The Quest for Perfection

How to Ensure Your Officiating Evaluation Program Is the Best It Can Be

By Doron Scharf, Ref360

Worldwide interest in sports is at an all-time high, with an ever-growing audience attending games and watching on television and other devices. As exposure continues to soar, all aspects of sports are being monitored very closely, and officiating is no exception. High-definition video, increased camera angles and improved slo-mo have taken referee scrutiny to new heights. The frequency of controversial calls erupting has increased substantially — at least according to coaches, fans and the media.

With this in mind, leagues are stepping up their efforts to ensure they have the most effective officials. While identifying and recruiting top talent is a crucial element of success, having the tools to properly evaluate and train league officials is an equally vital aspect of maintaining a high-functioning group of experts. To this end, having an effective Officiating Evaluation Program (OEP) is of the utmost importance.

Creating and managing an effective OEP is not easy. Leagues must carefully consider their program's evaluation measures and process, feedback loop, analysis and reporting. The following best practices and approaches have been proven to help leagues make their OEPs the best they can be.

Evaluation Measures

While all OEPs capture the accuracy measures of the officials, leagues must decide carefully *what* to measure *and* the relative values of those measurements. When designing an OEP, it is important to closely consider the following:

- **Events to Assess:** Although the data explosion that has penetrated the sports world has yielded a far greater number of data elements to assess, not all OEPs are collecting and evaluating that advanced data. Leagues risk missing key criteria, which will distort the accuracy of their referee evaluations. If advanced data is being collected for evaluation, it should be cohesive, organized and easy to access.

- **Evaluation Scoring System:** While many calls are assessed for accuracy, they are not all of equal importance. Figure 1 below presents three grading considerations to help leagues address a call's importance so they can develop fair and representative rating systems.
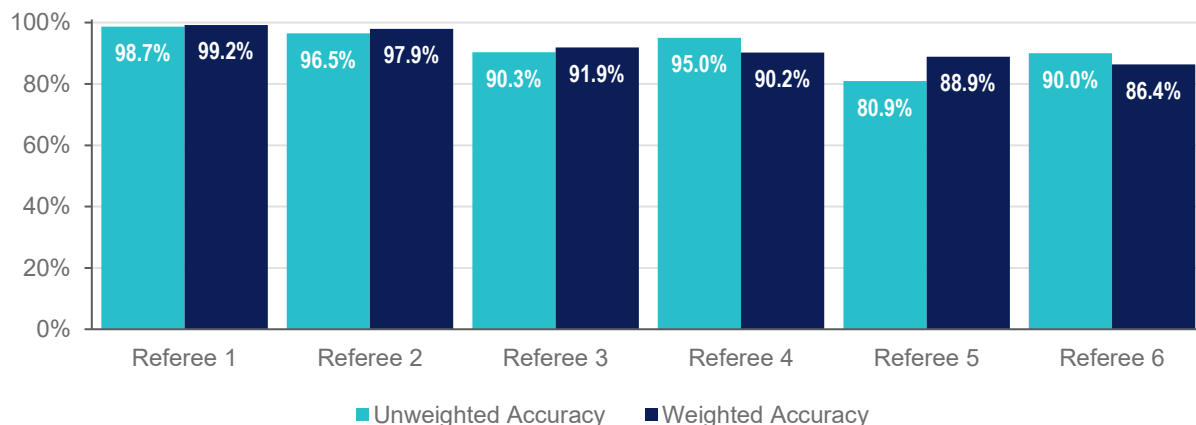
Figure 1: Assessing a Call's Importance

Consideration	Question to Ask
Incorrect vs. Missed Call	Is an incorrect whistle worse than not blowing a whistle when an infraction occurs?
Game Impact of Call	Should an error at a key moment in the game or on a game-changing event (*e.g.,* red card, pass interference in the end zone) have a greater impact on the rating score?
Difficulty of Call	Should the results on the rating score be more severe when a referee makes an error on an easy call as compared to making an error on a call that is more difficult for the referee to judge?

Source: Segal

Once the league has considered the relative importance of the call, it must calibrate the call's weighting so it is not improperly skewed, which would lead to misleading results. Many leagues have adopted adjusted scoring systems to reflect varying call impact. Figure 2 below compares referees' accuracy, on both an unweighted and a weighted basis.

Figure 2: Unweighted vs. Weighted Accuracy

	Referee 1	Referee 2	Referee 3	Referee 4	Referee 5	Referee 6
Unweighted Accuracy	98.7%	96.5%	90.3%	95.0%	80.9%	90.0%
Weighted Accuracy	99.2%	97.9%	91.9%	90.2%	88.9%	86.4%

Source: Segal

- **Qualitative Measures:** While accuracy is likely the most important measure of referee performance, non-accuracy criteria are also critical for success. Measures such as game management, professionalism and teamwork should be assessed and woven into the OEP, as applicable. This can provide an important opportunity to stress the less tangible forms of behavior referees need to succeed. Figure 3 below presents a sample formula that combines the accuracy score and qualitative measures to assess referee performance.

Figure 3: Grading Algorithm

Accuracy 70%

Qualitative

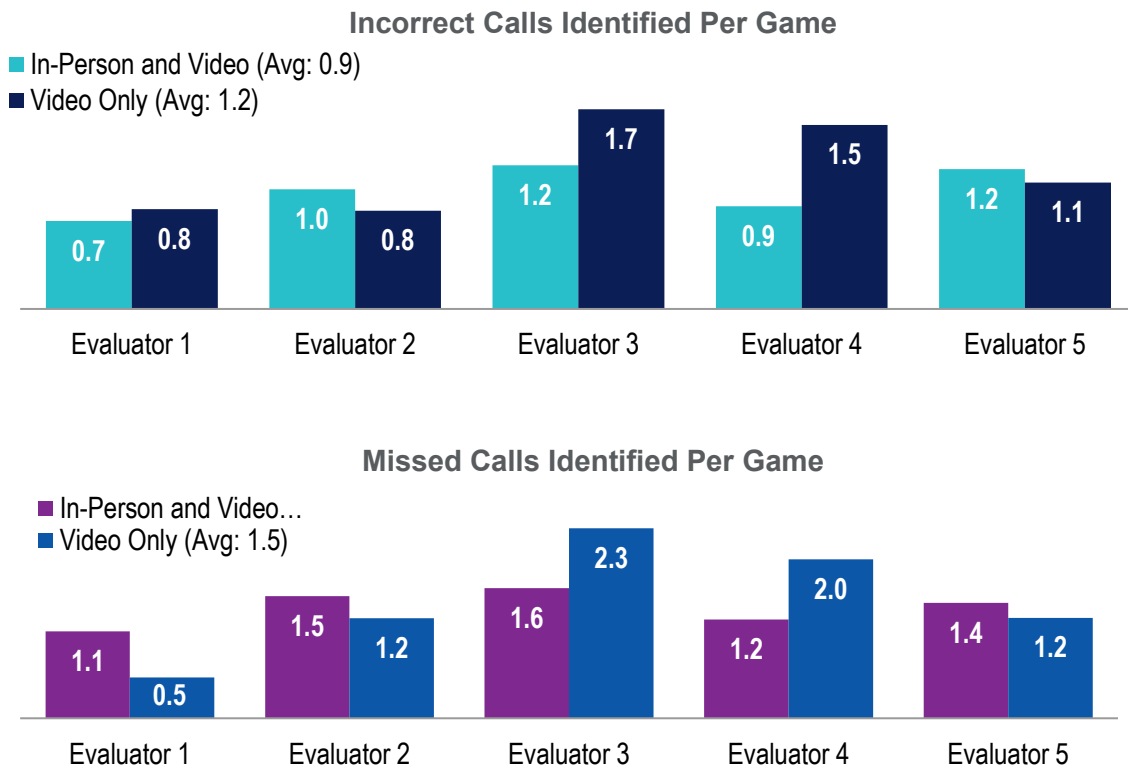**Game Management 10%**

Professionalism 10%

Teamwork 10%
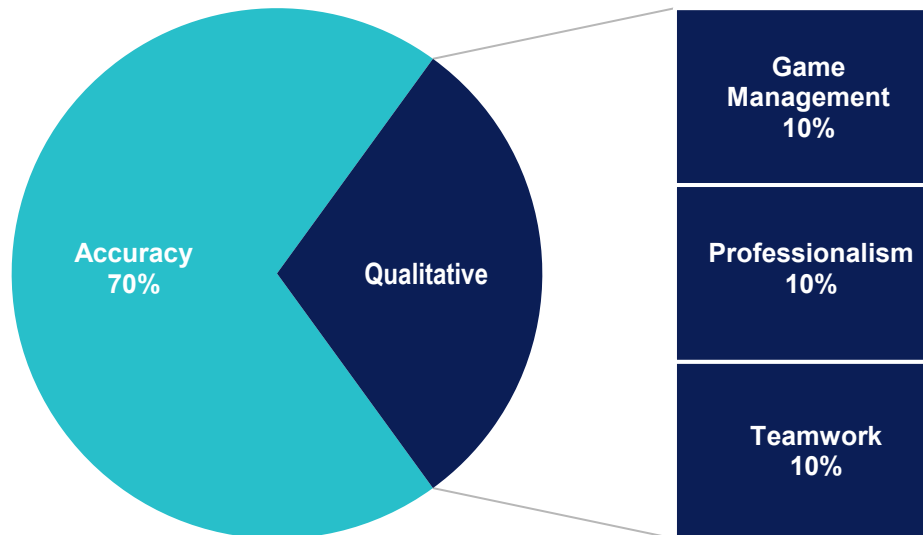
Source: Segal

Evaluation Process

The process by which games are evaluated is critical to attaining valuable and consistent results. Often, the process that has been in place for years has not been reviewed or modified to reflect the availability of updated media or systems. It is important to consider the following factors:

- **Evaluator's Game Attendance:** Although most assessments are conducted via video review, does it make a difference if the evaluator attends the game? While the evaluator may be able to capture a few items that would not be visible on video, the environment could have an impact on the assessment. In fact, Segal's analysis has shown that attending a game in person prior to video review does not necessarily improve the quality of the review. As illustrated by the graphs in Figure 4 on the next page, the opposite can often be true.

Figure 4: Impact of Evaluator Game Attendance

Incorrect Calls Identified Per Game

- In-Person and Video (Avg: 0.9)
- Video Only (Avg: 1.2)

Evaluator	In-Person and Video	Video Only
Evaluator 1	0.7	0.8
Evaluator 2	1.0	0.8
Evaluator 3	1.2	1.7
Evaluator 4	0.9	1.5
Evaluator 5	1.2	1.1

Missed Calls Identified Per Game

- In-Person and Video…
- Video Only (Avg: 1.5)

Evaluator	In-Person and Video	Video Only
Evaluator 1	1.1	0.5
Evaluator 2	1.5	1.2
Evaluator 3	1.6	2.3
Evaluator 4	1.2	2.0
Evaluator 5	1.4	1.2

Source: Segal

- **Levels of Review:** Typically, one evaluator will review game tape and capture information from the game. However, an additional layer of review from a second party for at least a subset of the game events will ensure the review process is not skewed by any one individual's point of view. A best practice is to hold a calibration session with all the evaluators to review the key game calls. The group dynamic will improve the consistency of the individual evaluators. The graphs in Figure 5 on the next page demonstrate that a group review will bring out more consistency in the evaluation.

Figure 5: Impact of Calibration Session

Errors Per Game Before Calibration

Errors Per Game After Calibration
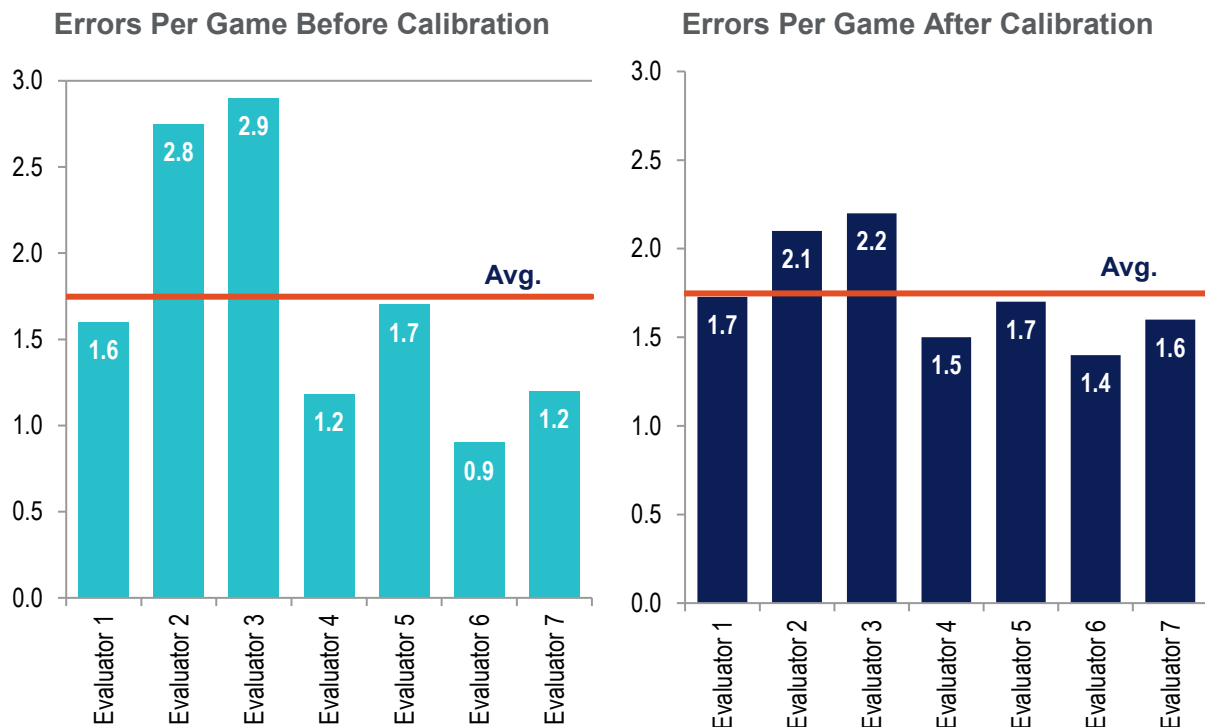
Source: Segal

- **Data Management:** Because there is a large amount of information to capture for any OEP, it is critical that evaluators are not encumbered by time-consuming data-collection methods that could affect their assessments. Customized, easy-to-use and flexible data-capture software is crucial to ensuring that the evaluators' hard work can be put to maximum use. This way, evaluators can spend their time on assessment, feedback and training, rather than data input and computer machinations. In addition, it is crucial that data is passed from one party to another in a safe and secure manner. As such, it is essential that data is centrally housed on a protected website rather than passed from party to party via emails and spreadsheets.

Feedback Loop

A crucial aspect of any OEP is the feedback that is gleaned from the system and provided to the referees. Clear and specific feedback can help the referees both understand their strengths and identify areas that need improvement and/or could benefit from training.

Although some leagues select the information they give their referees, many share all the commentary that is generated. The key benefit of this approach is that referees that have all the details will have a clear understanding of what they need to focus on for improvement. Moreover, it can also provide an opportunity for dialogue among the referees and evaluators. The referees may appreciate a chance to share their thoughts on the situation. While the evaluators and the referees will not always agree, having a chance to state their opinions in their discussion can help the referees internalize the feedback, leading to a positive impact on their performance.

Analysis and Reporting

A well designed OEP can generate a wealth of valuable data (see the sidebar "More than Referee Assessment"); however, attempting to crystallize findings from so much data can be difficult and daunting. The success of the program hinges on properly analyzing the trends and patterns and breaking them down into concise findings. Only then will it be clear which officials are exceeding or falling short of expectations and what specific areas of training the league needs to emphasize.

The three charts presented in Figures 6, 7 and 8 illustrate how the information from an OEP can be presented in ways that make the analysis of key aspects of referee performance clear and easy to understand.

More than Referee Assessment

OEPs have uses beyond referee assessment. The information can be used to help evaluate rule changes or hot topics. It can also be valuable in reports to management.

Leagues may want to consider sharing some information with the teams and, in certain circumstances, even with the media. Regardless, having statistical evidence to help answer questions and counter (or confirm) reports can be highly desirable, even if the data is not shared.answer questions and counter (or confirm) reports can be highly desirable, even if the data is not shared.

Figure 6

Individual Referee Summary
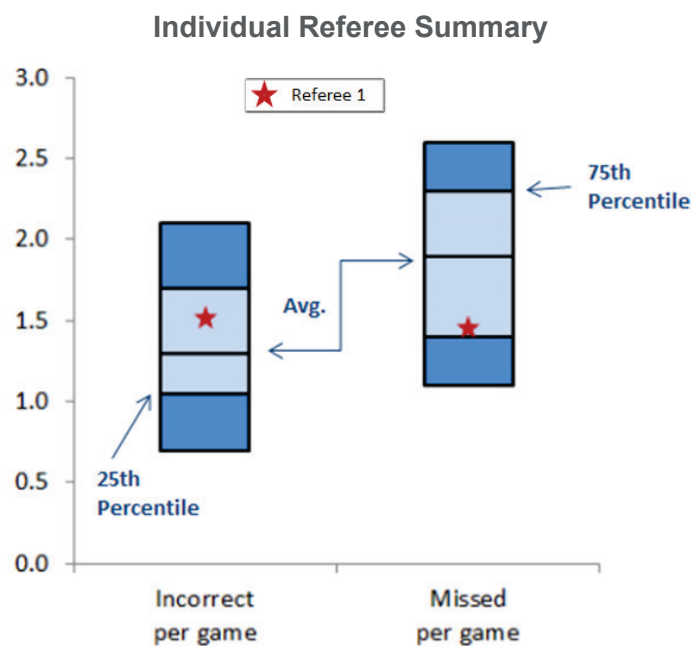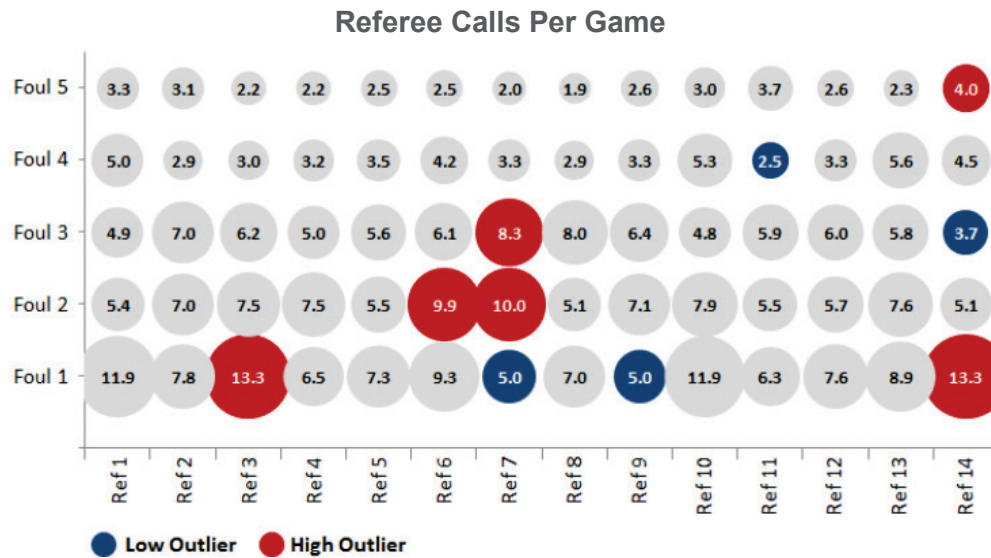

Referee 1 makes more incorrect calls than average, but is better than average in missed calls per game.

Source: Segal

Figure 7

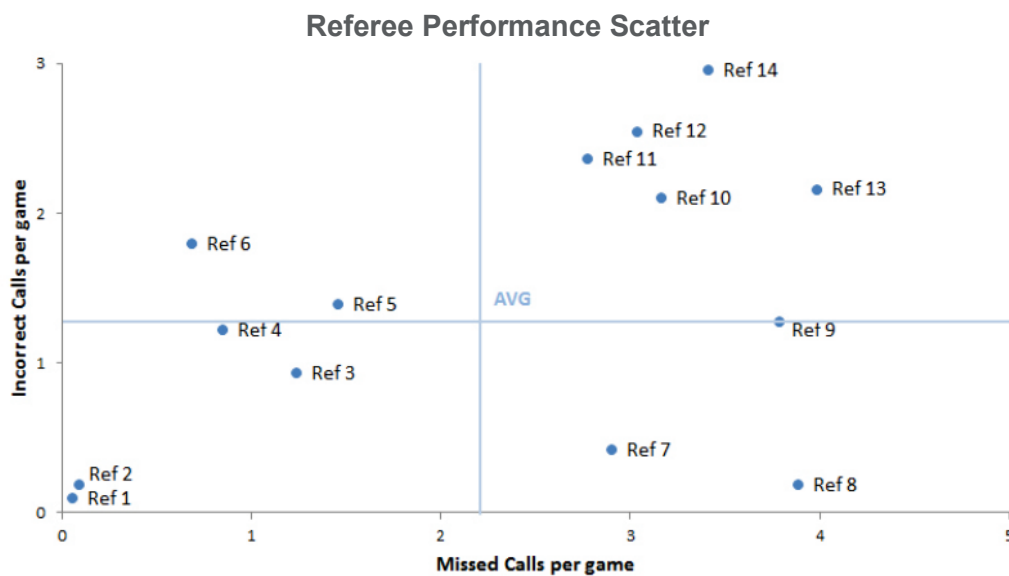### Referee Calls Per Game

	Ref 1	Ref 2	Ref 3	Ref 4	Ref 5	Ref 6	Ref 7	Ref 8	Ref 9	Ref 10	Ref 11	Ref 12	Ref 13	Ref 14
Foul 5	3.3	3.1	2.2	2.2	2.5	2.5	2.0	1.9	2.6	3.0	3.7	2.6	2.3	4.0
Foul 4	5.0	2.9	3.0	3.2	3.5	4.2	3.3	2.9	3.3	5.3	2.5	3.3	5.6	4.5
Foul 3	4.9	7.0	6.2	5.0	5.6	6.1	8.3	8.0	6.4	4.8	5.9	6.0	5.8	3.7
Foul 2	5.4	7.0	7.5	7.5	5.5	9.9	10.0	5.1	7.1	7.9	5.5	5.7	7.6	5.1
Foul 1	11.9	7.8	13.3	6.5	7.3	9.3	5.0	7.0	5.0	11.9	6.3	7.6	8.9	13.3

● Low Outlier ● High Outlier

The above chart details referee
calls per game by foul type.

Source: Segal

Figure 8

Referee Performance Scatter

Referee 8 has very few incorrect calls per
game but is hesitant with his whistle.

Source: Segal

Conclusion

The increased focus on officiating accuracy makes it crucial that leagues maintain the highest standards. A comprehensive, easy-to-use, secure and wide-reaching OEP is a key element in this endeavor. It can help leagues decide what measures to evaluate and how to evaluate them, review and improve game evaluation processes, maximize the use of feedback, and generate reports that allow leagues to make full use of the data they gather. In the end, a successful OEP will lead to vastly improved officiating.

About the author:

Doron Scharf, Senior Vice President and Actuary, is a member of Segal's
Sports Practice leadership team, and consults to many major professional sports leagues. se
He heads up Segal's Ref360 program, a comprehensive approach to understanding officiating performance. He can be reached at DScharf@segalco.com or ref360@segalco.com.
More information about Ref360's approach, data analysis and reporting, and software system can be viewed at ref360.com.