THE UNIVERSITY OF CHICAGO

MEASURING ABILITY IN SPORTS PERFORMANCE:

THE CASE OF BASEBALL

A THESIS SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES

AND THE DEPARTMENT OF EDUCATION

IN CANDIDACY FOR THE DEGREE OF

MASTER OF ARTS

DEPARTMENT OF EDUCATION

BY

PATRICK BRADLEY FISHER

CHICAGO, ILLINOIS

OCTOBER 1993

ACKNOWLEDGEMENTS

This thesis is the direct result of a conversation I had with my oldest brother, William, at our parents' house during Christmas, 1985. I asked him to tell me about his graduate work at the University of Chicago. He chose to tell me about measurement theory in terms of something dear to my heart - baseball. The one person who might be even more important to this work than William is my mentor and thesis advisor Professor Benjamin D. Wright. His measurement wisdom and insight are peerless; any contribution that this thesis makes to better measurement of sports abilities is due to the clarity Professor Wright has brought to measurement issues in general. I also thank my wife, Kimberly, my parents, William and Grace Fisher, my grandfather, Edward Castelein, and other family and friends for their encouragement and support in seeing this project through to completion. I'd also like to acknowledge Ned Colletti of the Chicago National League Ball Club, Inc. for his insight into baseball's management needs.

Finally, I want to say that masculine pronouns are used throughout this thesis because of the fact that male major league baseball players remain in the overwhelming majority. Though the text may therefore have a sexist ring to ears accustomed to more inclusive language, this male bent is not expressive of any attitude of my own, but should of course be read as a commentary on the situation as it exists.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

The purpose of this thesis is to raise and pursue briefly some questions regarding the measurement of sports performance. The first question that needs to be raised has to do with what measurement is. The word 'measure' has 40 different meanings, according to the Random House Dictionary of the English Language. When someone uses a ruler to establish the length or height of something, the most common and most rigorous definition of measurement has been invoked. However, people begin to deal with measurement in various vague ways before they realize it; thus the second question that needs to be raised has to do with why anyone should care about measuring sports performance. Many people who talk about sports sooner or later discuss the value of one player over another; these discussions are, in effect comparisons of subjective measures of the players' abilities. Sports talks, discussions, debates and arguments can and do go on forever, largely because the

participants do not agree on what counts as valid evidence in support of a measure of ability; the problem is that they do not know they disagree on this basic level.

Thus, analysts of sports performance have tried to clarify for many years which players are better and why, but, sadly, have made little progress in establishing facts that a majority can agree on as a basis for comparison. The often very imaginative work done in this area applies traditional statistical manipulations to a wide variety of numbers derived from virtually every conceivable playing situation. But this work, despite its frequently high level of technical virtuosity, makes some fundamental, but usually unexamined, assumptions about the numbers being used to represent player abilities. The most basic assumption is that the numbers represent an invariant continuum of player ability - that is, that they add up to something. As many (Thurstone, 1928; Guttman, 1950; Luce & Tukey, 1964; Loevinger, 1965; Rasch, 1960; Wright, 1968, 1977; Andrich, 1988; and others) who have worked in educational and psychological measurement have said, such an assumption ought actually be a requirement, an hypothesis that is tested and substantiated before statistics are derived from the data.

Rasch measurement models (Rasch, 1960; Wright, 1968, 1977; Wright & Masters, 1982; Andrich, 1988) test the hypothesis that numbers really represent quantities (this hypothesis is known as the quantitative hypothesis (Michell, 1990)), and thereby provide objective measurement; Rasch's models have contributed to American educational testing since 1960, when Rasch gave a series of lectures at the University of Chicago and the University of California, Berkeley.

Georg Rasch, a Danish mathematician, was commissioned by his country's government to design a test that would better measure the progress made by mathematics students. Rasch realized that, in order to measure more and less ability, it is necessary to draw out a continuum of greater and lesser performance, using specific criteria for what will count as an observation of some amount of ability to perform. The key concept is that the criteria (test items, in the case of educational testing) must stay in the same order of difficulty no matter which person is observed, and the persons must stay in the same order of ability no matter which test item's responses are examined.

Data that confirm the quantitative hypothesis and make measurement objective are not difficult to put together. Since his attendance at Rasch's 1960 Chicago lectures, Benjamin D. Wright has guided applications of Rasch's measurement theories to models for measuring everything from theatrical performances as rated by judges to certification testing for medical professions to attitudes and, now, sports performance. Rasch measurement is now routinely employed by school boards and testing agencies around the world, but there are many unexamined areas of daily life ripe for applications of Rasch's measurement models, such as economic forecasting, personnel evaluation, market research, and performance in sports; this thesis will focus on baseball performance, in particular.

The Problem of Measuring Baseball Players' Performance

Objective measurement in sport is lacking, particularly in baseball. There are several different methods used to categorize baseball players according to their abilities, and none is consistent with any other. Players are measured by different rulers and subsequent

comparisons of abilities are flawed. Some analysts configure new methods, claiming that new degrees of accuracy and insight follow from their application. But none of these methods are scientific. Their basic features give them away. No standard error of the ability measure is given. No analysis of the validity of the data analyzed is given. Reliability statistics are unheard of. The word and reputation of the analyst are all the reader has to go on for credibility.

For instance, the batting average is a very popular, but nonetheless flawed, method for measuring the ability of baseball players. The batting average is a statistic confounded by so many factors that comparison of the numbers is virtually meaningless. Better measures are needed in order to compare players without the influence of those factors. Objective measurement – measurement that tests and substantiates the hypothesis that the variable measured is quantitative – is needed in baseball in order to derive the needed improvements to baseball performance measures. Objective measurement standards would require that the basis for measures of player performance be only the ability of the player in question and the

difficulty of the situation posed. Influence from outside factors, such as the count, the ball park, the umpire, the wind speed, or the inning, could cause the quantitative hypothesis to be falsified, not confirmed. If this should happen, we could then move on to other, qualitative kinds of analysis, and stop playing with numbers that do not actually represent quantities, and so do not mean anything.

The potential for falsification of the quantitative hypothesis in the measurement of baseball playing ability is, in fact, a well known topic of discussion among baseball enthusiasts. Two baseball writers for the Chicago Tribune, Jerome Holtzman and Alan Solomon, wrote in the May 22, 1988, edition of the paper that they are skeptical about the value of batting averages:

> Holtzman: "Here's why batting averages are for the birds: When the Oakland A's and [Boston] Red Sox met last week, their team averages were almost identical, .273 for the A's vs. .272 for the Red Sox. Of much more importance is that the A's had scored 58 more runs, led in home runs 42-17 and in total bases 563-460. Also, Mark McGwire had hit 11 home runs, which was as many as the entire starting Boston lineup."

Solomon: "What do batting averages mean? During the eight games (May 11-19) of the [Chicago] Cubs' scoring slump when they totaled only 10 runs, the team batting average fell only nine points, to .253 from .262. That's a slide, but hardly a plummet. On May 11, Rafael Palmeiro was hitting .345; after the May 19 game, he was at .350. Others: Dave Martinez, .263, .244; Ryne Sandberg, .240, .236; Andre Dawson, .328, .307; Mark Grace, .280, .296; Vance Law, .297, .284; Shawon Dunston, .261, .241. Some higher, some lower, no major collapses from the high-scoring days to the no-scoring days. Timing was everything. No big hits."

These words of these two writers exhibit a common attitude of baseball fans toward the batting average as a measure of player or team ability. Holtzman drives home the point with his statement, "Of much more importance is that the A's had scored 58 more runs..." The Red Sox seem to have suffered from a problem similar to that afflicting the Cubs - "Timing was everything. No big hits." These articles appeared about six weeks into the 1988 season - approximately 30-36 games had been played. To be outscored by 58 runs in a season so

young is extraordinary, that being the equivalent of losing by almost two runs in every ball game.

Most analysts today admit that an influence is exerted on player performance by each ball park because of uncontrolled variations in physical dimension and climatic condition; unfortunately, the ball park is one of the few uncontrolled variables acknowledged by these analysts. The problem is that there are more than 30 other factors which exert small amounts of influence on each at-bat. How much these other factors influence each plate appearance needs to be assessed and statistically controlled if there is to be any basis for comparing players' performances.

An example of the kind of fruitless debates that erupt over variations in player performance will bring home the point. Every few years offensive production dramatically increases, as happened in 1987 and 1993, for example. Scott Smith, marketing services manager for Rawlings, the manufacturer of major league baseballs, was called upon to state in a September 13, 1993 Sports Illustrated article (Kurkjian, 1993) that "The baseball has not

changed in any way." He was motivated to point this out in response to people who insist that the ball is 'juiced' as the only means of accounting for the increased offensive production. Smith also says "It's somewhat funny. But people accuse the ball when there's an increase in offense and they neglect other elements, such as the human factor, wind, ballparks. There are so many variables." He is correct.

For instance, during the 1992 and 1993 seasons many people noticed the extent to which the weather affected games played at Wrigley Field in Chicago. Wrigley has a southwest to northeast orientation, meaning that if a straight line were drawn from home plate to center field it would go from southwest to northeast. Given the directions of the prevailing winds, Wrigley's orientation gives nature a part to play in determining the number of runs in a game. On warm summer days when the wind is blowing generally from the south it is also blowing out toward the fences in Wrigley Field. After a cold front blows through and changes the direction of the wind, the wind will be blowing in off Lake Michigan and into the batters' faces.

Many home runs in Wrigley Field have been directly attributed to the wind blowing out to left field and just as many have been kept in the ball park, allowing the outfield players to catch the ball and make an out. During the summer of 1992 the wind blew in quite frequently and run production was down markedly; the opposite was true during the past summer of 1993. The wind blew out frequently and consequently run production increased again. The nuances of Wrigley Field make a resounding argument for measuring the impact of wind and ballpark on a ball game, and for removing such factors from comparisons of player and performances.

Another explanation for the increased offense of 1993 is with regard to expansion. The National League expanded to 14 teams from 12. This translates to at least 20 new pitchers in the major leagues; 20 pitchers with no experience at the major league level or pitchers whose best days are behind them. This allows for the veteran offensive players to take advantage of the untested rookies or past their prime veterans.

Many authors, most notably Pete Palmer and John Thorn (1985), Bill James (1988), and, most recently, Mike Gimbel (1992), have tried to measure baseball performance, but none of them have provided a scientific basis for proceeding as they do. Scientific evidence would make it possible to establish 1) the reliability of the measures in terms of the error of measurement; 2) the validity of the measurement model for the intended purpose; and 3) the extent to which the data realize that purpose - do the data supply the evidence needed to say we are measuring what we purport to measure?

In track and field the methodology of performance assessment relies on the tape measure and stop watch: definitive tools of measurement. All runners, regardless of age and ability, are comparable based on the times and distances run; wind and wet weather can be factors influencing performance, but these are very few compared to the baseball situation. A helpful wind speed is explicitly discounted in the long jump, for instance. Field event athletes (except shot putters and discus throwers - these two events' projectiles have varying weights as the athlete gets older/stronger) also can be compared regardless of age and ability

because all performances are measured quantitatively – in terms of an amount that stays constant no matter where or when it is measured. In no other sport has there ever been equal comparability. My effort here is aimed at beginning to show how baseball players' abilities can be made comparable based on what they do at the plate under the conditions they find themselves in.

Play as a Test of Skill

This thesis asserts a fundamental role to a basic similarity that exists between academic performance and sports performance: both involve the expression of abilities that can be quantitatively measured by means of tests that provide data confirming the quantitative hypothesis. We can elaborate the connection by imagining a mathematics or reading test of 50 items. Each item has a different level of difficulty and each person taking the test has a different level of ability. A game of baseball is also a test in that it is

easy to consider each hitter's at bat as an item; then the batter is the person taking the test, and the pitcher and the opposing team are administering it.

The difficulty of the test depends on the speed and accuracy of the pitcher's throws, and of the players' abilities to field the ball. A batter typically responds to three to five test items (at bats) per game, and the pitcher will administer 27-40 items (batters faced) per game. For a batter to get the item right he needs to advance the runner(s) or get on base; when this happens, the batter was more able than the test item was difficult. For the batter to make an incorrect response to the test item, he must strike, ground, or fly out, and also will not advance any runner(s) who might be on base.

Because of the wide variety of variables involved, every at bat is different and has a different level of difficulty. In this analysis the offensive part of the game involves four independent variables (inning, outs, lead/deficit, event at bat) and one dependent variable (advancing runners); these variables are the focus of the following analysis.

METHOD

Software

Of the many computer software programs written for Rasch measurement applications, FACETS (Linacre, 1988) is among the most innovative and original. The FACETS program was originally written in order to include the severity or leniency of judges who rate the performance of persons taking an item on a test as a factor in professional certification examinations (Linacre, 1989). The FACETS measurement model allows the researcher to specify which of the various component factors in a measurement situation will have a predictable influence on the expression of ability; these factors are then removed from the final estimate of ability.

For instance, a severe or lenient judge, one who may have prejudged the capabilities of an examinee on the basis of sex, race, religion, familiarity, etc, will be more likely to assign inordinately low or high ratings than he or she would for another examinee. Most usually, however, judges rate consistently, no matter whether they are consistently more severe or

more lenient. The point is that the variation in the judges' ratings cannot be allowed to affect the examinees' measures (Lunz et al, 1990). When judge severity and leniency is included as a facet in a measurement model, along with item difficulty and person ability, the judge facet is factored out of the ability facet in the same way that the item facet is. The facets measurement model is not limited to three facets (judge, person, item), but can incorporate any number of factors that might exert confounding influence on the estimation of ability.

Overall Research Program Design

When using the FACETS program in conjunction with baseball, any factors which influence the game are modeled as separate facets in the same way that the judges were for the program's original application. Here are the separate facets for measuring baseball playing ability:

Personal	Environmental	Game (general)	Game (specific)
Team	Temperature	Team	Lead/Deficit
Height	Skies	Manager	Pitcher (l/r)
Weight	Precipitation	Home/Away	On base

Bats (l/r)	Wind speed	Plate umpire	Outs
Throws (l/r)	Wind direction	Stadium	Strikes
Age	Day/Night	Grass/turf	Balls
Def. pos.	Day of the week	Level of play	Event at bat
Bat. pos.	Month	Condition of the Field	Advances
	Year	Dome/outside	

Listed above are 34 factors (facets of baseball playing ability) that may have even a slight impact on an at-bat. The point in mentioning them is to lay out a broad research program; the following analysis will provide an example of a facets analysis that can be used as a foundation for building this research program.

Some of the 34 facets could be specified in even more detail; for instance, the on base facet should specify whether each base is occupied, not just lump the three of them together. More information would be included in the analysis if each base were assigned a dichotomous code (1=occupied, 2=not occupied) than if the number of men on base were simply recorded (no one on base=1, one man on base=2, etc.).

Most of these facets have not been tested for the amount of impact, if any, they have on a game. The difficulty and expense of obtaining the data has been prohibitive, and not

every facet will be needed for particular measurement problems. For instance, the year facet would not show any impact until there were several years of data being analyzed. Information on the year of play would become useful when comparing players of past and present. The year facet would spell out which year was the easiest or the most difficult in which to play. A trend may develop showing that a certain era, maybe the 'dead ball era,' was the most difficult time to be an outstanding offensive player.

The plate umpire must be included as a facet in any attempt to measure baseball playing ability because he has a direct impact on the interpretation of each pitch, and so on the count, the at-bat, and the game. To arrive at satisfactory measures of the umpires' severity or leniency in making calls, the ball and strike count of each at bat is needed. That information is currently unavailable, but future research must include it if only to evaluate the effect of the fact that each umpire has his own strike zone. Rarely does an umpire go by the book in the definition of the strike zone. The rule book definition of the strike zone is that it should extend from the armpits to the top of the knee, and be as wide as home plate. In

practice, however, it usually ranges from that rule book definition, which is considered large, to a smaller zone delimited by the letters on the players' uniform and the top of his thigh, which is also often somewhat wider than the width of the plate.

It is easy to see why every pitcher and batter needs to get to know every umpire's strike zone. The pitcher and batter know then what they have to do at the plate in order to succeed. Some umpires may give the corner of the plate if the pitcher has been accurate to that point in the game. The opposite is also true; the umpire may not give the corner if the pitcher has been erratic and inconsistent.

Measures of umpire severity will provide some insight into the consistency of the umpires, which could:

1. help the umpire become more consistent;
2. help managers include the umpire's calling characteristics in their decisions;
3. be used to evaluate or certify umpires; and
4. reveal specific biases in an umpire's calls.

The promise these four points hold for improving umpiring in particular and baseball in general is supported by analogous work done in the areas of clinical pathologist certification (Lunz et al, 1990), the assessment of motor and process skills in medical rehabilitation (Fisher, 1993; Fisher & Fisher, 1993), and in studies of novice, expert, and buff aesthetic judgment (Myford, 1989).

Lacking the pitch by pitch information needed to apply a complete facets model to the measurement of baseball playing ability, the at bat will be taken as the level of analysis. A model of offensive playing ability at this level of analysis must include bases on balls, sacrifices, being hit by pitched balls, and reaching base by defensive player's interference or error, as well as hits and RBIs. Any legitimate way a player reaches base or advances runners has to be included in the model as a productive component.

Linacre (1989) describes a general form for many-faceted measurement as:

$$\text{Log} \left(P_{nijk} \; / \; P_{nijk-1} \right) = B_n - D_i - C_j - F_k$$

where

P_{nijk} is the probability of examinee n, when rated on item i by judge j, being awarded a rating of k;

P_{nijk-1} is the probability of examinee n, when rated on item i by judge j, being awarded a rating of k-1;

B_n is the ability of examinee n;

D_i is the difficulty of item i;

C_j is the severity of judge j; and

F_k is the extra difficulty overcome in being observed at the level of category k, relative to category k-1.

This four facet model is easily expanded to allow more facets to be included; this model also specifies a single rating scale common to all items and judges, but the FACETS software can handle as many different rating structures as are needed. The only requirement is that the additional facets be advantageous or disadvantageous effects that are to be

removed from the estimation of ability. Codes for the proposed 34-facet design will now be specified.

Personal facets codes

All codes begin at 1. The lowest height, weight, and age is the starting point for coding. For example, if the shortest player to be analyzed is 65", then 65" = 1, 66" = 2, 67" = 3, etc. The same principle applies to weight and age. There is a numbering system for the defensive positions in the field; 1 = pitcher, 2 = catcher, 3 = first baseman, 4 = second baseman, 5 = third baseman, 6 = shortstop, 7 = leftfielder, 8 = centerfielder, and 9 = rightfielder. The numbering system for the batting order is one through nine and has no relation to the numbering of the defensive positions. Any player can lead off (1), any player can bat clean up (4) and any player can bat last (9).

Proposed Codes for the Personal Facet

Height, number of inches

Weight, number of pounds

Age, number of years

Bats, right = 1, left = 2

Throw, right = 1, left = 2

Defensive position, traditional number for the position

Batting order, number in the order

Environmental facets codes

The information needed to devise a rating scale for the sky condition was obtained in a conversation with WGN-TV (Chicago) meteorologist Tom Skilling (August 10, 1993, 3 p.m.). He indicated that scattered clouds will occupy up to 3/8ths of the sky, broken clouds take up from 3/8ths to 6/8ths and overcast is 7/8ths and the whole sky. He also mentioned the Beaufort scale for wind speed. The Beaufort goes well past the limited scale outlined below, however, since it is rare that ball games are played under conditions when the wind is stronger than 30 mph. Each of the first seven categories has a name; calm, light air, light

breeze, gentle breeze, moderate breeze, and fresh breeze and strong breeze. After that the descriptions are more for tropical storms. Sustained winds of higher than 30 mph on land are highly unusual (Williams, 1992).

Proposed Codes for the Environmental Facets

Temperature, cold = <61 (1), cool = 61 - 70 (2), warm = 71 - 80 (3), very warm = 81-90 (4), hot = 90+ (5)

Skies, clear = 1, scattered clouds = 2, broken clouds = 3, overcast = 4

Precipitation, none = 1, mist/drizzle = 2, rain = 3

Wind speed, under 1 mph = 1, 1 - 3 mph = 2, 4 - 7 mph = 3, 8 - 12 mph = 4, 13 - 18 mph = 5, 19 - 24 mph = 6, 25 - 31 mph = 7, 32 mph+ = 8

Wind direction, north = 1, northeast = 2, east = 3, southeast = 4, south = 5, southwest = 6, west = 7, northwest = 8

Day/night, day = 1, night = 2

Day of the week, Sunday = 1, Monday = 2,...Saturday = 7

Month, April = 1, May = 2, June = 3, July = 4, August = 5, September = 6, October = 7

Year, coded as needed; oldest year analyzed would be coded 1

Game (general) facets codes

These general factors for each game are coded according to how many games or how long of a period is being analyzed. For purposes of long term analysis, each team could be referred to in a more general sense in order to encompass their past aliases. The Oakland Athletics, for instance, were previously the Kansas City A's, and originally the Philadelphia A's.

The effect of the manager on the team and players' performance is a subject of great debate. Some say that managers do not matter; all of them know and use the same strategies in managing a baseball game. That may be close to true; however, a manager's role as psychologist in dealing with 25 egos and keeping the players up for a whole season playing at their best is a different matter that is no less a part of baseball team management. The Hall of Fame broadcaster Harry Carey said that deceased former New York Yankee manager Billy Martin would be his manager "first, last and every time." Martin was the type of manager who could take care of 25 players and extract the best from them. How much of

a difference a manager makes on a team's performance is an empirical question that cannot be answered without looking at the data, as future research in this area must.

Proposed Codes for the Game (general) Facets

Opponent, each league alphabetized and coded 1 - 14

Manager, team analysis - manager that was fired = 1, new manager = 2.

Long term analysis - All managers for that era alphabetized and coded 1 - n

Home/away, home = 1, away = 2

Grass/turf, grass = 1, turf = 2

Dome/outside, dome = 1, outside = 2

Stadium, correspond to opponent, above

Condition of the field, 1 = wet, 2 = damp, 3 = dry

Home plate umpire, alphabetized for each league and coded 1 - n

Level of play, dependent upon the size and scope of analysis. If minor leagues on up are being analyzed then; Class A = 1, AA = 2, AAA = 3, Major Leagues = 4

Game (specific) facets codes

The details of the specific game played are recorded in these codes. The codes for these facets supply the data that spearhead the effort aimed at replacing the batting average with a better method of measuring player ability. During the broadcast of a game, announcers will often provide detailed information on how a particular batter has fared against the opposing team and pitcher. This is a well intentioned but incomplete and potentially misleading acknowledgement that player performance will vary, depending at least in part on who the opposition is. The analysis proposed here, in contrast, will systematically and comprehensively adjust measures of player ability according to all of the uncontrolled variables in the situation included in the measurement model.

Proposed Codes for the Game (specific) Facets

Lead/deficit, 1 = down 2 or more runs, 2 = down 1 run, 3 = tied, 4 = up 1 run, 5 = up 2 or more runs

Pitcher, alphabetized and coded 1 - n

Pitcher, right = 1, left = 2

First base, unoccupied = 1, occupied = 2

Second base, unoccupied = 1, occupied = 2

Third base, unoccupied = 1, occupied = 2

Outs, no outs = 1, one outs = 2, two outs = 3

Strikes, no strikes = 1, one strike = 2, two strikes = 3

Balls, no balls = 1, one ball = 2, two balls = 3, three balls = 4

Event at bat, double play = 1, strikeout = 2, foul out = 3, fly out = 4, ground out = 5, fielder's choice = 6, sacrifice = 7, sacrifice fly = 8, reached on an error = 9, walk, HBP, IW = 10, infield single = 11, single = 12, double = 13, triple = 14, home run = 15

Data

The data were obtained from Sports Team Analysis & Tracking System, Inc., (STATS, Inc.). The data covers three teams from the American League (A.L.) for one week of games from each of four months during the 1991 championship season. The teams selected are the Minnesota Twins, California Angels, and Cleveland Indians. The games included in these data were played in the weeks of April 8-15, May 13-20, June 17-23, and July 22-28.

The teams were chosen with regard to their place in the standings at the conclusion of the season. The premise for the selection of these three teams was the notion that winning

and losing teams will have proportionate variation in offensive production. However, any baseball fan realizes that pitching (a defensive component of play) has at least as big a role in winning games as batting does. Thus, the final season standings from which the three teams were chosen does not entirely match up with the won-loss records from the four weeks of the season sampled.

Minnesota had the best record in the A.L. for the season at 95 - 67, a .586 winning percentage, California finished the season at 81 - 81, a .500 winning rate and Cleveland had the worst record in the A.L. at 57 - 105, a .352 success rate. In the four weeks of sample data, Cleveland lost games at a ratio similar to its final standing (9 - 17, .346). California and Minnesota, though, switched places. The Angels won 18 and lost 10 of the sample games (.642), and the Twins went 12 - 13 (.480).

The data supplied by STATS, Inc., represented 81 games involving 235 players. Many of these players appeared too few times to obtain a reliable ability estimate, so they

were dropped from the analysis; data from 127 players and 3724 at bats are reported here.

Coding

The factors (player, inning, out, lead, and event at bat) to be included as facets in the measurement model were coded using numbers starting at one; the FACETS program allows zero to be used only as a dependent variable code. Players were given a code number as they were added to the data file. Innings were not recoded in that there is no zero inning. The first inning = 1; the second inning = 2, etc. Outs, however, were recoded so that 0, 1, and 2 outs were codes of 1, 2, and 3, respectively. Lead was recoded from an actual count of runs as follows:

Down two runs or more	= 1
Down one run	= 2
Tied	= 3
Up one run	= 4
Up two runs or more	= 5

Event at bat was coded as follows:

Double play	= 1

Strikeout	= 2
Foul out	= 3
Fly out	= 4
Ground out	= 5
Fielder's choice	= 6
Sacrifice	= 7
Sacrifice Fly	= 8
Reached on error	= 9
Walk	= 10
Infield Single	= 11
Single	= 12
Double	= 13
Triple	= 14
Home run	= 15

Infield singles were originally included in the analysis, but since there was only one in the data, it was recoded to a single. Intentional walks and batters hit by pitches (HBP) were coded as walks, as there is nothing different from a walk that occurs when an intentional walk or HBP occurs. There were no triple plays so they were not assigned a code.

The coding for the various ways to advance runners proceeded according to what was accomplished during the player's at bat, and with an eye toward what other opportunities were left for the next batter. The middle column of the following table does not include the player who just batted.

CURRENTLY ON BASE	ON BASE AFTER AT BAT*	CODE
NO ONE	NO ADVANCE	0
1	2,3	1
2	3	1
12	13,23	2
13	23	2
1,2,3	H	3
12,13	1H,2H,3H	4
23	2H,3H	4
123	12H	4
12,13,23	HH	5
123	13H, 23H, 1HH, 2HH, 3HH	5
123	HHH	6

* - Not including the batter

Incorporating that player's event at bat into the analysis would result in the dependent variable not being mutually exclusive of the factors designed to describe it and the results would reflect the confusion that would be inherent in the data. Making the event at bat the dependent

variable is a portion of a three part plan that will help show which players produce when it is needed most.

RESULTS

The all facet summary page of the FACETS output (Table 1) can be used in a diagnostic manner or for predicting the outcome of the next at-bat. The group letters represent several players. Because of the quantity of names included in the analysis, the names were grouped, lettered, and then placed on the map where the individuals were originally located. This was done so that the map would fit on one page.

All Facet Summary

Table 1 shows an overall picture of the measurement situation, allowing one to see how the facets relate to one another. Each facet is quantified in the same unit of measurement as the others, so each column in Table 1 is comparable to the others in terms of the relative vertical positions of their elements. For instance, the innings facet,

having the lowest overall average scale value, thus makes the least overall contribution to productivity. Although any baseball fan would rightly guess that outs would be made (obviously), runs scored, and advances made in any given inning of a series of games (as is shown by the fact that the elements of the outs, difference, and event at bat facets have higher calibrations than the innings do), Table 1 also shows that the players most likely to produce those actions are those with measures above -1.0. The most able and productive players are at the top of Table 1 in the leftmost column, and the most productive events at bat are also at the top of the table further to the right.

Table 1 shows that, removing the inordinate influences of the inning, outs, and lead/deficit, a player such as Bob Geren or Randy Milligan has about a 50-50 chance of advancing runners by means of a sacrifice; the odds of their event at bat being a sacrifice fly or a home run are lower than 50-50 and the odds of any event below their measure (about .75) occurring increases the further down the page one looks. Thus, foul outs, strike outs, double plays, and fly outs are both the least productive and most often occurring events at

bat, which virtually anyone might have guessed; what virtually no one would have been able to say is how much less productive and more likely these events generally are after removing the uncontrolled effects of inning, number of outs, and the lead or deficit situation.

The event at bat column in Table 1 shows SF (sacrifice fly) and S (sacrifice) being the easiest and third most productive ways to advance runners. The data does not show failed sacrifice flies or sacrifices; what is more, one can conceive of positively and negatively failing sacrifices, neither of which are included in traditional baseball scoring. If a negative sacrifice failure occurs when a batter grounds or flies out and the runners do not advance, a positive failure occurs when no out occurs and the batter gets on base. Hence, anytime a sacrifice shows up in the data it is successful. Negative sacrifice failures disguise themselves as a fly out, a fielder's choice, a strikeout, a foul out or perhaps even a double play. Positive sacrifice failures appear in the record as hits or errors. In order to better measure the event's usefulness in the game failed sacrifice attempts must be counted as well as successful.

Currently available statistics are not kept on failed attempts. Having these data would lead to a decrease in the number of events which disguise the unsuccessful S and SF attempts.

Player Measures

The players measurement report (Table 2) provides, in addition to the basic measure values shown in Table 1, error (reliability) and data quality (validity) statistics on the players in this analysis. Starting on the left side of the table, the columns contain the following information:

Score - The sum of all the ratings the player received for each time at bat.

Count - The number of times the player appeared at the plate.

Average - The player's average rating for each at bat.

Calib logit - This is the player's calibration (ability) measured in logits (log odds unit)

Model error - The model error as estimated for each player

Infit mnsq - The information-weighted mean square fit statistic - a kind of data quality index

Std - A standardized version of the infit

Outfit mnsq - The outlier-sensitive mean square fit statistic - another data quality index

Std - A standardized outfit

Num - Number each player was assigned in order to be measured.

The logit calibration is the measure of player offensive ability that should replace the batting average. Each measure is reliable with its range of error. The measure indicates how much more or less able a player is than an inning, out, lead/deficit situation, or event at bat is difficult. The fit statistics enable the analyst to detect low quality data, and to re-examine the data for entry or coding errors, or for an extraordinarily fine or poor performance, all of which enhances the validity of the measurement system.

In Table 2's average column, it would seem that some players are out of order; some have a higher average raw rating than other players who have the same or lower logit

calibrations. An example of this is Candy Maldonado, who had an advance average of 1.3 and right above him is Kent Hrbek who averaged .5. The players have similar measures (.13) but yet Maldonado has a .8 advantage over Hrbek on average. This shows that Hrbek advanced runners in more difficult situations than did Maldonado. What Maldonado did was advance runners when it was easier to do so.

The overall separation and reliability statistics for these measures (shown in Table 2) are somewhat low. The overall results of the analysis are meaningful enough, and similar enough to the results of other FACETS analyses, though, to assert with confidence that the problem is simply one of sufficient statistical power. The players in this analysis are all major league players, and a relatively small number of players and games were analyzed. By including players of various levels of ability (high school, college, the minors) these two statistics would increase to levels more indicative of the utility of this mode of analysis.

Player Standardized Residuals

Table 3 shows that only one player has a negative standardized residual (SR). Randy Milligan has a -2 SR and a -3.4 residual for his at bat in which he hit a single in the third inning with one out and his team was up by one run. He has the second highest overall ability of all players in this analysis. He also has the only expected score above 2.0 for a single at-bat. Given his ability, and the fact that this occurred in the game's second easiest inning, during its easiest out, and when the lead was most favorable (+1), it stands to reason that the model projects him at a higher expected score. No one was on base for him to advance; therefore he couldn't have had a higher rating. This situation is where having Milligan at bat with runners on base could be very advantageous. There were 32 other at-bats in this analysis that met these exact criteria. Of those, 25% (8) were successful in advancing runners and only two batters drove in any runs.

No player performed inordinately well or poor enough to earn more than five mentions in this table. Two players, Chuck Knoblauch and Mike Pagliarulo, both of Minnesota, had five misfitting at bats, which are about 6% of Knoblauch's and 9% of Pagliarulo's at bats. They

were both unexpectedly productive in situations involving the latter two thirds of the game when their team was either behind or tied with the opposition. Seven players had four misfitting at bats, and the rest of the players had three or less. All of the misfitting ratings except for one (Milligan) reflect performances that were better than expected.

Outs Calibrations

A first glance at the outs measurement report in Table 4 makes one wonder whether the analysis makes any sense; is it really easier to advance runners with one or two outs than it is with none out? With a little thought, though, it becomes apparent that it should be easier for runners to score when there are some outs because outs occur only after some batters have been to the plate; the inning is further along, so there is greater likelihood of there being base runners, and these are more likely to be closer to home than when they got on base in the first place. It is commonplace to sacrifice an out for the sole purpose of advancing a runner into scoring position (i.e. to second or third base) where a single can bring in the run. A runner scoring from first base on anything other than a triple or a home run is not easy. A long double can do it for most base runners and almost any double can do it for the fastest players.

The substantive value of these considerations does not amount to much, however. The calibration values for the difficulty of playing with 0, 1, and 2 outs are very near one another; even though each condition is about two errors of measurement away from the adjacent

conditions, the error is shrunk by the large number of player, inning, lead/deficit, and event at bat interactions taking place with each number of outs. The overall impact of this facet on the measures is probably negligible.

Event at Bat Calibrations

Table 5 shows again that sacrifice flies and sacrifices are among the events at bat most successful at advancing runners. Double play (DP), strikeout (K) and foul out (Foul) have no calibrations because the data coding hinges on productivity and no base advancement occurs in association with these events. Not one player made an advance when the batter hit into a double play, struck out or fouled out. However, that is not to say that it is impossible to do. If a player hit a long foul ball that was caught for an out any base runner can tag up and advance to the next base as long as it is not occupied. This does not happen often since most ball parks do not have enough room along the sidelines for long foul balls to be caught and still leave time for a player to advance safely.

In the event of a double play, runners could advance if the situation allowed. For a runner to advance on a double play, though, there must be no outs and more than one runner on base. Since making two outs at once often causes the inning to end, advances are not common in this situation.

It is impossible to advance as a result of an ordinary strikeout. However, it is possible to advance as a result of a dropped third strike in which case the batter (when first base is unoccupied) and the runners can advance as far as the circumstances allow. But this is a relatively rare event, and did not occur in any of the games included in the present data.

DISCUSSION

Objective measurement can improve the understanding, management, and enjoyment of baseball in at least seven ways.

1. Basis for comparing players across playing levels - amateur through professional

Ned Colletti, the former Media Relations Director and the current Vice President, Baseball Administration for the Chicago Cubs, said that the one thing Major League Baseball is always searching for is a way to better predict minor league players' potential for success at the Major League level. Refined measurement techniques are exactly what is needed to make such prediction possible. Although data on players from various points in their careers when they played at different levels is not yet available, the application can be briefly described.

As has already been pointed out, the FACETS computer program adjusts ability measures according to the amount of influence exerted on them by various factors. In this instance, level of play (high school, college, A, AA, AAA, MLB) is the dependent variable, and the player's performance and the factors (inning, score, umpire, skies, temperature, day/night etc.) under which those games/at-bats were performed are the independent variables. A database on all professional ball players' career histories, from high school through college, the minors and the majors, containing ability measures adjusted for uncontrolled variation in all

of the independent variables, could be used by professional teams to evaluate a high school or college player's performance. Given the availability of this relatively pure ability measure, which has at least the nine or ten most influential facets factored out of it, the possibility of scientifically testing the hypothesis that a player should be drafted arises.

2. Restructuring of free agent compensation scale

The slotting of free agents as A, B, or C level players is determined by an ordinal ranking system that does not say anything about amounts of ability. All players are ordered by count from first on down in five offensive categories over two seasons: plate appearances, batting average, on base percentage, home runs and runs batted in. The top 30% of free agents are then dubbed type A, the next 30% were type B, and the rest type C. This system was devised to determine compensation for any team who lost a free agent to another team. If the player was a type A, his new team must compensate the his old team with a first round draft pick.

This ranking system results in a situation in which the playing abilities of the free agents might fluctuate dramatically from season to season, with no associated change in their negotiating position; A players one year might be B the next, or vice versa. Furthermore, the abilities of the top 15% of the players within the A slot, for instance, might be markedly greater than those in the bottom 15% of the A slot, and the bottom 15% may not be statistically different from the B players. These and the myriad of other inordinate kinds of variation that can occur in the use of ordinal rankings indicate that teams and players may not receive just compensation from the current method of dealing with free agents; a system based on measured amounts of ability adjusted for uncontrolled variation in playing situations would provide the basis for a fairer system.

3. Provide legitimate pay-for-play scale

The idea of pay for play has been around for a few years because many times players sign contracts for large amounts of money only to fail to perform as expected, a factor that also bears on the preceding point concerning the free agent compensation scale. The Chicago White Sox recently tried to implement a pay for play scale (PFPS) but the players rejected it because they feel it will deny them the larger amounts of money that a straight contract will pay. If the player performed to the pay for play contract's specifications he would earn approximately the same amount as with a straight contract. But since a consistently great offensive performance in baseball is virtually impossible to guarantee, there are very few players who would accept a PFPS.

There is, however, the fact that the major league ball clubs are experiencing a continuing problem of spiraling salaries and shrinking revenue. The salaries have increased dramatically in the past five years. Television contract revenue also increased immensely in that same time. Now, however, television revenue growth has ceased and is now decreasing.

Baseball franchise owners will soon start to see a decline in the amount of money they receive from TV contracts. The owners' desire to have a championship ball club leads them to spend the millions of dollars on the players and a PFPS is one way to reel in the multimillion dollar salaries and pay only for winners, not losers.

But in addition to helping control costs, a PFPS would impact players in many ways. Many of today's players who are in their 'free agent year', the last year of their contract after which they can try for a multi-year, multimillion dollar contract, forget about the team concept and play for themselves. These players try to hit homers and drive in runs in order to pad their own statistics, which in turn theoretically leads to more money. With a PFPS in place, the carrot on the player's stick is a measure of performance based on prior years' and adjusted for the conditions that the player finds himself in. If a player performs at a predetermined level, as adjusted for the inordinate influence of umpires, weather, number of outs, etc., he would be compensated accordingly. All players would have a base salary which

would still leave them quite comfortable and all that is done during the season would be a bonus.

4. Compare today's players with players of years ago

The arguments are made everyday that Babe Ruth played during the dead ball era so he is the greatest of all time or that Henry Aaron hit the most home runs against much better competition, so he is the greatest of all time. The arguments will continue until a sound measurement system brings those discussions to an end by providing a standard basis of comparison that everyone understands and agrees on. When was the easiest time to be a 'great' player: the 20s and 30s? The mid 40s to the late 50s? Or the middle 60s to the mid 70s? Or are today's players head and shoulders above all the others who ever played the game?

Objective ability measures could settle those arguments and many more. The independent variable for this analysis would be the year. The FACETS output would reveal

which years were the easiest and most difficult in which to perform well. If a player performed extraordinarily well during a year in which it was very difficult to do so, would not that player be perceived as a great player? Also, if most years during that player's career were difficult years and the player continued to play well, would not he have to be considered among the greatest ever?

Being able to adjust players' measures according to the difficulty of the competition and the variability of the playing conditions is necessary for us to compare the players of past years with each other and with current players. The fact that most of these players will never have played against one another will not matter as long as players who played in the same the years and under the same conditions provide the necessary links in the database.

5. Adjusting player measures according to the leniency or severity of the scout rating

Another major motivation for writing this thesis is the need to evaluate the ratings that professional sports scouts make of players. This measurement situation is the one closest in structure to the one involving judge-awarded ratings that provoked the development of the

FACETS computer program. Scouts are likely to be biased in one way or another just by the sheer nature of their own athleticism or by the needs of the organization. The only type of scouting that is comparable across the board is that which is done with a quantitative measuring device, as can be the case with running, jumping, weight lifting, etc.

But when a baseball scout says that athlete A has great range and exceptional quickness, what is he comparing this players' performance against? What does he have as a basis of comparison but his own observations and impressions as these have built up over the years? Another scout observing athlete A may say his range is only good, and his quickness is only average. But, again, what is the basis for saying this? Nothing but *this* scout's history of observations made during the tenure of his job. The point is not to discredit the work of scouts, but to provide an additional tool to be used when analyzing their observations.

The question is how lenient or severe a scout is as a judge of baseball talent, skill and ability. The answer to this question requires that her ratings of various athletes be consistent over time. In other words, is this scout going to say that athlete A possesses great quickness

on one day, then the next week see athlete B and say that B is of average quickness when he is of a quickness similar to A's? As long as a scout's ratings are consistent, measurement of player abilities can be adjusted to account for differences in the severity or leniency of the scout's ratings.

The problem is that reliable ratings from different judges are usually defined in terms of agreement, even though consistency provides the better criterion. That is, consistency is usually ignored in the determination of reliability, because people think they want raters to agree on the rating that should be assigned for a specific level of performance. This is almost always an impossible goal to strive for; but even though raters will not always assign the same rating for the same performance, they are usually very consistent in assigning higher ratings for better performances, and lower ratings for worse performances. This consistency is far more important to reliable measurement than agreement is, especially in light of the possibility of adjusting measures according to the severity of the rater (Linacre, 1989; Myford, 1989).

6. Team comparison; past and present

There are several ways by which historic teams might be compared; further research is needed to evaluate the pros and cons of each. The issue is too complex to address adequately here, but, given the availability of data, will be raised in the future.

7. Assist with on-field management decisions

The FACETS offensive player analysis makes it possible for coaches and managers to outline more precise game plans than those currently in use. As unexpected situations arise in a game, the manager can consult the computer or a table of past performance to see who would best fit the situation at hand. Batters who have histories of unexpectedly high performance in clutch situations could be identified and put in play at the specific time they are needed. Knoblauch's and Pagliarulo's unexpected performances (shown in Table 3) make them good candidates for batting order substitution in late game situations with a man on base and the Twins are behind.

A basic flaw with virtually every system intended to measure baseball player abilities is the incapacity to credit players who sacrifice their turn at bat for the good of the team. Because of the difficulty of scoring a run from first base unless the next batter hits a triple or a home run, a common strategy employed by managers is to have a batter make an out in such a way as to move the runners already on base into better position to score (i.e. to second and/or third). Conceiving of events at bat as test items allows us to include any event that occurs in a model of playing ability. Sacrifices often make the difference between winning and losing, and can account for up to 20-30 plate appearances for a particular batter in a season. Successful sacrifices, the term used in baseball to describe the event, are a positive occurrence and should be counted for the batter. Currently, Major League Baseball does not know how to handle sacrifices statistically other than as a count, so they are not included in any assessment of the player. Sacrifices are not even mentioned in the 1988 Elias Baseball Analyst, not even for notoriously good bat handlers who undoubtedly had many sacrifices. Although the traditional approach to baseball performance assessment does not

include a conceptual basis for integrating the contribution of sacrifice hits into the player's ability measure, the quantitative measurement approach described in this thesis does.

IMPLICATIONS

Measures of Team Performance

Many of the factors influencing player ability measures also fit well in analyses of team performance. The general tendencies or the conditions at the beginning of the game would be used in order to define the factors which would describe the dependent variable. In this case the dependent variable is the outcome of the game, who won and who lost. How a team performs on artificial turf as opposed to grass could prove interesting. Does an umpire call games unfavorably for a certain team or two? Or, does he favor a team? How do windy conditions affect a game? A team? Do some teams perform well regardless of the temperature and do some falter in the early and late season under cooler conditions? Partial and qualitative answers to these questions can be obtained by traditional means, but the whole story is not coming out. Future research employing the FACETS analysis will make it possible for baseball management, players, and fans alike to follow more closely the details of their favorite team and players.

Scouting: A Tennis Example

Among the many applications of multi-faceted measurement models for sports information analysis, one of the most far reaching is the effect it will have on scouting. Scouting is done on all levels of athletics, from high school to the major leagues. Colleges scout possible recruits and their future opponents. The pros scout the most. Professional baseball scouts all the other teams, the minor leagues, colleges, semi-pro leagues, and high schools. Professional basketball and football scout each other, the minor professional leagues and colleges. Multi-faceted measurement can help by turning the subjective ratings of various scouts into meaningful measures and assessing the scouts' severity or leniency in assigning ratings, but even the more common two-faceted (items and persons) measurement approach is useful.

For example, one person's ratings of professional tennis players were analyzed using the BIGSTEPS computer program (Wright & Linacre, 1992), which tests the quantitative hypothesis for two-faceted measurement models. George Lott, the late Wimbledon and U.S.

Open doubles champion, Hall of Fame member, and DePaul coach, rated the top players in the world for an article that appeared in the sports section of the Chicago Tribune on August 30, 1988 (Jauss, 1988). Lott, with 68 years of experience, playing and observing, was arguably the world authority on the subject of the best ever in tennis. He rated the top 20 men of all time on ten items on a scale of 0-10.

The BIGSTEPS analysis was done to reveal the story behind the ratings, not to undo or challenge Lott's rating of the players. Lott uses the ratings he assigned simply to produce an ordinal ranking of the players. But if the ratings are of a scientific quality high enough to confirm the quantitative hypothesis, Lott's ratings will say how much better or worse than each other these players are. This is where objective measurement can offer a much more detailed exposition of the information contained in ratings than a mere ranking can.

The BIGSTEPS program output on the ratings, measures, errors, and fit statistics is in Table 6. The count column shows the sum of Lott's ratings, the test column shows the number of items each player was rated on, and the measure column provides the measure of

ability. The error column is a basic guide for interpreting what amount of difference in the measures means anything. The smaller the error, the more accurate the measure. The measures range from -.36 to 2.81, across almost 3.2 units, with an average error of about .40, meaning that there are about eight ranges of error encompassed by the players' abilities. If we conservatively define a statistically significant difference as at least three error ranges (Wright & Masters, 1982), then Lott's ratings of the tennis players has produced a little more than two statistically distinct groups. Given the very small number of persons and items rated, this is a good beginning towards a system for measuring tennis player abilities.

But that is not all of the information to be gained from this analysis. The outfit and infit columns in Table 6 tell some interesting stories. A high positive infit or outfit reflects unexpected performance. Given the ratings on the other items, and the consistent ordering of those items across the players, a high fit statistic indicates noise in the data that is associated with the presence of ratings that are either higher or lower than expected. For instance, Henri Cochet (at the bottom of Table 6) has standardized outfit and infit statistics nearing 2.0 (two

standard deviations above the mean). His ratings on difficult items (overhead, volley, and anticipation/quickness) were better than expected given the low ratings he received on easier items (forehand and first serve).

The fit statistics enable one to make distinctions between players with the same measures; Ellsworth Vines, Ivan Lendl, Lew Hoad and Rod Laver all scored 25 and were tied for third best, but their fit statistics range from Laver's low -1.6 to Hoad and Lendl's -.8 and -.7 to Vines' 0.0. Laver's data is most predictable, almost too predictable, with the ratings progressing steadily from high to low as the difficulty of the items changes from easy to hard. The other three players' ratings have more of the bits of noise that are expected to occur at any transition across the items from one rating category to another. It may be that Laver has a steadier, more consistent and controlled game than the other players, a distinction that may be important to make when comparing him to the other players with the same ability measure.

Further examination of the data reveals that Jack Kramer, who has the highest fit statistics in Table 6, had a remarkable second serve, according to Coach Lott. Kramer was

the only player besides John McEnroe to score a 10 on the quality of his second serve. Kramer misfit because of this 10. Given the difficulty of the second serve and Kramer's lower scores on easier items (backhand, baseline, and anticipation/quickness), his high rating for second serve is surprising; the fit statistic might be indicating a special strength of Kramer's, or it might be detecting an anomaly in Lott's ratings. More ratings of the same players would enable application of a facets model to the data and an evaluation of the consistency of Lott's ratings relative to other experts.

Table 7 shows the calibration values for the criteria Lott applied in rating the tennis players. There is not much difference between the highest and lowest calibrations, and the error is high relative to that range, so the differences in the players' abilities do not separate the items out into distinct levels of skill. Because the players rated are the best of the best, they are strong in all areas, so their ratings vary across the available categories hardly at all. Players who are less consistent and well-rounded would rate very low in their weak spots and much higher on their strengths; this sort of variation would open up the number of distinctions

made among these criteria for judging tennis skill. These considerations constitute a hypothesis that must be tested against more data.

One item on the scale that misfit, volley, is of special interest. This item has the second highest calibration, meaning that it is among the most difficult tests of tennis skill and players can expect to receive some of their lowest ratings on this item. Two players, however, received inordinately low ratings on it. Bjorn Borg scored a five on this item, and Bill Tilden scored a seven. Borg was the only player rated to receive a five on any item, meaning that this is either a very serious weakness for a player in the top 20 to have, or Lott's rating might be just a little off. Similarly, one has to wonder how Tilden could be the player with the second highest measure and still be assigned a rating of 7 on any item.

Football Scouting

In an August 30, 1988 article about the Chicago area's top high school football players (Bell, 1988), Chicago Sun-Times writer Taylor Bell discussed the ratings of the 30 best prospects as evaluated by recruiters on eight items. On a scale of 1 to 5, the players were rated on their size, speed, drive, agility, grades, attitude, strength/reaction time, and potential.

Once again, the utility of this approach reveals itself even with a small amount of data. Table 8 shows the items the players were rated on, in calibration order. Size, agility, and potential are the areas the recruiters assigned the highest ratings for, and strength/reaction time, grades, and speed are the areas they assigned lower ratings for. Two items, size and grades, have high fit statistics, meaning that the players' ratings on these items do not vary consistently with their ratings on the other items; a player can be large or small, or have good grades or not, independent of his ratings on the other items. Potential, however, with its low negative fit statistics, is highly dependent on the other items' ratings; it varies in a rigid lock

step with the ratings on the other items since its purpose seems to be to summarize the player's overall ability.

The summary statistics for this data set are acceptable (item separation = 1.76, reliability = .76, person separation = .54, reliability = .23), considering that only the top 30 high school players from the Chicago area were rated. Many teams have as many as 50 players and there are probably 100 high schools in the area. So out of approximately 5000 football players, only the top 30 were evaluated. The separation and reliability of the measures would probably go up considerably with the addition of only the next 100 players. Item separation and reliability would probably also rise, but the validity of the instrument as a whole is compromised by the uncontrolled variation introduced by size and grades; furthermore, the strength/reaction time item was flawed in its design because the recruiters were asked to rate linemen for strength, and other positions, like defensive backs, for reaction time. This ambiguity in the item makes it impossible to generalize its ratings across the players. The best scenario for accurate evaluation would be to split the strength/reaction time

item in two, and then apply a facets model in which a players' position, size, and grades are controlled and separately calibrated.

CONCLUSION

This thesis presents a demonstration of how tests of the quantitative hypothesis will prove to be useful to persons who must apply sports information and evaluation in their day to day decisions. College recruiters, professional teams of all sports can use the information this type of analysis provides to build the winning program that fans, players, alumni, and investors want to see. Those who continue to play with numbers that do not really represent quantitative amounts of more and less ability and skill do so at their own risk.

TABLES

Table 1. All Facet Summary.

Measr	Players			Inning	Outs	Differenc	Event at bat	Cat
2	MORE PRODUCTIVE PLAYERS TOWARD TOP						MOST PRODUCTIVE INNINGS, OUT AND LEAD/DEFICIT SITUATIONS, & EVENTS TOWARD TOP SF HR	(6) 5
1	B.J. Surhoff Bob Geren Alvin Davis Mike Aldrete Group A Group B Group C	Randy Milligan Bobby Rose	Brook Jacoby				Sac T D E S	 ---
0	Group D Group E Group F Group G Max Venable Group H Andy Allanson Mike Huff	Mike Devereaux Dante Bichette		1 2 3 5 8 9 4 6 7	1 2 0	+1 +2 -2 -1 0	 FC BB GO	4 --- 3 --- 2
-1	Mike Greenwell Randy Bush							--- 1
-2	LEAST PRODUCTIVE PLAYERS TOWARD BOTTOM						LEAST PRODUCTIVE FACETS TOWARD BOTTOM	
-3							FO	
-4	A. Powell Jody Reed	B. Williams Pat Sheridan	Bob Melvin				DP Foul K	(0)

| Measr | Players | | | Inning | Outs | Differenc | Event at bat | Cat |

Metric maintained by + or |. Spacing expanded to print all elements.

Table 2. Players Measurement Report (ordered by mN).

Score	Count	Average	Calib Logit	Model Error	Infit MnSq	Infit Std	Outfit MnSq	Outfit Std	Num	Players
18	18	1.0	0.71	0.21	2.3	2	2.2	1	166	B.J. Surhoff
30	26	1.2	0.62	0.17	1.4	1	1.1	0	50	Randy Milligan
13	13	1.0	0.59	0.24	0.5	-1	0.4	0	43	Bob Geren
12	8	1.5	0.53	0.31	0.3	-1	0.2	0	60	Bobby Rose
29	25	1.2	0.50	0.18	0.7	-1	0.5	0	13	Alvin Davis
9	8	1.1	0.45	0.27	1.0	0	0.6	0	201	Mike Aldrete
32	37	0.9	0.44	0.15	1.0	0	0.5	0	203	Brook Jacoby
10	16	0.6	0.40	0.24	0.9	0	1.7	0	96	Junior Ortiz
20	23	0.9	0.39	0.18	1.9	2	2.6	1	171	Jim Gantner
8	10	0.8	0.38	0.27	1.0	0	0.6	0	138	Carlos Quintana
17	20	0.9	0.37	0.19	0.6	-1	0.8	0	176	Darryl Hamilton
36	33	1.1	0.35	0.14	0.9	0	0.6	0	115	Dave Henderson
57	72	0.8	0.32	0.11	1.0	0	2.4	1	27	Chili Davis
17	22	0.8	0.32	0.19	1.2	0	1.2	0	35	Mel Hall
4	10	0.4	0.32	0.36	0.4	0	0.2	0	91	Kurt Stillwell
8	14	0.6	0.31	0.26	0.7	0	0.3	0	47	Dwight Evans
22	30	0.7	0.30	0.21	0.5	-1	0.3	0	155	Cecil Fielder
7	10	0.7	0.29	0.28	1.6	1	1.0	0	135	Jack Clark
34	45	0.8	0.27	0.13	1.4	1	4.6	2	22	Donnie Hill
12	21	0.6	0.26	0.26	0.7	0	0.4	0	112	Walt Weiss
8	18	0.4	0.26	0.36	1.4	0	5.7	1	173	Bill Spiers
8	13	0.6	0.25	0.26	0.8	0	0.4	0	16	Tracy Jones
43	51	0.8	0.25	0.12	1.2	1	0.7	0	42	Dave Gallagher
17	27	0.6	0.24	0.19	1.5	1	1.1	0	108	Mike Gallego
19	22	0.9	0.23	0.19	1.1	0	2.7	1	49	Joe Orsulak
7	10	0.7	0.23	0.27	1.5	1	0.9	0	136	Tony Pena
4	12	0.3	0.23	0.37	0.6	0	0.2	0	149	Tom Brunansky
4	11	0.4	0.23	0.37	0.7	0	0.5	0	163	John Shelby
43	68	0.6	0.22	0.12	0.9	0	1.9	1	208	Carlos Baerga
9	17	0.5	0.21	0.31	0.6	0	0.2	0	37	Jesse Barfield
20	32	0.6	0.20	0.18	0.8	0	0.6	0	15	Pete O'Brien
18	34	0.5	0.20	0.18	1.2	0	0.8	0	48	Cal Ripken
7	16	0.4	0.20	0.28	0.4	-1	0.3	0	114	Willie Wilson
10	17	0.6	0.20	0.23	0.8	0	0.8	0	178	Robin Yount
12	25	0.5	0.20	0.21	0.9	0	3.1	1	214	Turner Ward
22	21	1.0	0.20	0.17	0.8	0	0.5	0	231	C. Martinez (cle)
55	96	0.6	0.18	0.11	1.1	0	0.5	0	7	Gary Gaetti
35	68	0.5	0.18	0.13	1.2	0	0.5	0	24	Dan Gladden
11	23	0.5	0.17	0.22	1.2	0	3.0	1	61	Ron Tingley
41	59	0.7	0.16	0.14	0.9	0	0.5	0	213	Albert Belle

TABLE CONTINUED ON NEXT PAGE

TABLE 2 (CONTINUED FROM PREVIOUS PAGE)

Score	Count	Average	Calib Logit	Model Error	Infit MnSq	Infit Std	Outfit MnSq	Outfit Std	Num	Players
30	48	0.6	0.15	0.14	1.0	0	0.6	0	6	Lance Parrish
12	20	0.6	0.15	0.23	0.7	0	0.6	0	20	Greg Briley
29	27	1.1	0.15	0.16	1.0	0	0.7	0	105	Terry Steinbach
15	21	0.7	0.15	0.21	0.9	0	0.4	0	109	Ernie Riles
8	12	0.7	0.15	0.26	0.5	0	0.5	0	168	Franklin Stubbs
64	93	0.7	0.13	0.09	1.1	0	1.0	0	1	Luis Polonia
23	33	0.7	0.13	0.17	1.1	0	1.7	1	8	Junior Felix
40	77	0.5	0.13	0.13	1.0	0	0.9	0	30	Kent Hrbek
12	9	1.3	0.13	0.27	0.8	0	0.5	0	222	Candy Maldonado
8	12	0.7	0.12	0.25	0.7	0	0.5	0	106	Jamie Quirk
12	22	0.5	0.11	0.21	1.1	0	0.8	0	170	Willie Randolph
21	29	0.7	0.11	0.17	1.0	0	0.5	0	206	Mark Lewis
10	22	0.5	0.10	0.26	1.4	0	0.5	0	164	Rob Deer
66	99	0.7	0.09	0.10	1.1	0	0.6	0	3	Wally Joyner
24	32	0.8	0.09	0.19	0.7	0	9.0	4	116	Jose Canseco
7	20	0.3	0.09	0.28	0.8	0	0.6	0	160	Lloyd Moseby
5	12	0.4	0.08	0.31	1.5	0	1.1	0	147	Ellis Burks
19	29	0.7	0.08	0.18	0.7	0	0.4	0	158	Travis Fryman
11	23	0.5	0.08	0.23	1.5	1	0.8	0	175	Greg Vaugh
46	79	0.6	0.07	0.11	0.9	0	1.6	0	5	Dave Parker
4	12	0.3	0.06	0.36	0.8	0	0.9	0	38	Randy Velarde
14	33	0.4	0.06	0.19	0.8	0	0.6	0	104	Harold Baines
18	28	0.6	0.06	0.22	0.7	0	0.4	0	153	Mickey Tettleton
17	30	0.6	0.06	0.21	0.6	0	0.4	0	207	Jerry Browne
33	62	0.5	0.05	0.13	1.0	0	0.5	0	2	Luis Sojo
37	71	0.5	0.05	0.13	0.9	0	2.3	1	28	Brian Harper
12	32	0.4	0.05	0.21	1.0	0	0.6	0	156	Lou Witaker
5	13	0.4	0.04	0.32	0.6	0	0.6	0	142	Wade Boggs
19	35	0.5	0.03	0.17	1.0	0	0.5	0	33	Don Mattingly
6	15	0.4	0.03	0.29	1.1	0	0.6	0	64	David Segui
4	13	0.3	0.03	0.37	0.6	0	0.4	0	196	Mark Whiten
24	58	0.4	0.03	0.15	1.2	0	0.5	0	200	Chris James
61	94	0.6	0.02	0.10	1.0	0	1.3	0	4	Dave Winfield
16	18	0.9	0.01	0.24	0.8	0	0.8	0	36	Matt Nokes
12	36	0.3	0.01	0.21	0.8	0	0.3	0	215	Sandy Alomar Jr
34	84	0.4	0.00	0.12	1.2	0	0.6	0	25	Chuck Knoblauch
5	11	0.5	-0.01	0.31	0.6	0	0.3	0	14	Jay Buhner
27	63	0.4	-0.01	0.15	1.0	0	0.6	0	209	Felix Fermin
16	20	0.8	-0.04	0.25	0.6	0	0.3	0	226	L. Gomez

TABLE CONTINUED ON NEXT PAGE

TABLE 2 (CONTINUED FROM PREVIOUS PAGE)

Score	Count	Average	Calib Logit	Model Error	Infit MnSq	Infit Std	Outfit MnSq	Outfit Std	Num	Players
41	83	0.5	-0.05	0.12	1.1	0	1.8	1	26	Kirby Puckett
8	29	0.3	-0.05	0.26	0.3	-1	0.3	0	34	Kevin Maas
9	16	0.6	-0.05	0.29	1.1	0	0.5	0	54	Bill Ripken
19	54	0.4	-0.05	0.16	1.3	0	1.1	0	97	Mike Pagliarulo
11	31	0.4	-0.06	0.21	0.9	0	0.4	0	10	Harold Reynolds
10	39	0.3	-0.07	0.24	2.3	1	2.6	1	32	Steve Sax
18	44	0.4	-0.08	0.19	1.1	0	0.6	0	157	Tony Phillips
30	82	0.4	-0.10	0.13	0.8	0	0.7	0	9	Dick Schofield
11	19	0.6	-0.10	0.25	1.0	0	0.3	0	102	Pedro Munoz
5	14	0.4	-0.10	0.32	0.5	0	0.6	0	202	Joel Skinner
18	47	0.4	-0.11	0.17	1.1	0	0.9	0	100	Greg Gagne
4	19	0.2	-0.11	0.36	0.7	0	0.2	0	212	Beau Allred
10	28	0.4	-0.14	0.23	0.7	0	0.4	0	99	Al Newman
11	31	0.4	-0.15	0.25	0.8	0	0.8	0	162	Milt Cuyler
16	59	0.3	-0.16	0.19	1.4	0	1.2	0	29	Shane Mack
5	11	0.5	-0.17	0.38	0.5	0	0.5	0	31	Pat Kelly
12	15	0.8	-0.17	0.24	0.5	0	0.4	0	174	Dale Sveum
6	24	0.3	-0.21	0.30	0.4	0	0.5	0	107	Mark McGuire
10	33	0.3	-0.23	0.23	0.7	0	0.5	0	11	Edgar Martinez
6	19	0.3	-0.24	0.29	1.8	1	0.8	0	40	Hensley Meulens
8	39	0.2	-0.25	0.26	1.8	1	0.7	0	210	Alex Cole
6	30	0.2	-0.31	0.30	0.2	-1	0.1	0	39	Alvaro Espinoza
8	20	0.4	-0.32	0.28	0.6	0	0.4	0	45	Roberto Kelly
4	18	0.2	-0.32	0.40	0.3	-1	0.2	0	159	Akan Trammell
6	30	0.2	-0.35	0.31	1.8	1	4.3	1	113	Rickey Henderson
8	29	0.3	-0.35	0.26	1.0	0	1.9	0	165	Paul Molitor
2	17	0.1	-0.36	0.57	0.6	0	0.6	0	55	Brady Anderson
3	24	0.1	-0.40	0.45	0.5	0	0.2	0	103	Gene Larkin
2	12	0.2	-0.40	0.54	0.3	0	0.1	0	161	Dave Bergman
2	16	0.1	-0.43	0.56	0.4	0	0.4	0	17	Dave Valle
2	17	0.1	-0.50	0.56	1.2	0	1.0	0	62	Max Venable
7	25	0.3	-0.54	0.31	0.5	0	0.3	0	46	Mike Devereaux
3	14	0.2	-0.56	0.44	0.4	0	0.3	0	56	Chris Hoiles
4	23	0.2	-0.57	0.37	2.0	1	0.9	0	18	Omar Vizquel
1	10	0.1	-0.57	0.84	0.6	0	0.6	0	59	Sam Horn
2	10	0.2	-0.58	0.53	0.4	0	0.3	0	86	George Brett
3	22	0.1	-0.63	0.45	0.4	0	0.2	0	98	Scott Leius
3	25	0.1	-0.65	0.46	0.5	0	0.4	0	12	Ken Griffey Jr.
1	11	0.1	-0.74	0.82	0.5	0	0.2	0	154	Andy Allanson

TABLE CONTINUED ON NEXT PAGE

TABLE 2 (CONTINUED FROM PREVIOUS PAGE)

Score	Count	Average	Calib Logit	Model Error	Infit MnSq	Infit Std	Outfit MnSq	Outfit Std	Num Players
2	18	0.1	-0.79	0.58	0.5	0	0.5	0	179 Dante Bichette
3	31	0.1	-0.84	0.48	0.6	0	0.7	0	211 Mike Huff
1	14	0.1	-1.10	0.87	0.7	0	0.4	0	146 Mike Greenwell
1	24	0.0	-1.24	0.92	0.7	0	0.2	0	101 Randy Bush
0	11		Minimum						41 Pat Sheridan
0	16		Minimum						57 Bob Melvin
0	15		Minimum						140 Jody Reed
0	15		Minimum						229 B. Williams
0	1		Minimum						235 A. Powell

Score	Count	Average	Calib Logit	Model Error	Infit MnSq	Infit Std	Outfit MnSq	Outfit Std	Num Players
15.4	29.8	0.5	-0.00	0.27	0.9	0.0	0.9	0.2	Mean of Count:
14.3	21.6	0.3	0.34	0.15	0.4	0.7	1.2	0.7	S.D. 127

RMSE 0.31 Adj S.D. 0.12 Separation 0.39 Reliability 0.13

Table 3. Misfitting ratings.

Cat	Step	Exp.	Resd	StRes	Num	Players	Nu	Inning N	Outs N	Difference in score	Nu	Event at bat
1	1	0.0	1.0	9	5	Dave Parker	1	1	2 1	5 +2	4	Fly out
5	5	1.4	3.6	2	4	Dave Winfield	5	5	3 2	4 +1	12	Single
4	4	0.8	3.2	2	9	Dick Schofield	6	6	2 1	4 +1	12	Single
5	5	1.3	3.7	2	4	Dave Winfield	7	7	2 1	5 +2	13	Double
1	1	0.1	0.9	2	9	Dick Schofield	6	6	1 0	4 +1	5	Ground Out
4	4	0.3	3.7	4	100	Greg Gagne	6	6	1 0	1 -2	12	Single
3	3	0.2	2.8	5	8	Junior Felix	2	2	1 0	4 +1	5	Ground Out
1	1	0.0	1.0	9	22	Donnie Hill	4	4	2 1	5 +2	4	Fly out
3	3	0.2	2.8	4	96	Junior Ortiz	5	5	1 0	1 -2	5	Ground Out
4	4	1.0	3.0	2	25	Chuck Knoblauch	5	5	3 2	1 -2	12	Single
5	5	1.3	3.7	2	7	Gary Gaetti	6	6	3 2	5 +2	12	Single
2	2	0.2	1.8	3	27	Chili Davis	8	8	1 0	1 -2	5	Ground Out
4	4	0.2	3.8	6	30	Kent Hrbek	8	8	2 1	1 -2	5	Ground Out
2	2	0.2	1.8	3	208	Carlos Baerga	9	9	1 0	1 -2	5	Ground Out
1	1	0.1	0.9	2	32	Steve Sax	1	1	1 0	3 0	5	Ground Out
1	1	0.1	0.9	2	36	Matt Nokes	2	2	1 0	3 0	5	Ground Out
4	4	0.4	3.6	4	35	Mel Hall	5	5	2 1	5 +2	5	Ground Out
4	4	0.6	3.4	2	6	Lance Parrish	9	9	2 1	1 -2	6	Fielder's ch
4	4	0.8	3.2	2	1	Luis Polonia	3	3	1 0	3 0	12	Single
4	4	0.4	3.6	3	48	Cal Ripken	1	1	1 0	2 -1	6	Fielder's ch
4	4	0.8	3.2	2	54	Bill Ripken	3	3	3 2	2 -1	12	Single
2	2	0.3	1.7	2	8	Junior Felix	6	6	3 2	3 0	10	Base on balls
5	5	1.1	3.9	2	3	Wally Joyner	9	9	3 2	3 0	12	Single
4	4	0.8	3.2	2	8	Junior Felix	3	3	2 1	5 +2	6	Fielder's ch
2	2	0.3	1.7	2	49	Joe Orsulak	4	4	1 0	1 -2	6	Fielder's ch
4	4	0.2	3.8	6	49	Joe Orsulak	6	6	2 1	1 -2	5	Ground Out
4	4	0.2	3.8	7	61	Ron Tingley	8	8	1 0	5 +2	5	Ground Out
4	4	0.3	3.7	4	27	Chili Davis	1	1	2 1	3 0	5	Ground Out
2	2	0.3	1.7	2	160	Lloyd Moseby	7	7	1 0	5 +2	6	Fielder's ch
2	2	0.3	1.7	2	97	Mike Pagliarulo	2	2	2 1	3 0	10	Base on balls
4	4	0.7	3.3	2	157	Tony Phillips	2	2	3 2	3 0	12	Single
4	4	0.6	3.4	3	156	Lou Witaker	7	7	1 0	5 +2	12	Single
5	5	1.3	3.7	2	24	Dan Gladden	2	2	3 2	3 0	12	Single
5	5	1.4	3.6	2	24	Dan Gladden	4	4	3 2	3 0	13	Double
4	4	0.9	3.1	2	26	Kirby Puckett	4	4	3 2	5 +2	12	Single
3	3	0.5	2.5	2	165	Paul Molitor	6	6	2 1	1 -2	13	Double
3	3	0.4	2.6	2	166	B.J. Surhoff	6	6	2 1	3 0	5	Ground Out
5	5	1.0	4.0	2	24	Dan Gladden	8	8	1 0	3 0	13	Double
4	4	0.1	3.9	9	26	Kirby Puckett	1	1	1 0	3 0	5	Ground Out
4	4	0.4	3.6	3	97	Mike Pagliarulo	4	4	1 0	3 0	12	Single
4	4	0.2	3.8	6	29	Shane Mack	4	4	1 0	4 +1	6	Fielder's ch

TABLE CONTINUED ON NEXT PAGE

TABLE 3 (CONTINUED FROM PREVIOUS PAGE)

Cat	Step	Exp.	Resd	StRes	Num	Players	Nu	Inning	N	Outs	N	Difference in score	Nu	Event at bat
1	1	0.0	1.0	9	173	Bill Spiers	8	8		2	1	3	0	4 Fly out
1	1	0.1	0.9	2	101	Randy Bush	9	9		2	1	3	0	12 Single
4	4	0.4	3.6	4	166	B.J. Surhoff	1	1		1	0	3	0	5 Ground Out
4	4	0.8	3.2	2	166	B.J. Surhoff	2	2		2	1	5	+2	5 Ground Out
3	3	0.5	2.5	2	25	Chuck Knoblauch	9	9		2	1	1	-2	6 Fielder's ch
2	2	0.0	2.0	9	116	Jose Canseco	1	1		1	0	3	0	4 Fly out
2	2	0.4	1.6	2	25	Chuck Knoblauch	8	8		2	1	1	-2	10 Base on balls
4	4	0.1	3.9	9	113	Rickey Henderson	8	8		2	1	1	-2	5 Ground Out
1	1	0.1	0.9	2	97	Mike Pagliarulo	7	7		1	0	2	-1	5 Ground Out
1	1	0.0	1.0	9	208	Carlos Baerga	2	2		2	1	4	+1	4 Fly out
4	4	1.0	3.0	2	208	Carlos Baerga	1	1		1	0	3	0	12 Single
4	4	0.7	3.3	2	203	Brook Jacoby	1	1		1	0	5	+2	10 Base on balls
5	5	1.4	3.6	2	136	Tony Pena	7	7		3	2	1	-2	13 Double
4	4	0.5	3.5	3	147	Ellis Burks	8	8		2	1	1	-2	6 Fielder's ch
2	2	0.3	1.7	2	142	Wade Boggs	3	3		1	0	3	0	6 Fielder's ch
1	1	0.1	0.9	2	210	Alex Cole	6	6		2	1	2	-1	5 Ground Out
1	1	0.1	0.9	2	146	Mike Greenwell	7	7		3	2	5	+2	12 Single
1	1	0.1	0.9	2	17	Dave Valle	2	2		2	1	4	+1	5 Ground Out
4	4	0.1	3.9	8	214	Turner Ward	2	2		1	0	2	-1	5 Ground Out
1	1	0.1	0.9	3	211	Mike Huff	3	3		1	0	2	-1	10 Base on balls
2	2	0.3	1.7	2	15	Pete O'Brien	4	4		1	0	3	0	10 Base on balls
4	4	0.4	3.6	4	18	Omar Vizquel	4	4		2	1	4	+1	12 Single
2	2	0.3	1.7	2	202	Joel Skinner	4	4		2	1	1	-2	10 Base on balls
1	1	0.1	0.9	2	20	Greg Briley	7	7		1	0	5	+2	5 Ground Out
1	1	0.1	0.9	2	209	Felix Fermin	1	1		1	0	2	-1	5 Ground Out
5	5	1.4	3.6	2	200	Chris James	3	3		3	2	4	+1	12 Single
5	5	1.6	3.4	2	208	Carlos Baerga	4	4		3	2	5	+2	12 Single
1	1	0.1	0.9	3	113	Rickey Henderson	9	9		1	0	1	-2	5 Ground Out
1	1	0.1	0.9	2	107	Mark McGuire	8	8		1	0	5	+2	5 Ground Out
1	1	0.1	0.9	2	211	Mike Huff	3	3		2	1	3	0	10 Base on balls
4	4	1.0	3.0	2	106	Jamie Quirk	4	4		3	2	1	-2	12 Single
5	5	1.3	3.7	2	200	Chris James	5	5		3	2	3	0	13 Double
0	0	3.4	-3.4	-2	50	Randy Milligan	3	3		2	1	4	+1	12 Single
4	4	0.6	3.4	3	25	Chuck Knoblauch	5	5		1	0	1	-2	12 Single
2	2	0.4	1.6	2	46	Mike Devereaux	9	9	c	1	0	1	-2	7 Sacrifice
1	1	0.1	0.9	2	55	Brady Anderson	1	1		1	0	1	-2	5 Ground Out
5	5	0.5	4.5	4	26	Kirby Puckett	2	2		2	1	5	+2	6 Fielder's ch
5	5	1.5	3.5	2	30	Kent Hrbek	2	2		3	2	5	+2	12 Single
1	1	0.1	0.9	2	59	Sam Horn	6	6		2	1	1	-2	10 Base on balls
4	4	0.5	3.5	3	50	Randy Milligan	8	8		2	1	1	-2	5 Ground Out

TABLE CONTINUED ON NEXT PAGE

TABLE 3 (CONTINUED FROM PREVIOUS PAGE)

Cat	Step	Exp.	Resd	StRes	Num	Players	Nu Inning N	Outs N	Difference in score	Nu Event at bat
4	4	0.7	3.3	2	64	David Segui	7 7	2 1	2 -1	12 Single
4	4	0.5	3.5	3	97	Mike Pagliarulo	9 9	1 0	2 -1	12 Single
4	4	0.3	3.7	4	29	Shane Mack	4 4	1 0	2 -1	12 Single
5	5	1.1	3.9	2	25	Chuck Knoblauch	4 4	3 2	4 +1	12 Single
2	2	0.3	1.7	2	26	Kirby Puckett	1 1	1 0	3 0	6 Fielder's ch
2	2	0.3	1.7	2	31	Pat Kelly	2 2	2 1	4 +1	10 Base on balls
5	5	0.2	4.8	8	32	Steve Sax	2 2	2 1	4 +1	5 Ground Out
5	5	1.2	3.8	2	33	Don Mattingly	2 2	3 2	5 +2	9 Error
3	3	0.5	2.5	2	45	Roberto Kelly	4 4	3 2	5 +2	12 Single
5	5	0.8	4.2	3	40	Hensley Meulens	7 7	2 1	5 +2	13 Double
5	5	1.1	3.9	2	135	Jack Clark	5 5	1 0	1 -2	12 Single
5	5	1.1	3.9	2	175	Greg Vaugh	1 1	1 0	4 +1	12 Single
5	5	1.4	3.6	2	170	Willie Randolph	1 1	1 0	5 +2	13 Double
2	2	0.2	1.8	3	9	Dick Schofield	3 3	2 1	1 -2	5 Ground Out
5	5	1.4	3.6	2	7	Gary Gaetti	5 5	3 2	1 -2	12 Single
1	1	0.1	0.9	2	179	Dante Bichette	5 5	1 0	5 +2	10 Base on balls
2	2	0.1	1.9	5	165	Paul Molitor	5 5	1 0	5 +2	5 Ground Out
4	4	0.4	3.6	3	171	Jim Gantner	5 5	2 1	5 +2	5 Ground Out
2	2	0.3	1.7	2	9	Dick Schofield	8 8	3 2	1 -2	10 Base on balls
2	2	0.2	1.8	3	62	Max Venable	6 6	2 1	5 +2	10 Base on balls
5	5	1.3	3.7	2	5	Dave Parker	1 1	2 1	3 0	9 Error
2	2	0.1	1.9	4	162	Milt Cuyler	2 2	2 1	1 -2	5 Ground Out
5	5	0.9	4.1	3	1	Luis Polonia	6 6	3 2	1 -2	12 Single
4	4	0.2	3.8	6	1	Luis Polonia	5 5	2 1	1 -2	5 Ground Out
4	4	1.0	3.0	2	5	Dave Parker	8 8	3 2	1 -2	12 Single
2	2	0.3	1.7	2	3	Wally Joyner	1 1	1 0	3 0	10 Base on balls
5	5	0.2	4.8	8	4	Dave Winfield	2 2	2 1	5 +2	5 Ground Out
6	6	2.0	4.0	2	1	Luis Polonia	9 9	3 2	5 +2	13 Double
4	4	0.8	3.2	2	158	Travis Fryman	7 7	3 2	1 -2	12 Single
5	5	0.2	4.8	9	22	Donnie Hill	9 9	1 0	3 0	5 Ground Out
4	4	0.9	3.1	2	157	Tony Phillips	9 9	2 1	2 -1	12 Single
5	5	1.3	3.7	2	164	Rob Deer	9 9	2 1	3 0	12 Single
3	3	0.5	2.5	2	105	Terry Steinbach	6 6	1 0	3 0	12 Single
1	1	0.1	0.9	2	104	Harold Baines	6 6	1 0	4 +1	5 Ground Out
5	5	1.4	3.6	2	209	Felix Fermin	3 3	3 2	1 -2	14 Triple
5	5	0.5	4.5	4	108	Mike Gallego	8 8	2 1	3 0	10 Base on balls
3	3	0.5	2.5	2	11	Edgar Martinez	3 3	1 0	5 +2	12 Single
4	4	0.6	3.4	3	209	Felix Fermin	3 3	1 0	1 -2	12 Single
1	1	0.1	0.9	2	209	Felix Fermin	7 7	2 1	1 -2	5 Ground Out
5	5	1.5	3.5	2	200	Chris James	5 5	2 1	5 +2	12 Single

TABLE CONTINUED ON NEXT PAGE

TABLE 3 (CONTINUED FROM PREVIOUS PAGE)

Cat	Step	Exp.	Resd	StRes	Num	Players	Nu	Inning N	Outs N	Difference in score	Nu	Event at bat
5	5	0.6	4.4	3	210	Alex Cole	2	2	1	-2	12	Single
2	2	0.4	1.6	2	5	Dave Parker	8	8	0	+2	6	Fielder's ch
1	1	0.1	0.9	2	38	Randy Velarde	9	9	0	-2	5	Ground out
2	2	0.1	1.9	4	4	Dave Winfield	1	1	0	0	5	Ground out
5	5	1.3	3.7	2	42	Dave Gallagher	6	6	1	-2	12	Single
4	4	0.5	3.5	3	32	Steve Sax	6	6	0	+2	12	Single
4	4	0.9	3.1	2	42	Dave Gallagher	8	8	0	-2	12	Single
4	4	1.0	3.0	2	22	Donnie Hill	8	8	0	-2	12	Single
5	5	1.4	3.6	2	2	Luis Sojo	2	2	1	-1	13	Double
4	4	0.9	3.1	2	2	Luis Sojo	7	7	2	+2	12	Single
3	3	0.6	2.4	2	36	Matt Nokes	9	9	0	-2	12	Single
2	2	0.3	1.7	2	178	Robin Yount	7	7	0	-2	6	Fielder's ch
5	5	1.2	3.8	2	28	Brian Harper	7	7	2	+2	13	Double
3	3	0.2	2.8	4	171	Jim Gantner	1	1	0	0	5	Ground out
4	4	0.5	3.5	3	171	Jim Gantner	3	3	1	+1	5	Ground out
3	3	0.3	2.7	3	176	Darryl Hamilton	4	4	1	+2	5	Ground out
1	1	0.1	0.9	2	97	Mike Pagliarulo	8	8	0	-2	5	Ground out
1	1	0.0	1.0	9	27	Chili Davis	1	1	0	-1	4	Fly out
1	1	0.1	0.9	3	165	Paul Molitor	7	7	0	+2	5	Ground out
2	2	0.3	1.7	2	170	Willie Randolph	9	9	0	+2	10	Base on balls
2	2	0.3	1.7	2	157	Tony Phillips	2	2	2	0	10	Base on balls
3	3	0.5	2.5	2	100	Greg Gagne	5	5	0	0	12	Single
1	1	0.0	1.0	9	28	Brian Harper	8	8	0	+2	4	Fly out

Cat	Step	Exp.	Resd	StRes	Num	Players	Nu	Inning N	Outs N	Difference in score	Nu	Event at bat
0.5	0.5	0.5	-0.0	-0.0	Mean of Count:							
1.2	1.2	0.8	0.9	1.0	S.D.	3724						

Table 4. Outs Measurement Report (ordered by mN).

Score	Count	Average	Calib Logit	Model Error	Infit MnSq	Infit Std	Outfit MnSq	Outfit Std	N Outs
797	1214	0.7	0.12	0.03	1.1	1	1.3	1	2 1
660	1192	0.6	0.06	0.03	1.0	0	0.4	-3	3 2
499	1318	0.4	-0.18	0.03	0.9	0	1.5	2	1 0
Score	Count	Average	Calib Logit	Model Error	Infit MnSq	Infit Std	Outfit MnSq	Outfit Std	N Outs
652.0	1241.3	0.5	0.00	0.03	1.0	0.0	1.1	0.1	Mean of Count:
121.8	55.0	0.1	0.13	0.00	0.1	0.9	0.5	2.6	S.D. 3

RMSE 0.03 Adj S.D. 0.12 Separation 3.93 Reliability 0.94

Table 5. Event at bat Measurement Report (ordered by mN).

Score	Count	Average	Calib Logit	Model Error	Infit MnSq	Infit Std	Outfit MnSq	Outfit Std	Nu	Event at bat
149	39	3.8	1.31	0.12	0.3	-3	0.4	-2	8	Sac Fly
185	52	3.6	1.16	0.09	0.5	-3	0.5	-2	15	Home Run
30	25	1.2	0.57	0.13	0.2	-4	0.4	-2	7	Sac
33	25	1.3	0.45	0.13	1.2	0	1.0	0	14	Triple
281	213	1.3	0.42	0.04	1.1	1	1.0	0	13	Double
851	823	1.0	0.29	0.03	1.0	0	1.0	0	12	Single
52	53	1.0	0.27	0.10	0.9	0	0.8	0	9	Error
60	128	0.5	-0.12	0.09	1.2	0	1.4	1	6	Fielder's ch
144	418	0.3	-0.23	0.06	0.7	-2	0.7	-1	10	Base on balls
163	941	0.2	-0.68	0.06	1.5	3	1.4	2	5	Ground out
8	1007	0.0	-3.44	0.35	1.2	0	1.2	0	4	Fly out
0	132		Minimum						1	Double play
0	647		Minimum						2	Strikeout
0	85		Minimum						3	Foul out

Score	Count	Average	Calib Logit	Model Error	Infit MnSq	Infit Std	Outfit MnSq	Outfit Std	Nu	Event at bat
139.7	327.7	1.0	-0.00	0.11	0.9	-0.6	0.9	-0.4	Mean of Count:	
214.1	354.7	1.2	1.22	0.08	0.4	2.2	0.3	1.6	S.D.	14

RMSE 0.14 Adj S.D. 1.21 Separation 8.74 Reliability 0.99

Table 6. Tennis Ratings by George Lott - Players in Measure Order ANDRICH (R) MODEL RASCH ANALYSIS VER. 2.04
Oct 26 13:12 1993 INPUT: 20 PERSONS 10 ITEMS ANALYZED: 20 PERSONS 10 ITEMS 6 CATEGORIES

PERSON STATISTICS -- MEASURE ORDER

NUM	COUNT	TEST	MEASURE	ERROR	MNSQ INFIT	MNSQ OUTFT	PTBIS	NAME	M G
14	93	10	2.81	.49	1.26 .7	1.24 .7	-.59	John McEnroe	
19	87	10	1.59	.43	1.50 1.2	1.48 1.1	.37	Bill Tilden	
8	85	10	1.23	.42	.65 -.8	.66 -.8	-.45	Lew Hoad	
11	85	10	1.23	.42	.42 -1.6	.41 -1.6	.16	Rod Laver	
13	85	10	1.23	.42	.69 -.7	.68 -.7	.23	Ivan Lendl	
18	85	10	1.23	.42	.97 .0	.97 .0	-.34	Ellsworth Vines	
2	84	10	1.06	.42	1.04 .2	1.03 .2	.07	Don Budge	
7	84	10	1.06	.42	.51 -1.2	.51 -1.2	-.15	Pancho Gonzales	
20	84	10	1.06	.42	.59 -.9	.61 -.9	.38	Bobby Riggs	
10	83	10	.89	.41	1.86 1.7	1.84 1.7	-.30	Jack Kramer	
17	83	10	.89	.41	1.01 .2	1.01 .2	.53	Fred Perry	
9	81	10	.56	.41	1.50 1.1	1.51 1.1	.29	Bill Johnson	
16	80	10	.39	.40	.50 -1.1	.50 -1.1	.38	John Newcombe	
1	78	10	.08	.39	1.58 1.2	1.54 1.1	.54	Bjorn Borg	
4	78	10	.08	.39	1.25 .6	1.27 .7	.67	Jimmy Connors	
12	78	10	.08	.39	.90 -.1	.91 .0	.41	Rene Lecoste	
15	78	10	.08	.39	.90 .0	.89 -.1	-.08	Illie Nastase	
5	77	10	-.07	.39	.40 -1.4	.41 -1.4	-.10	Roy Emmerson	
6	77	10	-.07	.39	.39 -1.5	.39 -1.5	-.05	Neale Frazier	
3	75	10	-.36	.37	1.81 1.5	1.75 1.4	-.21	Henri Cochet	

Table 7. Tennis Ratings by George Lott - Items in Calibration Order ANDRICH (R) MODEL RASCH ANALYSIS VER. 2.04
Oct 26 13:12 1993 INPUT: 20 PERSONS 10 ITEMS ANALYZED: 20 PERSONS 10 ITEMS 6 CATEGORIES

ITEMS STATISTICS -- CALIBRN ORDER

NUM	COUNT	SAMPLE	CALIBRTN	ERROR	MNSQ	INFIT	MNSQ	OUTFT	PTBIS	NAME	M	G
2	158	20	.49	.28	1.07	.3	1.09	.4	.73	SECOND SERVE		
8	158	20	.49	.28	.70	-.9	.73	-.8	.44	OVERHEAD		
5	160	20	.33	.28	1.64	1.7	1.57	1.6	-.05	VOLLEY		
9	162	20	.17	.29	1.08	.3	1.06	.3	.21	ANTICIPATION/QUICKNESS		
4	163	20	.09	.29	.86	-.3	.84	-.4	.37	BACKHAND		
1	165	20	-.07	.29	1.09	.4	1.09	.4	.65	FIRST SERVE		
7	165	20	-.07	.29	1.00	.1	1.05	.3	-.20	BASELINE		
3	168	20	-.33	.30	.95	-.1	.93	-.1	.35	FOREHAND		
10	169	20	-.42	.30	.57	-1.5	.58	-1.5	.27	TENNIS BRAIN		
6	172	20	-.69	.30	.87	-.3	.87	-.4	.20	MENTAL TOUGHNESS		

Table 8. 1988 High School Football Prospects - Items in Calibration Order ANDRICH (R) MODEL RASCH ANALYSIS
Oct 26 11:27 1993 INPUT: 30 PERSONS 8 ITEMS ANALYZED: 30 PERSONS 8 ITEMS 4 CATEGORIES

ITEMS STATISTICS -- CALIBRATION ORDER

NUM	COUNT	SAMPLE	CALIBRTN	ERROR	MNSQ	INFIT	MNSQ	OUTFT	PTBIS	NAME	M	G
7	99	30	.78	.28	.82	-.7	.79	-.8	.24	STRENGTH/REACTION TIME		
5	102	30	.55	.27	1.89	2.9	1.91	2.9	-.18	GRADES		
2	108	30	.11	.27	1.10	.5	1.09	.4	-.14	SPEED		
3	109	30	.04	.27	.76	-1.0	.75	-1.0	.14	DRIVE		
6	110	30	-.03	.27	.66	-1.6	.66	-1.6	.41	ATTITUDE		
8	110	30	-.03	.27	.53	-2.3	.53	-2.3	.47	POTENTIAL		
4	113	30	-.24	.27	.69	-1.4	.70	-1.4	.11	AGILITY		
1	126	30	-1.19	.28	1.62	2.4	1.59	2.3	-.33	SIZE		

REFERENCES

Andrich D. 1988. Rasch Models for Measurement. Sage Quantitative Applications in the Social Sciences Vol. 68. Newbury Park, CA: Sage University Papers.

Bell, Taylor. 1988. Chicago area's top 30 football prospects for 1988. Chicago Sun Times. August 30.

Fisher AG. 1993. Development of a functional assessment that adjusts ability measures for task simplicity and rater leniency. In M. Wilson (Ed.), Objective measurement: Theory into Practice (vol. 2). Norwood, NJ: Ablex.

Fisher WP, Fisher AG. 1993. Applications of Rasch analysis to studies in occupational therapy. *Physical Medicine and Rehabilitation Clinics of North America: New Developments in Functional Assessment 4:551-569.* C. Granger & G. Gresham, Eds. Philadelphia: W.B. Saunders.

Gimbel M. 1992. Mike Gimbel's Baseball Player & Team Ratings - 1992 Edition. New York: Boerum Street Press.

Guttman L. 1950. The basis for scalogram analysis. In Measurement and Prediction. Ed. S. A. Stouffer et al. New York: John Wiley & Sons.

Holtzman J. 1988. On baseball. Chicago Tribune, May 22.

James B. 1988. The Bill James Baseball Abstract. New York: Ballantine Books.

Jauss, Bill, 1988. After 7 decades, he gives out grades. Chicago Tribune, August 30.

Kurkjian T. 1993. The Big Bang. Sports Illustrated. Vol 79. No. 11(Sept. 13): 32-40.

Linacre JM. 1989. Many-facet Rasch Measurement. Chicago: MESA Press.

Linacre JM. 1988. Facets Rasch Analysis Computer Program. Chicago: MESA Press.

Loevinger J. 1965. Person and population as psychometric concepts. Psychological Review 72(2): 143-155.

Luce RD, Tukey JW. 1964. Simultaneous conjoint measurement: A new kind of fundamental measurement. Journal of Mathematical Psychology 1(1): 1-27.

Lunz ME, Wright BD, Linacre JM. 1990. Measuring the impact of judge severity on examination scores. Applied Measurement in Education 3/4:331-345.

Michell J. 1990. An introduction to the logic of psychological measurement. Hillsdale, NJ: Lawrence Erlbaum.

Myford CM. 1989. The nature of expertise in aesthetic judgement: Beyond inter-judge agreement. (Doctoral dissertation, University of Chicago). *Dissertation Abstracts International*, 50, 3562A.

Palmer P, Thorn J. 1985. The Hidden Game of Baseball: A Revolutionary Approach to Baseball and its Statistics. Garden City, New York: Doubleday.

Palmer P, Thorn J. 1989. Total Baseball. New York: Warner Books, Inc.

Rasch G. 1960. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen: Danmarks Paedogogiske Institut; reprint, with Foreword and Afterword by Benjamin D. Wright, Chicago: University of Chicago Press, 1980.

Siwoff S, Hirdt S, Hirdt P. 1988. The 1988 Elias Baseball Analyst. New York: Macmillan Publishing Company.

Solomon A. 1988. On the Cubs. Chicago Tribune, May 22.

Thurstone LL. 1928. Attitudes can be measured. American Journal of Sociology 33: 529-554. Reprinted in L. L. Thurstone. The Measurement of Values. Chicago: University of Chicago Press, Midway Reprint Series, 1959.

Williams J. 1992. The Weather Book: An Easy-to-Understand Guide to the USA's Weather. New York: Vintage Books.

Wright BD. 1968. Sample-free test calibration and person measurement. In Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton: Educational Testing Service.

Wright BD. 1977. Solving measurement problems with the Rasch model. Journal of Educational Measurement 14(2): 97-116.

Wright BD, Linacre JM. 1991. BIGSTEPS Rasch Analysis Computer Program. Chicago: MESA Press.

Wright BD, Masters G. 1982. Rating Scale Analysis. Chicago: MESA Press.