named "Retrieve/Initialize Database".
3. Prompt the user to place the coin on the scanner and press <enter>.
4. After the <enter> key has been pressed, scan the coin.
5. Store the coins image in the database.
6. Prompt the user to place the coin on the scale, weigh it and enter the weight.
7. Store the coins weight in the database.
8. The use case ends.

Name: Check Coin for Fake
Description: Check an individual coin against a database of known fakes. **Preconditions**: A user is eligible to use this system
Post conditions:

- The Coin has been checked for fakes.

Basic Course of Action

1. The user wants to use the system.

2. If the coin has not been scanned and weighed, perform the use case named "Scan and Weigh Coin".
3. Check the coins image in the database against the known fakes in the database.
4. The use case ends.

Name: Determine type, grade and other information
Description: Check the coin image to determine the grade and type
Preconditions: A user is eligible to use this system
Post conditions:

- The grade of the coin has been determined

Basic Course of Action

1. The user wants to use the system.
2. If the coin has not been checked for fakes, perform the use case named "Check for fakes".
3. Determine the type, grade and other information.
4. Store in the database
5. The use case ends.

Name: Get Value
Description: Check the coin image to get monetary value for the coin
Preconditions: A user is eligible to use this system
Post conditions:

- Get the value for the coin

Basic Course of Action

1. The user wants to use the system.
2. If the coin has not been graded, perform the use case named "Determine type, grade and other".
3. Determine the value.
4. Store in the database
5. The use case ends

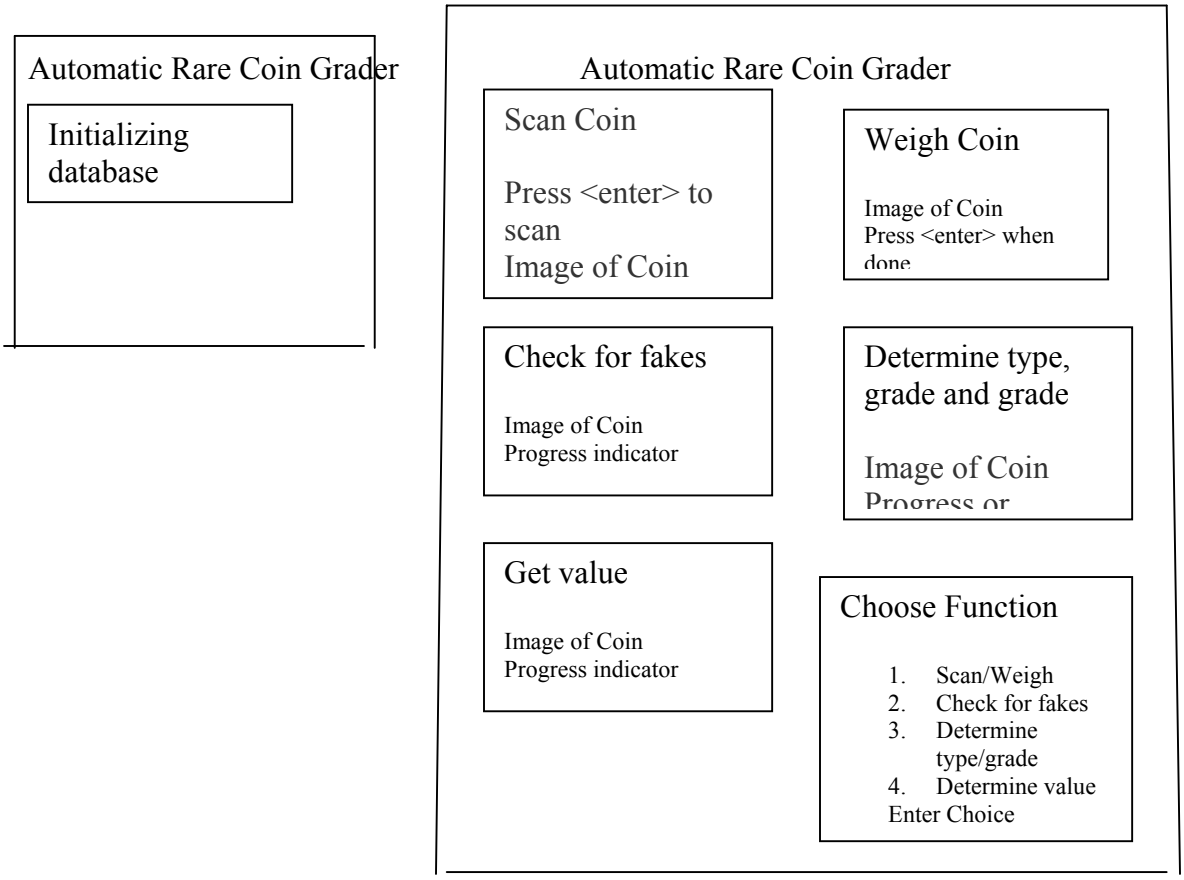
2.4. Use Case Scenarios

A user wants to grade the coins
Description: The user puts one coin at a time on the Automatic Rare Coin Grader to determinates the values
Steps

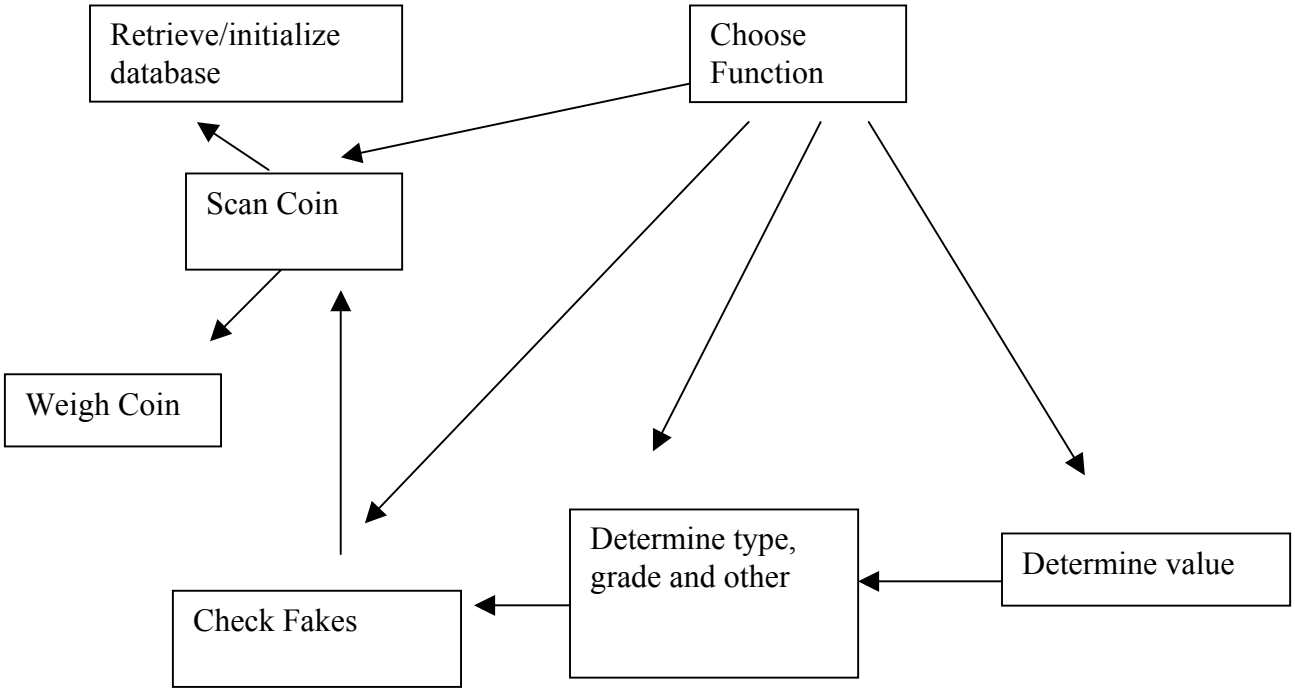
1. The user wants to grade coins
2. The user initializes the system
3. The user scans and weighs a single coin
4. The system will record all information about the coin such as the coin image and weight etc
5. The system will prompt the user to choose to check for a fake, or determine grade, date, etc., or to determine value. The user chooses to determine the value of the coin.

6. The system determines the information about the coin using stored images and information in the database.
7. The system reports the information to the user and records the information in the database.
8. The system prompts the user to examine another coin or to check for more information about this coin.
9. The user decides to grade another coin. Return to step 3
10. After the user grades several coins the scenario ends.

2.5. Use Interface Prototype Model



2.6. Use Interface Flow Diagrams



Appendix C

Coin Grading Domain Information

This appendix provides background information on the specific domain terminology of coin collecting and grading. It is meant to serve as a reference for readers that may not have a background in the collectibles field.

Appendix C1

Coinage Terminology

AG (About Good) Abbreviation. See definition below.

ANA (American Numismatic Association) A non-profit organization founded in 1888 for the advancement of numismatics.

ANACS (American Numismatic Association Certification Service) A service established by the ANA in 1972 to authenticate coins; its mission was broadened in 1979 to include coin grading as well. The ANA subsequently sold the grading portion of the service to Amos Press Inc. Of Sidney, Ohio, which operates it today under the acronym ANACS, although those letters no longer are shorthand for the full original name.

AU (Almost Uncirculated or About Uncirculated) Abbreviation. See definition below.

About Good The grade AG-3. The grade of a coin that falls short of Good. Only the main features of the coin are present in this grade. Peripheral lettering, date, stars, etc., sometimes are worn away partially.

Almost Uncirculated (alt. About Uncirculated) The term(s) corresponding to the grades AU-50, 53, 55, and 58. A coin that at first glance appears Uncirculated but upon closer inspection has slight friction or rub.

alteration A coin that has a date, mint mark, or other feature that has been changed, added, or removed, Usually to simulate a rarer issue.

artificial toning Coloring added to the surface of a coin by chemicals and/or heat. Many different methods have been employed over the years.

attributes The elements that make up a coin's grade. The main ones are marks (hairlines for Proofs), luster, strike, and eye appeal.

BN (Brown) A PCGS grading suffix used for copper coins that meet Brown standards. See definition on next page.

BU (Brilliant Uncirculated) Abbreviation. See definition on next page

bag friction Coin-on-coin friction that is the result of coins rubbing against each other in a bag. See "coin" friction, toll friction.

bag mark A generic term applied to a mark on a coin from another coin; it may, or may not, have been incurred in a bag.

bag toning Coloring acquired from the bag in which a coin was stored. The cloth bags in which coins were transported contained sulfur and other reactive chemicals. When stored in such bags for extended periods, the coins near and in contact with the cloth often acquired beautiful red, blue, yellow, and other vibrant colors. Sometimes the pattern of the cloth is visible in the toning; other

times, coins have crescent-shaped toning because another coin was covering part of the surface, preventing toning. Bag toning is seen mainly on Morgan silver dollars, though occasionally on other series.

basal state The condition of a coin that is identifiable only as to date mint mark (if present), and type; one year-type coins may not have a date visible.

basal value The value base from which Dr. William H. Sheldon's 70-point grade/price system started; this lowest grade price was one dollar for the 1794 large cent upon which he based his system.

blank The flat disk of metal before it is struck by the dies and made into a coin. See planchet.

blended A term applied to an element of a coin (de- sign, date, lettering, etc.) that is worn into another element or the surrounding field.

branch mint One of the various subsidiary government facilities that struck, or still strikes, coins.

brilliant A coin with full luster, unimpeded by toning, or impeded only by extremely light toning.

Brilliant Uncirculated A generic term applied to any coin that has not been in circulation. It often is applied to coins with little "brilliance" left, which properly should be described as simply Uncirculated.

bronze An alloy of copper, tin, and zinc, with copper the principal metal.

Brown The term applied to a copper coin that no longer has the red color of copper. There are many "shades" of brown color-mahogany, chocolate, etc. (abbreviated as BN when used as part of a grade).

buckled die A die that has "warped" in some way, possibly from excess clashing, and that produces coins which are slightly "bent." This may be more apparent on one side and occasionally apparent only on one side.

bulged die A die that has clashed so many times that a small indentation is formed in it. Coins struck from this die have a "bulged" area.

burnishing A process by which the surfaces of a planchet or a coin are made to shine through rubbing or polishing. This term is used in two contexts--one positive, one negative. In a positive sense, Proof planchets are burnished before they are struck-a procedure done originally by rubbing wet sand across the surfaces to im-part a mirrorlike finish. In a negative sense, the surfaces on repaired and altered coins sometimes are burnished by various methods. In some instances, a high-speed drill with some type of wire brush attachment is used to achieve this effect.

burnishing lines Lines resulting from burnishing, seen mainly on open-collar Proofs and almost never found on close-collar Proofs, These lines are incuse in the fields and go under lettering and devices.

business strike (alt. regular strike) A regular-issue coin, struck on regular planchets by dies given normal preparation. These are coins struck for commerce that the government places into circulation.

bust The head and shoulders of the emblematic Liberty seen on many U.S. coins. See Capped Bust and Draped Bust.

cabinet friction Slight disturbance seen on coins (usually on the obverse) that were stored in wooden cabinets used by early collectors to house their specimens. Often a soft cloth was used to wipe away dust, causing light hair- lines or friction.

cameo A term applied to coins, usually Proofs and prooflike coins, that have frosted devices and lettering that contrast with the fields. When this is deep, the coins are said to be "black-and-white" cameos. Occasionally, frosty coins have "cameo" devices that do not contrast as dramatically with the fields. Specifically applied by PCGS to 1950 and later Proofs that meet cameo (CAM) standards.

carbon spot A spot seen mainly on copper and gold coins, though also found occasionally on nickel U.S. coins (which are 75 percent copper) and silver coins, (which are 10 percent copper). Carbon spots are brown to black spots of oxidation that range from minor to severe--some so large and far advanced that the coin is not graded because of environmental damage. See copper spot.

cartwheel The pleasing effect seen on some coins when they are rotated in a good light source. The luster rotates around like the spokes of a wagon wheel. A term applied mainly to frosty Mint State coins, especially silver dollars, to describe their luster. Also, a slang term for a silver dollar.

cast blanks Planchets made by a mold method, rather than being cut from strips of metal.

cast counterfeit A replication of a genuine coin usually created by making molds of the obverse and reverse, then casting base metal in the molds. A seam is usually visible on the edge unless it has been ground away.

cent A denomination valued at one-hundredth of a dollar, struck continuously by the U.S. Mint since 1793 except for 1815.

chasing A method used by forgers to create a mint mark on a coin. It involves heating the surfaces and moving the **metal** to form the mint mark.

choice An adjectival description applied to a coin's grade, e.g., choice Uncirculated, choice Very Fine, etc. Used to describe an especially attractive example of a particular grade.

circulated A term applied to a coin that has wear, ranging from slight rubbing to heavy wear.

clad A term for any of the modem "sandwich" coins that have layers of different alloys.

clashed dies Dies that have been damaged by striking each other without a planchet between them. Typically, this imparts part of the obverse image to the reverse die and vice versa.

clash marks The images of the dies seen on coins struck from clashed dies. The obverse will have images from the reverse and vice versa.

cleaned A term applied to a coin whose original surface has been removed. The effects may be slight or severe, depending on the method used.

clipped A term for an irregularly cut planchet. A clip can be straight or curved, depending upon where it was cut from the strip of metal.

clogged die A die that has grease or some other contaminant lodged in the recessed areas. Coins struck from such a die have diminished detail, sometimes completely missing.

close(d) collar The edge device, sometimes called a collar die, that surrounds the lower die. Actually, open and close collars are both closed collars, as opposed to segmented collars. The close collar imparts either reeding or a smooth, plain edge.

coin Metal formed into a disk of standardized weight and stamped with a standard design to enable it to circulate as money authorized by a government body.

coin" friction Coin-on-coin friction imparted to coins when they rub together in rolls or bags and a light amount of metal is displaced. See also *roll friction* and *bag friction.*

collar A device placed around the lower die to prevent excessive spreading or later to impart reeding or devices to the edge.

commercial grade A grade that is usually one level higher than the market grade; refers to a coin that is "pushed" a grade, such as an EF/AU coin (corresponding to 45+) sold as AU-50.

consensus grading The process of determining the condition of a coin by using multiple graders.

contact marks Marks on a coin that are incurred through contact with another coin or a foreign object. These are generally small, compared to other types of marks such as gouges. See bag marks.

copper spot (stain) A spot or stain commonly seen on gold coinage, indicating an area of copper-concentration that has oxidized. Copper spots or stains range from tiny dots to large blotches.

copy Any reproduction, fraudulent or otherwise, of a coin.

copy dies Dies made at a later date, usually showing slight differences from the originals. Examples include the reverse of 1804 Class 11 and III silver dollars and 1831 half cents with the Type of 1840-57 reverse. Also used to denote counterfeit dies copied directly from a genuine coin.

corrosion Damage that results when reactive chemicals act upon metal. When toning ceases to be a "protective" coating and instead begins to damage a coin, corrosion is the cause. Usually confined to copper, nickel, and silver regular issues, although patterns in aluminum, white metal, tin, etc., also are subject to this harmful process.

counterfeit Literally, a coin that is not genuine. There are cast and struck counterfeits and the term is also applied to issues with added mint marks, altered dates, etc.

counting-machine mark A dense patch of lines caused by the rubber wheel of a counting machine where the wheel was set with insufficient spacing for the selected coin. Many coins have been subjected to counting machines-among these are Mercury dimes, Buffalo nickels, Walking Liberty half dollars, Morgan and Peace **dollars,** and Saint-Gaudens double eagles.

cud An area of a coin struck by a die that has a complete break across part of its surface. A cud may he either a *retained cud*, where the faulty piece of the die is still retained, or a *full cud*, where the piece of the die has fallen away. Retained cuds usually have dentil detail, while full cuds do not.

cupro-nickel Any alloy of copper and nickel. Now used primarily in referring to the modern "sandwich" coinage. The Flying Eagle and early Indian Head cents, nickel three-cent pieces, and nickel five-cent pieces also are cupro-nickel coins.

date The numerals on a coin representing the year in which it was struck. Restrikes are made in years subsequent to the one that appears on them.

deep cameo A term applied to coins, usually Proofs and prooflike coins, that have deeply frosted devices and lettering that contrast with the fields--often called "black- and-white" cameos. Specifically applied to those Proofs dated 1950 and later that meet deep cameo (DCAM) standards.

deep mirror prooflike A term applied to any coin that has deeply reflective mirrorlike fields, especially Morgan dollars. Those Morgan dollars that meet PCGS standards are designated deep mirror prooflike (DMPL).

denomination The value assigned by a government to a specific coin.

dentils (alt. denticles) The toothlike devices seen around the rim on many coins. Originally, these were somewhat irregular; later, they became much more uniform-the result of better preparatory and striking machinery.

design type A specific motif placed upon coinage which may be used for several denominations and subtypes, e.g., the Liberty Seated design type used for silver coins from half dimes through dollars and various subtypes therein.

device Any specific design element. Often refers to the principal design element, such as the head of Miss Liberty.

device punch A steel rod with a raised device on the end used to punch the element into a working die. This technique was used before hubbed dies became the norm.

die A steel rod that is engraved, punched, or hubbed with devices, lettering, the date, and other emblems.

die alignment The condition in which the obverse and reverse dies are aligned properly and therefore strike a coin evenly. When the dies are out of alignment, two things can happen: If the dies are out Of parallel, weak- ness may be noted in a quadrant of the coin's obverse and the corresponding part of the reverse; and if the dies are spaced improperly, the resultant coins may have overall weakness.

die break A defect in a die that has cracked during use; if not removed from service, such a die may break eventually. See also die crack.

die crack (alt. die break) A defect in a die. When dies crack, the coins struck from those dies have raised, irregular lines ranging from very slight to very large@-- some of them quite irregular. When a

die breaks apart totally, the break will result in a full-or retained-cud, depending upon whether the broken piece falls from the die.

die line (alt. die scratch) A polish line on a die; being incuse, it results in a raised line on the coins struck with that die.

die rust Rust that has accumulated on a die that was not stored properly. Often such rust was polished away, so that only the deeply recessed parts of the die still exhibited it. A few examples are known of coins that were struck with extremely rusted dies-the 1876-CC dime, for one.

die state A readily identified point in the life of a coinage die. Often dies clash and are polished, crack, break, etc., resulting in different stages of the die. These are called die states. Some coins have barely distinguishable die states, while others go through multiple distinctive ones.

die variety A coin that can be linked to a given set of dies because of characteristics possessed by those dies and imparted to the **coin** at the time it was struck. In the early years of U.S. coinage history, when dies were made by hand engraving or punching, each die was slightly different. The coins from these unique dies are die varieties and are collected in every denomination. By the 1840s, when dies were made by hubbing and therefore were more uniform, die varieties resulted mainly from variances in the size, shape, and positioning of the date and mint mark.

die wear Deterioration in a die caused by excessive use. This may evidence itself on coins produced with that die in a few indistinct letters or numerals or, in extreme cases, a loss of detail throughout the entire coin. Some coins, especially certain nickel issues, have a fuzzy, indistinct appearance even on Uncirculated examples.

dime (alt. disme) A denomination, one-tenth of a dollar, that has been struck from 1796 **to** date, with the exception of only a few years.

ding Slang term for a small to medium-size mark on a coin. See rim ding.

dipped A term applied to a coin that has been placed in a commercial "dip" solution, a mild acid wash that removes the toning from most coins. Some dip solutions employ other chemicals, such as bases, to accomplish a similar result. The first few layers of metal are removed with every dip, so coins dipped repeatedly will lose luster, hence the term *overdipped*.

dipping solution Any of the commercial "dips" available on the market, usually acid-based.

dollar The basic unit, along with the eagle, for the currency of the United States. At the time of the Mint Act of 1792 establishing the U.S. coinage system, a dollar was comparable to the Central European thaler and the Spanish silver peso of 8 reales, containing nearly an ounce of silver. Its legal value in silver and gold was modified over the years on a number of occasions. Today, it is a fiat-money unit with no fixed legal equivalent in silver or gold; rather, its purchasing power rises or falls, in a general way, in accordance with fluctuations in the market value of gold.

doubled die A die that has been struck more than once by a hub in misaligned positions, resulting in doubling of design elements. Before the introduction of hubbing, the individual elements of a coin's design were either engraved or punched into the die, so any doubling was limited to a specific element. With hubbed dies, multiple impressions are needed from the hub to make a single die with adequate detail. When shifting occurs in the alignment between the hub and the die, the die ends up

with some of its features doubled--then imparts this doubling to every coin it strikes. The coins struck from such dies are called doubled-die errors--the most famous being the 1955 Doubled Die Lincoln cent.

double eagle Literally two eagles, or twenty dollars. A twenty-dollar U.S. gold coin issued from 1850 through 1932. (One double eagle dated 1849 is known and is part of the National Numismatic Collection at the Smithsonian Institution. Nearly half a million examples dated 1933 were struck by the U.S. Mint, but virtually all were melted when private gold ownership was outlawed that year, and currently federal officials claim it is illegal to own any specimens that survive.)

double-struck A condition that results when a **coin is** not ejected from the dies and is struck a second time. Such a coin is said to be double-struck. Triple-struck coins and other multiple strikings also are known. Proofs are usually double-struck on purpose in order to sharpen their details; this is sometimes visible under magnification.

drift mark An area on a coin, often rather long, that has a discolored, streaky look. This is the result of impurities or foreign matter in the dies. One theory is that burnt wood was tolled into the strips from which the planchets were cut, resulting in these black streaks.

EAC (Early American Coppers) Abbreviation for an association made up of collectors who specialize in early U.S. copper coins, especially large cents, which they generally collect by Sheldon numbers.

EF (Extremely Fine or Extra Fine) Abbreviation. See definition on next page.

eagle A gold coin with a face value of ten dollars. Along with the dollar, this was the basis of the U.S. currency system from 1792 until 1971. No U.S. gold coins were struck for circulation after 1933, and all gold coins issued prior to that time were recalled from circulation.

edge The third side of a coin. It may be plain, reeded, or ornamented--with lettering or other elements raised or incuse.

edge device A group of letters or emblems on the edge of a coin. Examples would be the stars and lettering on the edge of Indian Head eagles and Saint-Gaudens double eagles.

electrotype A duplicate coin created by the electrolytic method, in which metal is deposited into a mold made from the original. The obverse and reverse metal shells are then filled with metal and fused together-after which the edges sometimes are filed to obscure the seam.

elements For the purposes of this book, the various components of grading. In other numismatic contexts, this term refers to the various devices and emblems seen on coins.

emission sequence The order in which die states are struck.

engraver The person responsible for the design and/or punches used for a particular coin.

envelope toning A term applied to toning that results from storage mainly in 2 x 2 manila envelopes; most paper envelopes contain reactive chemicals.

environmental damage Damage to a coin that results from exposure to the elements. This may be minor, such as toning that is nearly black, or major--as when a coin found in the ground or water has

severely pitted surfaces. PCGS does not grade coins with more thaw very minor environmental damage.

eroded die See worn die.

Extremely Fine (alt. Extra Fine) The term corresponding to the grades EF-40 and EF-45. Coins in these grades have nearly full detail with only the high points worn and the fields lightly rubbed; often, luster still clings in protected areas.

eye appeal The element of a coin's grade that "grabs" the viewer. The overall look of a coin.

F (Fine) Abbreviation. See definition below.

FB (full bands) A PCGS grading suffix used for Mercury dimes that meet the standards for full bands. See definition below.

FBL (full bell lines) A PCGS grading suffix used for Franklin half dollars that meet the standards for full bell lines. See definition on next page.

FH (full head) A PCGS grading suffix used for Standing Liberty quarters that meet the standards for a full head. See definition on next page.

FR (Fair) Abbreviation. See definition below.

FS (full steps) A PCGS grading suffix used for Jefferson nickels that meet the standards for full steps. See definition on next page.

Fair The adjective corresponding to the grade FR-2. In this grade, there is heavy wear with the lettering, devices, and date partially visible.

fantasy piece A term applied to coins struck at the whim of Mint officials. Examples include the 1868 large cent Type of 1857 and the various 1865 Motto and 1866 No Motto coins.

fasces A Roman symbol of authority used as a motif on the reverse of Mercury dimes. It consists of a bundle' of rods wrapped around an ax with a protruding blade. The designation full bands refers to fasces on which there is complete separation in the central bands across the rods.

field The portion of a coin where there is no design- generally the flat part (although on some issues, the field is slightly curved).

finalizer A PCGS grader who, before computers were used for this task, compared his own grade with those of other graders and determined the final grade. The verifier replaced the finalizer after PCGS began inputting the grades by computer.

Fine The adjective corresponding to the grades F-12 and 15. In these grades, most of a coin's detail is worn away. Some detail is present in the recessed areas, but it is not sharp.

first strike A coin struck early in the life of a die. First strikes sometimes are characterized by striated or mirrorlike fields if the die was polished. Almost always fully or well struck, with crisp detail.

flat luster A subdued type of luster seen on coins struck from worn dies. Often these coins have a gray or other- wise dull color that makes the fields seem even more lackluster.

flip A clear plastic holder, usually pliable, that is used to store an ungraded ("raw") coin. Flips are used to house coins when they are submitted to PCGS, but are not recommended for long-term storage if they contain polyvinyl chloride, or PVC. Care should be taken with PVC-free flips, as they usually are very brittle and can damage the delicate surfaces of a coin. (See PVC.)

flip rub Discoloration, often only slight, on the highest points of a coin resulting from contact with a flip. On occasion, highly desirable coins sold in auctions have acquired minor rub from being examined repeatedly in flips by eager bidders.

flow lines (alt. stress lines) Lines, sometimes visible, resulting when the metal flows outward from the center of a planchet as it is struck. "Cartwheel" luster is seen when light is reflected from these radial lines.

focal area The area of a coin to which a viewer's eye is drawn. An example is the cheek of a Morgan dollar.

friction Slight wear on a coin's high points or in the fields.

frost A crystallized-metal effect seen in the recessed areas of a die, thus the raised parts of a coin struck with that die. This is imparted to dies by various techniques, such as sandblasting them or pickling them in acid, then polishing the fields, leaving the recessed areas with frost.

frosty luster The crystalline appearance of coins struck with dies that have frost in their recessed areas. Such coins show vibrant luster on their devices and/or surfaces; the amount of crystallization may vary.

full bands (alt. full split bands) A term used to describe the central bands of the fasces on a Mercury dime's reverse when they are fully separated. The FB designation indicates an unusually sharp strike and is highly desired by collectors. To qualify for this designation, a coin can have no disturbance of the separation.

full bell lines A condition in which the lower set of lines on the Liberty Bell on Franklin half dollars are fully visible. Very slight disturbance of several lines is acceptable. Full bell lines (FBL) indicate sharpness of strike and seldom are seen on coins of certain dates.

full head A term used to describe Miss Liberty's head on Standing Liberty quarters when the helmet on her head has full detail. The FH designation can apply to both Type I and Type II coins, but the criteria are different. See chapter 5 for full-head standards.

full steps A term applied to Jefferson nickels when 51/2 or 6 steps are fully defined in the portrait of Monticello, Thomas Jefferson's home, on the reverse.

G (Good) Abbreviation. See definition below.

gem Adjectival description applied to Mint State and Proof-65 coins. It also is used for higher grades and as a generic term for a superb coin.

Good. The adjective corresponding to the grades G-4 and G-6. Coins in these grades usually have little detail but outlined major devices. On some coins, the rims may be worn to the tops of some letters.

grade The level of preservation of a particular coin. For circulated coins, this is mainly a function of wear, whereas marks, luster, strike, and eye appeal are the principal elements used to determine the grades of Mint State and Proof coins.

grader An individual who evaluates the condition of coins.

grading The process of numerically quantifying the condition of a coin. Before the adoption of the Sheldon numerical system, coins were given descriptive grades such as Good, Very Good, Fine, and so forth.

grading standards The rules, descriptions, and conventions applied to the grading of coins. PCGS has written standards as well as a grading set representing these standards.

hairlines Fine cleaning lines found mainly in the fields of Proof coins, although they sometimes are found across an entire Proof coin as well as on business strikes.

half cent The lowest-value coin denomination ever issued by the United States, representing one-two hundredth of a dollar. Half cents were struck from 1793 until the series was discontinued in 1857.

half dime (alt. half dime) A coin denomination, one twentieth of a dollar, struck from 1792 until it was discontinued in 1873.

half dollar A coin denomination, one half of a dollar, struck nearly continuously since 1794.

half eagle A coin denomination, valued at five silver dollars, struck from 1795 until 1916 and again in 1929. Half eagles were recalled, as were all U.S. gold coins, in 1933.

hammer die (alt. hammered die) The upper die-- usually the obverse, although on some issues with striking problems the reverse die was placed on top to improve striking quality. In early minting terminology, a hammer die was literally that: the lower die was fixed in a tree stump, the planchet was placed on top of it, and the upper die then was placed upon the planchet and struck with a hammer.

haze A cloudy film, original or added, seen on both business-strike coins and Proofs. This film can range from a light, nearly clear covering with little effect on the grade to a heavy, opaque layer that might prevent the coin from being graded.

high-end A term applied to any coin at the upper end of a particular grade. See premium quality.

high relief A condition in which the design elements of a coin stand out dramatically above the flat fields. This three-dimensional effect is achieved through the use of dies with deeply recessed devices and multiple strikings. The term high relief is applied specifically to Saint-Gaudens double eagles with the Roman numerals date MCMVII (1907). Coins with high relief often do not strike up properly in a single blow from the dies, leading to weak central detail.

holder Any device for housing a coin. PCGS holders are made of hard, inert plastic and are sonically sealed and tamper-resistant.

holder toning Any toning acquired by a coin as a result of storage in a holder. Mainly refers to toning seen on coins stored in Wayte Raymond-type cardboard holders which contained sulfur and other reactive chemicals. Sometimes vibrant, spectacular reds, greens, blues, yellows, and other colors are seen on coins stored in these holders.

hub Minting term for the steel device from which a die is produced. The hub is produced with the aid of a portrait lathe or reducing machine and bears a "positive" image of the coin's design-that is, it shows the design as it will appear on the coin itself. The image on the die is “negative" a mirror image of the design.

impaired Proof A Proof coin that grades less than PR-60; a circulated Proof. See mishandled Proof.

incomplete strike A coin that is missing design detail because of a problem during the striking process. The incompleteness may be due to insufficient striking pressure or improperly spaced dies.

incuse design The intaglio design used on Indian Head quarter eagles and half eagles. These coins were struck from dies which had fields recessed, so that the devices--the areas usually raised--were recessed on the coins themselves. This was an experiment to try to deter counterfeiting and improve wearing quality.

iridescence A "glow" displayed by a coin, often gleaming through light pastel colors.

lamination A thin piece of metal that has nearly become detached from the surface of a coin. If this breaks off, an irregular hole or planchet flaw is left.

large cent A large copper U.S. coin, one-hundredth of a dollar, issued from 1793 until 1857, when it was replaced by a much smaller cent made from a copper-nickel alloy. The value of copper in a large cent had risen to more than one cent, requiring the reduction in weight.

legend A phrase that appears on a coin--for instance, UNITED STATES OF AMERICA.

lettered edge A coin edge that displays an inscription or other design elements, rather than being reeded or plain. The lettering can be either incuse (recessed below the surface) or raised. Incuse lettering is applied before a coin is struck; the Mint did this with a device called the Castaing machine. Raised lettering is found on coins struck with segmented collars; the lettering is raised during the minting process, and when the coin is ejected from the dies, the collar "falls" apart, preventing the lettering from being sheared away.

lettering The alphabet characters used in creating legends, mottos, and other inscriptions on a coin, whether on the obverse, reverse, or edge.

Liberty The symbolic figure used in many U.S. coin designs.

Liberty Cap The head of Miss Liberty, with a cap on a pole by her head, used on certain U.S. half cents and large cents.

Liberty Head The design used on most U.S. gold coins from 1838 until 1908. This design was first employed by Christian Gobrecht, with later modifications by Robert Ball Hughes and James Longacre. Morgan dollars and Barber coinage sometimes are referred to as Liberty Head coins.

liner A coin that is on the cusp between two different grades. A 4/5 liner is a coin that is either a high-end MS/PR-64 or a minimum, standard MS/PR-65. See highend and premium quality.

lint mark A repeating depression on a coin, usually thin and curly, caused by a thread that adhered to a die during the coin's production. Lint marks are found primarily on Proofs. After dies are polished, they are wiped with a cloth, and these sometimes leave tiny threads.

loupe A magnifying glass used to examine coins. Loupes are found in varying strengths or "powers."

luster ***(alt. lustre)*** In numismatics, the amount and strength of light reflected from a coin's surface; original mint bloom. Luster is the result of light reflecting on a coin's flow lines, whether those are visible or not.

lustrous A term used to describe coins that still have original mint bloom.

MS (Mint State) Abbreviation. See definition on next page.

major variety A coin that is easily recognized as having a major difference from other coins of the same design, type, date, and mint. See minor variety.

market grading A numerical grade that matches the grade at which a particular coin generally is traded in the marketplace.

marks Imperfections acquired after striking. These range from tiny to large hits and may be caused by other coins or foreign objects.

master die The main die produced from the master hub. Many working hubs are prepared from this single die. See master hub, working die, and working hub.

master hub The original hub created by the portrait lathe. Master dies are created from this hub.

Matte Proof An experimental Proof striking, produced by the U.S. Mint mainly from 1907 to 1916, which has sandblasted or acid-pickled surfaces. These textured surfaces represented a radical departure from brilliant Proofs, having even less reflectivity than business strikes.

medal press A high-pressure coining press acquired by the U.S. Mint, probably in 1858, to strike medals, patterns, restrikes, and regular-issue Proofs.

metal stress lines Radial lines, sometimes visible, that result when the metal flows outward from the center of the planchet during the minting process. See flow lines.

milling machine (alt. upsetting machine) The mechanical device to which planchets are fed to upset their rims. These are sometimes referred to as Type II planchets, and this process makes it easier to strike the higher rims associated with close-collar dies. This results in longer die life.

milling mark A mark that results when the reeded edge of one coin hits the surface of another coin. Such contact may produce just one mark or a group of staccatolike marks. See reeding mark.

minor variety A coin that has a minor difference from other coins Of the same design, type, date, and mint. This minor difference is barely discernible to the unaided eye. The difference between a major variety and a minor variety is a matter of degree. See major variety.

mint A coining facility.

mintage The number of coins struck at a given mint during a particular year.

mint bloom Original luster that is still visible on a coin. See *luster* and lustrous.

mint mark The tiny letter (s) stamped into the dies to denote the mint at which a particular coin was struck.

Mint State The term corresponding to the numerical grades MS-60 through MS-70, used to denote a business- strike coin that never has been in circulation. A Mint State coin can range from one that is covered with marks (MS-60) to a flawless example (MS-70).

mishandled Proof A Proof coin that has been circulated, cleaned, or otherwise reduced to a level of preservation below PR-60. See impaired Proof.

mottled toning Uneven toning, usually characterized by splotchy areas of drab colors.

motto An inscription on a coin-especially IN GOD WE TRUST, which first appeared on the 1864 two-cent piece and now is required on all U.S. coinage.

mutilated A term used to describe a coin that has been damaged to the point where it no longer can be graded.

new A term for a coin that never has been in circulation.

nickel Popular term for a five-cent piece struck in cupro-nickel alloy (actually 75 percent copper, 25 percent nickel).

numerical grading A system for grading coins that uses the Sheldon 1-70 scale; it is employed by PCGS and others.

obverse The front, or heads side, of a coin. Usually the date side.

off center A term for a coin struck on a blank that was not properly centered over the anvil (or lower) die. To be graded by PCGS, a coin can be no more than 5 percent off center.

open collar Its name notwithstanding, a closed collar that surrounded the anvil (or lower) die used in striking early U.S. coins on planchets whose edges already had been lettered or reeded. An open collar was a restraining collar that made it easier to position a planchet atop the lower die, and also sometimes kept the planchet from expanding.

original A term used to describe a coin that never has been dipped or cleaned, or a coin struck from original dies in the year whose date it bears. See *restrike.*

original roll Coins in fixed quantities wrapped in paper and stored at the time of their issuance. The quantities vary by denomination, but typically include 50 one-cent pieces, 40 nickels, 50 dimes, 40 quarters, 20 half dollars and 20 silver dollars. U.S. coins were first shipped to banks in kegs, later in cloth bags, and still later in rolls. Silver and gold coins stored in such rolls often have peripheral toning and untoned centers. Obviously, coins stored in rolls suffered fewer marks than those in kegs or bags.

original toning Color acquired naturally by a coin that never was cleaned or dipped. Original toning ranges from the palest yellow to extremely dark blues, grays, browns, and finally black.

overdate A coin struck with a die on which one date is engraved over a different date. With few exceptions, the die overdated is an unused die from a previous year. Sometimes an effort was made to polish away evidence of the previous date. PCGS will not recognize a coin as an overdate variety unless the overdate is visible.

over-mint mark A coin struck with a die on which one mint mark is engraved over a different mint mark. In rare instances, branch mints returned dies that already had mint marks punched into them; on occasion, these were then sent to different branch mints and the new mint punched its mint mark over the old one. Examples in, crude the 1938-D/S Buffalo nickel and the 1900-0/CC Morgan dollar.

PVC (polyvinyl chloride) Abbreviation. A chemical used in coin flips to make them pliable.

PVC damage A film, usually green, left on a coin after storage in flips that contain PVC. During the early stage, this film may be clear and sticky.

PVC flip Any of the various soft coin flips that contain PVC.

pattern A test striking of a coin produced to demonstrate a proposed design, size, or composition (whether adopted or not). Patterns often are made in metals other than the one proposed; examples of this include aluminum and copper patterns of the silver Trade dollar.

pedigree A listing of a coin's current owner plus all known previous owners.

penny In American numismatics, slang for a one cent coin.

peripheral toning Light, medium, or dark coloring around the edge of a coin.

plain edge A flat, smooth edge seen mainly on small-denomination coinage. See *lettered edge* and *reeded edge.*

planchet (alt. flan) A blank disk of metal before it is struck by a coining press, transforming it into a coin. Type I planchets are flat. Type II planchets have upset rims imparted by a milling machine to facilitate striking in close collars. See blank.

planchet defects Any of the various abnormalities found on coin blanks. These include drift marks, laminations, clips and so forth.

planchet flaw An irregular hole in a coin blank, often the result of a lamination that has broken away.

planchet striations Fine, incuse lines found on some Proof coins, though rarely on business strikes, usually the result of polishing blanks to impart mirrorlike surfaces prior to striking. *See adjustment marks, burnishing lines, die striations and roller marks*.

plated A term used to describe a coin to which a thin layer of metal has been applied--for example, gold plated copper strikings of certain U.S. pattern coins.

plugged A term used to describe a coin that has had a hole filled, often so expertly that it can be discerned only under magnification.

polished coin A coin that has had some type of commercial polish, jeweler's rouge, or similar substance applied to its surface. Polished coins are not graded by PCGS.

polished die A die that has been basined to remove clash marks or other die injury. In a positive sense, Proof dies were basined to impart mirror like surfaces, resulting in coins with reflective fields. See *basining*.

Poor The grade PO- 1. A coin with a readable date and mint mark (if present), but little more, barely identifiable as to type. (One-year type coins do not require a readable date to qualify for this grade.)

portrait lathe (alt. transfer lathe) A mechanical device that reduces a cast (or galvano) of a design to coin size. From this miniature model, a hub and dies are fabricated. See Janvier reducing machine.

premium quality A term applied to coins that are the best examples within a particular grade. See high-end.

presentation striking A coin, often a Proof or an exceptionally sharp business strike, specially struck and given to a dignitary or other person.

press Any of the various coining machines. Examples include the screw press and the steam-powered knuckle press.

pristine A term applied to coins in original, unimpaired condition. These coins typically are graded MS/PR-67 and higher.

Proof (alt. proof) A coin typically struck with specially prepared dies on a specially prepared planchet. Ordinarily, Proofs are given more than one blow with the coining dies and struck with presses operating at slower speeds and higher striking pressure. Because of this extra care, Proofs usually exhibit much sharper detail than circulation-quality coins, or business strikes. PCGS applies the term Proof (PR) to special U.S. coins struck in 1817 and later. Similar coins struck prior to 1817 are recognized by PCGS as specimen strikes (SP).

Proof dies Specially prepared dies, often sandblasted or acid-pickled, that are used to strike Proof coins. Often, the fields are highly polished to a mirrorlike finish, while the recessed areas are left "rough"; on coins struck with such dies, the devices are frosted and contrast with highly reflective fields. Matte, Roman, and Satin Proof dies are not polished to a mirror-like finish.

prooflike A term used to describe the rnirror-like surfaces on business-strike coins that resemble Proofs, particularly Morgan silver dollars. Morgan dollars that meet PCGS standards for prooflike quality are designated PL. See *deep mirror prooflike*.

Proof-only issue A coin struck only in Proof, with no business-strike counterpart.

Proof set A coin set containing Proof issues from a particular year. A few sets contain anomalies such as the 1804 dollar and eagle in 1834 Proof presentation sets.

punch A steel rod with a device, lettering, date, star, or some other symbol on the end which was sunk into a working die by hammering on the opposite end of the rod.

quarter dollar A U.S. coin denomination, one-fourth of a dollar, first struck in 1796, minted sporadically until 1840, then issued virtually without interruption to the present.

quarter eagle A U.S. gold coin denomination, two and-a-half dollars or one-fourth of an eagle, first struck in 1796, issued sporadically thereafter, and discontinued as a regular issue in 1929.

questionable toning Coloring on a coin that may not be original. After a coin is dipped or cleaned, any subsequent toning, whether acquired naturally or induced artificially, will look different from the original toning. PCGS will not grade coins with questionable color.

RB (Red and Brown) A PCGS grading suffix used for copper coins that meet the standards for designation as Red and Brown. See definition below.

RD (Red) A PCGS grading suffix used for copper coins that meet the standards for designation as Red. See definition below.

rare A general term, often overused, to describe the availability of a particular coin. Truly rare coins usually have fewer than 50 to 75 specimens known; however, the term often is used to describe a coin such as the 1856 Flying Eagle cent, of which several thousand examples exist.

rarity scale A system for denoting the relative rarity of different coins. As with the 1-70 numerical grading system, the most commonly used rarity scale is credited to Dr. William H. Sheldon, who adapted it from a nineteenth-century rarity scale. The system uses the terms R-1 through R-8, with an R-8 coin having one to three examples known and an R- I coin having 1,251 or more examples known.

raw Numismatic slang for a coin that has not been encapsulated by a grading service.

real Numismatic slang for genuine coin. See *counterfeit and alteration*.

Red A copper coin that still retains 95 percent or more of its original mint bloom or color. PCGS will award this designation (RD) if a coin has only slight mellowing of color, but not beyond that.

Red Brown A copper coin that has from 5 to 95 percent of its original mint color remaining (RB).

reeded edge (alt. milled edge, **reeding)** The grooved notches on the edges of some coins. These were first imparted by the Mint's edge-lettering machine, later in the minting process by the use of a close collar, which sometimes is described as the third die or collar die.

reeding mark (s) A mark or marks caused when the reeded edge of one coin hits the surface of another coin. The contact may leave just one mark or a series of staccato-like marks. See milling mark.

relief The height of the devices of a particular coin design, expressed in relation to the fields.

replica A copy, or reproduction, of a particular coin.

restrike A coin struck later than indicated by its date, often with different dies. Occasionally, a different reverse design is used, as in the case of restrike 1831 half cents made with the reverse type used from 1840-1857.

retoned A term used to describe a coin that has been dipped or cleaned and then has reacquired color, whether naturally or artificially.

reverse The back, or tails side, of a coin. Usually opposite the date side.

rim The raised area around the edges of the obverse and reverse of a coin. Pronounced rims resulted from the introduction of the close collar, first used in 1828 for Capped Bust dimes.

rim ding Numismatic slang for a mark or indentation on the rim of a coin.

ring test A test used to determine whether a coin was struck or is an electrotype or cast copy. The coin in question is balanced on a finger and gently tapped with a metal object-a pen, another coin, and so on. Struck coins have a high-pitched ring or tone, while electro-types and cast copies have none. This test is not infallible; some struck coins do not ring because of planchet defects such as cracks or gas occlusions.

rolled edge Numismatic slang for rim. This has become part of the vernacular because of the Rolled Edge Indian Head eagle.

roller marks (alt. roller lines) Marks imparted to a planchet by the giant steel rollers used in reducing sheets of coinage metal to the proper thickness. During this procedure, the strips of metal are passed through the rollers several times, if necessary, in order to achieve the desired thickness. With each succeeding pass, the rollers are placed closer together until the proper thickness is attained. Sometimes pieces of metal became imbedded in the rollers, then impart parallel grooves or lines to the metal strips. When planchets are cut from these strips, some display lines called roller marks or roller lines. Most of these disappear during the minting process, but some make it through to the finished coins. Lines of this kind are fairly common on some coins, such as 1902-S Morgan dollars.

roll friction Minor displacement of metal, mainly on the high points, seen on coins stored in rolls. See "coin" *friction* and *bag friction*.

Roman finish An experimental Proof surface used mainly on U.S. gold coins of 1909 and 1910. This is a hybrid surface with more reflectivity than Matte surfaces but less than brilliant Proofs. The surface is slightly scaly, similar to that of Satin Proofs.

rub (alt. rubbing) A numismatic term for slight wear,. often referring to just the high points or the fields. See friction.

Saints (alt. Saint-Gaudens) Slang term for U.S. double eagles struck from 1907-1933, which were designed by the illustrious sculptor Augustus Saint-Gaudens at the behest of President Theodore Roosevelt.

Satin finish Another of the experimental Proof surfaces used on U.S. gold coins after 1907. The dies were treated in some manner to create the silky surfaces imparted to the coins.

satin luster Fine, silky luster seen on many coins, especially copper and nickel issues. Almost no "cartwheel" effect is seen on coins with this type of luster.

scratch A detracting line that is more severe than a hairline. The size of the coin determines the point at which a line ceases to be viewed as a hairline and instead is regarded as a scratch; the larger the coin, the greater the tolerance. A heavy scratch may result in a coin not being graded.

screw press The first type of coining press used at the U.S. Mint. Invented by Italian craftsman Donato Bramante, this press had a fixed anvil (or lower) die, with the hammer (or upper) die being attached to a rod with screw-like threads. When weighted arms attached to the rod were rotated, the screw mechanism quickly moved the rod with the die downward, striking the planchet placed into the lower die. The struck coin then was ejected and the process was repeated.

sea-water damage Damage incurred by a coin through exposure to sea water, as with coins found in sunken treasure ships. The copper in gold and silver alloys reacts with seawater and, over time, is "eaten" away. The surfaces of these coins appear dull and lightly pitted. PCGS will not grade such coins unless the damage is extremely minor.

second toning Any toning, natural or artificial, that results after a coin is dipped or cleaned. This second toning is seldom as attractive as original toning, though some coins "take" second toning better than others.

semi-prooflike A term used to describe a coin that has some mirror-like surface mixed with satin or frosty luster. Reflectivity is obscured on such a specimen, unlike the reflectivity on prooflike and deep mirror prooflike coins.

Sheldon scale The l-to-70 grading system devised for U.S. large cents by Dr. William H. Sheldon and adopted for all coins by the coin industry. PCGS based its grading system on this scale with certain refinements, such as the inclusion of intermediate grades (61, 62, 64, and so on).

"shiny" spots Areas on Matte, Roman, and Satin Proofs where the surface has been disturbed. On brilliant Proofs, dull spots appear where there are disturbances; on textured-surface coins such as Matte, Roman, and Satin Proofs, these disturbances create "shiny" spots.

silver-clad The composition of Kennedy half dollars struck from 1965 to 1970, whose overall content is 40 percent silver and 60 percent copper. These are commonly referred to as silver-clad halves because two outer layers containing primarily silver are bonded to a core made primarily of copper.

silver dollar The silver coin that served as a comer- stone of U.S. currency from 1792 until 1964. (The gold eagle played a similar role from 1792 until 1971.) First struck in 1794 and issued sporadically thereafter until 1935. Defined as so many grains of silver, changing with market conditions. See *dollar and Trade dollar*.

slab Numismatic slang for the holder in which a coin is encapsulated by a grading service. The coin contained therein is said to be *slabbed.*

slide marks (alt. album slide marks) Detracting lines imparted to a coin when the plastic slides in a coin album rub across the high points of the coin as the slides are inserted or removed.

slider A term used to describe an AU coin that looks, or can be sold as, Uncirculated. Occasionally used in reference to another grade; a slider EF coin, for example, would be a VF/EF coin that is nearly EF.

Spark-erosion strike A coin made from spark-erosion dies. These are characterized by the telltale "pimples" noted mainly on the areas in relief.

split grade A situation where separate and different grades are assigned to the obverse and reverse of the same coin. A coin graded MS-63/65, for example, has an MS-63 obverse and an MS-65 reverse. Since reverses often have higher grades than obverses, split grading has been employed to infer a superior coin. The obverse grade is more important than the reverse; thus, an MS-63/65 coin would be graded MS-63 by PCGS.

splotchy toning Color that is uneven, both in shade and in composition.

standard silver The official composition of U.S. silver coinage, set by the Mint Act of 1792 at approximately 89 percent silver and 11 percent copper, later changed to 90 percent silver and 10 percent copper--the composition seen in most U.S. silver coins.

star A device used as a motif on many U.S. coins. On the earliest U.S. coins, thirteen stars were depicted, representing the thirteen original colonies/states. As new states were admitted into the Union, more stars were added; up to sixteen appeared on some coins. Adding stars for each state was impractical, however, so the number was reduced to the original thirteen. Exceptions include the forty-six stars, later forty-eight stars, around the periphery of Saint-Gaudens double eagles, reflecting the number of states in the Union at the time those coins were issued.

steam-powered press A coining press driven by a steam-powered engine. This type of press, more powerful than its predecessors, was installed in the United States Mint in 1836, replacing the hand and horse-powered screw presses.

stella A term applied to the experimental four-dollar gold coins struck by the U.S. Mint in 1879-1880. So named for the large star on the coins' reverse.

stock edge A counterfeit edge collar used for various dated fakes. These have the same repeating characteristics.

striations (alt. striae) Incuse polish lines on a die which result in raised lines on coins struck with that die. These are usually fine, parallel lines, though on some coins they are swirling and still others have crisscross lines. Planchet striations are burnishing lines not struck away by the minting process.

strike n. The completeness, or incompleteness, of a coin's detail. **v.** The act of minting a coin.

striking n. A coin. **adj.** The process of minting.

strip The flat metal, rolled to proper thickness, from which planchets are cut.

struck A term used to describe a coin produced from dies and a coining press.

struck copy A replica of a particular coin made from dies not necessarily meant to deceive.

struck counterfeit A fake coin produced from false dies.

surface preservation The condition of the surface of a coin. On weakly struck coins, this is a better indicator of grade than is the coins detail.

surfaces The entire obverse and reverse of a coin, though often used to mean just the field areas.

sweating A procedure in which coins are placed in a bag and shaken vigorously to knock off small pieces of metal. Later these bits of metal are gathered and sold, producing a profit as the coins are returned to circulation at face value. Mainly employed with gold coins, leaving their surfaces peppered with tiny nicks.

tab toning Toning often seen on commemorative coins which were sold in cardboard holders with a round tab. Coins toned in these holders have a circle in the center and are said to have tab toning.

target toning (alt. bullet toning) Toning that resembles an archery target, with deeper colors on the periphery often fading to white or cream color at the center.

technical grading A method of grading that takes into account only the surfaces of a coin, ignoring luster, strike, and eye appeal. The surfaces of a particular coin may be nearly mark-free, so its technical grade may be MS-65 or higher. But if it is poorly struck and has flat luster, its market grade--the grade assigned by PCGS--may be only MS-63.

three-cent piece (alt. trime) A U.S. coin denomination used for two separate nineteenth-century series--one struck in copper-nickel, the other in silver. The silver three-cent piece was first struck in 1851, in debased silver, to facilitate the purchase of postage stamps; the composition was changed to standard silver in 1854, and production continued until 1873. The nickel three-cent piece was introduced in 1865 and was issued until 1889.

Three-dollar piece A U.S. gold coin struck from 1854 until 1889.

thumbed A term used to describe a coin that has been doctored in a specific way to cover marks, hairlines or other disturbances. Often associated with silver dollars, it actually is used on many issues, mainly business strikes. The thumb is rubbed lightly over the disturbances, and the oils in the skin help to disguise any problems.

tissue toning Color, often vibrant, acquired by coins stored in original Mint paper. Originally, this was fairly heavy paper; later, very delicate tissue. Sometime during the nineteenth century, the Mint began wrapping Proof coins, and occasionally business strikes, in this paper. The paper contained sulfur; as a result, the coins stored in it for long periods of time-acquired blues, reds, yellows and other attractive colors.

toning (alt. patina) The color seen on many coins. Infinite numbers of shades, hues, and pattern variations are seen, depending upon how, where, and how long a coin has been stored. Every coin begins to tone the second it leaves the dies, as all U.S. coins contain reactive metals in varying degrees.

tooling mark A line, usually small and fine, found on both genuine and counterfeit coins. On genuine coins, such lines result when Mint workmen touch up dies to remove remnants of an overdate or

other unwanted area. On counterfeits, they often appear in areas where the die was flawed and the counterfeiter has attempted to "fix" the problem.

Trade dollar A U.S. silver coin, slightly heavier than the regular silver dollar, issued from 1873 until 1885 and intended specifically to facilitate trade in the Far East- hence its name. Trade dollars were made with a marginally higher silver content than standard silver dollars in an effort to gain acceptance for them in commerce.

transfer die A die created by sacrificing a coin for a model.

transitional issue A coin struck after a series ends, such as the 1866 No Motto issues. A coin struck before a series starts, such as the 1865 Motto issues. A coin struck with either the obverse or the reverse of a discontinued series, an example being the 1860 half dime With Stars. A coin struck with the obverse or reverse of a yet-to-be-issued series, an example being the 1859 Stars half dime with the Legend-type reverse.

two-cent piece A U.S. coin struck from 1864 until 1873, in part to facilitate the purchase of postage stamps, whose price had been reduced to two cents each.

type A variation in design, size, or metallic content of a specific coin design. Examples include the Small and Heraldic Eagle types of Draped Bust coinage, Large-Size and Small-Size Capped Bust quarters, and the 1943 Lin- coin-cent struck in zinc-coated steel.

Uncirculated A term used to describe a coin that has never been in circulation, a coin without wear. See *brilliant Uncirculated, Mint State, and new*.

used A term used to describe a coin that has light to heavy wear or circulation. See circulated, friction, and rub.

variety A coin of the same date and basic design as another but with a discernible difference. PCGS recognizes all major varieties; there are thousands of minor varieties, most of which have significance only to specialists in the particular series.

Very Fine The term corresponding to the grades VF-20, 25, 30, and 35. This has the broadest range of any circulated grade, with nearly full detail on some VF-35 coins and less than half on some VF-20 specimens.

Very Good The term corresponding to the grades VG-8 and VG-10. In these grades, between Good and Fine, a coin shows slightly more detail than in Good, usually with full rims except on certain series such as Buffalo nickels.

weak strike A term used to describe a coin that does not show intended detail because of improper striking pressure or improperly aligned dies.

whizzing The process of mechanically moving the metal of a lightly circulated coin to simulate luster. Usually accomplished by using a wire brush attachment on a high-speed drill.

wire edge (alt. knife edge, wire rim) The thin, knife-like projection seen on some rims created when metal flows between the collar and the die.

wonder coin Numismatic slang for a coin whose condition is particularly superb.

working die A die prepared from a working hub and used to strike coins.

working hub A hub created from a master die and used to create the many working dies required for coinage.

worn die (alt. eroded die) A die that has lost detail because of extended use. In earlier periods of U.S. coinage history, dies were often used until they wore out, became excessively cracked, or broke apart. Coins struck from worn dies often appear to be weakly struck, but no amount of striking pressure will produce detail on a coin that does not exist on the die.

XF (Extremely or Extra Fine) An abbreviation sometimes used interchangeably with EF. See definitions for Extra Fine and Extremely Fine.

Glossary Source: Sahara Coins LLC [61]

Appendix C2

Standard Grades of Collectibles

Standard Grades & Condition of Coins:

- **Poor, filler or cull (P1)** - barely recognizable, may contain holes
- **Fair (F2)** - very heavily worn; major portions may be completely smooth
- **About Good (AG)** - heavily worn; date may be barely discernable
- **Good (G)** - Coin will be heavily worn, but the main design and legend will be visible. Lettering may be worn smooth. May be dull or faded areas.
- **Very Good (VG)** - Still well worn but more of the rim will be evident. Design and legend will be clear but worn flat. Lacks specific details.
- **Fine (F)** - Medium to heavy wear but even overall. The design becomes clearer and details begin to appear. Some letters within the design will be apparent.
- **Very Fine (VF)** - A visibly nicer coin. High spots will show light, even wear. Various major features are visible. Lettering is all readable.
- **Extra Fine (XF)** - Slight wear will show on the highest points of the main devices. Words are sharp and easily readable. All details are clearly defined.
- **AU 50** - Slight traces of wear on the highest points of the coin; may be dull with some evidence of luster under any toning.
- **AU 53** - Just slightly better than an AU 50 with a little more luster visible. Eye appeal begins to make a difference between the AU grades.
- **AU 55** - An obviously nicer coin than an AU 50 with no major difficulties. More luster shines through the surfaces.
- **AU 58** - This is oftentimes called a slider as it will appear to many observers to be uncirculated. Just the faintest wear on the highest points of the coin. Luster should be quite evident, although some toning can be apparent. Usually coins with poor eye appeal will not make the AU 58 grade.
- **MS 60** - Mint State indicates a coin that has no wear and is uncirculated. It may have numerous bagmarks and/or be toned. MS 60 is the lowest quality of an uncirculated coin.
- **MS 61** - An uncirculated coin that is just slightly better than MS 60. However, no question that it is uncirculated. Whereas, some may debate over the merits of a coin being MS60 because of the excessive bagmarks, the MS61 should be more desirable.
- **MS 62** - This coin should be a much cleaner specimen than an MS 60, yet, just slightly better than an MS 61. There should be fewer bagmarks as the coin takes on more attractive features.
- **MS 63** - This is the grade that many collectors feel is the most collectible in numismatics. Prices are typically reasonable compared to higher grades and the coin should have at least an average strike and eye appeal, with minimal distracting marks.
- **MS 64** - This is the grade where prices in many series begin to increase dramatically. For this reason the coin will begin to show fewer marks and the strike will be the

strongest yet. No primary distractions that will draw your eye. A near-gem coin with just a few tiny marks or weakness in strike to keep it from a higher grade.

- **MS 65** - This is the gem category. Coin should be fully struck with eye appeal. Either brilliant or toned but there should not be any unsightly marks or color that negates eye appeal. Any marks should be very minor in appearance. Prices spread out even further.
- **MS 66** - A coin that just jumps out at you as being nicer than an MS 65. The main devices on either side should have no more than very minor ticks and the fields should be cleaner than that of an MS 65.
- **MS 67** - A superior coin that has no major distractions to speak of. The fields should be near flawless with just the slightest contact on the main device. This coin should emit a look of satisfaction from the viewer. Prices increase further especially for coins with short supplies and strong demand.
- **MS 68** - A difficult grade to determine by most experts. When does a coin become MS 68 but is not quite MS69 or 70? A very superior coin with maybe just a minor tick on either side keeping it from perfection.
- **MS 69** - This is a coin that should create a gasp when viewed. There should be no imperfections to the naked eye. With a magnifying glass a minor mark or impediment may be visible.
- **MS 70** - A perfect coin with no imperfections seen with a magnifying glass. There should be no marks whatsoever; the coin must look like it just left the Mint. Very unusual in early coins as the mint did not have the quality they do today. Modern coins have been given this exalted grade although there is debate whether coins can be perfect.

Source: Numismatic Interactive Network LLC [73]

Standard Grades and Condition of Stamps:

Stamps fall into two major categories: Mint (unused) and Used (cancelled). The final grade of a mint stamp has three principal components: centering, soundness and eye appeal. Mint stamps are additionally qualified by the condition of the gum, and whether or not the stamp has been hinged. The final grade of a used stamp is likewise determined by its centering, soundness and eye appeal (including an adjustment for the cancellation.) Gum condition obviously does not apply to used stamps.

- **Superb:** Perfect in all respects. The finest quality. A rare grade.

- **Extremely Fine or Extra-Fine:** Close to perfect. Design is well-centered. Margins are even all around. Designs of even the earliest issues are well clear of the perfs on all sides. Imperforates have even margins that are wider than usual for that particular issue. Cancels are light and neat. Condition: Rich, bright color. Clean. Perfs intact with no faults. Many early stamps are never seen in this condition.

- **Very Fine:** Design is balanced and well-centered. There are ample margins, though not necessarily perfectly even. Imperforates have three normal sized margins. Cancels are light and neat. Mint have OG. Condition: Rich, bright color. Clean. Perfs intact with no faults. Grade used for most catalog values.

- **Fine/Very Fine:** Design is "slightly" off-center, or may be off-center either horizontally or vertically but not both. Design is well clear of the perfs. Imperforates have two normal size margins, and design does not touch the edge. Cancels do not detract from the design of the stamp. Mint have LH or HH, depending on the age of the issue with no faults.

- **Fine:** Design is "noticeably" off-center both horizontally and vertically. The design barely misses the perfs, but they do not cut into the design. Early issues have perfs or separations that may cut into the design. Imperforates have thin margins. Cancels may be heavier than usual, perhaps even obscuring the design. Mint have LH or HH, depending on the age of the issue with no faults.

- **Good or Average:** Design is off-center and perfs may cut into the design. Cancel is heavy and obscures the stamp's design. No tears or thin spots. Lowest collectible grade.

- **Poor:** Design is off center and the perforations cut far into the design. Cancellation is thick and heavy, smeared, blurred. Generally not suitable for a collection.

Sources: Professional Stamp Experts - Collectors Universe [71] and The Glassine Surfer Stamp Collecting [49]

Standard Grades and Condition of Baseball Cards:

- **10.0 GEM MINT -** A perfect card with four perfectly sharp corners, sharp focus and full original gloss. No staining of any kind, only slight print spots allowed under magnification. Centering must be within 60/40 or better on front and 75/25 on the back.

- **9.5 MINT+ -** A card that looks perfect to the naked eye with sharp corners, sharp focus and full original gloss but may show subtle wear under magnifications. Centering must be within 60/40 or better on front and 75/25 on the back.

- **9.0 MINT -** A near perfect card with only the slightest wear visible, corners must still be sharp with no surface wear. Original color and gloss with only minor print imperfections. Centering must be within 60/40 to 65/35 or better on front and 80/20 or better on back.

- **8.0 NM-MT -** A card with a slight touch of wear, barely noticeable, on one or two corners. May show minor color and print imperfections, slight wax staining allowed on back. Centering must be within 65/35 to 70/30 or better on front and 90/10 or better on back.

- **7.0 NM -** A card upon close inspection shows one fuzzy corner or two or more corners with slight touching. May have minor print spots, slightly rough edges or minor surface wear. Wax staining allowed on back. Centering must be 70/30 to 75/25 or better on front and 90/10 or better on back.

- **6.0 EX-MT -** A card that shows a few minor flaws, some corner wear visible with slight surface and edge wear. printing imperfections may be found but must still have overall appeal with nice color and gloss. Centering must be 80/20 or better on front.

- **5.0 EX -** A card that shows several minor flaws with noticeable corner wear. Will show some surface wear and printing imperfections. May show slight loss of color and gloss. Centering must be 85/15 or better on front.

- **4.0 VG-EX -** A card with frayed, layered or slightly rounded corners. A small crease may be detected upon close inspection with surface wear and edge wear noticeable, but modest. Centering must be 85/15 or better on front.

- **3.0 VG -** A card that shows noticeable wear on corners, edges and surface. May have a crease or surface scratching but nothing extreme. Will show loss of color and gloss. Centering must be 90/10 or better on front and back.

- **2.0 GOOD -** A card with considerable wear. Will have creases, surface scuffing and overall noticeable flaws.
- **1.0 PR-FR -** A card with severe wear. Will have badly rounded corners, creasing and show noticeable abuse.

Source: Advanced Grading Specialists [3]

Standard Grades and Condition of Comic Books:

- **MINT (MT) CONDITION:** Near perfect in every way. Only the most subtle bindery or printing defects are allowed. Cover is flat with no surface wear. Cover inks are bright with high reflectivity and minimal fading. Corners are cut square and sharp. Staples are generally centered and clean with no rust. Cover is generally centered and firmly secured to interior pages. Paper is supple and fresh. Spine it tight and flat.

- **NEAR MINT (NM) CONDITION:** A nearly perfect copy with only minor imperfection allowed.

- **FINE (FN) CONDITION:** An exceptional, above average copy that show minor wear but still is relatively flat, clean and glossy with no subscription crease or brown margins. Typical defects include: light spine wear, minor surface wear, a light crease, minor yellowing/tanning to interior pages. Compared to a VF, cover inks are beginning to show a significant reduction in reflectivity, but the comic remains highly collectible and desirable.

- **VERY GOOD (VG) CONDITION:** Generally, this condition represents the average used comic book, one that has not been taken care of. These comics show moderate wear but eye appeal has not been reduced to the point that the comic ceases to be collectible. One or two minor markings on the cover or minor spine roll are allowed. These comics may be lightly creased along the spine or other edges, or have a minor piece or pieces missing.

- **GOOD (GD) CONDITION:** Comics in this condition have all pages and covers, although there may be small rips or tears. These still very readable copies are commonly creased, scuffed, abraded, and soiled. Paper quality is low, but not brittle.

- **FAIR (FR) CONDITION:** Comics in this condition are very heavily read and soiled, but still complete. Damaged beyond collectible for many collectors but not all, fair comics typically bring 30 to 50 percent of the good price, but can command more where better condition copies are not available at affordable prices.

- **POOR (PR) CONDITION:** Comics in this condition have an aggregate of defects so extensive as to render them all but uncollectible in most cases. Poor comics are often severely stained, abraded, deface, or otherwise damaged so as to be almost unreadable, but even a poor copy can be sought by readers an even collectors when better condition copies are not available at affordable prices. Lone Star Comics has in the past described poor copies as "looking like a truck ran over them," and this is a description that is often apt.

Source: Gottawiz.com [32]

Appendix C3

Third Party Grading Companies

June 12, 2002 marked the thirty-year anniversary that ANACS began accepting coins for authentication, certifying them as genuine, and producing photo certificates of authenticity. Grading would not begin until 1978. In the years since then many other companies have come and gone. Here is the most recent list of companies. Approximately 22 of these companies are currently active.

- A-Mark has been around since at least the 1960's but their slabs only date from 1983 to 1988.
- ACCUGRADE also known as ACG or ASA-ACCUGRADE. 1984 - Date
- ACGS (American Coin Grading Service) 2002
- ACGS (Australian Coin Grading Service) 1998? - Date
- AGA (American Grading Association) 1986? - 1987
- AGS (Advanced Grading Specialists) 1997? - 1998
- AGS (American Grading Service.) Unknown
- ANAAB (American Numismatic Association Authentication Bureau) 1990 - Date
- ANACS (American Numismatic Association Certification Service) 1972 - Date
- ANICS (American Numismatic Institute Certification Service) 1987
- ARC (American Rare Coin and Collectors Inc) 1986
- ASA (Accugrade Sports Authentication.) 1990 (As a coin grading firm.)
- BLANCHARD An "in house" slabbing service by Blanchard Co. 1986 - 1987?
- CGA (Currency Grading Association) 2000 - Date
- CGC (Currency Grading Corporation). 2001 - Date
- CGCGS (Certi-Graded Coin Grading Service) Unknown

- COMPUGRADE New Orleans based research and development firm begun by Jim Diffenthal in late 1990 that intended to slab computer graded coins. Late 1991.
- CTGS (Coin and Token Grading Service) 1993 - Date
- DCGS (Digital Coin Grading Service) June 2001 - Date
- DCGS (Distinctive Coin Grading Service) Now NSCGS unknown, may never have issued slabs
- FCS (Federal Certification Service) 1986 - 1989
- FTGS (First Token Grading Service) 2000 - 2001
- GCS (Global Certification Service) Nov 2001 to Date?
- GLGS
- GSA (General Services Administration) 1972 - 1980
- HALLMARK Grading service begun by Bowers and Merena in 1987. Closed 1991
- HOLT (T J Holt company)
- ICCS (International Coin Certification Service) ? - Date
- ICG (Independent Coin Grading) Late 1998 - Date
- ICGS (International Coin Grading Service) 1986? - 1988
- ICI (International Coin Investments) Late 80's is all that is known.
- IGA (Independant Grading Association) 1986 - 1987
- IGS (Independent Grading Service) Unknown
- INGS (International Numismatic Grading Service) 1987- ?
- INL (International Numismatic Laboratories) 1986
- INS (International Numismatic Society Authentication Bureau) 1975 - 1992?
- IRI (Investment Rarities Incorporated) 1986? - Date?
- LCG (Laser Coin Grading) 2001?
- MCGC (Modern Coin Grading Company) 2001 - ?
- MonExpert (MonExpert Grading Service) May? 2002 - ?

- MTB (Mantra, Tordella & Brookes Inc) 1986 - ?
- NCCA (National Coin Certification Association?) 1987?
- NCG (Numismatic Coin Grading Service) Mid 2000 - Early 2001
- NCG (Company name not known) Existed briefly during November of 2000.
- NCI (Numismatic Certification Institute) 1984 - 1988?
- NECA (Numismatic Error Collectors of America) mid 1960's
- NES (Numismatic Evaluation Service) was a photo-certificate company.
- NGC (Numismatic Guaranty Corporation of America) 1987 - Date
- NNCS (National Numismatic Certification Service) 1987?
- NSCGS (New Standard Coin Grading Service) 2002
- NTC (NumisTrust Corporation) May 2001 - Date
- Numex 2002
- Paramount (Redfield Dollars) 1976 - 1978?
- PCGS (Professional Coin Grading Service) 1986 - Date
- PCI (Photo-Certified Coin Institute) 1986 - Date
- PCS (Preferred Customer Service) 2001?
- PHOTO-SEAL
- PNGL (Professional Numismatic Grading Laboratories) Mid to late 80's
- PPGS (Peoples Professional Grading Service) 2002 - Date
- SEGS (Sovereign Entities Grading Service) 1998 - Date
- SILVERTOWNE Certificate issued by Leon Hendrickson's company in Winchester IN. 1986 - 1988
- TCGS (Twenty-first Century Grading Service) Early 2002 -
- TCTS (Tom's Coins Technical Services) 1999 - Nov 2000
- Tulving (Hannes Tulving)
- USGA (U S Grading & Authentication) Early 90's?
- USGCO (United States Grading Company) Unknown

- USGCS (U. S. Grading and Certification Service) 1986? - 1989
- USRCCT (U S Rare Coin Certification & Trading Co Inc.) late 80's
- USTI (United States Tangible Investment) Mid to late 90's?
- WCG (World Class Grading)
- WCGS (World Coin Grading Service) 2002

Source: Michael Schmidt - Collectors Universe Message Board [63]

Appendix C4

Lincoln Cent Features

Features which should be examined when determining the grade of a Lincoln Cent

Obverse

1. Lettering – top "In God We Trust" aka Motto
2. Lettering – left "Liberty"
3. Lincoln's Outline
4. Date
5. Mintmark
6. Coat - Folds and detail in upper part of coat
7. Coat - Folds and detail in lower part of coat into rim
8. Facial - Eye
9. Facial - Hair
10. Facial – Cheek
11. Facial – Forehead
12. Facial - Jaw
13. Facial – Ear
14. Facial - Ear Lobe
15. Facial – Mouth
16. Facial - Nose

Reverse (Wheat Ear Type)

17. Wheat stalks
18. Lettering – top "E Pluribus Unum"
19. Lettering – mid center "ONE CENT"
20. Lettering – bottom center "United States of America"

Appendix C5

Mintmark Locations

Listing of each type of U.S Coins and where to find their Mint Marks

U.S. Mints & The Mint Marks

- (P) – Philadelphia
- (D) – Denver
- (S) - San Francisco
- (W) - West Point
- (C) for Charlotte, North Carolina (on gold coins only).
- (CC) for Carson City, Nevada.
- (D) Dahlonega, Georgia-used only for gold coins minted from 1838-1861.
- (0) for New Orleans, Louisiana...
- (D) Showing the coin was minted at Dahlonega, Georgia as the Denver Mint was not yet in operation (On some gold coins minted from 1838 to 1861, you may find the letter D)

Specific Series Mint Information:

- Half Cents-All Half Cents were minted at the Philadelphia Mint, thus they had no mint mark

- Large Cents-All Large Cents were minted at the Philadelphia Mint, thus no mint mark.

- Flying Eagle Cents-All Flying Eagle Cents were minted at the Philadelphia Mint, thus no mint mark.

- Indian Cents-on only Two years, 1908 & 1909, under the wreath on the back of the coin.

- Lincoln Cents-on the front, under the date.

- Two Cent Piece-All Two Cent Pieces were minted at the Philadelphia Mint, thus no mint mark.

- Three Cent Pieces-All Three Cent Pieces were minted at the Philadelphia Mint (no mint mark), except 1851 which were minted in New Orleans (O) mint mark on the back to the right of the Roman Number III).

- Shield Nickels-All of the Shield Nickels were minted at the Philadelphia Mint, thus no mint mark.

- Liberty Nickels-All Liberty Nickels were minted at the Philadelphia Mint (no mint mark), except the 1912 that may have a D or S on the back to the left of the word CENTS.

- Buffalo Nickels-found on the back under the words FIVE CENTS.

- Jefferson Nickels-1938-1964 on he back to the right of the building..
  1968 to present day on the front near the date. 1942-1945 (The War Years) Mint Mark above the Dome of the Monticello Building.

- Half Dimes-on the back either in or below the wreath.

- Bust Dimes-All bust Dimes were minted at the Philadelphia Mint, thus no mint mark.

- Seated Dimes-on the back either in or just below the wreath.

- Barber Dimes-on back centered on the bottom near rim.

- Mercury Dimes-on back bottom left of the fasces.

- Roosevelt Dimes-1946-1964 on the back, the bottom left of the torch. 1968 and up on front above the date.

- Twenty Cent Pieces-on back under the Eagle.

- Bust Quarters-All Bust Quarters were minted at the Philadelphia Mint, thus no mint mark.

- Seated Quarters-on back under the Eagle above Quarter Dollar.

- Barber Quarters-on back under the Eagle above Quarter Dollar.

- Standing Liberty Quarters-on front small mint mark above date just to the left.

- Washington Quarters-1946-1964, on the back, centered under the Eagle..
  1968 and up, on front to the right of the hair ribbon.

- Bust Half Dollars-on the front above the date.

- Seated Half Dollars-on the back just below the Eagle (above HALF DOLLAR).

- Barber Half Dollars-on the back just below the Eagle (above HALF DOLLAR).

- Standing Liberty Half Dollars-1916 & some 1917 on the front just below TRUST. 1917-1947, on back, lower left just below branch.
- Franklin Half Dollars-on back, centered above the Liberty Bell beam.
- Kennedy Half Dollars-1964 on back to the left of the olive branch near claw. 1968 and up, on front centered above date near neck.
- Bust Dollars-on the back under the Eagle.
- Liberty Seated Dollars-on the back under the Eagle.
- Morgan Dollars-on the back under the Eagle.
- Peace Dollars-on back , left side at tip of Eagles Wing.
- Trade Dollars-on back under the Eagle.

Source: Disk Works of South Jersey [66]

Appendix C6

Historical Overview of Coin Collecting and United States Coinage

Early Collections

Coin collecting dates back to approximately the late seventh century BC. The classical authors Pliny and Plutarch referred to famous art collections that included special coins renowned for their artistic qualities and signed by well-known artists. In the fourteenth century, ancient coins received serious attention from scholars and collectors alike. Francesco Petrarca (1304-1374) of Florence is reported to have traveled frequently to Rome to buy ancient coins depicting Roman emperors.

Enthusiasm for coin collecting increased between the fourteenth and sixteenth centuries as it was during this period that the foundations of some of the most famous collections were established. An early renowned collector of ancient coins was the French King Louis XIV (1638-1715 [27] Louis, who regarded himself as a patron of the arts, appointed several advisors to acquire entire collections of ancient cameos, engraved gems, and coins on his behalf.

The Motivation of Collectors

Author Ray Wyman, Jr. [78] describes coin collectors as having one of the three basic personalities: the incidental collector, the investor collector, and the professional collector.

- The incidental collector collects, or hoards, coins and spare change in the dresser drawer, coffee cans, bottles or some other large container. Ultimately the incidental collector is faced with a choice of either getting serious about their accumulation or cashing the coins at the local bank or supermarket.

- The investor usually graduates from the ranks of the incidental collector and sometimes comes about because of advice received by his/her financial planner to diversify their assets into rare collectibles.

- The professional becomes interested and fascinated by rare coins and the potential for profit making.

Editor Jim Davis of the Elgin Coin Club writes, “For as long as people have been collecting coins one of the strongest motivations is to get the best example of a certain coin they can afford” [22]. Collectors will often purchase lower quality and low cost coins to ‘fill holes’ in their collections when they are starting out with the ultimate hope of upgrading their collections when better specimens as they become available. As collectors gain more knowledge in the area of grades in their domain they seek out better quality items and become more discriminate in their grading requirements.

Early American Mintage:

The first coins struck in the United States were silver NE threepence, sixpence, and shilling pieces made in Massachusetts in 1652 as Massachusetts challenged England's ban on colonial coinage. The colony struck a series of silver coins, including the Pine Tree Shilling. [75] Even though they were produced for many years all Pine Tree shillings were dated 1652 in the event that England ever found out about this illegal coinage, Massachusetts could make the claim that it had not produced any coins since 1652. From 1652 until the 1790s, numerous individuals, state governments, and merchants issued coins. Vermont, Connecticut, New Jersey, and Massachusetts each issued copper coins in the 1780's, which were struck from hand-engraved dies. [66]

In 1787 the Constitution gave Congress exclusive power to coin money and in 1792 Congress passed its first coinage act authorizing the United States Mint on April 2, 1792. As Philadelphia was then the nation's capital, the first mint was erected there at 7th and Arch Streets and by March 1793, it delivered its first circulating coins of 11,178 copper cents. [50]

Modern US Coinage and Mints:

From 1792 to present day the U.S. Mint has expanded and contracted operations as needed by establishing regional branch mints near where supplies of precious medals sources were discovered and as demand has called for.

The Philadelphia Mint began striking coins for circulation in 1792. It was entrusted with making the nation's circulating legal tender coinage in both precious metals, copper, and nickel.

The U.S. Mint then opened a branch mint in New Orleans, Louisiana in 1838 and struck official coins for the United States, the Southern Confederacy, and Mexico. The location of this branch corresponded to the influx of foreign gold and silver into the port of New Orleans. [7]

After 1838, branches of the United States Mint were opened in Charlotte, North Carolina and Dahlonega, Georgia as gold was discovered in the Southern States of Georgia, North Carolina, and Alabama.

In 1854, The San Francisco Mint was established in response to the gold discovery in California at Sutter's Mill. [76]

A massive silver find in Nevada at the Comstock Lode paved the way for the creation of the Carson City Mint in 1870. This mint struck gold pieces and Morgan Silver Dollars. The coins produced from this are highly sought after by modern day collectors.

In 1906, The Denver Mint was established to initially strike gold found in the Colorado Rockies into official coins for general circulation. [7] The Denver facility currently strikes billions of the circulating U.S. coins found in modern circulation.

In the 1984, the U.S. Mint opened the most recent mint at West Point, N.Y. to strike American Eagle Silver Dollars, Gold Bullion Coins, and other modern commemoratives. The

coins struck at this facility are produced for collectors and not intended for general circulation.

Timeframe	Mint	Mintmark
1793 to date	Philadelphia	No mint mark until recent years when 'P' was used on some coins.
1838 – 1861	Charlotte	C
1838 – 1861	Dahlonega	D
1838 – 1909	New Orleans	O
1854 to date	San Francisco	S
1870 – 1893	Carson City	CC
1906 to date	Denver	D
1984 to date	West Point	W

Table C6.1 – U.S. Modern Day Mints

The number of branch mints has contributed to the diversity in varieties that exist within series of coins that collectors seek to accumulate. Within a year of a single series of coins there may have been specimens produced at as many as 5 of the branch mints. The Morgan Silver Dollar is a good example of this situation as there were a number of years when these dollar coins were produced at Philadelphia, New Orleans, San Francisco, Carson City and Denver. In these years collectors would need to accumulate a specimen bearing the date and the mintmark from each appropriate branch mint in order to have collected the entire series for the year.

There have been no concrete rules on the coinage of series as well. The Treasury Department makes decisions on which mint will strike which coins on an annual basis based on the money supply, the demand for new coins and the cost of producing them. For instance, the Lincoln Cent series has run from 1909 to the present day. During many years of its' run, coins were generally produced in the Philadelphia, San Francisco and Denver mints. However, in certain years the mint did not produce Lincoln Cents at certain branch mints. Limited examples of these gaps in production include: No cent coinage in the San Francisco in 1922, 1956 thru 1967 and no cent coinage in Denver from 1965 – 1967.

Still the Mint may decide to produce a coin at a mint for a limited period to commemorate a special occasion, such as an anniversary, or to test the feasibility of producing new coinage at a branch mint. Modern day examples of this include the coinage of the 1996-W Roosevelt Dime and the 1995-W American Eagle Silver Dollar. Both of these coins were produced exclusively for collectors and not intended to circulate, although they could technically circulate, as they are legal currency. The challenge to the collectors of each of these series is that they have an additional coin to collect with an unusual mintmark.

Appendix D

Sample Coins Used in all Experiments

Coin ID	Image	Yearmm	Machine Expert Grade	Hi-Low Grade	Test 1	Test 2	Test 3
01	COIN11	1919D	8-VG	4-G	MG	HL	NG
02	COIN03	1911	6-G	12-F	MG	HL	NG
03	COIN19	1941	12-F	8-VG	MG	HL	NG
04	COIN21	1946	53-AU	60-MS	MG	HL	NG
05	COIN60	1951D	15-F	10-VG	MG	HL	NG
06	COIN61	1952	45-EF	55-AU	MG	HL	NG
07	COIN39	1944	53-AU	45-EF	NG	MG	HL
08	COIN46	1946S	60-MS	63-MS	NG	MG	HL
09	COIN32	1954	12-F	6-G	NG	MG	HL
10	COIN82	1968D	63-MS	65-MS	NG	MG	HL
11	COIN31	1953S	6-G	10-VG	NG	MG	HL
12	COIN23	1947S	12-F	20-VF	NG	MG	HL
13	COIN74	1959	45-EF	30-VF	HL	NG	MG
14	COIN16	1935S	25-VF	30-VF	HL	NG	MG
15	COIN40	1944D	55-AU	50-AU	HL	NG	MG
16	COIN25	1949	40-EF	50-AU	HL	NG	MG
17	COIN17	1940	12-F	10-VG	HL	NG	MG
18	COIN22	1946S	53-AU	60-MS	HL	NG	MG
19	COIN60	1951D (05)	15-F	10-VG	NG	MG	HL
20	COIN39	1944 (07)	53-AU	45-EF	MG	HL	NG

Where

MG = Machine Grade provided
HL = Hi-Low (or misleading) Grade provided
NG = No Grade provided

Appendix E

Detailed Results from the machine grading experiments

MACHINE TEST A

50 Images against the remaining 55 images in system

ID	Year	NumGrade	Text	Machine Grade	Difference
2	1910	6	06-G	8	2
3	1911	6	06-G	6	0
5	1913	4	04-G	6	2
7	1915D	6	06-G	6	0
8	1916	12	12-F	12	0
9	1917D	6	06-G	10	4
11	1919D	8	08-VG	8	0
12	1923	6	06-G	6	0
14	1925	40	40-EF	40	0
16	1935S	25	25-VF	20	5
17	1940	12	12-F	25	13
20	1941D	25	25-VF	30	5
21	1946	53	53-AU	53	0
22	1946S	53	53-AU	58	5
26	1949S	6	06-G	6	0
27	1950	12	12-F	12	0
34	1955	25	25-VF	30	5
35	1956	40	40-EF	53	13
36	1958D	53	53-AU	45	8
41	1944D	55	55-AU	53	2
42	1944S	50	50-AU	53	3
48	1947	25	25-VF	25	0
49	1947S	15	15-F	15	0
50	1948	55	55-AU	58	3
54	1949D	12	12-F	15	3
57	1950D	25	25-VF	25	0
59	1951	40	40-EF	40	0
60	1951D	30	30-VF	45	15
63	1952D	35	35-VF	35	0
66	1953D	53	53-AU	53	0
69	1955	50	50-AU	50	0
74	1957D	55	55-AU	55	0
75	1958	53	53-AU	45	8
76	1959	45	45-EF	35	10
77	1960D	53	53-AU	55	2
79	1963	63	63-MS	60	3
80	1966	64	64-MS	61	3
82	1968D	62	62-MS	55	7
83	1969D	63	63-MS	60	3
85	1971	65	65-MS	65	0

87	1975	58	58-AU	60	2
88	1953	50	50-AU	50	0
91	1948	61	61-MS	58	3
92	1979	64	64-MS	60	4
93	1944	60	60-MS	60	0
94	1934	30	30-VF	25	5
96	1926	20	20-F	25	5
97	1960	60	60-MS	45	15
98	1961D	45	45-EF	53	8
102	1951	15	15-F	20	5
				Totals	**171**

Summary - Test A

# Coins Tested	50	
# Coins matched in test	20	40.00%
# Coins not matched in test	30	60.00%
Average Grade	36.580	
Average Machine Grade	36.960	
Standard Deviation	4.126	
Average Grade Variance	5.70	Points
Grading Scale	70	Points
Variance % (Incorrect Grades)	8.14%	
Level of Accuracy	91.86%	

MACHINE TEST B

40 Images against the remaining 65 images in system

ID	Year	NumGrade	Text	Machine Grade	Difference
2	1910	6	06-G	6	0
3	1911	6	06-G	6	0
5	1913	4	04-G	6	2
7	1915D	6	06-G	6	0
8	1916	12	12-F	12	0
9	1917D	6	06-G	8	2
11	1919D	8	08-VG	8	0
14	1925	40	40-EF	40	0
20	1941D	25	25-VF	25	0
21	1946	53	53-AU	53	0
22	1946S	53	53-AU	58	5
26	1949S	6	06-G	8	2
27	1950	12	12-F	12	0
34	1955	25	25-VF	30	5
42	1944S	50	50-AU	53	3
48	1947	25	25-VF	15	10
50	1948	55	55-AU	53	2
54	1949D	12	12-F	15	3
57	1950D	25	25-VF	15	10
59	1951	40	40-EF	40	0
63	1952D	35	35-VF	35	0
66	1953D	53	53-AU	53	0
69	1955	50	50-AU	50	0
74	1957D	55	55-AU	55	0
75	1958	53	53-AU	40	13
76	1959	45	45-EF	40	5
77	1960D	53	53-AU	55	2
79	1963	63	63-MS	60	3
80	1966	64	64-MS	60	4
82	1968D	62	62-MS	55	7
83	1969D	63	63-MS	60	3
85	1971	65	65-MS	65	0
87	1975	58	58-AU	60	2
91	1948	61	61-MS	58	3
93	1944	60	60-MS	58	2
94	1934	30	30-VF	25	5
96	1926	20	20-F	25	5
97	1960	60	60-MS	60	0
98	1961D	45	45-EF	55	10
102	1951	15	15-F	20	5

	Totals	**113**

Summary - Test B

# Coins Tested	40	
# Coins matched in test	16	40.00%
# Coins not matched in test	24	60.00%
Average Grade	36.975	
Average Machine Grade	36.450	
Standard Deviation	3.350	
Average Grade Variance	4.71	Points
Grading Scale	70	Points
Variance % (Incorrect Grades)	6.73%	
Level of Accuracy	93.27%	

MACHINE TEST C

20 Images against the remaining 85 images in system

ID	Year	NumGrade	Text	Machine Grade	Difference
2	1910	6	06-G	6	0
5	1913	4	04-G	4	0
7	1915D	6	06-G	6	0
14	1925	40	40-EF	35	5
20	1941D	25	25-VF	25	0
22	1946S	53	53-AU	55	2
27	1950	12	12-F	15	3
34	1955	25	25-VF	25	0
42	1944S	50	50-AU	50	0
54	1949D	12	12-F	15	3
59	1951	40	40-EF	45	5
74	1957D	55	55-AU	55	0
79	1963	63	63-MS	63	0
80	1966	64	64-MS	64	0
91	1948	61	61-MS	60	1
94	1934	30	30-VF	25	5
96	1926	20	20-F	25	5
97	1960	60	60-MS	61	1
98	1961D	45	45-EF	55	10
102	1951	15	15-F	20	5
				Totals	**45**

Summary - Test C

# Coins Tested	20	
# Coins matched in test	9	45.00%
# Coins not matched in test	11	55.00%
Average Grade	34.300	
Average Machine Grade	35.450	
Standard Deviation	2.789	
Average Grade Variance	4.09	Points
Grading Scale	70	Points
Variance % (Incorrect Grades)	5.84%	
Level of Accuracy	94.16%	

MACHINE TEST D

Run each coin through the machine-based system

ID	Year	NumGrade	Text	Machine Grade	Difference
1	1909	40	40-EF	35	5
2	1910	6	06-G	6	0
3	1911	6	06-G	8	2
4	1912	4	04-G	6	2
5	1913	4	04-G	4	0
6	1914	6	06-G	6	0
7	1915D	6	06-G	6	0
8	1916	12	12-F	12	0
9	1917D	6	06-G	6	0
10	1918	8	08-VG	10	2
11	1919D	8	08-VG	10	2
12	1923	6	06-G	6	0
13	1924	25	25-VF	25	0
14	1925	40	40-EF	35	5
15	1926	8	08-VG	8	0
16	1935S	25	25-VF	25	0
17	1940	12	12-F	10	2
18	1940S	8	08-VG	8	0
19	1941	12	12-F	10	2
20	1941D	25	25-VF	25	0
21	1946	53	53-AU	53	0
22	1946S	53	53-AU	55	2
23	1947S	12	12-F	12	0
24	1948	53	53-AU	53	0
25	1949	40	40-EF	40	0
26	1949S	6	06-G	6	0
27	1950	12	12-F	15	3
28	1950D	40	40-EF	45	5
29	1950S	25	25-VF	25	0
30	1952	40	40-EF	40	0
31	1953	12	12-F	12	0
32	1953S	6	06-G	8	2
33	1954	12	12-F	15	3
34	1955	25	25-VF	25	0
35	1956	40	40-EF	40	0
36	1958D	53	53-AU	53	0
38	1942D	25	25-VF	30	5
39	1942S	12	12-F	12	0
40	1944	53	53-AU	53	0
41	1944D	55	55-AU	55	0

42	1944S	50	50-AU	50	0
43	1945	53	53-AU	53	0
44	1945D	15	15-F	12	3
45	1945S	12	12-F	12	0
46	1946	63	63-MS	62	1
47	1946D	60	60-MS	60	0
48	1947	25	25-VF	15	10
49	1947S	15	15-F	15	0
50	1948	55	55-AU	58	3
51	1948D	12	12-F	12	0
52	1948S	4	06-G	4	0
53	1949	52	52-AU	58	6
54	1949D	12	12-F	15	3
55	1949S	58	58-AU	58	0
56	1950	15	15-F	15	0
57	1950D	25	25-VF	15	10
58	1950S	53	53-AU	52	1
59	1951	40	40-EF	45	5
60	1951D	30	30-VF	30	0
61	1951S	15	15-F	15	0
62	1952	45	45-EF	45	0
63	1952D	35	35-VF	30	5
64	1952S	35	35-VF	35	0
65	1953	50	50-AU	50	0
66	1953D	53	53-AU	53	0
67	1954	50	50-AU	58	8
68	1954D	12	12-F	12	0
69	1955	50	50-AU	53	3
70	1955D	50	50-AU	50	0
71	1956	50	50-AU	55	5
72	1956D	53	53-AU	53	0
73	1957	45	45-EF	45	0
74	1957D	55	55-AU	55	0
75	1958	53	53-AU	40	13
76	1959	45	45-EF	45	0
77	1960D	53	53-AU	53	0
78	1962	55	55-AU	53	2
79	1963	63	63-MS	63	0
80	1966	64	64-MS	64	0
81	1968	61	61-MS	63	2
82	1968D	62	62-MS	62	0
83	1969D	63	63-MS	61	2
84	1970	65	65-MS	65	0
85	1971	65	65-MS	65	0
86	1974	58	58-AU	58	0
87	1975	58	58-AU	60	2
88	1953	50	50-AU	50	0
89	1937	10	10-VG	8	2

90	1961	62	62-MS	63	1
91	1948	61	61-MS	60	1
92	1979	64	64-MS	62	2
93	1944	60	60-MS	60	0
94	1934	30	30-VF	30	0
95	1927	10	10-VG	10	0
96	1926	20	20-F	25	5
97	1960	60	60-MS	61	1
98	1961D	45	45-EF	45	0
99	1968D	50	50-AU	50	0
100	1945	58	58-AU	55	3
101	1934	12	12-F	15	3
102	1951	15	15-F	20	5
103	1955	40	40-EF	35	5
104	1952	45	45-EF	45	0
105	1942	8	08-VG	8	0
106	1968	55	55-AU	58	3
				Totals	**157**

Summary - Test D

# Coins Tested	105	
# Coins matched in test	54	51.43%
# Coins not matched in test	51	48.57%
Average Grade	34.867	
Average Machine Grade	34.952	
Standard Deviation	2.454	
Average Grade Variance	3.08	Points
Grading Scale	70	Points
Variance % (Incorrect Grades)	4.40%	
Level of Accuracy	95.60%	

Machine Test E

Samples coins run through the machine-based system

Coin	Yearmm	Expert # Grade	Expert Text Grade	Machine # Grade	Diff
01	1919D	8	8-VG	8	0
02	1911	6	6-G	6	0
03	1941	12	12-F	12	0
04	1946	55	53-AU	53	2
05	1951D	12	15-F	15	3
06	1952	40	45-EF	45	5
07	1944	53	53-AU	53	0
08	1946S	58	60-MS	60	2
09	1954	12	12-F	12	0
10	1968D	63	63-MS	63	0
11	1953S	4	6-G	6	2
12	1947S	15	12-F	12	3
13	1959	40	45-EF	45	5
14	1935S	25	25-VF	25	0
15	1944D	55	55-AU	55	0
16	1949	40	40-EF	40	0
17	1940	12	12-F	12	0
18	1946S	50	53-AU	53	3
19	1951D (05)	12	15-F	15	3
20	1944 (07)	53	53-AU	53	0
				Totals	28

# Coins Tested	20	
# Coins matched in test	11	55.00%
# Coins not matched in test	9	45.00%
Average Coin Grade	31.250	
Average Machine Grade	32.150	
Standard Deviation	1.759	
Average Grade Variance	3.11	Points
Grading Scale	70	Points
Variance % (Incorrect Grades)	4.44%	
Level of Accuracy	95.56%	

Appendix F

Detailed Results from the Online grading experiments

Appendix F1

All Grades and All Coins

Summarized Online Grading Results - All Graders & All Coins

Coin # 01 **Year 1919D**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	37	8	9.486	1.486	3.636	3	25	30	4	188

Coin # 02 **Year 1911**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	36	6	7.361	1.361	2.518	3	20	7	19	190

Coin # 03 **Year 1941**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	36	12	23.611	11.611	10.077	3	45	30	1	185

Coin # 04 **Year 1946**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	36	53	60.722	7.722	3.557	50	65	56	7	153

Coin # 05 **Year 1951d**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	36	15	24.194	9.194	11.226	1	45	4	33	179

Coin # 06 **Year 1952**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	36	45	49.917	4.917	4.676	35	62	21	34	161

Coin # 07 **Year 1944**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	36	53	51.972	1.028	5.209	30	63	23	41	152

Coin # 08 **Year 1946s**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	35	60	61.057	1.057	2.663	45	65	43	8	159

Coin # 09 **Year 1954**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	35	12	33.743	21.743	11.948	8	55	18	8	184

Summarized Online Grading Results - All Graders & All Coins

Coin # 10 **Year 1968d**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	35	63	61.171	1.829	3.090	50	65	25	15	170

Coin # 11 **Year 1953s**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	35	6	18.943	12.943	10.491	6	55	13	6	191

Coin # 12 **Year 1947s**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	34	12	24.441	12.441	11.900	6	55	7	13	184

Coin # 13 **Year 1959**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	34	45	32.324	12.676	13.060	6	60	17	5	182

Coin # 14 **Year 1935s**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	34	25	18.824	6.176	6.150	4	40	4	15	185

Coin # 15 **Year 1944d**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	34	55	54.971	0.029	3.176	45	63	19	13	172

Coin # 16 **Year 1949**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	34	40	48.529	8.529	7.713	6	65	20	4	180

Coin # 17 **Year 1940**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	34	12	21.206	9.206	9.212	4	45	9	7	188

Coin # 18 **Year 1946s**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	32	53	62.281	9.281	2.787	55	66	41	0	151

Summarized Online Grading Results - All Graders & All Coins

Coin # 19 **Year 1951d**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	32	15	19.531	4.531	6.187	6	40	5	16	171

Coin # 20 **Year 1944**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
TOTALS	32	53	51.031	1.969	7.210	15	64	10	33	149

Appendix F2

Grading Results by Grade Shown

Summarized Online Grading Results by Grades shown to Graders

Coin # 01 **Year 1919D**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	12	8	7.500	0.500	1.443	4	12	7	3	62
ALL NOGRADE	15	8	12.133	4.133	4.736	4	25	16	1	73
ALL HILOW	10	8	7.900	0.100	1.446	3	12	7	0	53

Coin # 02 **Year 1911**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	12	6	4.667	1.333	1.179	3	6	3	12	57
ALL NOGRADE	14	6	7.929	1.929	1.586	3	12	3	6	75
ALL HILOW	10	6	9.800	3.800	3.628	4	20	1	1	58

Coin # 03 **Year 1941**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	12	12	15.750	3.750	6.103	3	35	16	0	56
ALL NOGRADE	14	12	29.786	17.786	10.171	12	45	10	1	73
ALL HILOW	10	12	24.400	12.400	10.413	12	40	4	0	56

Coin # 04 **Year 1946**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	12	53	57.667	4.667	3.508	50	65	22	3	47
ALL NOGRADE	14	53	62.643	9.643	2.844	55	65	26	2	56
ALL HILOW	10	53	61.700	8.700	2.369	58	64	8	2	50

Coin # 05 **Year 1951d**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	12	15	14.750	0.250	4.907	1	30	2	16	54
ALL NOGRADE	14	15	36.429	21.429	9.583	10	45	0	8	76
ALL HILOW	10	15	18.400	3.400	7.351	8	40	2	9	49

Coin # 06 **Year 1952**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	12	45	45.417	0.417	4.573	35	62	11	10	51
ALL NOGRADE	14	45	52.286	7.286	4.868	35	62	4	16	64
ALL HILOW	10	45	52.000	7.000	3.742	40	58	6	8	46

Coin # 07 **Year 1944**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	10	53	52.900	0.100	2.625	50	63	6	5	49
ALL NOGRADE	12	53	52.417	0.583	4.890	35	63	15	16	41
ALL HILOW	14	53	50.929	2.071	5.639	30	62	2	20	62

Coin # 08 **Year 1946s**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	10	60	63.000	3.000	1.897	60	65	8	1	51
ALL NOGRADE	11	60	58.636	1.364	4.033	45	64	20	4	42

Summarized Online Grading Results by Grades shown to Graders

Coin # 08 **Year 1946s**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL HILOW	14	60	61.571	1.571	1.355	55	65	15	3	66

Coin # 09 **Year 1954**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	10	12	28.500	16.500	10.094	10	45	5	0	55
ALL NOGRADE	11	12	43.182	31.182	5.750	35	55	3	8	55
ALL HILOW	14	12	30.071	18.071	12.731	8	53	10	0	74

Coin # 10 **Year 1968d**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	10	63	62.500	0.500	1.204	60	65	5	2	53
ALL NOGRADE	11	63	58.091	4.909	3.933	50	64	15	5	46
ALL HILOW	14	63	62.643	0.357	1.630	58	65	5	8	71

Coin # 11 **Year 1953s**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	10	6	16.900	10.900	9.669	6	40	4	2	54
ALL NOGRADE	11	6	24.727	18.727	11.355	12	55	3	4	59
ALL HILOW	14	6	15.857	9.857	8.262	6	40	6	0	78

Coin # 12 **Year 1947s**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	10	12	19.200	7.200	7.060	10	35	3	2	55
ALL NOGRADE	11	12	34.000	22.000	13.889	12	55	2	9	55
ALL HILOW	13	12	20.385	8.385	8.013	6	40	2	2	74

Coin # 13 **Year 1959**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	13	45	38.077	6.923	7.692	20	50	6	1	71
ALL NOGRADE	10	45	24.400	20.600	15.497	6	53	5	0	55
ALL HILOW	11	45	32.727	12.273	11.282	10	60	6	4	56

Coin # 14 **Year 1935s**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	13	25	17.462	7.538	6.194	4	30	0	4	74
ALL NOGRADE	10	25	16.100	8.900	6.007	6	40	3	6	51
ALL HILOW	11	25	22.909	2.091	3.857	12	30	1	5	60

Coin # 15 **Year 1944d**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	13	55	55.154	0.154	2.406	50	61	2	6	70
ALL NOGRADE	10	55	58.000	3.000	2.449	50	63	7	3	50
ALL HILOW	11	55	52.000	3.000	3.766	45	63	10	4	52

Summarized Online Grading Results by Grades shown to Graders

Coin # 16 **Year 1949**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	13	40	47.154	7.154	5.776	40	58	4	1	73
ALL NOGRADE	10	40	49.100	9.100	9.311	6	65	6	0	54
ALL HILOW	11	40	49.636	9.636	5.014	45	58	10	3	53

Coin # 17 **Year 1940**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	13	12	17.308	5.308	5.930	10	30	3	2	73
ALL NOGRADE	10	12	25.900	13.900	8.139	4	45	5	2	53
ALL HILOW	11	12	21.545	9.545	10.498	8	45	1	3	62

Coin # 18 **Year 1946s**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	12	53	61.917	8.917	2.900	55	65	15	0	57
ALL NOGRADE	10	53	63.700	10.700	1.792	60	66	11	0	49
ALL HILOW	10	53	61.300	8.300	2.900	55	65	15	0	45

Coin # 19 **Year 1951d**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	10	15	19.700	4.700	5.711	10	35	2	5	53
ALL NOGRADE	10	15	17.300	2.300	4.605	8	30	1	8	51
ALL HILOW	12	15	21.250	6.250	6.435	6	40	2	3	67

Coin # 20 **Year 1944**

Test Type	Graders	Machine Grade	Average Grade	Difference	St Dev	Low Grade	High Grade	SQ Positive	SQ Negative	SQ No Effect
ALL REAL	10	53	50.600	2.400	3.544	40	63	4	8	48
ALL NOGRADE	12	53	52.917	0.083	6.020	30	63	1	18	53
ALL HILOW	10	53	49.200	3.800	10.265	15	64	5	7	48

Appendix F3

Subjective Qualities Considered most important by expert Graders

Subjective Qualities Considered most important by expert graders

		Aesthetic Appeal			Color			Toning			Defects			Strike			Planchet		
Coin	Year	Positive	Negative	No Effect	Positive	Negative	No Effect	Positive	Negative	No Effect	Positive	Negative	No Effect	Positive	Negative	No Effect	Positive	Negative	No Effect
01	1919D	8	2	27	10	1	26	0	0	37	0	1	36	7	0	30	5	0	32
02	1911	1	6	29	3	4	29	0	1	35	0	5	31	3	3	30	0	0	36
03	1941	8	0	28	9	0	27	0	0	36	1	1	34	9	0	27	3	0	33
04	1946	14	1	21	20	2	14	1	1	34	2	2	32	12	1	23	7	0	29
05	1951d	0	9	27	0	13	23	0	2	34	0	6	30	3	1	32	1	2	33
06	1952	4	10	22	9	6	21	1	3	32	1	12	23	4	2	30	2	1	33
07	1944	3	16	17	6	6	24	3	5	28	1	13	22	8	0	28	2	1	33
08	1946s	14	1	20	16	1	18	1	3	31	2	0	33	7	3	25	3	0	32
09	1954	3	3	29	4	3	28	0	1	34	2	1	32	6	0	29	3	0	32
10	1968d	4	5	26	7	6	22	2	3	30	3	0	32	5	1	29	4	0	31
11	1953s	3	2	30	3	2	30	0	0	35	0	2	33	4	0	31	3	0	32
12	1947s	1	4	29	2	3	29	0	1	33	1	3	30	2	0	32	1	2	31
13	1959	5	2	27	5	1	28	2	0	32	1	0	33	1	2	31	3	0	31
14	1935s	0	6	28	1	3	30	0	0	34	0	4	30	2	1	31	1	1	32
15	1944d	4	3	27	5	5	24	3	3	28	2	2	30	3	0	31	2	0	32
16	1949	5	0	29	3	3	28	4	0	30	1	1	32	5	0	29	2	0	32
17	1940	3	1	30	2	0	32	0	0	34	1	5	28	1	0	33	2	1	31
18	1946s	12	0	20	13	0	19	3	0	29	1	0	31	7	0	25	5	0	27
19	1951d	2	4	26	0	4	28	0	0	32	0	7	25	1	0	31	2	1	29
20	1944	1	11	20	3	7	22	1	7	24	1	6	25	3	0	29	1	2	29
	Totals	**95**	**86**	**512**	**121**	**70**	**502**	**21**	**30**	**642**	**20**	**71**	**602**	**93**	**14**	**586**	**52**	**11**	**630**

References:

[1] "Comics Guaranty Web Site",2002, http://cgccomics.com/index.asp

[2] Advocate, W.T., "Robert Parker's wine rating system",2001, http://search.cometsystems.com/search.php?qry=wine+scoring+system&gg_ca=kwmxdIcKRFYC:www.wine-advocate.com%2Finfo%2Flegend.html&origin=csearch

[3] AGS, "Advanced Grading Standards for Collectible Cards",2003, http://www.advancedgrading.com/ags.html

[4] AmericanNumismatic, "Official A.N.A. Grading Standards for United States Coins". Fifth Edition ed, ed. K. Bressett and A. Kosoff. 1996, Colorado Spings, CO: AmericanNumismatic Association.

[5] Amspacher, B., "1933 $20 is MS65! Market Continues to Show Widespread Strength", Newport Beach, CA, PCGS, Professional Coin Grading Service,2002, Newport Beach, CA, http://www.pcgs.com/articles/article_view.chtml?artid=3464&universeid=313

[6] Anderson, J., "Cognitive Psychology and Its Implications". 4th ed. 1995, New York: W.H, Freeman and Company.

[7] Austin, "US Mint Research - A History of the United States Mint",2001, http://www.acoins.com/usmint.html

[8] Bassett, R., et al. "Development of an Automated Coin Grader: A Progress Report". in *MASPLAS '02*. 2002. Pace University.

[9] Baumgart, J., "Compugrade Demonstration at the 1991 Chicago ANA show", N. rec.collecting.coins, Editor. 2001, rec.collecting.coins.

[10] Bellis, M., "The History of Comic Books", in *Inventors.Com*. 2001. p. 1.

[11] Biederman, I., "Recognition-by-components: A theory of human image understanding". Psychological Review, 1987: p. 115-147.

[12] Bischof, W.F. and T. Caelli, "Visual Learning of Patterns and Objects." IEEE Transactions on Systems, Man and Cybernetics, 1997(27): p. 907-917.

[13] Brachman, R. and Z. Lemnios, "The IPTO Renaissance: Cognitive Computing", AT&T Labs Research,2002, http://sl.cometsystems.com/r?u=http%3A%2F%2Fwww.fdis.org%2F02%

[14] Buhmann, J.M., T. Caelli, and F. Esposito, "Machine Learning in Computer Vision", Berlin, Germany, 2000, Berlin, Germany, http://www.mlnet.org/news/2000-09/ml-in-cv.htm

[15] Campbell, M., J. Hoane, and F.-h. Hsu, "Deep Blue". Elsevier Science Publishers Ltd., 2002: p. 57 - 83.

[16] Casti, J.L., "Searching for Certainty - What Scientists Can Know About the Future". 1991, New York, NY: William Morrow and Co.

[17] Cha, S.-H.C. and S. Srihari, "On Measuring The Distance Between Histograms". Pattern Recognition, 2001.

[18] Chin, B., "PCGS Computerized Coin Grading",1999, http://groups.google.com/groups?hl=en&lr=&ie=UTF-8&selm=11f733ec.9f21659a%40usw-ex0102-012.remarq.com

[19] Cook, J., "A Brief Introduction To Sports Card Collecting", in *Collecibles.com*. 2002. p. 5.

[20] D'Ambra, C., "How can I protect my collection from loss by fire or theft?" 2001, http://www.telesphere.com/ts/coins/faq2.html#PROTECT

[21] Data, R., "Browser Statistics", 2003, http://www.w3schools.com/browsers/browsers_stats.asp

[22] Davis, J., "Better than uncirculated?"2001, http://www.prairienet.com/coins/ecc/ecc0112.htm

[23] DeLosh, E.L. and P.S. Merritt, "PATTERN RECOGNITION", Colorado State University - Cognitive Psychology Laboratory,1997-2000, http://lamar.colostate.edu/~bclegg/PY453/pattern_recognition.pdf

[24] Dix, A., et al., "Human-Computer Interaction". 2nd ed. 1998: Prentice Hall.

[25] Estes, S., "Grading is Subjective Opinion",2003, http://www.steveestes.com/opinion.html

[26] Finlay, J. and A. Dix, "An introduction to Artificial Intelligence". 1996, London: UCL Press.

[27] France, B.n.d., "History of Coins, medals and antiquities",2000, http://www.bnf.fr/site_bnf_eng/connaitrgb/monnaiegb.htm

[28] Gallivan, P., et al., "Pervasive Computing: Automated Rare Coin Grader",2002, http://www27.brinkster.com/golddigger4/Project.html

[29] Ganz, D., "Computer Grading - Compu-Grade",1996, http://groups.google.com/groups?hl=en&lr=&ie=UTF-8&selm=54k1jo%241j1%40newsbf02.news.aol.com

[30] Goel, R., "Review: Machine Learning", London, South Bank University,1997, London, http://www.scism.sbu.ac.uk/inmandw/review/ml/review/rev6542.html

[31] Goldsborough, R., "Electronic Coin Collector's Guide", 2003, http://www.rg.cointalk.org/guide/slabs.html

[32] Gottawiz.com, "Comic Book Grading Standards",2003, http://www.gottawiz.com/Comics/Comics_Grading.htm

[33] Halperin, J., "CoinGrading.com". 1999, Dallas, TX: Heritage Capital Corporation.

[34] Halperin, J., "Heritage Comics Website",2002, http://www.heritagecomics.com

[35] Halperin, J., "Heritage Rare Coins and Currency",2002, http://www.heritagecoin.com

[36] Hickmott, B., "CompuGrade Input - 5 Places to the right of the decimal", R. Bassett, Editor. 2002.

[37] Johnson, R., "COINSHEET Numismatic Directory", 2002, http://www.goldsheetlinks.com/unceagle.htm

[38] Jörgenson, L. "Is Visualization Struggling under the Myth of Objectivity?" in *Visualization '95*. 1995. Online: IEEE.

[39] Lee, W. "Medical Imaging 1998: Image Display". in *SPIE*. 1998. San Diego, CA: The International Society for Optical Engineering.

[40] Locke, M., "Professional Grading Accuracy". 1995, Locke, Mike.

[41] Lu, G. and J. Phillips, "Using Perceptually Weighted Histograms for Colour-based Image Retrieval", Churchill Vic 3842, Australia, GSCIT, Monash University,1998, Churchill Vic 3842, Australia, http://www.gscit.monash.edu.au/~guojunl/icsp98-2.pdf

[42] Lueke, J., "The problems with computer coingrading". 2002.

[43] Lupton, C.J., F.A. Pfeiffer, and N.E. Blakemen, "Economic Impact of Pre-sale Fiber Measurements on Prices Paid for Wool". Sheep & Goat Research Journal, 1993. **9**(1): p. 35-37.

[44] Mallory, J., "Computer System Standardizes Rare Coin Grading 07/09/91". 1991, Mallory, Jim of CompuGrade.

[45] Marr, D., "Vision: a computational investigation into the human representation and processing of visual information". 1982, San Francisco: W.H. Freeman.

[46] Marus, B., "What you don't know about the Rare Coin Market can cost you",2002, http://www.coinmag.com/article_rarecoinmarket1.html

[47] Miller, G.A., "The magical number seven, plus or minus two: Some limits on our capacity for processing information." Psychological Review, 1956. **63**: p. 81-97.

[48] Miller, S., "The Grading Challenge Polls",2002, http://www.thestujoecollection.com/grade/grade.htm

[49] Mills, M., "Stamp Collecting for Beginners and Philatelists", The Glassine Surfer Stamp Collecting,2001, http://www.glassinesurfer.com/f/gsgrading.shtml

[50] Mint, U., "History of the United States Mint",2002, http://www.usmint.gov/about_the_mint/mint_history/

[51] Nagy, G. and J. Zou. "Interactive Visual Pattern Recognition". in *International Conference on Pattern Recognition*. 2002: Electrical, Computer and System Engineering, Rensselaer Polytechnic Institute.

[52] Numistrust, "The Coin Grading Process",2002, http://www.numistrust.com/index.shtml

[53] Payne, J.W., "Contingent Decision Behavior." Psychological Bulletin, 1982. **92(2)**: p. 382-02.

[54] PCGS, "History of Grading",2001, http://www.pcgs.com/coinguidetext/display_chapter.chtml?chapter=grading&rnk=1&qry=coin+grading+history%3f

[55] Poor, J., et al., "Objective Verses Subjective Measures of Water Clarity in Hedonic Property-Value Models". Land Economics,, 2001(November 2001).

[56] Power, D., "Coin Grading Expert Systems". 1994, newsgroup: rec.coin.collecting.

[57] Puto, C.P., "The Framing of Buying Decisions." Journal of Consumer Research, 1987. **3**: p. 301-305.

[58] Reed, B., "COMPUGRADE technology",2001, http://groups.google.com/groups?q=reed+compugrade&hl=en&lr=&ie=UTF-8&selm=20011205215750.12739.00000988%40mb-cu.aol.com&rnum=1

[59] Riedl, A., "The Psychology of Learning", Munchen, Technische University Munchen,2002, Munchen, http://www.lrz-muenchen.de/~riedlpublikationen/pdf/kuw3psychologylearning.pdf

[60] Rueckl, J., "Cognitive Psychology Psych 256 Course Notes",2000, http://sp.uconn.edu/~ps256is1/syllabus.html

[61] Sahara, C., "Coin Grading Glossary",2002, http://209.10.209.244/grading/glossary.htm

[62] Sahara, C., "I-70 Grading Scale",2002, http://209.10.209.244/grading/grading.htm

[63] Schmidt, M., "It was thirty years ago today!"2002, http://forums.collectors.com/messageview.cfm?catid=26&threadid=99514

[64] Sheldon, W.H., W. Breen, and D. Paschal, "Penny Whimsy". 1976: Quarterman Publications, Incorporated.

[65] Sims, A., "Comments about THIRD PARTY GRADING COMPANIES (SLABBING)", Alamogordo, New Mexico, CAMACS,2001, Alamogordo, New Mexico, http://www.camacs.com/coin_grading_services.html

[66] Sleepy, "Early American Coins",1998, http://www.collectorscorner.org/coin.html#Early

[67] Sperling, G., "Disseratation: The information available in brief visual presentations". 1960.

[68] Tarr, M.J., "Visual Pattern Recognition". 2000, Encyclopedia of Psychology.

[69] Travers, S.A., "The Fundamentals of Counterfeit Detection", in *Coin Grading and Counterfeit Detection*. 1997, Random House, Inc.: New York. p. 225.

[70] Umbaugh, S.E., "Computer Vision and Image Processing". 1999, Upper Saddle River, NJ: Prentice Hall.

[71] unknown, "The Components of Grading Stamps",2003, http://www.psestamp.com/components.chtml

[72] unknown, "Dictionary.com",2003, http://dictionary.reference.com/search?q=Objective

[73] unknown, "General Grading Standards for U.S. Coins",2002, http://www.numismedia.com/fmv/grades.shtml

[74] unknown, "The Internet Encyclopedia of Philosophy",2002, http://www.utm.edu/research/iep/about.htm

[75] unknown, "Money In Colonial Times",2003, http://www.phil.frb.org/education/colonial.html

[76] unknown, "US Mint History from 1792",2002, http://rarecoins.org/mint_history.htm

[77] Wandell, B.A., "Foundations of Vision", Sinauer Associates,1995,

[78] Wyman, R.J., "A Companion to Rare Coin Collecting". 1997, HeavyPen Productions.

[79] Yeoman, R.S., "A Guide Book of United States Coins 2003". 56th ed, ed. K. Bressett. 2002, New York: St. Martin's Griffin.