B.A.S.E.S.

(Baseball Activity Scoring and Evaluation System)

Michael Sussna

B.A.S.E.S.

INTRODUCTION

WHAT IS B.A.S.E.S.?

B.A.S.E.S. is a baseball activity scoring and evaluation system. It provides a way to measure player performance. In that aim to quantify contribution it is similar to many traditional methods and others being developed. Where it differs is in its approach to the problem. That approach will be explained as the discussion progresses.

B.A.S.E.S. is a complete system, containing several complementary components. First, there are the underlying concepts and principles behind the B.A.S.E.S. approach. The key concept is that of looking at the significant activity in a baseball game in terms of bases gained or lost by the offense. Second, there is a method for measuring how much effect a player's activity has on a given play, whether offensively or defensively, when a player does measurably affect play. Third, given how well players have done by situation in batting, baserunning, pitching, and fielding, there is a set of procedures for determining how good that level of performance really is. Techniques are presented for comparing performance and combining performance measures across situations. Various statistics are developed by type of activity. Fourth, there is a set of forms and procedures that can be used to record the data required for B.A.S.E.S. analysis. Fifth and finally, there is a body of guidelines on a file system and the logic needed to store, retrieve, and process all of the data and information appropriate (not presented here).

HOW WAS IT DEVELOPED?

THE PROBLEM WITH EXISTING STATISTICS

During the 1970's, as the system was being developed, there were traditional statistics and some new ones. Although the traditional statistics were interesting, there seemed to be something wrong when one knew that someone was a star yet their league rank in key departments was not stellar, or when someone ranked high but were not really that great. Similarly, why was it so hard to pick the most valuable player most years if there were so many statistics?

The introduction of the new statistics should have helped solve these problems but didn't. They, too, were interesting but just added to the list of partial yardsticks. There was still no comprehensive yet natural way to gauge performance, especially on offense.

What were the traditional statistics? Batting average, home runs, total bases and slugging percentage, runs batted in and runs scored, stolen bases, and so on. What were the new statistics? On base percentage, batting average with runners in scoring position, game winning runs batted in, and so on.

At the same time that B.A.S.E.S. was taking shape, a new generation of approaches was being explored. These other methods, culminating in the Linear Weights system, are amply described in Thorn and Palmer's *The Hidden Game of Baseball* (Doubleday, Garden City, N.Y., 1984). The book also amply critiques traditional statistics as well as more recent ones.

It will be claimed here that although the Linear Weights system is ingenious and the book is generally fascinating and a must for fans, the answer lies elsewhere. Why that is so will better be seen soon, as what might just be that comprehensive yet natural approach we opined about earlier unfolds.

A SOLUTION WAS PERCEIVED

An early handle on the offensive star problem was to add runs scored to runs batted in and divide by at bats. R + RBI roughly reflects both power and speed, an important requirement for any kind of overall offensive performance measure. Significantly, what led to this being looked at was the fact that both Hank Aaron and Willie Mays both continued to do well in these two departments in their later years although their names sank out of the spotlight in the league ranks in batting average. Yet one knew that they were still stars.

A slightly more refined version of R + RBI divided by AB would divide by plate appearances instead. This takes into account that activities where no at bat is charged, such as walks, can contribute to both scoring and driving in runs. And, this statistic is appealing because it is very similar to batting average -- the numerator for a star might top 200 while the denominator is very similar to at bats.

And, since scoring runs is the heart of success on offense, such "run involvement" is very close to a true measure of contribution. In the sense that scoring runs and driving in runs are the "two halves" of runs, this measure is ideal. But runs are not composed of halves. They are composed of quarters, being built base by base. This is pointed out by the fact that many important plays consist of just moving runners up. Also, scoring runs is often passive, the runner scoring merely due to the efforts of the batter or some other party, once they are on base.

So for these reasons, achievement on offense was looked at in terms of bases rather than runs. By doing so, the varied aspects of contribution on offense measured individually by existing statistics are unified with a common denominator. Hits, walks, sacrifices, steals, and any other actions advancing runners can all be seen in terms of bases contributed.

A base hit, which shows up in batting average, total bases and slugging percentage, on base percentage, and home runs when a four-base hit, is counted here in terms of how many bases it contributes to the team's total. Exactly how such counting is done is the subject of a later section entitled "Scoring." Bases on balls, which affect on base percentage, also add to the team's base total. Likewise with sacrifices, sacrifice flies, and stolen bases. Double plays grounded-into figure in as well. Even some activities which fall through the cracks between existing statistics yet which truly affect the team's fortunes are easily seen to fit in here. Examples of these are runner-advancing ground balls and runner-advancing fly outs other than sacrifice flies.

A natural result of this way of looking at activity on offense is a complete system for measuring contribution. In addition, even the relatively good statistics for pitching and fielding, earned run average and fielding percentage respectively, can be enhanced by the B.A.S.E.S. approach.

A DIFFERENT APPROACH THAN LINEAR WEIGHTS'

Granting for the moment that B.A.S.E.S. does all this as claimed, why doesn't Linear Weights have the answer? Why are the two systems so different although both are based on a deep analysis and employ such techniques as normalization? Because Linear Weights works at the run level instead of the base level, works back from all-time totals to individual events, and assigns fixed weights to items like singles, doubles, walks and so on.

The B.A.S.E.S. philosophy is in agreement with the Linear Weights (LWTS) philosophy that runs are the key to the game. Success in a game is of course winning it. Winning a game consists of having scored more runs than the opponent has at the time that the game ends. So, success at the game level depends on scoring as many runs as possible while holding the other team to as few runs as possible, or at least in being able to score runs or keep them from scoring as required.

The problem is that while it is appropriate to measure team success in terms of runs (as well as wins, etc.), it is not appropriate to apply that same approach at the level of the individual player. The only time that a player can be said to have single-handedly "created" a run is when they hit a home run. There is no time when it can be said unequivocably, in other words when there is a measurable contribution, that a defensive player has "prevented" a run. It *is* perfectly appropriate however to speak in terms of a player contributing a number of *bases* through various offensive activities. On defense, either a player makes a play that they are expected to or they commit an error, a negative "contribution." There is no positive measurable contribution possible on defense. Rating fielding plays as to how great they are is not practicable.

LWTS attempts to quantify a player's "runs created" based on all-time totals of teams' runs scored, for and against, and the all-time counts of singles, doubles, walks and so on. Over such a massive statistical survey, the relative proportions of singles to doubles to triples to walks etc., with respect to runs, settle into a stable pattern. Using this foundation, LWTS can say that on the average a single contributes X amount towards a run, a double contributes Y amount, and so on. The problem is that we also know that individual singles are not all equal, nor are all doubles, etc, in actual impact. For example, a single with no one on base is not at all the same in importance as a single with the bases loaded where two runners score and the runner from first goes to third. So, to say that an average single is worth .46 runs is fine, but we can't consider any particular single as average. Some actual singles are worth one base (.25 runs), some are worth two bases (.50 runs), the batter advancing themself and another runner one base each, on up to some singles being worth six bases (1.50 runs) as in the bases loaded example above.

So, LWTS derives weights for the particular types of activities it considers important, such as singles, doubles, walks and so on, based on the average worth of each type of activity. It then applies those weights to the actual counts of singles, doubles, walks and so on that a player accumulates. This then results in the number of runs that the player can be thought of as having "created." Similar concepts are used for pitching and fielding.

B.A.S.E.S. does not work at the run level, does not necessarily seek all-time proportions, nor does it use fixed weights for each activity type. Although LWTS is far more sophisticated and successful than slugging percentage in gauging individual offensive contribution, it still follows in that line of thinking, however distant a descendant it may be, by giving fixed weights to a static set of items. It is in its own way a modern approach to solving the problem of how to count true total bases. B.A.S.E.S. is another such approach. Let us look at the specifics of the B.A.S.E.S. system by beginning with a discussion of "true total bases."

TRUE TOTAL BASES

INTRODUCTION

In order to get a handle on quantifying a player's effect on a play, we'll look at the traditional notion of total bases. This will help us develop our approach to measuring effect, on offense. A natural byproduct of this thinking will be a way to quantify effect, on defense, as well.

Most would agree that the term "total bases" as it has been traditionally used is a misnomer. In spirit it is supposed to be the total number of bases that a player has contributed to the team through batting. But, it only covers base hits, and treats all singles equally, all doubles equally, and so on, rather than take into account the actual effect of a given hit in a particular situation. "True total bases" should take that actual effect into account. Is there anything else that should be taken into account?

We might expect to reflect in some fashion the effects of walks, stolen bases, sacrifices, and sacrifice flies, as they are recognized tangible contributions an offensive player can be considered responsible for, and therefore worthy of credit for. By this criterion, being hit by a pitch, receiving an intentional walk, or advancing due to an error should not be credited to the offensive player(s) benefitting thereby. Any other efforts that advance the team's cause, as measured in bases possessed, should also be counted in. This would include moving runners up on groundouts or on flyouts other than sacrifice flies. On the other side of the ledger, efforts setting the team's cause back should also be reflected. On offense this would include being put out on the bases and grounding into double plays having a negative net effect.

On defense, one kind of measurable effect is due to errors. Other important activities by a player on defense, i.e. ones having a measurable effect, are fielder's or catcher's interference, wild pitches and passed balls, hit batsmen, and balks.

The actual impact of each of these events can be measured as the number of bases that the offense benefits by them. The sum of these effects for a defensive player would be analogous to total bases on offense.

Remarkably enough, B.A.S.E.S. addresses all of these requirements. In addition, each activity's effect, whether positive, negative, or neither, is captured in direct proportion to its actual impact.

Just one digression before we get to the question of how "true" total bases can be counted.

Since the term "total bases" has already been appropriated, we are faced with the problem of finding another name for what really should be called that. One possibility is to use "true total bases"; that certainly describes matters. An equally accurate label, and a wonderfully short one, is just plain "bases"!

Bases can be counted by working with three concepts:

1. bases possessed by the offense at a particular moment in a time at bat;

2. the change in offense bases possessed due to a play;

3. assignment of responsibility for each part of that change to a particular party.

In B.A.S.E.S., assignment is made in terms of whole bases only.

OFFENSE BASES POSSESSED

BASES POSSESSED VS. BASES OCCUPIED

First, let's look at bases possessed by the offense at a particular moment in a time at bat. An important distinction needs to be made right off the bat. "Bases possessed" does not mean the same thing as "bases occupied." Possession of a base has a value to the offense, in proportion to how much of the way around the bases that base is. Occupation of first base represents possession of one fourth of a run, or one base, as it is one fourth of the way around the bases. Similarly, occupation of second base, since it is two quarters of the way home, is worth two bases. Occupation of third base is possession of three bases, and each run scored is worth four bases.

So, when we speak of contribution (or damage) in terms of bases the cause has been advanced by (or set back by), we're talking about possession rather than occupation.

THE IMMEDIATE VICINITY OF A PLAY

We are interested in the accumulation of bases by a team and by players, but for the purposes of figuring out how many bases a player has added to or subtracted from the team's total on a given play, we will narrow our focus to the immediate vicinity of that play. That vicinity is the span of time from the start of the play to the end of the play. Conditions at the start of a play can be described in terms of which bases are occupied, if any. Any runs that have scored, whether in previous innings or in this inning, are ignored. So if it is asked "How many bases does the offense possess?" at the start of a play being analyzed, it'll be understood that we're just interested in the bases that the runners on base right now, if any, represent. Likewise, when we're talking about bases possessed by the offense after a play, we'll be limiting our discussion to any runners now on base and any runners scoring on the play.

Just to get some practice with this notion, try fielding the following questions. Before a play, how many bases does the offense possess with just a runner on first? One. Runners on first and second? Three. Second and third? Five. Bases loaded? Six. And so on.

THE CHANGE IN BASES POSSESSED

INTRODUCTION

Now let's look at the change in offense bases possessed due to a play.

There are two kinds of change in offense bases possessed: good, and bad. Good change is advance of one or more runners from where they started to some succeeding base. Bad change is forfeiture of one or more bases safely possessed. For our purposes, the batter will be considered to safely possess home base, which allows describing all play results in terms of bases gained or lost, without having to bring in outs. Of course, both kinds of change can occur in the same play.

ADVANCE AND FORFEITURE

The advance of a runner from one base to the next can be thought of as a gain of one base by the offense. This is consistent with the counting of bases possessed. For example, if a runner on second base, representing a possession of two bases by the offense, moves to third base, representing possession of three bases, the change is plus one base possessed, arithmetically speaking.

The forfeiture of a base safely possessed is the loss of "n" bases, where "n" is the value of the base lost. In other words, forfeiture of second base, for example, is equivalent to a loss of two bases. This, too, is consistent. Technically, a batter being put out, where they have not first safely reached base, will be considered forfeiture of home base, or "minus zero bases." Intuitively, this is having used up a chance for progress. We'll just say "zero bases."

The net result of any advances or forfeitures on a non-inning-ending play will equal the bases safely possessed after the play minus the bases possessed before the play. This fact provides a handy way of checking one's arithmetic.

TWO PERSPECTIVES ON THE CHANGE

For example, with a runner on first and none out the batter singles to left and the runner stops at second. Two runners each advanced one base and no runners forfeited possession of any base. So, looking at what happened in detail, by runner, the play resulted in +1 and +1, or +2 bases total. Looking at the play from the other perspective, the bases safely possessed after the play are first and second, which amounts to a team possession of three bases, while the bases possessed before the play, first base, equals one base. Subtracting before from after, or one from three, yields a net change of +2 bases. The math checks.

Let's take another example, this time one involving forfeiture. With a runner on first and none out the batter grounds into a double play. This is equivalent to two forfeitures -- forfeiture of first base and forfeiture of home base. Forfeiture of first base amounts to -1 base, and forfeiture of home is -0. The ledger reading counterclockwise from the batter is minus zero bases and minus one base, for a total of minus one base. From the other perspective, after = zero (none on), before = one, and after minus before equals minus one. Once again, either way you look at it you get the same answer.

What happens when a play involves the third out? Does it makes sense to talk about bases safely possessed after the play? No. But we can still talk about individual advances or forfeitures.

Needless to say, there are as many forfeitures as there are outs made on an inning-ending play, as on any other play. Advances are possible, but not when the third out is a force out or a batted ball caught before it hits the ground. Force outs here include plays where a runner is tagged out on the way to a base they have been forced to try for, even when the force has been removed during the play.

An example of an advance on a play involving the third out is a runner scoring on a two-out single where the batter is thrown out trying for second. Just as the run counts, the base advance counts.

ASSIGNING RESPONSIBILITY FOR CHANGE

INTRODUCTION

The assignment of responsibility for changes in offense base possession, whether credit or blame, is an essential element of B.A.S.E.S. For each base advance or forfeiture, an attempt to identify the party most responsible is made. They get the credit or blame.

THE AGENTS

There are six possible responsible parties, or "agents," as we will have occasion to call them. They are:

1. the batter
2. a runner
3. circumstances of offense
4. the pitcher
5. a fielder
6. circumstances of defense

BASES ASSIGNED AND EVENTS CHARGED

We'll also speak in terms of bases assigned and events charged. The number of bases contributed or negated by a particular party due to a particular action or circumstance are assigned to that party. Each such action or circumstance counts as one "event" charged to the account of that party.

EVENT TYPES AND CATEGORIES WITHIN TYPES

Events and associated bases collected by individuals are accumulated by type of event, i.e. batting, running, pitching, or fielding, and within type of event by setting or other appropriate subdivision.

There are 24 batting settings, one for each of the 24 settings a batter can possibly come to bat in (0, 1, or 2 out in combination with all 8 runner configurations). B.A.S.E.S. distinguishes six subdivisions of baserunning events: successful steals from first, second, or third; forfeitures of first, second, or third. Circumstances of offense events are grouped in the same three categories as are baserunning loss events. Pitching events are counted in the same way as batting events, as to settings that are distinguished. Fielding events are categorized by position, and within position by the configuration of baserunners at the time of the error. This makes for a possibility of 72 (9 x 8) fielding categories. Circumstances of defense events are grouped by what the configuration of baserunners is at the time, meaning eight subdivisions.

We're now ready to look at the scoring component of B.A.S.E.S., which will be explained with the help of numerous examples. Rules and guidelines for assigning bases and events will be given as we go, and often the considerations behind them will be included. Hopefully this will serve to reinforce the reader's confidence in the system.

SCORING

INTRODUCTION

In B.A.S.E.S., scoring concerns assigning responsibility for any base advance or forfeiture on a play, and recording that assignment so that it may be properly processed during performance evaluation. Each result of a play, whether a base advanced or forfeited, is accounted for. Each result's cause, whether activity or circumstance, is found. Call the causing activity or circumstance an "event." Since there is only one possible agent for any given event, determining the party responsible for an event follows automatically. That agent is assigned the change in bases possessed due to each event they are responsible for, and charged with one event.

Discussion of the recording of base and event assignment will be deferred until agent scoring settings and categories, and the playlog, are addressed in depth.

For now, discussion will be limited to how to tell which agent is to be assigned bases and events, and how many bases and events to assign, on any given play.

An important rule to keep in mind is the following:

If a party is not responsible for some development even if they are affected by it, then they are not assigned any bases nor are they charged with an event. So, for example, when a batter is hit by a pitch or receives an intentional base on balls, they are not charged with an event nor are they assigned any bases, since they did nothing.

Generally speaking, most plays involve the batter, and the batter is responsible for most of the change on most of those plays. Contribution by other agents comes on an exception basis.

BATTING EVENTS

The batter is responsible for each base advanced or forfeited by any runner on a batted ball or on a base on balls, except for any gains or losses due to non-batting events as listed below. Included in those exceptions are baserunning events, where for example the baserunner is put out when not in a force situation. The batter cannot be blamed for that kind of forfeiture. Note that it is still considered a force situation when a runner is trying for a base they were forced to try for when the ball was hit, even though the force has been removed during the play.

The batter is also held responsible when they strike out or are called out for interference, obstruction, or batting out of the batter's box.

Note that when a batter reaches base on an error on a batted ball, or on a escaped third strike, two events have occurred. The first event in each case is what should have amounted to an out by the batter, and as such is charged to the batter as if they had made the out. This results in one event being charged, and only as many bases being assigned as would have resulted without the benefit of the defensive mistake. The second event in each case is the activity allowing the batter to reach base, and as such is charged to the defensive party responsible. That defensive party is charged with one event and assigned as many bases as the offense was able to gain thanks to their misplay.

The batter is not given credit for their advance to first base on a fair grounder that is not ruled a base hit nor ruled an error, where no force out is recorded -- for example when the batter reaches on a fielder's choice and everyone is safe. An unsuccessful suicide squeeze play is a case where an out is recorded, but not a force out, where the batter reaches first on a fair grounder ruled a fielder's choice but should not get the credit for their advance to first. When a force out is recorded, normal scoring is in effect. That is, the batter gets the net change in bases due to their batted ball, after the effects on the play due to any events by other agents have been taken into account.

Why a fielder's choice where no force occurs has its own ruling can best be explained with an illustration.

Consider three very similar plays:

1. No outs and a runner on first, the batter gets an infield single, the runner moving to second on the play. After the play: 1st and 2nd.

2. No outs and a runner on first, the batter hits a grounder, the play is attempted at second but the runner is safe at second, the batter gets on via the fielder's choice. After the play: 1st and 2nd.

3. No outs and a runner on first, the batter hits a grounder that is booted, both runners are safe; an error is charged to the fielder. After the play: 1st and 2nd.

Now, let's look at the first example and the third example, before we try to deal with example 2. In the first case it is pretty straightforward assigning the batter two bases contributed, the advance of the runner one base and the advance of themself one base. In the third case it is also straightforward, given that a double play can not be presumed but a single out can if the error hadn't occurred, assigning one base to the batter and the other to the erring fielder. Now what do we do with the second case, the fielder's choice? It lies somewhere between the two others -- do we give the batter a base and a half? No. The batter still only gets credit for one base contributed; the other base is due to the fielder trying for the lead runner and failing, a fielder's choice. The batter reaching base is due to the defense, but it wouldn't be fair to dock the fielder one base as if they had made an error, either. This is an example of circumstances of defense -- when it is appropriate to assign the responsibility to the defense but not to "punish" an individual as if they had committed an error.

Now just to clarify the matter somewhat further, compare the play where both runners are safe on the fielder's choice versus an almost identical play where the runner is out at second. The batter still reaches on a fielder's choice, but a force occurred on the play. In that case, the net change in bases possessed by the offense on the play was zero; before = 1st, after = 1st. In that case, the batter gets zero bases by the ordinary ruling, and it seems the appropriate amount of contribution to assign. So remember, *this special rule only applies when the batter reaches on a fielder's choice and no force occurs.* The only difference is that we subtract one from the net change in bases so that the batter doesn't get credit for advancing themself -- after all, the only reason that they were able to reach base safely was because the fielder chose to try a play on a different runner -- had they tried for the play at first, we can presume that the batter would have been out -- otherwise the official scorer would not have ruled the batter's advance a fielder's choice, but a hit.

Other special rules are needed to deal with inning-ending plays where the third out is a force out. Say that the bases are loaded and there are two outs. If the batter grounds into a force out, it shouldn't matter which base the force occurred at, should it, as far as the net effect of the batter's activity and therefore how many bases to assign them? The net effect is the same whether the force was at home, third, second, or first. Similarly with double-force inning-ending double plays. If the forces were at home and third versus second and first, what difference does it make? The net effect is the same. And, with triple-force triple plays, ditto. Since there is no "after the play," it makes no difference which bases happened to be involved -- that only matters when the inning is still alive and we can look at where runners are "after the play" to measure the net effect of the batted ball.

So, what do we do with inning-ending plays with the third out a force out? Since on the one hand we can't look at the net change in bases possessed due to the play because there is no "after the play," and on the other hand we want to treat each of these force-finished double plays, for example, equally, we need to decide on an arbitrary but reasonable number of bases to assign to the batter for these batted balls.

Force outs with two outs are easy. They're just like forces of the batter when no one's on -- zero bases for forfeiture of home.

All inning-ending double-force double plays are scored as minus one base for the batter. All triple-force triple plays are scored as minus three bases for the batter. These scorings are equivalent respectively to removal of the batter and the runner on first, and the removal of the batter and the runners on first and second. This seems a reasonable equivalence since those plays are the most likely ways for such double forces or triple forces to occur.

Inning-ending plays involving a batted ball but ending in a runner being put out when not forced can be handled with normal scoring. That running event is subsequent to the batter's event. We can back up in time to just before the runner event with no impact on what the batter contributed. We can keep backing up to the last "batting" out, if any, on the play, and see what bases the runners safely possess, and treat that as if it is the setting "after the play" as far as what the batter did is concerned. Any runner event after that point in the play is a stand-alone event that the particular runner put out is responsible for.

Let's illustrate this last point about "backing up" in time. One example: With two outs and a runner on first, the batter singles. The runner is out trying for third. The runner was not forced to try for third, so their being put out is their own fault. Backing up to the situation immediately preceding the forfeiture event, the single caused the batter's advance to first and the runner's advance to second. Thus the batter gets credit for two bases contributed for their batting event. Another example: with one out and runners on first and second the batter singles. The runner from second is out at the plate, the runner from first makes third, and the batter is out trying for second on the throw home. Three events have occurred. The final event is the batter being put out trying for second. That event is not a force and is separate from the batting event. Backing up from that inning-ending event, the runner from second being put out at the plate is another non-force event, and is separate from the batting event. Backing up again, we finally arrive at the first event of the play, the batting event. The batter and both runners each moved up safely one base on the batted ball, and the batter thus gets credit for three bases contributed for that event. Each subsequent event during the play is treated separately.

We'll make one final note on batters to clarify that last example and which will lead us into our next topic, running events. Events assigned to the batter-as-runner once the batter-runner has safely reached base are considered baserunning events and are covered under that heading. For example, if a batter singles and is then put out trying to stretch the single into a double, that's two events, a batting event followed by a baserunning event.

BASERUNNING EVENTS

The baserunner is responsible for the bases gained or lost by the following two activities:

1. successfully stealing a base (+1 base for the runner)

2. forfeiting safe possession of a base when not forced, except when the force has been removed during the play (-n bases, where "n" is the base lost).

In general, for each of the following agents, assign the agent as many bases as were directly attributable to the event being charged to the agent.

CIRCUMSTANCES OF OFFENSE

Consider the responsible party "circumstances of offense" on a result that the offense has caused, when it is not appropriate to assign responsibility to an individual player on the offense. Examples are:

1. the runner from third base is out on a suicide squeeze play
2. a runner, running with the pitch, is doubled off base on a line drive
3. coach's interference

PITCHING EVENTS

1. a wild pitch
2. a hit batsman
3. a balk.

FIELDING EVENTS

1. an error
2. a passed ball
3. fielder's interference or obstruction.

CIRCUMSTANCES OF DEFENSE

Consider the responsible party "circumstances of defense" on a result that the defense has caused, when it is not appropriate to assign responsibility to an individual player on the defense. Examples are:

1. an intentional base on balls
2. a fielder's choice

Responsibility on unusual plays such as appeal plays and other technicality rulings should be assigned as most appropriate, whether to an individual or to a team, in the same spirit as set forth above.

SAMPLE PLAYS

Let's look at a systematic sample of plays and see base and event assignment in action. Starting with very simple plays, we'll work our way up through more and more complicated examples, until we've seen how even the most involved combinations of events can be handled.

In each of the following plays, one event is charged to the agent involved.

Single event, batter, none on, advance:

EVENT	BASES	AGENT
1. single	1	batter
2. double	2	batter
3. triple	3	batter
4. home run	4	batter
5. base on balls	1	batter
6. hit by pitch	1	pitcher
7. catcher's interference	1	fielder (catcher)

Single event, batter, none on, forfeiture:

EVENT	BASES	AGENT
8. any out where batter does not reach base safely before being put out	0	batter

At this point it will be useful to introduce some abbreviations:

1st = first base
2nd = second base
3rd = third base
r0 = the batter/runner
r1 = the runner from first base
r2 = the runner from second base
r3 = the runner from third base
-> = goes to

Single event, batter, runners on:

#	ON/OUT	PLAY	EVENT	BASES	AGENT
9.	1st / 0	single to left, r1 stops at 2nd	single	2	batter
10.	1st / 0	single to right, r1 -> 3rd	single	3	batter

#	ON/OUT	PLAY	EVENT	BASES	AGENT
11.	1st / 0	grounder, r1 -> 2nd, r0 forced	ground ball	1	batter
12.	1st / 0	grounder, r1 forced, r0 -> 1st	ground ball	0	batter
13.	1st / 0	grounder, double play	ground ball	-1	batter
14.	1st / 0	fly out, r1 -> 2nd	fly ball	1	batter
15.	1st / 0	bunt, r1 -> 2nd, r0 out at 1st	ground ball	1	batter
16.	1st / 0	strike out	strike out	0	batter
17.	1,2 / 0	single, bases loaded	single	3	batter
18.	1,2 / 0	single, r2 scores, r1 -> 2nd	single	4	batter
19.	1,2 / 0	single, r2 scores, r1 -> 3rd	single	5	batter
20.	1,2 / 0	grounder, r2 -> 3rd, r1 -> 2nd, r0 forced	ground ball	2	batter
21.	1,2 / 0	grounder, r2 -> 3rd, r1 forced, r0 -> 1st	ground ball	1	batter
22.	1,2 / 0	grounder, r2 forced, r1 -> 2nd, r0 -> 1st	ground ball	0	batter
23.	1,2 / 0	grounder, double play, r2 -> 3rd	ground ball	0	batter

#	ON/OUT	PLAY	EVENT	BASES	AGENT
24.	1,2 / 0	grounder, r2 forced, r1 -> 2nd, r0 out at 1st	ground ball	-1	batter
25.	1,2 / 0	grounder, r2 forced, r1 forced, r0 -> 1st	ground ball	-2	batter
26.	1,3 / 0	grounder, double play, r3 scores, r1 forced, r0 out at 1st	ground ball	0	batter
27.	2,3 / 0	single, r3 scores, r2 scores, r0 -> 1st	single	4	batter
28.	1,2,3/0	single, r3 scores, r2 -> 3rd, r1 -> 2nd, r0 -> 1st	single	4	batter
29.	1,2,3/0	single, r3 scores, r2 scores, r1 -> 2nd, r0 -> 1st	single	5	batter
30.	1,2,3/0	double, r3 scores, r2 scores, r1 scores, r0 -> 2nd	double	8	batter
31.	1,2,3/0	grand slam HR	home run	10	batter
32.	1,2,3/0	grounder, double play, r3 scores, r2 -> 3rd, r1 forced, r0 out at first	ground ball	1	batter
33.	1,2,3/0	strike out	strike out	0	batter

Single event, batter, inning-ending play:

#	ON/OUT	PLAY	EVENT	BASES	AGENT
34.	none/ 2	ground out	ground out	0	batter
35.	1st / 2	ground out	ground out	0	batter
36.	1st / 1	double play	double play	-1	batter
37.	1,2 / 1	double play	double play	-1	batter
38.	1,3 / 1	double play	double play	-1	batter
39.	1,2,3/1	double play	double play	-1	batter
40.	1,2 / 0	triple play	triple play	-3	batter
41.	1,2,3/0	triple play	triple play	-3	batter

Single event, runner:

#	BASE ON	PLAY	EVENT	BASES	AGENT
42.	1st	picked off	out on bases	-1	runner
43.	2nd	picked off	out on bases	-2	runner
44.	3rd	picked off	out on bases	-3	runner
45.	1st	steals 2nd	stolen base	1	runner
46.	2nd	steals 3rd	stolen base	1	runner
47.	3rd	steals home	stolen base	1	runner
48.	1st	caught stealing	out on bases	-1	runner
49.	2nd	caught stealing	out on bases	-2	runner
50.	3rd	caught stealing	out on bases	-3	runner

Single event, pitcher:

#	RUNNERS	PLAY	EVENT	BASES	AGENT
51.	1st	wild pitch	wild pitch	1	pitcher
52.	1st	balk	balk	1	pitcher
53.	1st,2nd	wild pitch	wild pitch	2	pitcher
54.	1,2,3	hit batsman	hit batsman	4	pitcher

Note: for multiple event plays, the play description is a narrative which may continue on subsequent lines. The events for a play are listed one to a line. Don't try to read across from one line of a play description to the same line of the list of events for the play.

Multiple event, none on, imminent forfeiture and subsequent advance:

#	PLAY	EVENT	BASES	AGENT
55.	strikeout and escaped third strike	strikeout escaped third strike	0 1	batter fielder *
56.	fly ball and n-base error by RF	fly ball error	0 n	batter fielder 9

*In this example the fielder is the pitcher if a wild pitch was ruled, and the catcher if a passed ball was ruled.

Multiple event, none on, advance and further advance:

#	PLAY	EVENT	BASES	AGENT
57.	m-base hit, n-base error by RF	hit error	m n	batter fielder 9

Multiple event, none on, advance and subsequent forfeiture:

#	PLAY	EVENT	BASES	AGENT
58.	hit, batter out on the bases	hit out on the bases	m -m	batter batter/ runner
59.	error by RF on fly ball, batter out on the bases	batted ball error out on the bases	0 n -n	batter fielder 9 batter/ runner
60.	hit, error by RF, batter out on the bases	hit error out on the bases	m n -(m+n)	batter fielder 9 batter/ runner

Multiple event, runners on:

#	ON/OUT	PLAY	EVENT	BASES	AGENT
61.	2nd / 0	single, r2 scores, r0 out at 2nd	single out on bases	3 -1	batter batter/ runner
62.	2nd / 0	single, r2 scores, r0 -> 2nd on the throw	single advance on throw	3 1	batter circ. of def.
63.	1st / 0	single, r1 out at 3rd, r0 -> 2nd on the throw	single out on bases advance on throw	2 -2 1	batter runner 1 circ. of def.

#	ON/OUT	PLAY	EVENT	BASES	AGENT
64.	1st / 0	grounder, both safe on error by SS	ground ball error	1 1	batter fielder 6
65.	1st / 0	grounder, both safe on fielder's choice	ground ball fielder's choice	1 1	batter circ. of def.
66.	1st / 0	stolen base, error on throw, r1 -> 3rd	stolen base error	1 1	runner 1 fielder 2
67.	1st / 1	grounder, r0 forced, r1 -> 2nd, r1 out on bases	ground ball out on bases	1 -2	batter runner 1
68.	1,3 / 0	grounder, r0 forced, r1 -> 2nd, r3 out at plate, r1 out at 3rd	ground ball out on bases out on bases	1 -3 -2	batter runner 3 runner 1

Multiple-event agents:

#	ON/OUT	PLAY	EVENT	BASES	AGENT
69.	none/ 0	single, out stretching	single out on bases	1 -1	batter batter/ runner
70.	1st / 2	single, r1 -> 3rd, r0 out stretching	single out on bases	3 -1	batter batter/ runner
71.	1st / 0	stolen base, r1 out trying for 3rd	stolen base out on bases	1 -2	runner 1 runner 1
72.	none/ 0	grounder, error on pickup by 3rd baseman, r0 -> 1st, error on throw by 3rd baseman, r0 -> 2nd	ground ball error error	0 1 1	batter fielder 5 fielder 5
73.	none/ 0	strikeout, passed ball, r0 -> 1st, error on throw by catcher, r0 -> 2nd	strikeout passed ball error	0 1 1	batter fielder 2 fielder 2

THE PLAY LAYOUT

For complicated plays, charting the action makes the assignment of bases and events to agents easier. One aid in such charting is the play layout. Refer to figure 1 below.

The legend across the top, "B0" through "B4," refers to the bases:

B0 = home base
B1 = first base
B2 = second base
B3 = third base
B4 = home base.

The legend down the left side, "R0" through "R3," refers to the runners:

R0 = the batter/runner
R1 = the runner from first base
R2 = the runner from second base
R3 = the runner from third base.

```
            B0         B1         B2         B3         B4

             * ------ * ------ * ------ * ------ *
R0           |        |        |        |
             |        |        |        |

             *        * ------ * ------ * ------ *
R1                    |        |        |
                      |        |        |

             *        *        * ------ * ------ *
R2                             |        |
                               |        |

             *        *        *        * ------ *
R3                                      |
                                        |
```

Figure 1.

Using the chart to lay out a play is simple. An advance is indicated on the row for the runner advancing by putting an arrowhead pointing right on the right end of the line connecting the base started from and the base advanced to. The cause of the advance is given as a brief description below the arrow. For example, say that the batter has advanced to first base on a single. That would look like this:

```
              B0          B1          B2          B3          B4

              * ------->* ------  * ------  * ------  *
   R0         | single  |         |         |
              |         |         |         |

              *         * ------  * ------  * ------  *
   R1                   |         |         |
                        |         |         |

              *         *         * ------  * ------  *
   R2                             |         |
                                  |         |

              *         *         *         * ------  *
   R3                                       |
                                            |
```

Figure 2.

For forfeitures the process is similar. On the row for the runner forfeiting, put an arrowhead pointing down at the bottom of the line descending from the base forfeited. Put a brief description for the cause of the forfeiture to the right of the arrow. For example, say that the runner on first is thrown out trying to steal second. That would look like this:

```
              B0          B1          B2          B3          B4

              * ------  * ------  * ------  * ------  *
   R0         |         |         |         |
              |         |         |         |

              *         * ------  * ------  * ------  *
   R1                   | caught  |         |
                        | stealg  |         |
                        V

              *         *         * ------  * ------  *
   R2                             |         |
                                  |         |

              *         *         *         * ------  *
   R3                                       |
                                            |
```

Figure 3.

By using the play layout it is easy to count bases contributed by an event, and to decide which agent is responsible for some outcome.

Here are a couple of examples of layouts for complicated plays.

#	ON/OUT	PLAY	EVENT	BASES	AGENT
74.	2nd / 0	single, r2 scores, r0 -> 2nd on the throw, error on throw, r0 -> 3rd	single advance on throw error	3 1 1	batter circ. of def. fielder

```
            B0       B1       B2       B3       B4

            * ------>* ------>* ------>* ------ *
R0          | single | throw  | error  |
            |        |        |        |

            *        * ------ * ------ * ------ *
 R1                  |        |        |
                     |        |        |

            *        *        * ------>* ------>*
R2                            | single | single
                              |        |

            *        *        *        * ------ *
R3                                     |
                                       |
```

Figure 4.

#	ON/OUT	PLAY	EVENT	BASES	AGENT
75.	1,2,3/0	single, r3 scores, r2 scores, r1 tries for 3rd on the throw home out in rundown, r0 -> 2nd on rundown	single rundown out on bases	5 1 -2	batter circ. of def. runner 1

```
            B0          B1          B2          B3          B4

            * ------>* ------>* ------ * ------ *
R0          | single |run     |        |
            |        |down    |        |

            *        * ------>* ------ * ------ *
R1                   | single | out on |
                     |        | bases  |
                              V

            *        *        * ------>* ------>*
R2                            | single | single
                              |        |

            *        *        *        * ------>*
R3                                     | single
                                       |
```

Figure 5.

EVALUATION

INTRODUCTION

The effects on the play-by-play due to the batter, a runner, the pitcher, or a fielder are the result of activities, whereas those of the two team-type agents are the result of circumstances.

We shall concentrate here on evaluating the activities: batting, baserunning, pitching, and fielding. We'll call each of these four departments "activity types." We will develop statistics measuring performance for each category within an activity type, across categories for an activity type, and across the two offense activity types. We will also develop statistics for circumstances of offense and for circumstances of defense.

OFFENSE

BATTING

Categories

Twenty-four different batting categories are distinguished in B.A.S.E.S. They are based on the setting at the time of an event, and there are twenty-four settings. A setting is simply the unique combination of outs and runners at the start of a play. We distinguish between settings because the potential for contributing bases differs between settings. For example, the chances of contributing one base with no one on and two outs are considerably less than with the bases loaded and no outs.

The fact that the potential for contributing bases depends on the setting is important because in B.A.S.E.S. contribution is counted in context. Events with like settings are grouped together. For example, all events with none on and no outs are counted together, all events with none on and one out are in a second group, and so on.

Keeping events from each setting by themselves and separate from those of any other setting allows accurate measure of a given contribution. Just looking at the raw number of bases contributed by a batter on a play does measure how much they did but not how well they did. How well they did depends on what they had to work with, which is the potential of the situation, and how good what they did with that is, compared to some standard.

So, a given batter will have up to twenty-four individual batting statistics, one for each setting that they have events in.

Average Bases Per Setting

The first step in deriving the batter's statistic for a given setting, the measure of their performance in that category, is to obtain their average bases in that setting. This is simply their total of bases contributed in the setting, divided by their number of batting events in the setting. For example, in the setting "none on and none out" a batter might have 100 events which contributed a total of 30 bases. That would yield a ratio of .300 average bases per event for that setting.

Let's give the twenty-four batting settings numbers for easy reference, where "S#" means "setting number*:

S#	BASES OCC.	OUT	S#	BASES OCC.	OUT	S#	BASES OCC.	OUT
1		0	9		1	17		2
2	1	0	10	1	1	18	1	2
3	2	0	11	2	1	19	2	2
4	3	0	12	3	1	20	3	2
5	1,2	0	13	1,2	1	21	1,2	2
6	1,3	0	14	1,3	1	22	1,3	2
7	2,3	0	15	2,3	1	23	2,3	2
8	1,2,3	0	16	1,2,3	1	24	1,2,3	2

So, expressing a batter's average bases for setting 1 as in the above example would look like:

$\frac{B_1}{E_1}$. In general it is expressed as $\frac{B_n}{E_n}$.

Average bases is interesting in itself but still doesn't tell us how well a batter has done unless we know how good that base average is.

One approach to determining how good a particular average bases figure is in a given setting is to have on hand a scale of measurement, and to grade each batter's figure according to that standard. B.A.S.E.S. does do that but the scale of measurement is not one that is fixed for all time, for a period of time covering many seasons, nor even for recent seasons. Instead the standard stands for only a single season, being replaced by a new one the next season. How is this possible, let alone manageable, and why is this a good idea? It's very simple.

League Rank of Average Bases

The standard of accomplishment for a batting setting, the scale against which each batter's performance in that setting is measured, is nothing more than the list of base averages for that setting for every batter in the league with events in that setting that season. The list is sorted so that the batter with the highest average bases for the setting is at the top of the list and the batter with the lowest average bases is at the bottom of the list. One's level of performance, how well one has done in a given situation, is thus determined with respect to one's peers' performance in that same situation.

Before we go on we'll just note that the standard actually changes during the season. In fact it changes daily. Technically, it actually changes with every batting event that occurs, play by play. This type of standard is the fairest, in that it does not measure a player's performance in terms of the level of performance of any other era, but rather in the correct context for that performance, the current season. It is also the most accurate in that no arbitrary values are associated with given values of average bases in a setting. Rather, the perfectly appropriate values, whatever they may happen to be at the moment, as determined by the actual performance of the players in the league with events in the setting, are dynamically derived, giving the most accurate possible scale for comparison.

Percentile of League Rank

But there is more that needs to be done. If we leave things as they are then we've got the player's rank in the league in the setting,

$$R\left(\frac{B_n}{E_n}\right).$$

This is all right if we're just interested in this setting, but in order to look at a player's overall batting performance across all twenty-four settings we need to translate the ranks into a more universal relative measure.

To illustrate the problem, suppose that a player ranks 10th in one setting and 20th in another. But suppose also that only 50 players have events in the first category while 100 have events in the second. The player ranks 10th out of 50 in the first category and 20th out of 100 in the second, so actually they rank the same in the two settings, having done equally well in them when compared with their peers.

So, it is apparent that we must convert a player's rank in a setting to a ratio. We do this by using the player's "percentile" in a setting. More precisely, it is the percentile of the rank of their average bases in the setting, expressed as three decimal digits. For example, the 98th percentile would be expressed as .980.

There is a fine distinction between how percentiles are determined in B.A.S.E.S. and how they might be determined normally. Normally, the highest possible percentile is the 99th. This happens because the range of possible scores is broken up into 100 intervals, the 0th percentile through the 99th. One's percentile is computed as the percent of scores that are lower than one's own. Since it can never be true that all scores are lower than one's own, it is impossible to be in the 100th percentile. So, if there were 100 scores and one's own were at the top of the list, with no one tied for first place, then one's percentile would be the 99th because one's score would be higher than 99% of the scores (99 out of 100).

Normally, even a perfect score on a test only results in being placed in the 99th percentile. And, even with a perfect score, if the percentage of scores less than it is only 98% or 97% and so on, because there aren't that many scores in the list or because there are other scores tied for first place, then the percentile is accordingly lower. For example, say that there are 100 scores and 10 are tied for the highest spot. Only 90 out of the 100 scores are lower, so the 10 scores at the top only get a percentile of 90. Or say that there are only 10 scores and that the top score is not in a tie. Since 9 out of 10 scores are lower, it is in the 90th percentile.

In B.A.S.E.S. the thinking is subtly different. The range of 100 intervals here is the 1st percentile through the 100th. 100 seems more appropriate than 99 for the highest score. 99 gives an impression of just falling short of perfection. It also seems appropriate to give top scorers the highest possible percentile, 100, rather than something less, even if they're in a tie or there aren't that many scores in a list. It is more rewarding to the players, who strive to be the best and to have the highest score in a category, to get the highest rating when they succeed in doing so. Any who tie for first place should share in the satisfaction of being in the 100th percentile, like getting 100 on a test.

To address these issues, then, the following method of computing percentiles is used in B.A.S.E.S.: for a given score or average, divide the number of scores that are higher in the list for that category by the number of scores in the list. Subtract that result from 1. This can be expressed as:

$$\text{percentile} = 1 - \frac{\text{\# higher in category}}{\text{\# in category}}.$$

To contrast this with the "normal" method, the normal formula is:

$$\text{percentile} = \frac{\text{\# lower in category}}{\text{\# in category}}.$$

Let's see how the B.A.S.E.S. method does what is claimed for it.

First, the top score in a list of 100, in sole possession of first place, would get:

$$1 - \frac{\text{\# higher}}{\text{\# in cat}} = 1 - \frac{0}{100} = 1\text{-}0 = 1.000 \text{ or the 100th percentile.}$$

Second, the sole top score in a list of only 10 scores:

$$1 - \frac{\text{\# higher}}{\text{\# in cat}} = 1 - \frac{0}{10} = 1\text{-}0 = 1.000 \text{ or the 100th percentile.}$$

Third, a top score tied with nine others in a list of 100:

$$1 - \frac{\text{\# higher}}{\text{\# in cat}} = 1 - \frac{0}{100} = 1\text{-}0 = 1.000 \text{ or the 100th percentile.}$$

The same three situations would have resulted in .990, .900, and .900 respectively, using the normal method.

Our statistic for a batting setting now looks like this:

$$\% \left(R \left(\frac{B_n}{E_n} \right) \right).$$

Weighted Percentiles

These rank percentiles allow us to compare a player's performance across settings. A percentile in one setting means the same thing in any other setting. For example, .800 is excellent in any setting, while .200 is poor in any setting. This is where normalization appears in B.A.S.E.S.

To get a player's overall batting rating we obviously need to combine all of the individual percentiles for the settings that they have events in. But we can't just average them all together indiscriminately. For example, it wouldn't be fair to simply average an .800 rating in a setting in which a batter has 100 events and a .200 rating in a setting in which they only have one event, yielding an overall average of .500, would it? The setting with more events should carry more weight. But how much weight? It should be weighed in proportion to its number of events. So, we weight each percentile by the number of events in its setting.

An individual term in the grand statistic we are building will therefore look like:

$$E_n \quad x \quad \% \left(R \left(\frac{B_n}{E_n} \right) \right).$$

Combined Weighted Percentiles

We're finally ready to derive the overall batting statistic. It is the sum of all of the weighted percentiles, divided by the total number of batting events. Let's illustrate this with a simple example.

Say that a batter has 10, 20, and 30 events in three different settings S1, S2, and S3 respectively. Say that the batter has performed at the .800, .600, and .500 levels in the three settings respectively. Their overall statistic is then:

$$\frac{(10 \text{ x } .800) + (20 \text{ x } .600) + (30 \text{ x } .500)}{10 + 20 + 30}$$

$$= \frac{8.00 + 12.00 + 15.00}{60} = \frac{35}{60} = .583.$$

BASERUNNING

Categories

Baserunning events are of two types: advances and forfeitures. The only kind of advance is a stolen base. Aren't there other ways that baserunning ability positively contributes to the team besides base stealing? And if so, why aren't these other contributions acknowledged here?

Yes, there are other ways that running skill helps. Why they aren't acknowledged here can be understood by looking at the kinds of things that they are. Besides stealing more bases and being caught stealing less frequently, a good baserunner will beat out more infield hits, break up more double plays, stretch more hits for extra bases, advance more bases by tagging up on fly balls, go from first-to-third and second-to-home on singles more often, and rattle the pitcher's concentration more than other baserunners will. Each of these contributions except for the last one can be measured in terms of bases contributed. Those that can be measured in bases either increase the runner's own batting statistics or else one of their teammates' batting statistics. Either way, the team benefits, which is important. So, in general, as we can see, such baserunning contribution does not go unacknowledged.

A baserunning forfeiture is any event where a runner is out on the bases when they are not forced to run. Examples of forfeitures are: being caught stealing, being picked off, being put out trying to stretch a hit for an extra base, being put out on a grounder when not forced, and being doubled off base as in when a runner fails to get back to their base on a line drive out or on a fly out.

An advance will always consist of plus one base. A forfeiture will always consist of minus "n" bases, where "n" is the base given up.

How many categories should we distinguish in order to divide up baserunning events by context? Stolen bases are advances to second base, third base, or home base. Forfeitures are relinquishments of first base, second base, or third base. We can go with six categories then, but there is an elegant way to think of them as derived from three "settings."

We can look at steals as advances from a base instead of to a base. A steal of second is an advance from first, a steal of third is an advance from second, and a steal of home is an advance from third. Looking at steals this way, all baserunning events can be considered either advances from or forfeitures of first, second, or third base. Each such "from-base" then has two categories of events associated with it, positive events and negative events (in baserunning there are no neutral events). This makes a total of three positive event categories and three negative event categories. The three positive event categories are: advances from first, second, or third. The three negative event categories are: forfeitures of first, second, or third. We'll call these categories gain categories and loss categories, respectively, and call their events gain events and loss events, respectively.

Why do we split up gains and losses into their own categories? Why don't we lump them together by from-base? Because the background against which to measure performance differs between gains and losses. It is natural to use "attempts" to measure steals against (or times caught stealing against). For losses in general, however, a different background is required. We'll see the thinking behind its recognition in more detail below, in the discussion of loss category ranking, but for now we'll just say that it relates to the number of times "on the bases."

Averages For Ranking

Gain Categories

For determining how well a player has done in stealing bases from a given base, their success rate is appropriate. The success rate is simply the ratio of stolen bases to attempted steals:

$$\frac{S_n}{A_n}, \text{ where n is the base started from.}$$

So, for example, the success rate for steals of second base, i.e. steals from first base, is:

$$\frac{S_1}{A_1}.$$

Since all statistics bei_1ng developed for determining relative performance in a category, i.e. the averages like average bases, can be stated with a number of bases in their numerator, and since the number of steals from a base is equal to the number of bases that those steals contributed, each steal contributing one base, we will express the ratio for steal success as:

$$\frac{B_n}{A_n}.$$

The higher the rate, or "average," the better. The list of averages for that category for the league is sorted so that the highest success rate is at the top. A player's rank in the list of averages for the category in the league is:

$$R\left(\frac{B_n}{A_n}\right).$$

Loss Categories

Determining how well a player has done in loss categories requires different thinking. In loss categories it is the failure rate rather than the success rate that we want to measure. For negative "contribution" to the team's cause we will penalize a player in proportion to their degree of failure rather than reward them in proportion to their degree of success, as in the gain categories and the batting statistics.

A higher failure rate is caused by more frequent forfeiture. Calculating frequency of forfeiture can make use of the number of loss events or the number of bases negatively "contributed" by those events, but what should the background against which these are measured be? We can't use the number of events in the denominator here. In baserunning the number of bases advanced or forfeited with respect to a given from-base is always +1 and -n respectively, where n is the from-base. So dividing the number of bases by the number of events in a loss category will always turn out the same:

$$\frac{B_n}{E_n} = \frac{E_n \times -n}{E_n} = -n.$$

Besides, it isn't appropriate for finding the failure rate in terms of total events anyway. But how should we think? Intuitively, given two runners with the same number of forfeitures of a base, the runner who has forfeited the base less often is the better runner, their failure rate being lower. But less often in terms of what? In terms of times in a position to forfeit the base -- in other words "times on the base." But how can we quantify "times on a base"? By counting safe possessions of each base.

Every time that a runner occupies a base after a play, that is safe possession of that base. Every time that a runner is out on the bases, the base that they just forfeited is the last base that they safely possessed, whether they started the play there or not.

Since we're counting safe possessions of each base in order to measure losses against a background, we don't count safe possessions of home other than as runs scored, since you can't forfeit home after circling the bases. Also, momentarily possessing a base on the way to another base does not count as safe possession for these statistical purposes. For example, if there is a runner on first base and the batter singles, with the runner from first going to third on the play, a safe possession of second base will not be counted for the runner from first. A safe possession of the base where they ended up, third base, will be counted. If, on the other hand, the runner from first was put out on the bases between second and third after having safely reached second base on the play, then a safe possession of second base would be counted. The forfeiture of second base would then have its proper perspective.

So, the denominator for our ratio for baserunning loss categories is the number of safe possessions of the base that the category covers. The numerator is the number of loss events.

For a given number of possessions, the fewer the loss events the better. This means that the failure rate will be minimized. So, if we were to sort a list of failure rates, we would have to be careful to get them in the right direction. We want the *worst* ratio at the top. *Why* we want the worst average at the top is another matter. Before we get to that let us continue the current train of thought by saying that for a given number of forfeitures, the more possessions the better. This also minimizes the failure rate, which is desirable. So our loss category ratio, or failure rate, is:

$$\frac{E_n}{P_n}.$$

The *lower* the ratio the better, here. Also, we want as *low* a rank as possible in these categories -- *the lower the rank the better*. That is so that the smallest possible amount will be subtracted when we subtract the weighted percentiles, which we will do. You see, here, where we are penalizing failure instead of rewarding success, we do things just the opposite of normally. It is desirable to minimize the number of loss events in a category, and it is also desirable to minimize their frequency per possession. So, the smaller the number of events, the lighter the weight on the category's statistic, the failure rate. And, the lower the failure rate, the less that there is to be weighted. We take the weighted percentile, based on the failure rate's ranking, and subtract it from the gain categories' accomplishments.

Why don't we rank failure rates from best to worst as in batting and gain categories, and add in the weighted percentiles as we do there? Because of an irony. If we weight by the number of events here, i.e. forfeitures, the better a runner does the lighter the weight would be. We don't want that. Taken to the extreme, a runner with no forfeitures would get no credit at all since their percentile would be weighted by zero. By subtracting, though, things make sense. Now, the fewer the forfeitures the lighter the weight, and the less that gets subtracted. This explains why we rank the failure rates from worst to best. The worse the rate the higher the rank and the percentile, therefore the more that there is to be weighted and deducted.

So, to minimize what gets subtracted, it behooves a player to minimize the weight (the number of loss events), and the percentile (rank in the category). The better the rate the lower it ranks, so the lower its percentile is. The rank can be expressed as:

$$R\left(\frac{E_n}{P_n}\right).$$

At this point we had better illustrate these concepts. Let us limit the discussion to losses of one particular base, say second base. Let us also limit our players to three, A, B, and C. First, let's look at failure rate. Suppose that all three of the players have safely possessed second base 100 times. Player A has forfeited second base 5 times, player B 10 times, and player C 15 times. A's failure rate is then 5/100 or .050, B's is 10/100 = .100, and C's is 15/100 = .150. Again to simplify things, suppose that these three are the only players with safe possessions of second base, so A is best, B comes next, and C is worst. Their ranks would be 1, 2, and 3 respectively if we ranked the best first, but we reverse the list when we rank. So, the list is C, B, A. Their percentiles are therefore C = 1 - (0/3) = 1 - 0 = 1.000, B = 1 - (1/3) = 2/3 = .667, and A = 1 - (2/3) = 1/3 = .333. Thus, the best runner, player A, has the lowest rank and percentile. This means that when we *subtract* player A's weighted percentile, the amount subtracted is minimal. We weight the percentile by the number of events, which for A is again minimal compared to the other players' event counts. This also minimizes the amount subtracted. Player A's weighted percentile is 5 x .333 = 1.667, player B's is 10 x .667 = 6.667, and player C's is 15 x 1.000 = 15.000. Thus, the amount subtracted for this loss category varies significantly between the players. Base forfeiture is always bad -- it's just a matter of how minimal a player's baserunning mistakes are that makes the difference, and which is reflected by subtracting a measure of their mishaps in their running and overall offense statistics.

Percentiles

For both gain and loss categories, percentiles are based on league rank as is done in batting evaluation:

$$\%\left(R\left(\frac{B_n}{A_n}\right)\right) \quad \text{or} \quad \%\left(R\left(\frac{E_n}{P_n}\right)\right) \quad \text{as appropriate.}$$

Note that a player with no successful steals of a given base has no events in the gain category for that base, no matter how many unsuccessful attempts at stealing the base that they've had.

Weighted Percentiles

Weighting of each category's percentile in order to compile cross-category statistics for both gain and loss categories uses the number of events in a category. In the gain categories, the number of events is the number of stolen bases. For the loss categories, the number of events is the number of forfeitures where the runner is liable. In order to distinguish between gain event counts and loss event counts when they are both present, as they will be in the overall statistics, we will mark the gain event and base counts with a plus sign and the loss event and base counts with a minus sign. So, the weighted percentiles for baserunning categories look like this:

$$E^{+}_{n} \times \%\left(R\left(\frac{B^{+}_{n}}{A_n}\right)\right)$$

$$\text{or} \quad E^{-}_{n} \times \%\left(R\left(\frac{E^{-}_{n}}{P_n}\right)\right) \quad \text{as appropriate.}$$

Combined Weighted Percentiles

As in batting, we can combine the individual statistics for baserunning to derive an overall rating. Combining the three gain statistics with the three loss statistics is actually easy. We add the three weighted gain percentiles and subtract from that the three weighted loss percentiles to form the numerator. We divide that figure by the sum of the six event counts:

$$\frac{E_1^+ \%R\frac{B_1^+}{A_1} + E_2^+ \%R\frac{B_2^+}{A_2} + E_3^+ \%R\frac{B_3^+}{A_3} - E_1^- \%R\frac{E_1^-}{P_1} - E_2^- \%R\frac{E_2^-}{P_2} - E_3^- \%R\frac{E_3^-}{P_3}}{E_1^+ + E_2^+ + E_3^+ + E_1^- + E_2^- + E_3^-}$$

Let's illustrate how this is much simpler than it looks by working with an example. Here is a table giving the figures in the six baserunning categories for a fair baserunner:

EVENT NAME	SYM BOL	# E	# B	# ATT	# POS	B/A	E/P	RANK	%	E x %
STEALS FROM FIRST	E_1^+	5	5	10	-	.500	---	70/150	.533	2.665
STEALS FROM SECOND	E_2^+	0	0	0	-	---	---	--	---	0.000
STEALS FROM THIRD	E_3^+	0	0	0	-	---	---	--	---	0.000
LOSSES OF FIRST	E_1^-	3	-3	-	10	---	.300	91/100	.100	0.300
LOSSES OF SECOND	E_2^-	2	-4	-	4	---	.500	41/ 50	.200	0.400
LOSSES OF THIRD	E_3^-	0	0	-	2	---	.000	21/ 25	.200	0.000

To determine their baserunning rating, or index, we add the weighted percentiles for the three gain categories, 2.665, 0, and 0, respectively, to get 2.665. We then subtract from that the weighted percentiles for the three loss categories, .300, .400, and 0, respectively, giving 2.665 - .700 = 1.965. This result is then divided by the total number of events for the six categories, 5 + 0 + 0 + 3 + 2 + 0 = 10, giving .197 rounded.

This is not an admirable figure in general. But shouldn't this fairly competent baserunner fare better than a poor score? Well, this fairly competent baserunner contributed a net of -2 bases to the team's cause by way of their baserunning, not a very desirable outcome. If they hadn't tried stealing any bases nor been caught on the bases at all, in other words played perfectly safe at all times, then their net contribution would at least have been better, zero bases contributed.

Let's look at the figures for a premier baserunner and see if at least they might be rewarded by a rating reflecting excellence:

EVENT NAME	SYM BOL	# E	# B	# ATT	# POS	B/A	E/P	RANK	%	E x %
STEALS FROM FIRST	+ E 1	75	75	85	-	.882	---	22/150	.853	63.975
STEALS FROM SECOND	+ E 2	35	35	40	-	.875	---	2/ 50	.960	33.600
STEALS FROM THIRD	+ E 3	5	5	7	-	.714	---	1/ 1	1.000	5.000
LOSSES OF FIRST	- E 1	20	-20	-	200	---	.100	84/100	.150	3.000
LOSSES OF SECOND	- E 2	8	-16	-	100	---	.080	31/ 50	.400	3.200
LOSSES OF THIRD	- E 3	2	-6	-	50	---	.040	28/ 30	.100	.200

The baserunning rating for this player is:

$$\frac{63.975 + 33.600 + 5.000 - 3.000 - 3.200 - .200}{75 + 35 + 5 + 20 + 8 + 2} = \frac{96.175}{145} = .663.$$

This is a good rating, in fact probably a very good rating considering that it spans six categories, and that it gets harder and harder to ascend above a .500 index the higher one goes even for one category let alone across multiple categories.

One other important point to note is that this very good index will weigh-in quite heavily when it is time to combine all offense category statistics because it represents 145 events. An intuitive corroboration of the fact that this runner has excelled is their net contribution in bases: +73. The fair runner, remember, only contributed -2 bases via baserunning. Their poor index will not weigh that heavily overall, though, because they only had 10 baserunning events.

COMBINING BATTING AND BASERUNNING STATISTICS

To derive a player's overall offense index, the twenty-four batting statistics are combined with the six baserunning statistics. The method of combination is the familiar one -- the numerator of the overall ratio is the sum of the 24 weighted batting percentiles and the three weighted gain baserunning percentiles, minus the three weighted loss baserunning percentiles. The denominator of the ratio is the sum of the thirty weights, the total number of offense events:

$$\text{OFFENSE INDEX} = \frac{\text{24 batting stats + 3 gain stats - 3 loss stats}}{\text{the sum of the thirty event counts}}.$$

CIRCUMSTANCES OF OFFENSE

We will briefly discuss statistics for circumstances of offense. Events of this type are rare. Also, there are more meaningful statistics for team performance such as won-lost percentage. For these reasons, these circumstances of offense statistics need not be emphasized very strongly.

Circumstances of offense activity is treated exactly the same as baserunning loss activity, since circumstances of offense events are always forfeitures. Three contexts are distinguished -- possessions of first base, second base, and third base. Ranking of teams is by average bases per possession. Percentile is based on rank. Percentiles are weighted by the number of events in the category. The weighted percentiles are subtracted (from zero) and that result is divided by the total number of events to produce the overall index.

DEFENSE

PITCHING

Categories and Averages

To measure pitching effectiveness we count up bases allowed to opponent batters. Added to that count are any bases that a pitcher has been directly responsible for giving to the offense via pitching events, i.e. via wild pitches, balks, and hit batsmen. Hand-in-hand with opponent batter bases are the twenty-four batting settings. Although the three kinds of pitching events are not sensitive to the number of outs as far as their potential impact, their impacts are affected by the configuration of runners, and since we need to distinguish the 24 settings anyway we'll categorize pitching events by on-and-out as well.

So, there are the same twenty-four categories in pitching analysis as there are for batting. The average used for ranking in a pitching category is bases per category event, whether the event is an opponent batting event or a pitching event. One might wonder why opponent baserunning events have been excluded. That is because they are for most practical purposes beyond a pitcher's control. A pitcher may affect the number of stolen bases against them, but that subtlety is better captured by looking at opponent gain events by themselves. So, the average used for ranking in a pitching category is:

$$\frac{PB_n + OBB_n}{PE_n + OBE_n}$$, where the following legend applies:

PB = pitching bases
PE = pitching events
OBB = opponent batting bases
OBE = opponent batting events.

Ranking and Percentiles

Since the fewer the bases allowed the better, the averages in a category are sorted with the lowest at the top. Percentiles are again based on league rank in a category:

$$\%\left(R\left(\frac{PB_n + OBB_n}{PE_n + OBE_n}\right)\right).$$

Weighting and Combining Percentiles

Category percentiles are weighted by the number of events in that category, and weighted percentiles are combined by adding them and dividing that sum by the sum of the weights:

$$\frac{(PE_1 + OBE_1) \times \%\left(R\left(\frac{PB_1 + OBB_1}{PE_1 + OBE_1}\right)\right) + \ldots + (PE_n + OBE_n) \times \%\left(R\left(\frac{PB_n + OBB_n}{PE_n + OBE_n}\right)\right)}{(PE_1 + OBE_1) + \ldots + (PE_n + OBE_n)}.$$

FIELDING

Categories

A fielding event, whether an error or some other activity by a fielder that contributes to the offense's cause, will vary in impact depending on the placement of baserunners. For that reason, as in batting evaluation, the eight possible configurations of runners are distinguished in the establishment of fielding performance categories.

There are therefore eight event categories per fielding position. Given the nine fielding positions, at eight categories per position there are a total of seventy-two fielding categories.

Ideally, the configuration of runners at the instant of the fielding event should be the one used to record the data under. Note that this configuration may differ from that at the start of a play.

Average For Ranking

Ranking in a fielding category is determined by the average bases per fielding chance for that category. "Chance" is used here in its traditional sense. This ratio may be expressed as:

$$\frac{B_n}{C_n}.$$

The list of averages in a fielding category for a league is sorted so that the lowest average is at the top of the list. This reflects our desire to rank the best fielders the highest, to get the highest percentiles. The fact that the fewer bases allowed per fielding chance the better will guarantee this desired result. The ranking is:

$$R\left(\frac{B_n}{C_n}\right).$$

Percentiles

Since the lower the average bases per fielding chance the higher the league rank, the lower that average the higher the percentile. Percentile is as always based on league rank in a category, with the highest percentile at the top of the list. Percentile for a fielding category is:

$$\%\left(R\left(\frac{B_n}{C_n}\right)\right).$$

Weighted Percentiles

We can prepare to combine percentiles across categories for one fielding position or across all fielding categories by weighting each category's percentile by the number of chances in the category:

$$C_n \times \% \left(R \left(\frac{B_n}{C_n} \right) \right).$$

Combined Weighted Percentiles

The overall fielding rating for a group of categories, e.g. categories for one position or for all positions, is the sum of the weighted percentiles for each category, divided by the sum of the weights:

$$\frac{C_1 \; \% R \frac{B_1}{C_1} \; + \; \dots \; + \; C_n \; \% R \frac{B_n}{C_n}.}{C_1 + \dots + C_n}$$

CIRCUMSTANCES OF DEFENSE

Circumstances of defense events, although team-oriented as are those of circumstances of offense, are not as rare. The statistics developed for circumstances of defense are therefore correspondingly of more interest.

There are eight categories distinguished here, one for each of the eight possible configurations of baserunners. Average bases per event determines ranking. Ranking is low-to-high since the fewer the bases per event the better. Percentile is based on rank. Weighting is by number of events. The weighted percentiles are combined by adding them and dividing the result by the sum of the weights.

SUMMARY TABLE

The following table summarizes the key information for the determination of performance evaluation indices:

STATISTIC TYPE	# OF CATEG.	RANK BY	AT TOP OF LIST	WEIGHT BY	ADD/ SUBTR
Batting	24	$\frac{\text{bases}}{\text{events}}$	high	# of events	+
Running - gain	3	$\frac{\text{bases}}{\text{attempts}}$	high	# of events	+
Running - loss	3	$\frac{\text{events}}{\text{possessions}}$	high	# of events	-
Running - combined	6	$\frac{\text{3 gain - 3 loss}}{\text{total events}}$			
Offense - combined	30	$\frac{24b + 3r^{+} - 3r^{-}}{\text{total events}}$			
Circumstances of Offense	3	$\frac{\text{events}}{\text{possessions}}$	high	# of events	-
Pitching	24	$\frac{PB + OBB}{PE + OBE}$	low	# of events	+
Fielding	72	$\frac{\text{bases}}{\text{chances}}$	low	# of chances	+
Circumstances of Defense	8	$\frac{\text{bases}}{\text{events}}$	low	# of events	+

OLD HONORS IN A NEW LIGHT

Given these new statistics, the traditional individual honors can be reviewed and revised to incorporate them. In each case below there should be a minimum number of events or chances in order for a player to qualify for eligibility.

The best offense index by fielding position for a league could determine the All-Stars.

The best offense index regardless of position for a league could determine "offensive" Most Valuable Player.

If pitching events (and opponent batting events) for a pitcher are separated into two sets, one for starting and one for relieving, two complementary performance indices would result. The best pitching index for starters for a league could determine the Cy Young Award winner, while the best pitching index for relievers for a league could determine the Fireman of the Year.

The B.A.S.E.S. version of pitching evaluation supplements earned run average with "earned base average." There are two important effects of this approach. First, there is really no need any longer to distinguish starters from middle relievers from late relievers. All pitching performance is captured in context, the twenty-four pitching settings, so all pitching has a common frame of reference regardless of "point in the game." Second, and intimately tied in with this, is the solution to the long-standing problem of a pitcher leaving runners on base when being relieved. Such pitchers have been at the mercy, statistically, of the bullpen. A reliever can negatively affect the departing pitcher's record without affecting their own record by allowing the inherited runners to score. In B.A.S.E.S. analysis the departing pitcher is only accountable for events occurring while they are in the game, while a relief pitcher is accountable for events once they enter a game. If a reliever comes in when there are runners on base, what happens from there is the reliever's responsibility. Since context is taken into account, the fact that there are runners on base is factored into the reliever's record automatically and appropriately.

The best fielding index by position for a league could determine the Gold Glove winners. Whereas with fielding percentage all errors are of equal weight, in B.A.S.E.S. errors (and other fielding events not even accounted for traditionally, for that matter) weigh in proportion to their actual impact (in bases that the offense benefits by).

Can B.A.S.E.S. statistics unequivocably determine a league's Most Valuable Player? Can they decide whether the player with the best offensive season in a league or the pitcher with the best pitching season in the league should get the MVP? No to both questions. They're still judgment calls. B.A.S.E.S. analysis can only help in the decision process, hopefully providing even better illumination on the subject than that given by previous measures.

PARK FACTOR

If one wishes to, one can derive and apply a "park factor" for each ballpark played in. Such a factor would reflect the relative ease or difficulty of accumulating bases by situation in a given park. Applying such a factor would adjust players' performance by situation by park to take into account that relative ease or difficulty.

The way to arrive at such a park factor is straightforward. For any given setting or category of events, compare the average performance in that category in that park in that season to the average performance in that category for all parks in that league in that season. Once the ratio of the average bases per event in the category for the given park versus for the league has been determined, all events occurring in that situation in that park can be weighted by the inverse of that ratio.

For example, suppose that for a "low-scoring" ballpark, for the situation where there is no one on and no one out the average bases per batting event is only 80% of what it is for the league as a whole. That's equivalent to 4/5. Invert the fraction to get 5/4 and multiply that times each player's average bases in that situation in that park in that season. The way to multiply by the inverse of a number, for example 80%, or .80, is to divide by the number itself. So, dividing a player's average bases by .80 is the equivalent of multiplying it by 5/4.

DATA CAPTURE

INTRODUCTION

Two forms can be used to capture play-by-play data so that it may readily provide B.A.S.E.S. information. The forms, the playlog and the lineup chart, are designed for ease of use, requiring the minimum of effort and technical knowledge of the B.A.S.E.S. system on the part of the scorekeeper. Nevertheless they can supply all of the data needed in order for a person or computer program to extract the various statistics derived in B.A.S.E.S. analysis.

The bulk of the data is recorded on the playlog. The lineup chart merely gives the key to tying game events to particular players. Refer to the sample forms in the section "An Example of a Series of Plays" for the following discussion.

THE LINEUP CHART

We'll talk about the lineup chart first, since it is the simpler of the two forms and will provide a lead-in to the playlog.

Both teams in a game get their own lineup chart. The chart is dated and identifies the team. If it is the second game that day then that is indicated. Reading down, there is the batting order. There is room for two extra batters per spot in the batting order. Below the batting order is room for several pitchers, useful when the designated hitter is in effect -- otherwise the pitchers would appear in whatever spot that they're in, in the batting order. Usually it's the ninth spot, so the ninth spot on down the chart can hold quite a few pitchers' names and entries.

The entries going across are: a player's uniform number, their fielding position (e.g., "2B" rather than "4"), and the point in the game at which they started playing that position. Points in a game are associated with particular plays. Plays are uniquely identified with a three-part ID: half-inning, inning, and play number. Half-inning is either "T" or "B", for "top" or "bottom." Inning is inning number. Play number is the play number that half-inning. For example, the first play in a half-inning is play number 1 for that half-inning. So, the unique play-ID for the first play in the top of the ninth inning is "T 9 1," the top of the ninth inning, play one.

Moving across the lineup chart we find two extra "as of" entries for each player. These allow for a player changing fielding positions during a game. By means of the lineup chart we can tell who a batter or a fielder is for any play during a game, given the play-ID and a batter or fielder number involved. The same reasoning applies if the pitcher is involved in a play.

The lineup chart accommodates keeping track of substitutions of any kind. Any activities by substitutes can be assigned properly to them whether they be relief pitchers, pinch hitters, pinch runners, or defensive replacements, since each player in the batting order, in a fielding position, or pitching is uniquely identified as of each play in a game.

THE PLAYLOG

THE DATA ITEMS

Besides carrying general information such as the date, the team, and whether it is the first or second game that day, the playlog contains the following data items, any of which could be relevant and therefore filled in, for each play:

PLAY-ID
  HALF-INNING
  INNING NUMBER
  PLAY NUMBER
FIELDING PLAY
BATTER INFORMATION
  BATTER NUMBER
  LEFT/RIGHT
  BASES +
  BASES -
  NEW BASE
RUNNER INFORMATION (4 OCCURRENCES)
  BATTER NUMBER
  BASES +
  BASES -
  NEW BASE
OFFENSE BASES -
PITCHER BASES +
FIELDER INFORMATION (2 OCCURRENCES)
  POSITION NUMBER
  BASES +
  RUNNERS
DEFENSE BASES +

HOW TO MAKE ENTRIES

Both teams in a game have a playlog. When a team is at bat, every play's activity is recorded on its playlog, one play per line.

Each play is identified by a play-ID. Half-inning, the "T" or "B" for "top" or "bottom," need only be entered for the first play of the game on offense for a team, since it will never change during a game. Inning number need only be filled in for the first play of each team-at-bat. Play number starts over with "1" each team-at-bat and goes up by one with each play in an team-at-bat.

The fielding play is entered more or less as is done traditionally. There are five spaces on a line for the fielder numbers on a play, so a 3-6-3 double play could be entered with the hyphens. Four or more players involved in a play would not fit if hyphens were used, so there are two choices: continue the play on as many lines as it takes to hold the fielder numbers with hyphens, or don't use hyphens. Even when not using hyphens, if there are more than five players involved, a rare event, then the fielding information could be continued on a new line. Although there is nothing special about the number five, it should be enough to hold the fielder numbers without hyphens for almost all plays, since all fielder numbers are a single digit, so that will be the recommended method. On the rare occasion when there are more than five players involved, the fielding information can be continued on the next line. The same play-ID would be used on any continuation lines.

The batter information sees the heaviest use, as the batter is involved in most of the activity in a game. Batter number is position in the batting order. Batter number is filled in whenever a batter has an event or reaches base due to some other agent. Left/right is an "L" or an "R" for whether the batter batted lefthanded or righthanded. This is filled in when a batter has a batting event. Bases+ is filled in with the number of bases contributed, when the batter has contributed a positive number of bases or zero bases on a play. Of course zero bases contributed is not the same as no involvement. In one case there is a batting event, whereas in the other case there is not. Bases- is filled in when the batter "contributes" a negative number of bases on a play. "New base," when filled in, is either a dash or a base number. A dash is entered when the batter is put out on the play before first reaching base safely. A base number is entered when the batter reaches base safely. Whichever base the batter ends up at is the one represented in "new base."

There are four runner information entries. Each contains four items: batter number, bases+, bases-, and new base. When a runner has a running event or simply ends up at a new base after a play, an entry is filled in for them. The four entries are filled in left-to-right, one for each runner needing information recorded due to a play. A convention for which runners to put in which entries is: capture the information for runners starting with the batter/runner and moving counterclockwise around the bases, using the leftmost runner entry and filling entries left-to-right. In other words, suppose that there is information to be captured about the runner on first and the runner on third. That means that the two runner information entries on the left would be used. The information for the runner from first would be leftmost, then the information for the runner from third would be next to it on its right. Examples coming shortly will help clarify this.

Following the runner entries is an entry for circumstances of offense. It consists of one item, bases-. Any bases due to circumstances of offense on a play would be entered here.

Next comes pitching bases+. Any bases caused to be advanced by the offense due to pitching events, i.e. hit batsmen, wild pitches, or balks, are entered here.

Next there are two fielding entries. Each contains three items: fielding position, bases+, and "runners." Fielding position is the traditional number standing for the defensive position of the fielder responsible for an event, for example "9" standing for the rightfielder. Bases+ is the number of bases that the offense benefitted due to the fielding event. "Runners" records the configuration of baserunners at the moment of the fielding event, when different from what it was at the start of the play. A handy one-digit shorthand can be used to stand for each of the eight configurations.It is based on the number of bases possessed by the offense given the placement of the runners:

RUNNERS ON	CODE
none on	0
1st	1
2nd	2
3rd	3_1
1st & 2nd	3_2
1st & 3rd	4
2nd & 3rd	5
bases loaded	6

The only ambiguity is when the total is three bases possessed. The two situations where that occurs, a runner on third or runners on first and second, are distinguished by the subscripts "1" and "2" respectively. Which subscript goes with which situation is easy to remember since the "1" and the "2" correspond to the number of runners in each situation. In other words, for three bases possessed with one runner on (a runner on third) the code is 3_1 while for three bases possessed with two runners on (runners on first and second) the code is 3_2 .

The "runners" item in a fielding information entry need only be filled in when the configuration of runners at the time of the fielding event has changed from what it was at the start of the play. If the configuration hasn't changed, then it can be deduced from the previous plays.

Finally, there is an entry for circumstances of defense, called defense bases+.

EXAMPLES OF SINGLE PLAYS

Now let's look at some examples. We'll use the same 75 plays that we used in the section on scoring. Please refer to those plays, as the following playlog entries correspond one-to-one with them. Note that the play numbering in the following 75 examples is not the standard way of numbering plays. It is only done this way for this set of examples. After seeing how various events are entered we'll work with a series of plays, to get a feel for the flow of action as it is actually captured on the playlog.

DATE:
TEAM:
GAME:

HALF INNING	INNING NUMBER	PLAY NUMBER	FIELDING PLAY	BATTER NUMBER	LEFT/RIGHT	BASES+	BASES-	NEW BASE	BATTER NUMBER	BASES+	BASES-	NEW BASE	BATTER NUMBER	BASES+	BASES-	NEW BASE	BATTER NUMBER	BASES+	BASES-	NEW BASE	BATTER NUMBER	BASES+	BASES-	NEW BASE	OFFENSE BASES-	PITCHER BASES+	FIELDER NUMBER	BASES+	RUNNERS	FIELDER NUMBER	BASES+	RUNNERS	DEFENSE BASES+
		1		4		1		1																									
		2		4		2		2																									
		3		4		3		3																									
		4		4		4		4																									
		5		4		1		1																									
		6		4				1																		1							
		7		4				1																									1
		8		4		0		-																									
		9		4		2		1	3			2																					

HALF INNING	INNING NUMBER	PLAY NUMBER	FIELDING PLAY	BATTER NUMBER / LEFT/RIGHT / BASES+ / BASES- / NEW BASE	BATTER NUMBER / BASES+ / BASES- / NEW BASE	BATTER NUMBER / BASES+ / BASES- / NEW BASE	BATTER NUMBER / BASES+ / BASES- / NEW BASE	BATTER NUMBER / BASES+ / BASES- / NEW BASE	OFFENSE BASES-	PITCHER BASES+	FIELDER NUMBER / BASE+ / RUNNERS	FIELDER NUMBER / BASE+ / RUNNERS	DEFENSE BASES+
		10		4 3 1	3 3								
		11		4 1 -	3 2								
		12		4 0 1	3 -								
		13		4 1-	3 -								
		14		4 1 -	3 2								
		15		4 1 -	3 2								
		16		4 0 -									
		17		4 3 1	3 2	2 3							
		18		4 4 1	3 2	2 4							
		19		4 5 1	3 3	2 4							
		20		4 2 -	3 2	2 3							
		21		4 1 1	3 -	2 3							
		22		4 0 1	3 2	2 -							
		23		4 0 -	3 -	2 3							
		24		4 1-	3 2	2 -							
		25		4 21	3 -	2 -							
		26		4 0 -	3 -	2 4							
		27		4 4 1	3 4	2 4							
		28		4 4 1	3 2	2 3	1 4						
		29		4 5 1	3 2	2 4	1 4						
		30		4 6 2	3 4	2 4	1 4						
		31		4 10 4	3 4	2 4	1 4						
		32		4 1 -	3 -	2 3	1 4						
		33		4 0 -									
		34		4 0 -									
		35		4 0 -									
		36		4 1-	3 -								
		37		4 0 -	3 -								
		38		4 0 -	3 -								
		39		4 1 -	3 -								
		40		4 3-	3 -	2 -							
		41		4 2-	3 -	2 -							
		42			3 1-								
		43			3 2-								
		44			3 3-								

HALF INNING	INNING NUMBER	PLAY NUMBER	FIELDING PLAY	BATTER NUMBER / LEFT/RIGHT / BASES+ / BASES- / NEW BASE	BATTER NUMBER / BASES+ / BASES- / NEW BASE	BATTER NUMBER / BASES+ / BASES- / NEW BASE	BATTER NUMBER / BASES+ / BASES- / NEW BASE	BATTER NUMBER / BASES+ / BASES- / NEW BASE	OFFENSE BASES-	PITCHER BASES+	FIELDER NUMBER / BASES+ / RUNNERS	FIELDER NUMBER / BASES+ / RUNNERS	DEFENSE BASES+
		45			31 2								
		46			31 3								
		47			31 4								
		48			3 1-								
		49			3 2-								
		50			3 3-								
		51			3 2					1			
		52			3 2					1			
		53			3 2	2 3				2			
		54			3 2	2 3	1 4			4			
		55		4 0 1							21		
		56		4 0 n							9n		
		57		4 m *							9n		
		58		4 m m	4 m-								
		59		4 0 n	4 n-						9n		
		60		4 m *	4 *-						9n		
		61		4 3 1	4 1-	3 4							
		62		4 3 2	3 4								1
		63		4 2 2	3 2-								1
		64		4 1 1	3 2						61		
		65		4 1 1	3 2								1
		66			31 3						21		
		67		4 1 -	3 2-								
		68		4 1 -	3 2-	2 3-							
		69		4 1 1	4 1-								
		70		4 3 1	4 1-	3 3							
		71			31 2	3 2-							
		72		4 0 2							51	511	
		73		4 0 2							21	211	
		74		4 3 3	3 4						71		1
		75		4 5 2	3 2-	2 4	1 4						1

* = m + n

AN EXAMPLE OF A SERIES OF PLAYS

Now we're ready to look at that series of plays. We'll use a playlog from an actual game, the New York Yankees vs. the Texas Rangers on July 26, 1983. We'll just look at the visiting team's playlog. Also, we'll leave out the fielding play entry for each play -- partly because these are nothing new, partly because the form will be less cluttered by numbers, but mainly because they weren't recorded. So, though we'd like to say what kind of outs occurred or at least who got assists and putouts, all that we can tell for sure are events and bases. If we were worried about fielding evaluation here, then we'd need that information. Similarly, in some cases here we won't be able to reconstruct whether a batter reached base on a single or a walk, as with plays T-1-1, T-2-2, T-3-2, T-4-6, T-5-2, T-6-1, and T-8-2. Not that we're supposed to be able to, of course, given the limited information that we've got.

Under official scoring conditions all appropriate entries would be filled in, and any other counts or statistics of interest not covered by B.A.S.E.S. could be recorded to supplement the playlog. An example of such a count would be runs batted in. Runs scored, although not explicitly recorded on the playlog, are easily derived from it. Just by simple inspection, any "new base" of "4" is a run scored.

Other counts and statistics can be derived from the playlog. Examples of such extraction are: determining the setting at the start of a play, and safe possessions of a base by each player. Suffice it to say for now that by keeping track of forfeitures and "new bases," as we do on the playlog, we can derive the number of outs and the occupied bases at the end of any given play, and therefore as of the start of the next play. Thus, settings for batting activity evaluation can be derived. Similarly, given recording of "new base" when a batter or runner ends up safely on a base, or the amount of forfeiture when a runner is put on the bases due to a running event, we can derive bases safely possessed.

So, on one hand we don't have the complete information that we would ordinarily have and cannot therefore give an exact play-by-play description of what transpired in the game (for the visitors), on the other hand, for our purposes at this point we don't really care. We just want to get a feel for the flow of events as captured on a playlog. For all of that, it is interesting to see how much we can reconstruct.

Please refer to the sample playlog and lineup chart for the Yankees below as we roughly describe the course of events recorded there:

Top of the first:
- Campaneris gets on 1st.
- Smalley fails to advance him.
- Campaneris is out stealing.
- Winfield is out to retire the side.

T2:
- Baylor makes an out.
- Piniella reaches first.
- Balboni singles, moving Piniella to 3rd.
- Mumphrey grounds into a double play.

T3:
- Cerone leads off with an out.
- Robertson gets on 1st.
- Campaneris forces Robertson at 2nd.
- Campaneris is out stealing.

T4:

Smalley reaches first.
Winfield doubles him to 3rd.
Baylor fails to advance the runners.
Piniella is intentionally walked to load the bases, Kemp goes in to run for Piniella.
Balboni hits a grand slam home run.
Mumphrey gets on 1st.
Cerone forces Mumphrey.
Robertson makes the third out.

etc.

T8:

Winfield is out.
Baylor gets on 1st.
Baylor goes to second due to the pitcher - probably on a wild pitch, less probably due to a balk.
Kemp walks.
Balboni fails to advance the runners.
Mumphrey singles, scoring Baylor, moving Kemp to 2nd.
Cerone ends the inning.

etc.

```
DATE: July 26, 1983
TEAM: New York Yankees
GAME: First
```

HALF INNING	INNING NUMBER	PLAY NUMBER	FIELDING PLAY	BATTER NUMBER / LEFT/RIGHT / BASES+ / BASES- / NEW BASES	BATTER NUMBER / BASES+ / BASES- / NEW BASES	BATTER NUMBER / BASES+ / BASES- / NEW BASES	BATTER NUMBER / BASES+ / BASES- / NEW BASES	BATTER NUMBER / BASES+ / BASES- / NEW BASES	OFFENSE BASES-	PITCHER BASES+	FIELDER NUMBER / BASES+ / RUNNERS	FIELDER NUMBER / BASES+ / RUNNERS	DEFENSE BASES+
T	1	1		1 1 1									
		2		2 0 -									
		3			1 1-								
		4		3 0 -									
	2	1		4 0 -									
		2		5 1 1									
		3		6 3 1	5 3								
		4		7 0 -	6 -								
	3	1		8 0 -									
		2		9 1 1									
		3		1 0 1	9 -								
		4			1 1-								

HALF INNING	INNING NUMBER	PLAY NUMBER	FIELDING PLAY	BATTER NUMBER	LEFT/RIGHT	BASES+	BASES-	NEW BASE	BATTER NUMBER	BASES+	BASES-	NEW BASE	BATTER NUMBER	BASES+	BASES-	NEW BASE	BATTER NUMBER	BASES+	BASES-	NEW BASE	BATTER NUMBER	BASES+	BASES-	NEW BASE	OFFENSE BASES-	PITCHER BASES+	FIELDER NUMBER	BASES+	RUNNERS	FIELDER NUMBER	BASES+	RUNNERS	DEFENSE BASES+
	4	1		2		1		1																									
		2		3		4		2	2			3																					
		3		4		0		–																									
		4		5				1																									1
		5		6		10		4	5			4	3			4	2			4													
		6		7		1		1																									
		7		8		0		1	7			–																					
		8		9		0		–																									
	5	1		1		0		–																									
		2		2		1		1																									
		3		3		0		–																									
		4		4		0		–																									
	6	1		5		1		1																									
		2		6		0		–																									
		3		7		0		–																									
		4		8		0		–																									
	7	1		9		0		–																									
		2		1		0		–																									
		3		2		0		–																									
	8	1		3		0		–																									
		2		4		1		1																									
		3							4			2														1							
		4		5		1		1																									
		5		6		0		–																									
		6		7		4		1	5			2	4			4																	
		7		8		0		–																									
	9	1		9		0		–																									
		2		1		0		–																									
		3		2		0		–																									

DATE: July 26, 1983
TEAM: New York Yankees
GAME: First

BATTER NO.	PLAYER NAME	UNIFORM NUMBER	POSITION	AS OF 1/2 INN.	AS OF INNING	AS OF PLAY NO.	POSITION	AS OF 1/2 INN.	AS OF INNING	AS OF PLAY NO.	POSITION	AS OF 1/2 INN.	AS OF INNING	AS OF PLAY NO.
1	Campaneris		2B	T	1	1								
2	Smalley		3B	T	1	1								
3	Winfield		LF	T	1	1								
4	Baylor		DH	T	1	1								
5	Piniella		RF	T	1	1								
	Kemp		RF	T	4	5								
6	Balboni		1B	T	1	1								
	Mattingly		1B	B	8	1								
7	Mumphrey		CF	T	1	1								
8	Cerone		C	T	1	1								
9	Robertson		SS	T	1	1								
P	Fontenot		P	T	1	1								

SOME ACTUAL RESULTS (1989 Addendum)

Some of the software needed to produce B.A.S.E.S. information has now been designed and written. The software was written to use play-by-play data as provided by Project Scoresheet (P.O. Box 12009, Lansing, MI 48901), so much of the discussion on data capture is not applicable here. Project Scoresheet provides game data for a team or league for a given season. The author chose the American League's 1986 season for this test of the B.A.S.E.S. methods. The programs only process batting data and that processing is somewhat simplified on the subtler, more complex determinations. Nevertheless, the author feels that the results of this information generation effort should come quite close to what a painstakingly accurate system would produce. Regretfully, baserunning information has not been generated at this time. Thus, overall offense contribution cannot be determined exactly. Again, because of the preponderance of batting events in proportion to all others, and because this somewhat simplified approximation is nevertheless a good one, the statistics shown should reflect fairly well how the players fared that season.

The following table thus contains the slightly simplified batting rankings for the entire American League for 1986:

LEAGUE BATTING SUMMARY

LEAGUE: American
SEASON: 1986

RANK	PLAYER	BASES TOTAL	EVENTS TOTAL	SUM OF WEIGHTED PERCENTILES	OVERALL PERCENTILE
1	Lollar	2	1	0.978571	0.978571
2	Grubb	275	241	202.176544	0.838907
3	Mattingly	710	730	605.174988	0.829007
4	Pasqua	319	326	260.139862	0.797975
5	Barfield	607	658	519.434692	0.789414
6	Puckett	587	712	547.558838	0.769043
7	Boggs W	604	679	514.625732	0.757917
8	Bell Geo	610	685	508.864197	0.742867
9	Parrish Ln	312	366	271.521667	0.741862
10	Phelps	408	429	317.488708	0.740067
11	Gaetti	571	651	477.383209	0.733308
12	Carter J	627	701	511.783691	0.730077
13	Bradley P	473	605	441.677063	0.730045
14	Brett	465	510	372.046570	0.729503
15	Lynn	412	453	325.384521	0.718288
16	Henderson R	529	697	500.572601	0.718182
17	Hall M	414	471	337.314148	0.716166
18	Hrbek	531	619	443.118256	0.715861
19	Parrish Lr	467	515	368.400665	0.715341
20	Griffey	174	218	154.793350	0.710061
21	Gibson K	467	511	362.543854	0.709479

RANK	PLAYER	BASES TOTAL	EVENTS TOTAL	SUM OF WEIGHTED PERCENTILES	OVERALL PERCENTILE
22	Yount	439	584	412.879852	0.706986
23	Walker G	279	310	217.793976	0.702561
24	Evans Dw	578	630	439.868744	0.698204
25	Bernazard	499	625	433.463348	0.693541
26	Trammell	523	644	445.731598	0.692130
27	Porter D	167	177	122.411942	0.691593
28	O'Brien P	550	631	435.508850	0.690188
29	Dwyer	166	186	127.709160	0.686608
30	Ripken	568	698	476.807770	0.683106
31	White F	504	613	418.683228	0.683007
32	Henderson D	327	426	290.176025	0.681164
33	Hassey	344	387	263.160950	0.680002
34	Downing	562	612	415.978607	0.679704
35	Joyner	577	664	451.203552	0.679523
36	Rice	608	684	460.960052	0.673918
37	Tartabull	528	575	385.730530	0.670836
38	Pagliarulo	447	551	369.386444	0.670393
39	Murray E	510	571	381.232605	0.667658
40	Davis A	467	549	366.158508	0.666955
41	Incaviglia	479	600	399.895050	0.666492
42	Whitt	350	430	285.284973	0.663453
43	Deer	471	540	357.200714	0.661483
44	Smalley	418	527	347.553345	0.659494
45	Evans Da	497	596	392.716156	0.658920
46	McDowell O	423	637	418.148499	0.656434
47	Stanley M	27	34	22.308802	0.656141
48	Easler	443	533	347.157257	0.651327
49	Jackson Re	420	503	327.502228	0.651098
50	Baines	493	607	392.712311	0.646972
51	Snyder C	357	433	280.034668	0.646731
52	Winfield	556	641	414.477600	0.646611
53	Davis Mike	398	530	341.421631	0.644192
54	DeCinces	485	566	364.172455	0.643414
55	Jones Ru	345	463	297.773529	0.643139
56	Johnson Cliff	321	389	249.838882	0.642259
57	Balboni	453	559	357.867523	0.640192
58	Whitaker	488	646	412.182587	0.638054
59	Coles	474	578	368.630463	0.637769
60	Grich	257	362	230.355911	0.636342
61	Sierra	335	407	258.757324	0.635767
62	Laudner	174	220	139.563339	0.634379
63	Tabler	387	502	317.813934	0.633096
64	Clark D	55	68	42.916950	0.631132
65	Tettleton	211	261	164.366150	0.629755
66	Jacoby	515	636	400.073975	0.629047
67	Thomas G	252	372	232.104172	0.623936
68	Presley	523	653	406.779236	0.622939
69	Murphy Dw	313	392	243.885757	0.622158

RANK	PLAYER	BASES TOTAL	EVENTS TOTAL	SUM OF WEIGHTED PERCENTILES	OVERALL PERCENTILE
70	Mulliniks	308	393	243.996902	0.620857
71	Hendrick	247	311	192.962738	0.620459
72	Brunansky	474	650	403.209473	0.620322
73	Howell Jk	151	175	108.440384	0.619659
74	Burleson	239	310	192.028870	0.619448
75	Buechele	353	507	313.999756	0.619329
76	Washington R	56	81	50.132401	0.618919
77	Baylor	513	644	397.767944	0.617652
78	Moseby	503	659	406.124969	0.616275
79	Fletcher	428	590	363.290436	0.615746
80	Bell J	21	16	9.822589	0.613912
81	Greenwell	35	40	24.446970	0.611174
82	Molitor	354	482	294.006012	0.609971
83	Traber	206	233	142.106049	0.609897
84	Lowry	132	170	103.463036	0.608606
85	Buckner	595	668	406.532196	0.608581
86	Manning	161	224	136.223511	0.608141
87	Seitzer	89	115	69.788391	0.606856
88	Browne J	27	24	14.510897	0.604621
89	Franco Ju	480	635	383.378723	0.603746
90	Sheets	314	359	216.313477	0.602544
91	Randolph	423	598	360.116882	0.602202
92	Washington C	95	142	85.483315	0.601995
93	Dodson	20	15	9.007699	0.600513
94	Dempsey	249	379	227.251770	0.599609
95	Smith Lo	396	558	334.334045	0.599165
96	Valle	68	60	35.919216	0.598654
97	Upshaw	473	655	391.551941	0.597789
98	Schofield	384	522	311.318848	0.596396
99	Woods A	37	32	19.025042	0.594533
100	Boston	150	221	131.043518	0.592957
101	Sax D	12	11	6.493312	0.590301
102	Slaught	267	336	198.040359	0.589406
103	Steinbach	19	16	9.416333	0.588521
104	Lombardi	38	40	23.480747	0.587019
105	Karkovice	83	108	63.317139	0.586270
106	Canseco	589	673	393.610657	0.584860
107	Butler	486	678	396.434875	0.584712
108	Cerone	149	241	140.629639	0.583525
109	Bush	325	396	230.703247	0.582584
110	Lansford	443	633	368.602356	0.582310
111	Kingery	139	221	128.360504	0.580817
112	Foster	35	54	31.334051	0.580260
113	Bradley S	188	233	135.055435	0.579637
114	Kittle	311	418	242.270538	0.579595
115	Thornton A	352	474	273.889313	0.577826
116	Lacy L	358	535	308.961578	0.577498
117	Lemon	329	446	256.331177	0.574734

RANK	PLAYER	BASES TOTAL	EVENTS TOTAL	SUM OF WEIGHTED PERCENTILES	OVERALL PERCENTILE
118	Beniquez	285	390	222.481583	0.570466
119	Roenicke G	141	164	93.185463	0.568204
120	Laga	38	49	27.801527	0.567378
121	Orta	287	358	202.153946	0.564676
122	Ward G	319	411	231.946854	0.564348
123	Hernandez L	21	23	12.978164	0.564268
124	Fernandez T	527	723	407.546356	0.563688
125	Bochte	351	470	264.636749	0.563057
126	Oglivie	298	378	212.560043	0.562328
127	Castillo C	173	216	121.241203	0.561302
128	Gedman	386	492	276.001007	0.560978
129	Brantley	57	113	63.326851	0.560415
130	Nokes	16	24	13.414411	0.558934
131	Willard	163	191	106.607170	0.558153
132	Leach R	228	263	146.681381	0.557724
133	Hoffman	14	26	14.441699	0.555450
134	Hairston	180	249	138.043045	0.554390
135	Sakata	31	38	21.009792	0.552889
136	Young Mike	318	421	231.478592	0.549830
137	Kingman	420	598	328.690430	0.549650
138	Heath	77	103	56.534767	0.548881
139	Hulett	339	551	301.288605	0.546803
140	Phillips	390	529	288.235321	0.544868
141	Griffin Alf	435	641	348.965515	0.544408
142	Gagne	358	518	281.748810	0.543917
143	Barrett	520	712	385.798370	0.541852
144	Bathe	60	111	59.997566	0.540519
145	Armas	351	450	240.745422	0.534990
146	Law R	251	339	180.844284	0.533464
147	Engle	51	93	49.503059	0.532291
148	Hill D	251	365	193.980560	0.531454
149	Craig Rd	9	12	6.309550	0.525796
150	Sheridan	166	257	134.626862	0.523840
151	Petralli	95	142	74.179192	0.522389
152	Riles	403	587	305.558716	0.520543
153	Pecota	19	33	17.117977	0.518727
154	Braggs	138	219	113.106514	0.516468
155	Moore C	204	262	135.239182	0.516180
156	Herndon	234	313	160.918961	0.514118
157	Paciorek	135	217	111.373871	0.513244
158	Berra	86	120	61.379456	0.511495
159	Morman	109	178	90.713821	0.509628
160	Bonilla B	190	268	136.498856	0.509324
161	Bergman	99	150	76.146866	0.507646
162	Salas	191	282	142.900574	0.506740
163	Nelson Rob	4	10	5.039717	0.503972
164	White D	33	57	28.695368	0.503428
165	Wilson W	415	665	334.584869	0.503135

RANK	PLAYER	BASES TOTAL	EVENTS TOTAL	SUM OF WEIGHTED PERCENTILES	OVERALL PERCENTILE
166	Harrah	244	339	170.348450	0.502503
167	Gantner	340	534	267.625916	0.501172
168	Sveum	263	355	177.749603	0.500703
169	Tolleson	364	544	270.608063	0.497441
170	Shepherd	39	73	36.306755	0.497353
171	Schroeder	136	231	114.399315	0.495235
172	Romine	24	39	19.246553	0.493501
173	Brookens	200	309	151.649002	0.490773
174	McRae	211	294	144.107117	0.490160
175	Ryal	26	33	16.152178	0.489460
176	Collins	304	474	229.870316	0.484958
177	Spilman	43	53	25.501860	0.481167
178	Motley	130	228	109.704453	0.481160
179	Lombardozzi	312	512	245.376572	0.479251
180	Pettis	431	626	299.400177	0.478275
181	Kearney	146	225	107.587349	0.478166
182	Reed J	103	184	87.747070	0.476886
183	Shelby	290	426	202.369720	0.475046
184	Cooper C	424	586	278.138763	0.474640
185	Wynegar	170	224	106.146080	0.473866
186	Quirk	144	234	110.784668	0.473439
187	Biancalana	113	209	98.835381	0.472897
188	Nixon O	76	110	51.725113	0.470228
189	O'Malley	132	199	93.446716	0.469581
190	Hatcher M	209	338	157.186829	0.465050
191	Moses	263	439	203.596039	0.463772
192	McGwire	40	57	26.416449	0.463446
193	Ready	45	89	41.082577	0.461602
194	Robidoux	152	214	98.657585	0.461017
195	Tillman	34	42	19.267069	0.458740
196	Meacham	111	182	83.212463	0.457211
197	Boone	323	502	229.488800	0.457149
198	Bando	202	289	132.049515	0.456919
199	Miller D	34	62	28.305361	0.456538
200	Rohn	9	11	5.015336	0.455940
201	Iorg G	257	350	159.116104	0.454617
202	Cangelosi	333	518	233.135147	0.450068
203	Brewer M	6	20	8.997698	0.449885
204	Householder	80	89	39.876781	0.448054
205	Narron	64	106	47.486813	0.447989
206	Garcia D	287	442	197.706390	0.447300
207	Sundberg	314	490	219.040405	0.447021
208	Owen S	383	590	262.997742	0.445759
209	Dodd	16	15	6.682254	0.445484
210	Calderon	104	173	76.825142	0.444076
211	Baker Dy	173	270	117.623032	0.435641
212	Gallego	29	40	17.235161	0.430879
213	Gruber	87	152	65.415497	0.430365

RANK	PLAYER	BASES TOTAL	EVENTS TOTAL	SUM OF WEIGHTED PERCENTILES	OVERALL PERCENTILE
214	Lyons S	170	272	115.983498	0.426410
215	Polidor	9	20	8.522404	0.426120
216	Stenhouse	28	34	14.360826	0.422377
217	Gerhart	40	75	31.566313	0.420884
218	Stefero	88	137	57.445042	0.419307
219	David	3	5	2.090408	0.418082
220	Jackson B	59	89	37.178722	0.417738
221	Wilfong	209	312	130.213348	0.417350
222	Espino	24	40	16.648491	0.416212
223	Hearron	15	26	10.797287	0.415280
224	Cruz Ju	153	256	106.189285	0.414802
225	Hengel	37	65	26.914692	0.414072
226	Salazar L	2	8	3.300529	0.412566
227	Rayford	119	228	93.530212	0.410220
228	Gross W	1	3	1.220016	0.406672
229	Jones Ri	28	39	15.714911	0.402946
230	Bell T	2	5	1.995198	0.399040
231	Felder	91	173	69.023041	0.398977
232	Diaz E	6	14	5.556915	0.396923
233	Bonilla J	187	313	124.061661	0.396363
234	Fields	39	47	18.561375	0.394923
235	Fisk	311	483	189.964386	0.393301
236	Tolman	25	43	16.787363	0.390404
237	Perconte	47	84	32.520355	0.387147
238	Reynolds H	258	483	186.214874	0.385538
239	Nichols	92	150	57.791573	0.385277
240	Mercado	51	111	42.699211	0.384678
241	Martinez B	105	185	70.710762	0.382220
242	Mullins	25	44	16.786247	0.381506
243	Romero E	141	261	99.428978	0.380954
244	Williams K	10	32	12.053420	0.376669
245	Wright G	57	110	41.265640	0.375142
246	Brower	2	9	3.372187	0.374687
247	Salazar A	187	311	116.113213	0.373354
248	Quinones R	189	333	124.022758	0.372441
249	Kunkel	9	13	4.805398	0.369646
250	Allanson	197	322	118.897499	0.369247
251	Fielder	59	89	32.686008	0.367259
252	Cochrane	26	67	24.590519	0.367023
253	Javier	71	130	47.278854	0.363683
254	Baker Dg	12	30	10.907923	0.363597
255	Skinner J	201	335	120.011086	0.358242
256	Wiggins	135	270	95.728806	0.354551
257	McGriff F	1	5	1.764750	0.352950
258	Beane	102	194	68.072220	0.350888
259	Fischlin	53	116	40.683014	0.350716
260	Lee M	38	85	29.638947	0.348693
261	Nelson R	8	12	4.148965	0.345747

RANK	PLAYER	BASES TOTAL	EVENTS TOTAL	SUM OF WEIGHTED PERCENTILES	OVERALL PERCENTILE
262	Yeager	81	145	49.734482	0.342996
263	Wilkerson	130	248	84.433342	0.340457
264	Bonnell	22	53	17.777943	0.335433
265	Hill M	8	20	6.611513	0.330576
266	Giles	4	11	3.635896	0.330536
267	Guillen	335	575	187.463852	0.326024
268	Adduci	7	13	4.206396	0.323569
269	Davidson	29	77	24.358782	0.316348
270	McLemore	1	6	1.893557	0.315593
271	Cowens	45	87	27.404909	0.314999
272	Paris	1	10	3.135668	0.313567
273	Sanchez A	8	17	5.330605	0.313565
274	Peters	17	44	13.795768	0.313540
275	Little	34	82	25.675230	0.313113
276	Castillo J	31	61	19.078466	0.312762
277	Jones L	19	54	16.481579	0.305214
278	Cotto	42	83	24.769464	0.298427
279	Pardo	19	51	14.798421	0.290165
280	Pryor	45	117	33.802608	0.288911
281	Sullivan M	67	130	37.371174	0.287471
282	Johnson Rondin	16	31	8.420527	0.271630
283	DeJesus	1	5	1.345366	0.269073
284	Harper B	18	41	10.995895	0.268193
285	Espinoza	19	45	11.913373	0.264742
286	Ramos D	39	109	28.075850	0.257577
287	Williams E	4	7	1.707018	0.243860
288	Stapleton	20	42	10.205849	0.242996
289	Zuvella	23	57	13.800583	0.242115
290	Gutierrez J	58	151	36.185989	0.239642
291	Jones B	11	23	5.430016	0.236088
292	Taylor D	0	2	0.426373	0.213186
293	Braggs G	3	12	2.330992	0.194249
294	Madison	1	8	1.530793	0.191349
295	Dawley	0	2	0.346097	0.173049
296	Jones Ro	2	21	3.414154	0.162579
297	Tarver	8	26	3.899567	0.149983
298	Henderson S	7	27	3.945394	0.146126
299	Fontenot	0	1	0.142857	0.142857
299	Hudler	0	1	0.142857	0.142857
301	Pittaro	3	21	2.764382	0.131637
302	Nichols C	0	5	0.596715	0.119343
303	Kiefer	0	6	0.548308	0.091385
304	Quinones	0	10	0.885432	0.088543

The following table shows the setting-by-setting information for a single player, Don Mattingly:

OUT	ON	BASES	EVENTS	AVG. BASES	RANK	PERCENTILE	EVENTS x PERCENTILE
0		84	127	0.661417	18	0.942953	119.755028
0	1	54	53	1.018868	70	0.746324	39.555145
0	2	14	15	0.933333	112	0.529661	7.944915
0	3	0	1	0.000000	100	0.302817	0.302817
0	12	30	11	2.727273	27	0.885463	9.740088
0	1 3	8	6	1.333333	84	0.536313	3.217877
0	23	12	7	1.714286	56	0.635762	4.450331
0	123	7	3	2.333333	74	0.537975	1.613924
1		71	118	0.601695	30	0.902027	106.439186
1	1	81	72	1.125000	80	0.718861	51.758007
1	2	45	29	1.551724	34	0.868526	25.187252
1	3	25	15	1.666667	26	0.883178	13.247663
1	12	32	17	1.882353	79	0.686747	11.674699
1	1 3	1	4	0.250000	163	0.232227	0.928910
1	23	11	5	2.200000	35	0.831683	4.158416
1	123	11	6	1.833333	111	0.481132	2.886793
2		90	137	0.656934	29	0.903448	123.772415
2	1	44	35	1.257143	55	0.807143	28.250000
2	2	28	21	1.333333	32	0.881679	18.515268
2	3	12	10	1.200000	25	0.893805	8.938053
2	12	30	16	1.875000	61	0.773585	12.377358
2	1 3	11	11	1.000000	114	0.531120	5.842324
2	23	1	5	0.200000	139	0.364055	1.820276
2	123	8	6	1.333333	120	0.466368	2.798206
		710	730				605.174988

We can learn many things by studying this table. Though Mattingly clearly had the best batting season of those players with a significant number of batting events, thus ranking first, his best rank in any individual batting setting was 18th. This affirms that overall contribution is spread across situations. Mattingly did very well in enough settings to make him first overall. Note the number of percentiles over .800 -- any time that you get .800 or better in a setting, you have done very well.

Not only did Mattingly have a lot of high percentile settings, but in many of those he had a lot of events. That means that those settings count heavily for him (and really, that he did very well over many opportunities). Conversely, in a few settings Mattingly had poor percentiles, but they were usually settings in which he only had a small number of events. Thus these bad showings don't weigh that heavily against him (and probably don't reflect what he would do over a large number of opportunities).

Looking at raw bases and events, one cannot conclude too much. One can see which settings Mattingly came up in more than in others, and get a rough idea of whether or not he was productive by looking at his average bases by setting. But average bases can be deceptive. Although Mattingly "only" had an average bases figure of about .661 with none on and no outs, compared with much higher average base figures for other settings, his performance in that setting was his best. His setting rank, 18th, was his best, but even this can be deceiving. If there were only 20 players with events in that setting, then 18th isn't very good at all. If, on the other hand, there were 250 players with events in that setting, then 18th is excellent. Can we tell how many players did have events in that setting just from the data in the table? Yes. By working backward from Mattingly's percentile of .942953 (very, very good) we can determine how many he was 18th out of. Recalling that percentile equals 1 - (# higher in category)/(# in category), we can substitute so that we get 0.942953 = 1 - 17/X. Note that since his rank was 18th, there had to be 17 players *higher* than Mattingly in the setting. This gives us 17/X = 1 - .942953. Therefore, 17/X = .057046, so X = 17/.057046 = 298.005, so there were 298 players with events in that setting.

All of this illustrates the key concepts of B.A.S.E.S. evaluation (in this case for batting): the base orientation, average bases, setting distinctions, ranking and percentile within setting, weighting percentiles, and combining weighted percentiles across settings to give the best picture of player performance.