

Una revisión de los sistemas de valoración de jugadores de baloncesto (III).

Discusión general

Martínez, Jose A.

Departamento de Economía de la Empresa.

Universidad Politécnica de Cartagena.

Facultad de Ciencias de la Empresa. Calle Real, 3. 30201. Cartagena. España.

Tel.: +34 968 32 57 76; Fax: +34 968 32 70 81.

E-mail: josean.martinez@upct.es

Cómo citar este artículo:

Este artículo es una versión del aceptado para publicación en la Revista Internacional de Derecho y Gestión del Deporte, y que previsiblemente se publicará en el número 12 de esa revista.

Martínez, J. A. (2010). Una revisión de los sistemas de valoración de jugadores de baloncesto (III).
Discusión general. *Revista Internacional de Derecho y Gestión del Deporte*, 12.

How to cite this paper:

This article is a draft of the accepted paper to be published in the number 12 of the Revista Internacional de Derecho y Gestión del Deporte.

Martínez, J. A. (2010). Una revisión de los sistemas de valoración de jugadores de baloncesto (III).
Discusión general. *Revista Internacional de Derecho y Gestión del Deporte*, 12.

Una revisión de los sistemas de valoración de jugadores de baloncesto (III).

Discusión general

A review of the basketball player evaluation metrics (III)

Overall discussion

Resumen

En la última parte de esta trilogía se discuten ventajas y limitaciones de los diferentes sistemas de valoración, incidiendo en diferentes críticas y debates que han intercambiado algunos de los principales analistas. Además, se realizan varios análisis estadísticos, con el fin de mostrar la relación existente entre diferentes sistemas, o sobre las asunciones de algunos de ellos. Se plantea, asimismo, una reflexión sobre el uso de los criterios de validación y el uso de información cualitativa, indicando la idoneidad de la perspectiva formativa sobre medición. Finalmente, se discute acerca de los incipientes retos futuros, tanto para los analistas y académicos, como para las diferentes competiciones.

Palabras clave: Baloncesto, estadística, sistemas de valoración de jugadores.

Abstract

In this third article of the trilogy on basketball evaluation metrics, strengths and shortcomings of the different evaluation systems are discussed. In addition, diverse statistical analyses are achieved, in order to study the relationship among different systems, and discussing some assumptions. Moreover, this paper provides a framework to discuss about the basis of these metrics, proposing several ideas related to validity criteria, formative measurement, and the use of qualitative information. Finally, implications for further academic and applied research are discussed.

Keywords: Basketball, statistics, player evaluation metrics

Introducción

Tras la recopilación exhaustiva realizada sobre los sistemas de valoración de jugadores de baloncesto que actualmente se utilizan (Martínez 2010a), así como su correspondencia con el uso en las principales competiciones del mundo y en los *fantasy games* (Martínez 2010b), es hora de discutir acerca de cuáles de esos sistemas son los más adecuados, mostrar el debate existente sobre las ventajas o inconvenientes de los más utilizados, y comentar los desafíos que se presentan en el futuro.

Esta tercera parte que completa esta trilogía de artículos pretende arrojar un poco de luz ante la previsible desorientación del decisor a la hora de considerar los alrededor de 200 sistemas descritos para evaluar y realizar predicciones sobre el desempeño de los jugadores. No se va poner punto y final al gran debate existente en los foros especializados, tanto académicos como no académicos, ya que ello es imposible, puesto que como se discutirá a continuación, no existe ningún sistema perfecto. No obstante, muchos de los sistemas utilizados en grandes competiciones de manera oficial son en principio mejorables, por lo que es de esperar que, al menos, esta investigación sirva para reflexionar a los encargados de decidir qué sistema utilizar en las diferentes ligas.

Para no duplicar información y alargar innecesariamente este artículo, debe leerse de la mano de la primera y la segunda parte comentada, con el fin de que el lector no se pierda entre esta pléyade de sistemas y siglas.

Valoración combinada: ponderaciones de categorías estadísticas o sistemas plus/minus

De todos los sistemas de valoración de jugadores repasados catalogados como “combinados” emergen dos categorías principales: (1) sistemas que se construyen en base a la ponderación de diferentes categorías estadísticas; (2) sistemas que reflejan el rendimiento del equipo cuando el jugador ha estado o no en cancha. En el primer caso, encontraríamos los sistemas más sencillos como la Eficiencia de Manley, TENDEX o las Valoraciones (hasta 9 diferentes, según la revisión realizada), así como los más complejos, como el PER, APER, WinScore, WinsProduced, Alternate WinScore, Estimated Wins Added, eWins, WinShares, etc. En el segundo caso, habría que considerar sistemas sencillos como Net Plus Minus, On Court / Off Court, o más complejos como el WINVAL, Adjusted Plus Minus o Regularized Adjusted Plus Minus.

El debate sobre la idoneidad de unos u otros sistemas ha sido en parte tratado por algunos de los autores de esos índices (Berri y Bradburi, 2010; Lewin y Rosnbaum, 2007; Winston, 2009). Básicamente, Winston (2009), critica abiertamente sistemas como Eficiencia o PER, por premiar precisamente la ineficiencia de los jugadores, que con unos porcentajes de acierto bajos pueden incrementar esos índices. Esta crítica es también realizada por Berri y Bradburi (2010). Bajo este prisma, los sistemas WinScore o WinsProduced, son más adecuados, ya que aunque también se basan en ponderaciones lineales, se premia a los jugadores a partir de un cierto rango de eficiencia (por encima del 50% en lanzamientos de dos puntos y del 33,3% en triples, cifras bastante más lógicas que el 30,4% y 21,4% del PER, o del 33% y

25% de Eficiencia) (Berri y Bradburi, 2010; Winston, 2009). En realidad esas críticas son extensibles a prácticamente todos los sistemas etiquetados como “Valoración”.

La diferencia entre los sistemas más usados de ponderaciones lineales, como todos los que se han repasado en las principales ligas del mundo y en los juegos de fantasía, y otros sistemas avanzados que también usan ponderaciones es el criterio seguido para su formación. Así, la mayoría de sistemas avanzados utilizan un criterio basado en la conversión de la producción del jugador a las victorias del equipo. Esta traducción del desempeño individual en victorias es la característica de esos sistemas llamados “wins”. De entre todos ellos, el propuesto por Berri (1999; 2008), WinsProduced y su versión simplificada WinScore, tiene una base econométrica, y es el único publicado en revistas académicas. No obstante, ha sido abiertamente criticado por Lewin y Rosenbaum (2007), y anteriormente por Oliver (2004). Las críticas de Lewin y Rosenbaum (2007) se fundamentan en el erróneo planteamiento teórico de la función de producción del jugador, donde se diferencia entre dos tipos de posesiones diferentes (ofensiva y defensiva), lo que les lleva a proponer su sistema Alternate WinScore, que modifica finalmente las ponderaciones lineales propuestas por Berri (1999). Rosenbaum y Berri han discutido ampliamente en foros especializados no académicos sobre este tema, pero no han llegado a ningún consenso al respecto (Berri y Bradburi, 2010).

Una forma de evaluar los sistemas “wins” es a través de su correlación. Goodman (2010), recopiló los valores de eWins, WS, EWA, WARP y D&PWP, de los mejores jugadores de la NBA en la temporada 2009/2010 acumulados hasta la jornada del 9 de Enero (Tabla 1), ordenados en función de la media geométrica de esos índices.

Tabla 1. Comparación de diferentes sistemas “wins”.

Jugador	eWins	WS	WARP	D&PWP	EWA	Media
L. James	8.1	8.8	12.1	17.1	14.1	11.7
K. Bryant	6.4	6.1	7.4	13.3	10.1	8.3
D. Wade	5.7	4.8	8.4	9.9	10.0	7.4
K. Durant	6.0	5.9	6.4	9.6	9.7	7.3
C. Bosh	5.7	5.5	8.0	6.8	9.7	7.0
B. Roy	5.1	5.9	6.2	9.2	8.3	6.8
S. Nash	4.3	5.4	7.8	9.4	8.2	6.7
T. Duncan	5.1	5.5	7.8	6.2	8.9	6.6
D. Howard	5.1	5.3	8.0	7.4	7.5	6.5
D. Nowitzki	5.5	5.4	5.7	8.1	7.9	6.4
C. Anthony	5.2	4.9	5.5	7.5	8.8	6.2
C. Paul	3.8	5.0	7.1	6.1	7.4	5.7
Z. Randolph	4.6	4.7	5.8	4.9	7.3	5.4
D. Williams	4.1	4.3	5.4	7.2	5.9	5.3
G. Wallace	4.4	5.0	5.6	5.3	5.4	5.1
R. Rondo	3.6	4.6	6.3	6.4	5.1	5.1
D. Lee	4.7	5.0	5.4	3.3	7.4	5.0
C. Boozer	5.1	4.3	5.0	4.7	5.7	4.9
J. Smith	4.0	4.1	6.7	4.0	6.0	4.8

J. Johnson	4.0	3.8	3.8	6.9	6.0	4.7
------------	-----	-----	-----	-----	-----	-----

Como puede comprobarse en la Tabla 1, estos cinco sistemas no concuerdan en sus valores ya que se basan en definiciones diferentes del concepto de victoria asignada a los jugadores. Se puede estudiar su asociación a través de las correlaciones de Pearson y Spearman, como muestra la Tabla 2.

Tabla 2. Correlaciones entre diferentes sistemas “wins”.

Jugador	eWins	WS	WARP	D&PWP	EWA
eWins		.69	.41	.63	.83
WS	.85		.57	.59	.77
WARP	.68	.79		.47	.66
D&PWP	.79	.82	.69		.72
EWA	.90	.87	.80	.82	

Las correlaciones de Pearson se muestran en el triángulo inferior y las de Spearman en el superior.

Las correlaciones de Pearson son mayores que las de Spearman, lo que indica que la variación entre el valor de los mejores jugadores de la liga es medianamente consistente en la mayoría de esos sistemas (tal vez con la excepción de WARP), aunque las clasificaciones finales obtenidas (el ranking) no produzca resultados tan similares, por lo que la asociación por rangos es más baja.

Es de destacar que esta forma de valorar a los jugadores depende de las victorias globales del equipo, es decir, un jugador tiene más probabilidad de destacar en estos índices si está en un equipo ganador (porque seguramente contribuirá decisivamente a esas victorias). Asimismo, a mayor número de minutos mayor valor, ya que la contribución individual del jugador a las victorias del equipo se incrementa. Si comparamos estos razonamientos con la formulación del WinScore de Berri (2008), encontramos que WinScore proporciona resultados independientes del desempeño del equipo, aunque no olvidemos que es sólo una simplificación del WinsProduced (Berri, 2008), el cual tiene en cuenta un ajuste en función de las estadísticas del equipo, de los contrarios y de la liga. La suma de los WinsProduced de los jugadores de un equipo predice de manera aceptable las victorias del equipo (Berri, 1999).

Como puede intuirse, las valoraciones más simples como Eficiencia y Valoración, etc., no son capaces de relativizar el desempeño del jugador en función de otras variables, es decir, son similares en su interpretación al WinScore. No obstante, TENDEX tiene la particularidad de que relativiza las aportaciones individuales en función de las globales de ambos equipos.

Los sistemas Plus/Minus también tienen numerosas limitaciones. Los más sencillos, como Net Plus Minus o On Court / Of Court, no corrigen el desempeño del jugador individual por el contexto en el que se produce, es decir, por la calidad de los compañeros o el contrario. Estas críticas han llevado a varios analistas a diseñar propuestas que corrigen esos factores situacionales, resultando en índices ajustados, como WINVAL, las diferentes versiones de Adjusted Plus Minus, o la más sofisticada de Regularized Plus Minus. Berri y Bradburi (2010) realizan una crítica sólida a este tipo de índices ajustados, basada en

los altos errores estándar obtenidos y en la carencia de estabilidad cuando se comparan de una temporada a otra, lo que es indicativo de falta de fiabilidad. Es cierto que se han intentado reducir esos errores estándar utilizando un mayor volumen de datos (varias temporadas), pero ello distrae del objetivo principal del índice: valorar al jugador por su desempeño en la temporada. La propuesta de regularización de Joe Sill es muy interesante, ya que reduce los errores estándar, pero a cambio paga el precio de obtener estimaciones sesgadas de los parámetros. Aunque ese *trade-off* entre sesgo y error de estimación puede ser deseable, los errores estándar obtenidos siguen siendo considerables. Recordemos que las comparaciones estadísticas en presencia de error deben realizarse utilizando un criterio probabilístico, como indican Martínez y Martínez (2010a), y no mediante simple comparación directa.

Valoraciones combinadas usadas oficialmente en las competiciones

Al margen del uso muy poco representativo del Net Plus Minus, las competiciones repasadas en la segunda parte de esta investigación utilizan las ponderaciones lineales simples, y que pueden calcularse prácticamente “de cabeza” al echar un vistazo al *box-score*. La simplicidad es una característica deseable en ciencia, y en este caso, también facilita la diseminación e interpretación de esos índices por parte de los aficionados. No obstante, hay que analizar si esos índices son correctos en su formación.

Los más utilizados son la Eficiencia y la Valoración. La Eficiencia es una versión del TENDEX en la que no se tiene en cuenta el ritmo del partido. Ya hemos comentado las críticas de Berri y Bradbury (2010) o Winston (2009) en referencia a que este índice premia a jugadores “ineficientes”, es decir, con bajos porcentajes de lanzamiento de campo. Esta limitación es fundamental para recomendar este índice para valorar a los jugadores, porque indica que malos tiradores pueden hipotéticamente ayudar a su equipo lanzando cada vez más, lo que es teórica y prácticamente cuestionable.

Al margen de ello, otros factores son susceptibles de discusión. Por ejemplo, en la Eficiencia, el valor de los lanzamientos de tiros libres fallados es el mismo que para el de tiros de campo. No es así en la versión original del TENDEX, cuya ponderación es de 0.5. Esa ponderación parece más adecuada, ya que alrededor el 44% de tiros libres resultan en finales de posesión (Kubatko, Oliver, Pelton y Rosenbaum, 2007) por lo que prácticamente en la mitad de los casos, fallar un tiro libre no supone una pérdida de posesión para el equipo, y no tendría que considerarse como un factor negativo¹. Además, no todos los tapones resultan en una recuperación de posesión, por lo que también podría corregirse su ponderación. Siguiendo esta línea argumental, como no todos los lanzamientos fallados son pérdidas de posesión, se podría ponderar el valor de esos lanzamientos y de los rebotes ofensivos y defensivos, tal y como hacen Lewin y Rosenbaum (2007) en su Alternate WinScore.

¹ Podría argumentarse, sin embargo, que existe un coste de oportunidad al fallar un tiro libre, y es el del posible aprovechamiento de la posesión si se hubiera lanzado de 2 o 3 puntos. Si esto se tuviera en cuenta, también habría que razonar de la misma forma con los lanzamientos de 2 puntos, en referencia a la oportunidad perdida de anotar de 3 puntos.

La Valoración, sistema utilizado en competiciones como la ACB² o Euroliga, es uno de los sistemas más criticables. Así, considerar los tapones en contra como un factor negativo es redundante, ya que cuando un jugador recibe un tapón ya se le considera como lanzamiento fallado. De igual manera, un jugador que realiza una falta personal en ataque ya se le atribuye un balón perdido, por lo que de nuevo existe redundancia. También se puede cuestionar el hecho de considerar las faltas recibidas como un factor positivo, ya que aquellas que implican lanzamiento de personal ya tienen su ganancia en ese aspecto, y aquellas que no se traducen en lanzamientos libres no producen un cambio de posesión, y por tanto tampoco debieran premiarse. Muchos entrenadores de principales equipos europeos están descontentos con este sistema³. Tanto Val2 como Val3, son mejoras de Valoración, pero siguen siendo más limitadas que Eficiencia.

En cuanto a Val4, es exactamente igual que Eficiencia salvo porque las asistencias cuentan el doble. La consideración de las asistencias como aspecto positivo del juego es también cuestión de debate (Berri, 2008). Una asistencia no produce un aprovechamiento de la posesión ni una pérdida de ésta, sino solamente es un elemento facilitador. Ese tipo de factores facilitadores son numerosos en baloncesto: bloqueos, intimidación, ayudas defensivas, etc., y ninguno de ellos se considera en el box-score. Además, la consideración o no de que un pase es asistencia es algo controvertida (ver Martínez y Martínez, 2010a), por lo que no es de extrañar los índices específicos que se están proponiendo para valorar esta categoría, como “Assits leading to” o “At Rim”. En cualquier caso, el peso de las asistencias derivado de manera formal por Berri (2008) o Lewin y Rosenbaum (2007) es de 0.5, por lo que es cuatro veces menor que el propuesto por Val4, sistema por tanto, que está bastante limitado.

El resto de Valoraciones (desde Val5, Val6, Val7 y Val9)⁴ proponen ponderaciones más diversas y complejas. No conocemos los criterios por los cuales esas ponderaciones son utilizadas, por lo que es difícil saber si han sido propuestas de manera subjetiva o bajo algún método más sofisticado. Aún así, podemos decir que Val5 tiene la novedad de ponderar con mayor importancia los rebotes ofensivos (algo que estaría en línea con el razonamiento de Lewin y Rosenbaum, 2007, y Schaller, 2001), pero sigue contando con las mismas limitaciones que Val2.

Val6 es un índice tomado del Fantasy Game de Sporting News, el cual fue subjetivamente elaborado, por lo que tampoco tiene una base sólida fundamentada. El sistema Val7 es el utilizado en la Liga de Venezuela, y tampoco existe más información en relación a su génesis. Sí que podemos decir que es un

² Este sistema fue introducido en España por Franco Pinotti, periodista, scouting y entrenador, en base al sistema utilizado en Italia. El objetivo era que fuese un índice simple de entender y fácilmente aplicable y objetivo de cara a cualquier jugador.

³ Jean-Luc Monscheau, entrenador del Nancy de la liga francesa, en comunicación personal con el autor en 2009, comentaba este hecho, y las valoraciones “ridículas” que obtenían a veces algunos jugadores con el sistema de la Euroliga. Monscheau, fue uno de los responsables de la instauración en la liga francesa del sistema de Eficiencia, a comienzos de los 90, algo innovador en Europa en aquellos años.

⁴ Excluimos a Val8, la utilizada por la liga de Kazajstán, que ha sido propuesta por los creadores del software de la liga, y que es exactamente igual a Eficiencia, salvo que los aspectos negativos no se restan, sino que son el divisor de los positivos. Rápidamente se ve que un jugador que no comete ninguna acción negativa en un partido tendría una valoración de infinito...

sistema que considera el valor relativo y no absoluto de los puntos, es decir, los puntos por partido, y que no da la misma importancia a los balones robados y perdidos. Esta última condición se antoja básica en cualquier sistema de ponderaciones lineales, ya que significan una recuperación o pérdida de posesión, es decir, la oportunidad de crear valor para el equipo, por lo que ambos deberían ser ponderados exactamente igual. Finalmente, Val9 es el sistema usado en la Liga de Rusia, del que tampoco conocemos más acerca de su creación, aunque cuenta con la novedad de incluir una categoría novedosa, como los pases para el lanzamiento del compañero, es decir, contar positivamente acciones de pase que no tienen por qué ser registradas como asistencia. De hecho, y como hemos comentado anteriormente, esta sería una idea atractiva, porque se evitaría que se premiara más a los jugadores que tienen tiradores eficientes en sus equipos.

No obstante, es muy interesante comprobar como aunque los diferentes sistemas varían, su correlación es prácticamente idéntica. La Tabla 3 muestra la asociación entre diferentes sistemas de ponderaciones lineales⁵, tomadas de las estadísticas de varios jugadores de la Liga ACB, con dilatadas trayectorias profesionales en esta competición. Se han considerado las estadísticas acumuladas en liga regular y play-offs de los siguientes jugadores: P. Prigioni, S. Vidal, T. Splitter, J. C. Navarro, C. Jiménez, C. Cabezas, G. Gabriel, B. Rodríguez, P. Jasen, F. Reyes, A. Mumbrú y R. Montañez, que componen un total de 235 registros. Estos doce seleccionados tienen diferentes características y juegan en diferentes posiciones y equipos, por lo que la heterogeneidad y representatividad de categorías estadísticas está fundamentada. Además, al incluir toda la carrera deportiva, se consideran temporadas de máximo rendimiento (muchos minutos en cancha) y temporadas menos brillantes como las primeras (menos minutos en cancha). De nuevo este factor contribuye a que la muestra sea lo más heterogénea posible. Ante la carencia de normalidad en los datos se ha calculado la correlación de Spearman (triángulo inferior) y la Tau b de Kendall (triángulo superior). Este último estadístico da una idea de la concordancia en la ordenación de la serie de datos, y en este caso sirve para mostrar diferencias en los rankings obtenidos.

Tabla 3. Asociación entre diferentes sistemas de ponderaciones lineales

Sistema	Val	Val2	Val3	Val4	Val5	Val7	Eff	WinSc	AWS	TPR	GmSc	SPcr	VIR	MMM
Val		.988	.935	.912	.955	.751	.934	.826	.866	.918	.947	.950	.944	.926
Val2	.999		.942	.916	.961	.753	.937	.822	.861	.916	.947	.949	.948	.929
Val3	.992	.994		.930	.963	.732	.956	.792	.821	.914	.929	.931	.959	.952
Val4	.989	.989	.992		.932	.710	.953	.796	.845	.929	.928	.939	.953	.955
Val5	.996	.997	.997	.993		.738	.967	.826	.852	.926	.942	.962	.959	.943
Val7	.907	.907	.891	.877	.897		.728	.678	.717	.731	.754	.737	.735	.731
Eff	.993	.993	.997	.996	.998	.888		.821	.851	.941	.941	.965	.958	.960
WinSc	.951	.950	.936	.939	.953	.864	.950		.885	.881	.865	.880	.792	.784
AWS	.972	.970	.954	.964	.968	.892	.966	.981		.881	.949	.941	.841	.833
TPR	.990	.990	.990	.993	.993	.891	.995	.951	.977		.949	.941	.927	.938
GmSc	.996	.996	.992	.993	.995	.905	.995	.942	.971	.996		.949	.954	.951
SPcr	.996	.996	.992	.994	.997	.897	.997	.961	.977	.995	.996		.944	.937

⁵ El cálculo de los índices mostrados está realizado sin normalizar por minutos o por posesiones, para poder comparar fielmente todos esos sistemas.

VIR	.995	.995	.997	.997	.997	.892	.997	.937	.962	.992	.997	.995		.965
MMM	.991	.992	.996	.997	.995	.889	.997	.932	.958	.994	.996	.993	.998	

Las correlaciones de Spearman se muestran en el triángulo inferior y la Tau b de Kendall en el superior.

Como puede deducirse de la Tabla 3, no existirían diferencias destacables en la aplicación de los diferentes sistemas, salvo el de Val7, es decir, el de la Liga de Venezuela, que aunque correlaciona altamente con el resto, produciría clasificaciones diferentes en diversos jugadores, lo que sería un cambio a tener en cuenta. Es lógico que la correlación sea más baja, ya que la fórmula es ciertamente distinta al resto, considerando los puntos por partido como *input*.

Sin embargo, el resultado más reseñable es el de la similitud entre Val y Eficiencia, siendo la primera teóricamente más débil en su fundamento que la segunda. Los resultados son casi idénticos, por lo que no existe una gran disociación por decantarse por Val en lugar de Eficiencia. Pero además, la correlación con los sistemas más sofisticados como WinScore, AWS, TPR o MMM es también muy alta; de nuevo las diferencias serían mínimas. Además, como Winston (2009) indica, análisis realizados en la NBA han mostrado que la Eficiencia y el GameScore (simplificación del PER de Hollinger) correlacionan 0.99, es decir, proporcionan información casi idéntica. Esa misma correlación es encontrada con estos datos de la Liga ACB. Recordemos que en todos estos casos las escalas son arbitrarias, y por ello la correlación es una medida de distancia atractiva para estudiar la similitud entre los diferentes sistemas. No obstante, el hecho de que exista una correlación casi perfecta entre las diferentes variables no significa que no se produzcan cambios importantes en las posiciones de los rankings al aplicarse un sistema u otro. Esta idea, que puede parecer contra intuitiva, es capital para entender el efecto de utilizar un sistema u otro a la hora de establecer clasificaciones de los jugadores en función de su rendimiento. Así, si nos fijamos en el índice de asociación ordinal de Kendall, los valores son más bajos que los obtenidos con la correlación de Spearman. Esto indica que, aunque las variables se “mueven” de manera prácticamente idéntica, se producirían cambios en la ordenación de los rankings, es decir, habría diferencias entre las clasificaciones de los jugadores al aplicar un sistema u otro, que podrían ser importantes. Aunque la Tau b de Kendall proporciona valores muy altos de concordancia ordinal, tal vez no muestra de forma intuitiva la verdadera relevancia de los cambios potenciales en los rankings. Por ello, se ha realizado un análisis heurístico de ese cambio, el cual se dispone en la Tabla 4.

Tabla 4. Cambios en los rankings (235 observaciones)

	Val - Eff	Val – WinScore	Val - AWS
Números de cambios en el ranking	223 (95%)	228 (97%)	223 (95%)
Promedio de cambios	5.83	14.8	11.4
Mediana de cambios	4	11	9
Números de cambios en el ranking (30 primeros)	26 (87%)	26 (87%)	26 (87%)
Promedio de cambios (30 primeros)	2.93	5.9	6.43
Mediana de cambios (30 primeros)	3	3	5
Números de cambios en el ranking (10 primeros)	7 (70%)	6 (60%)	7 (70%)
Promedio de cambios (10 primeros)	1.2	3	3
Mediana de cambios (10 primeros)	1	2	1

Las comparaciones dispuestas en la Tabla 4 se han realizado tomando como referencia Val, en relación a Eficiencia, WinScore y AWS. Así, con respecto a la Eficiencia, aunque la correlación de Spearman es 0.99 y la Tau b de Kendall es 0.934, se producirían un 95% de cambios en la ordenación del ranking de los 235 observaciones estudiadas. El promedio de cambios sería de 5.83, lo que indica que la media de los saltos de posición es de casi 6 puestos en el ranking. No obstante, como la distribución no es simétrica, la mediana proporciona información más rica sobre el centro de la distribución, en este caso 4 cambios. Bien es cierto que para los 30 primeros del ranking los cambios serían menos acusados, tendencia que se reforzaría en los 10 primeros. De este modo, el cambio de posiciones en los puestos privilegiados de la clasificación no sería tan drástico.

Con respecto a WinScore y AWS, los cambios en el ranking serían mucho más acusados, aunque de nuevo se suavizarían para los puestos más altos de la clasificación. No obstante, el dibujo que se obtendría sobre esos rankings sería obviamente diferente. Por tanto, las medidas de asociación pueden dar una impresión equivocada sobre la equivalencia de las medidas, al menos, el hecho de que sean tan altas esas asociaciones entre los diferentes índices no es sinónimo de inmovilismo en las clasificaciones. Más bien, habría que interpretarlas con prudencia, y no sacar conclusiones precipitadas sobre su equivalencia. No obstante, la aplicación del razonamiento postulado por Martínez y Martínez (2010a), es decir, la aproximación probabilística para crear los rankings, relativizaría mucho más esas divergencias encontradas en las clasificaciones producidas por los diferentes sistemas, incrementando de forma importante el índice de concordancia.

Sistemas de valoración defensiva y ofensiva

Muchos analistas coinciden en una premisa básica: el baloncesto es 50% defensa y 50% ataque (Winston, 2009). Bajo esta perspectiva, los *box-score* son injustos con la valoración de la aportación defensiva de los jugadores. De hecho, esta es una limitación que reconocen algunos de los creadores de índices que pretenden ser una valoración combinada (ofensiva y defensiva), como el PER (Holligner, 2005). Por ello se han propuesto diversos índices que tratan de cuantificar la aportación de los jugadores a nivel defensivo.

En verdad, la riqueza de las propuestas a este respecto es enorme. Desde la cuantificación del rendimiento de los jugadores que ha defendido el jugador en cuestión (CPER), la relativización de la calidad de esos oponentes en función del nivel de la liga (dQUAL), los puntos permitidos por el jugador mientras está en el campo en relación a otros factores (DWS), o los puntos que un equipo permite con el jugador en cancha frente a cuando está fuera (DPM). Existen además, propuestas que combinan varios de estos índices, en aras de minimizar las limitaciones de unos con las ventajas de otros, como DCS.

Estas propuestas más avanzadas necesitan de los datos del *play by play* y de la competición para poderse implementar, por lo que es muy difícil que puedan aparecer para resumir un *box-score*. También existen propuestas más simples, basadas en los tapones, balones recuperados, faltas provocadas en ataque o rebotes defensivos. Cabe destacar que los rebotes ofensivos no son considerados como acción defensiva, cuando se podría afirmar que es preceptivo hacerlo, ya que suponen una nueva jugada para el equipo (aunque muchos autores no consideren ello un cambio de posesión).

En cuanto a las valoraciones de las acciones ofensivas, de nuevo la riqueza de propuestas es sustantiva. Entre ellas, cabe destacar la contribución del jugador a las posesiones anotadas por el equipo (ISP) y las posesiones usadas por el jugador (UR). Estos índices dan una idea de la involucración ofensiva del jugador y su aportación relativa al juego de ataque del equipo. Además, de nuevo los sistemas Plus Minus consideran cómo afecta a la anotación del equipo que un jugador esté o no en cancha. Finalmente, también existen propuestas de combinación de varios de estos índices (OCS).

La explotación del *play by play* permite diferenciar entre diversos tipos de pase (no sólo las asistencias), así como categorizar los tipos de lanzamiento y asignarles una localización espacial. También permite conocer el tipo de pérdida de balón (forzada, no forzada). Esa localización en el espacio permite el conocimiento de la distribución de lanzamientos a canasta, por lo que se puede establecer también una normalización en función de la dificultad de anotar un lanzamiento, es decir, proponer índices que permitan diferenciar, por ejemplo, a dos jugadores con el mismo porcentaje de acierto en tiros de campo pero que realicen sus lanzamientos desde posiciones dispares en cuanto a dificultad. Esto es lo que consideran los índices SCAB y SHTAB.

Otros sistemas innovadores

Existen otros sistemas que no se encuadran dentro de la clásica distinción entre rendimiento ofensivo y defensivo, y que proporcionan información interesante. Así, la consistencia o regularidad en el rendimiento de los jugadores puede ser estudiada con diversos índices (varianza, coeficiente de variación, índice de Atkinson). De esta forma se puede analizar qué jugadores son más consistentes en su juego, lo que reduciría el riesgo en la predicción de su futura productividad (Boadvarsson y Brastow, 1998). No obstante, esos sistemas tienen el problema de basarse en la varianza de la distribución, lo que los hace muy sensibles a los casos atípicos. El índice de Atkinson, por otra parte, es usado para modificar los rankings estadísticos en función del riesgo que quiera asumir el decisor en valorar esa regularidad, pero puede proporcionar cambios en las clasificaciones que pueden ser cuestionables. Por ejemplo, ¿un jugador que anota 30 puntos por partido debería estar siempre por encima en el ranking que un jugador que anota 25, independientemente de su regularidad?

Siguiendo con el asunto de los rankings, Martínez y Martínez (2010a) realizan una propuesta de clasificación de jugadores basándose en una simple aproximación probabilística de su rendimiento. La propuesta realizada por estos autores sería aplicable a cualquier tipo de sistema que valore a los jugadores en una muestra de partidos, es decir, asumiendo que el valor esperado de la muestra de partidos

corresponde con el valor real (parámetro poblacional). Martínez y Martínez (2010a) muestran como esta perspectiva metodológica produce cambios importantes en los rankings de algunas categorías estadísticas, y se traduce en una relativización de las diferencias encontradas entre jugadores que ocupan posiciones diferentes en los rankings, haciendo que en muchas ocasiones no pueden considerarse diferentes los rendimientos de jugadores en posiciones divergentes en las clasificaciones.

En relación a los procedimientos de estandarización, tanto los z-scores como los índices de similitud que se basan en el análisis de distribuciones agregadas de jugadores, son muy interesantes para ofrecer comparaciones relativas. Se tiende a interpretar el valor de los valores normalizados como la distancia en desviaciones típicas al centro de la distribución. No obstante, hay que primero considerar cuál es la distribución de la población, en aras a poder interpretar de esa forma los valores estandarizados. Para ello, se necesita que la distribución sea normal, lo que no siempre es así. Por ejemplo, si tomamos los puntos por partido de los 444 jugadores que participaron en la temporada 2008/2009 en la NBA (Figura 1), se obtiene una distribución mucho más similar a una *power law*, que a una normal (dibujada en color rojo). Esto hace que la transformación de las puntuaciones a z-scores no siga tampoco una normal, y por tanto la interpretación no sea la deseable. Sin embargo, si vamos restringiendo esa distribución a aquellos jugadores que juegan al menos la mitad de los partidos en la competición (Figura 2), vemos como la distribución se va asemejando más a una normal. Si finalmente sólo consideramos aquellos jugadores que han jugado el 75% de los partidos, entonces prácticamente tenemos un modelo muy ajustado de distribución normal (Figura 3), con lo que la conversión a valores normalizados proporciona la información buscada. Por tanto, han de realizarse diversos filtrados de los datos para realizar ese tipo de conversiones, si es que se quiere seguir la interpretación prefijada sobre la estandarización.

Figura 1. Distribución de los puntos por partido de todos los jugadores de la NBA en 2008/2009

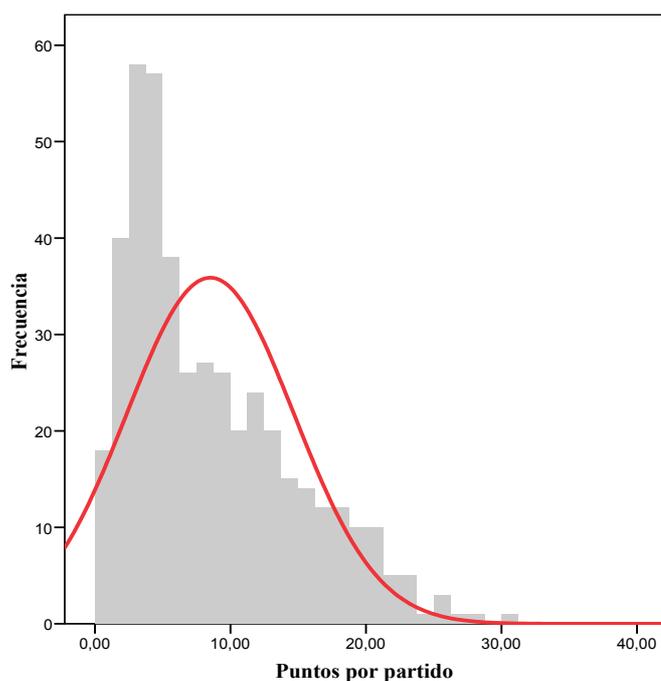


Figura 2. Distribución de los puntos por partido los jugadores de la NBA en 2008/2009 que han jugado al menos el 50% de partidos.

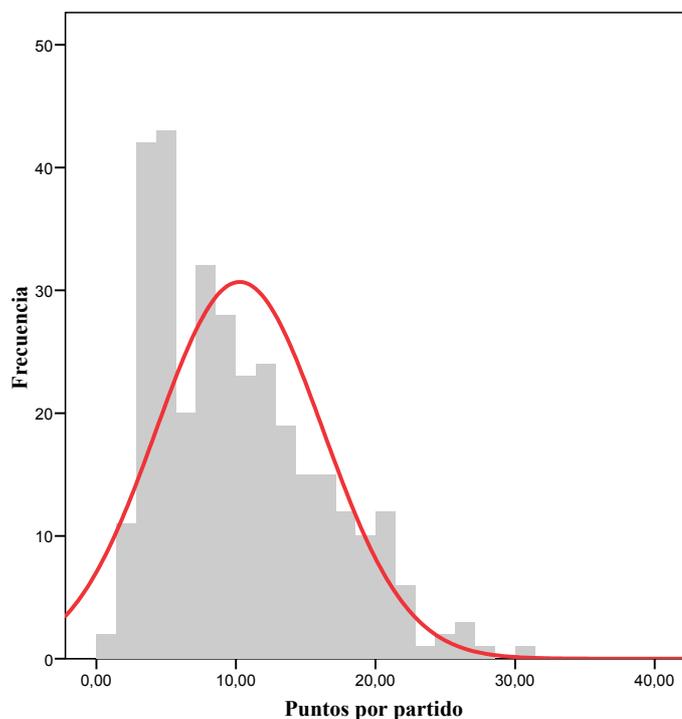
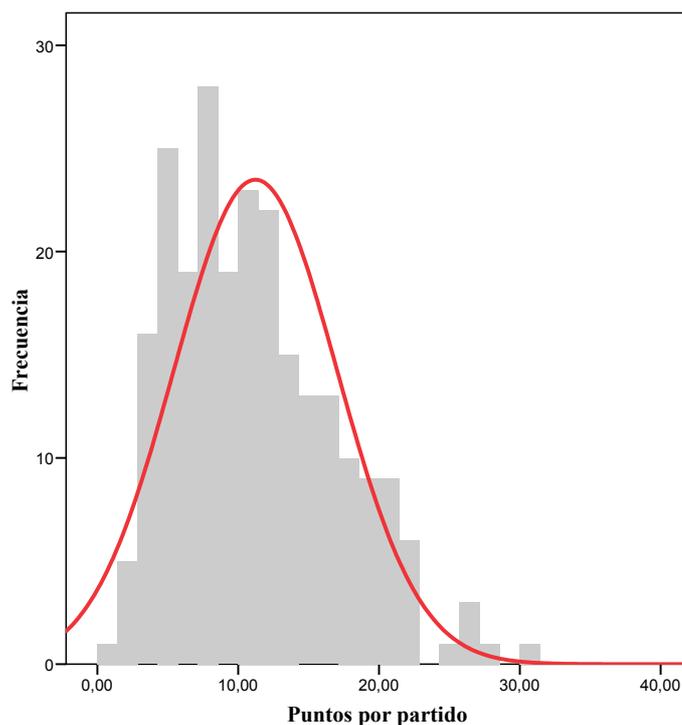


Figura 3. Distribución de los puntos por partido los jugadores de la NBA en 2008/2009 que han jugado al menos el 75 % de partidos.



Otra perspectiva en relación a la valoración de jugadores nace del concepto de la importancia de las acciones del juego. El valor de una acción de juego (una canasta, un rebote, etc.) no es el mismo si se consigue al comienzo del partido o con un marcador muy desigual, que si se hace con el partido igualado y en los últimos minutos. Al menos, esta es una perspectiva que comparten muchos analistas, que defienden que la dificultad de realizar las acciones de juego en los “momentos calientes” es mayor que en circunstancias con menos presión, por lo que debería premiarse a aquellos jugadores cuyas acciones se producen en los momentos de mayor dificultad. Así, se han creado diversos índices “clutch” donde se mide la producción de los jugadores en esos momentos clave. No obstante, la propuesta más importante es la realizada por www.binahoop.com, basada en matemáticas financieras, y del que su autor prefiere mantener el anonimato momentáneo. Este sistema, se podría decir que es una mejora formal de la inteligente idea del sistema SEDENA, creado en España, y el cual adolece del fundamento matemático deseable.

Las estadísticas de binahoop.com se denominan z-points, y proporcionan una visión muy diferente de los *box-score*. La Tabla 5 muestra un ejemplo del *box-score* del partido entre los Heat y los Lakers, del 4 de Diciembre de 2009 (tomado de www.basketball-reference.com y www.nba.com), donde los Lakers consiguieron la victoria por un punto y con un triple de Bryant en el último segundo. Tanto Bryant como Gasol realizaron un gran partido, donde ambos jugaron los mismos minutos. El ratio de anotación entre los dos fue del 1,5 a favor de Bryant, pero ambos jugadores fueron casi igual de Eficientes (ratio de 1,11 a favor de Bryant). Sin embargo, el Net Plus Minus indica que Gasol fue un 2,5 veces más valioso que Bryant para el equipo (cuando estuvo en cancha Gasol la diferencia fue de +10 mientras que cuando estuvo en cancha Bryant la diferencia fue de +4). Como vemos es un dibujo muy diferente el que retrata el Net Plus Minus. Pero si consultamos los z-points, veremos de nuevo un resultados extremadamente distinto; Bryant fue 2,61 veces más productivo que Gasol, sin duda, provocado por su acierto anotador en los momentos clave del partido.

Tabla 5. Comparación entre z-points y otros sistemas. Partido Heat-Lakers, 4-12-09

Lakers	Minutos	Puntos	Eficiencia	Net Plus Minus	z-points
K. Bryant	42:18	33	30	4	48,53
P. Gasol	42:18	22	27	10	18,58
R. Artest	41:06	9	8	-5	5,50
A. Bynum	32:33	19	20	-9	14,93
D. Fisher	31:31	11	12	4	7,19
L. Odom	20:51	4	6	-1	2,77
S. Brown	15:03	2	0	-4	1,15
J. Farmar	13:52	8	9	10	6,65
S. Vujacic	0:28	0	0	6	No calculado

Como puede comprobarse en la Tabla 5, el Net Plus Minus ofrece un dibujo del juego que puede resultar equivoco. Bynum también realizó un buen partido y contribuyó en anotación en los momentos clave. Sin embargo su Net Plus Minus fue de -9, peor que Artest, que jugó más, fue menos eficiente y menos

productivo en anotación. Winston (2009) ofrece varios ejemplos de la información sesgada que puede llegar a ofrecer este sistema.

Al igual que ocurría con las valoraciones ofensivas y defensivas, existen sistemas que son una combinación de varios, es decir, que pretenden ser más plurales, recoger las bondades de diferentes métodos con el fin de minimizar las limitaciones del sistema resultante. Como muestra está el S.C.O.R.E, el Composite Score o el Roland Rating, donde se combinan estadísticas Más Menos con otros sistemas de valoración ofensiva y defensiva. Este tipo de combinaciones son interesantes porque permiten recoger en un único índice varias propuestas de valoración, pero el principal problema es decidir cuál es la fórmula que relaciona esos sistemas para componer el resultante, que en el caso de esos sistemas los autores no especifican el porqué de su fórmula.

Medidas formativas y validación de sistemas

Más allá de las discusiones referentes a si un sistema es mejor que otro, existe un problema de base que debe ser tratado. Ese problema se refiere a la especificación de las medidas y a los criterios de validación de los sistemas. La cuestión fundamental es cómo se define lo que se quiere medir, y para ello se necesita establecer claramente la definición de cada variable susceptible de ser cuantificada. Y ese es el primer paso para establecer si un sistema es válido o no.

Nos remitimos a las reflexiones de Martínez y Martínez (2009b) sobre el concepto de validez y a los postulados de Martínez, Choi, Martínez y Martínez (2009) sobre la especificación de las medidas. Así, el valor de un jugador debe ser consistente con su definición. Para los que apoyan los sistemas “wins” el valor del jugador se refiere a la contribución individual a las victorias del equipo. De este modo las victorias son el criterio sobre el que optimizar un modelo que contenga las diferentes categorías estadísticas. No obstante, la especificación del modelo es fundamental, y ese modelo queda restringido por la información disponible (que en muchos casos sólo se refiere a las categorías del *box-score*). La predicción resultante es una variable distinta a la variable que refleja las victorias del equipo.

Sin embargo, también es posible establecer sistemas de medición a través de la delimitación del concepto que se quiere medir, es decir, de manera formativa, a través de una serie de indicadores que, en este caso, serían categorías estadísticas. Así, por ejemplo, si la versatilidad del jugador se define como su aportación en puntos, rebotes y asistencias, el resultado de esa combinación es una variable formativa compuesta por esas categorías o indicadores. En este caso, la versatilidad no es una variable distinta a la combinación de puntos, rebotes y asistencias, sino que se compone a través de ellas. Llegados a este punto, la cuestión estriba en establecer un consenso sobre las definiciones, a partir del cual poder construir sistemas de medición. Fórmulas de consenso son habituales en otras disciplinas, como en medicina, como muestran Colditz, Atwood, Emmons, Monson, Willett, Trichopoulos y Hunter (2000) en relación a la identificación de los factores causas de cáncer.

Hay que destacar que las variables criterio tienen que ser variables que no estén bajo cuestión, es decir, no se puede establecer como criterio para construir un sistema el Adjusted Plus Minus, tal y como, por ejemplo, realizan Lewin y Rosenbaum (2007) o Paine (2009), porque es un sistema controvertido y que no es más que otra propuesta, similar a las que se quieren evaluar comparándola con ella. Este es uno de los ejemplos clásicos de procedimientos inadecuados de validación (Martínez y Martínez, 2009a).

De manera ideal, deberían existir procedimientos experimentales para establecer la definición de las mediciones de un sistema. Así lo explican Martínez y Martínez (2009b) en relación al concepto de temperatura. En baloncesto, eso es inviable. No obstante, se pueden realizar razonamientos análogos a los realizados por Apgar (1953) y su famoso test para recién nacidos, donde a través de la observación y experiencia se seleccionan una serie de indicadores susceptibles de ser medidos (no tienen por qué ser todos los posibles, sino aquellos que son mensurables de manera práctica), y establecer una variable formativa (una combinación de esos indicadores). La bondad de ese sistema se puede posteriormente contrastar con el resultado real de la salud de los neonatos. El problema es que en baloncesto no está tan claro cuál puede ser esa variable de contraste. Sin embargo, el razonamiento de formación del índice es perfectamente trasladable. La propuesta de Jon Nichols, y su Talent Plus Style (TPS), podría ser un primer paso para seguir esta línea de razonamiento.

Algunos analistas defienden que las variables criterio o las variables de contraste deben ser opiniones subjetivas, es decir, que los resultados de un sistema de valoración deben coincidir con el “sentido común” del baloncesto. Bajo nuestra perspectiva, esta postura es muy cuestionable. Así, por ejemplo, Oliver (2004) critica el trabajo de Berri (1999) porque los resultados finales concluyen que D. Rodman es más productivo que M. Jordan, algo que es como una especie de “insulto” a la inteligencia de los que participan del mundo del baloncesto. Es lo que se suele denominar el “test de la risa”, es decir, cuando un sistema de valoración produce resultados inconsistentes con la sapiencia popular. Sin embargo, a nuestro entender, los sistemas de valoración estadística tienen su fundamento en precisamente romper las barreras de la subjetividad, y obtener resultados “invisibles” extraídos de la explotación de datos objetivos. Por tanto, si ese proceso de cuantificación está diseñado para mejorar la información proveniente de las opiniones subjetivas, no se debería establecer como variable criterio esas opiniones, ya que eso es precisamente lo que se quiere evitar cuando se produce la cuantificación. De este modo, si se cuestionan las métricas estadísticas porque no coinciden con las convenciones subjetivas acerca de la productividad, eficiencia o valor de los jugadores, entonces ¿para qué se necesitan esas estadísticas? El esfuerzo debe ir encaminado a construir sistemas métricos lo más consistentes desde el punto de vista teórico y metodológico, y no en empeñarse en contrastar sus resultados con esas opiniones subjetivas. Otros autores, como Lewin y Rosenbaum (2007), toman como criterio los salarios de los jugadores. Esa puede ser una variable más atractiva que las opiniones subjetivas, pero de nuevo no está exenta de un sesgo potencial importante. Así, por ejemplo, los salarios en la NBA dependen, entre otros, de factores tales como el límite salarial, las exenciones a ese límite, la habilidad de los agentes para la negociación o la experiencia de los jugadores en la liga (Falk, 2009). Ejemplos históricos de contratos que se podrían etiquetar como sobredimensionados los podemos encontrar en jugadores como P. Ewing, A. Mourning, J.

Howard o M. Finley, en relación a otros jugadores que coexistieron con ellos y que están considerados como más importantes y decisivos⁶. Propuestas como el Fair Salary Rating (FSR) podrían ayudar a analizar la idoneidad de esos salarios, algo en lo que también han trabajado otros autores como David Berri. Sin embargo, volvemos de nuevo al problema de si esos sistemas manejados para valorar la idoneidad de los salarios son los más adecuados. Así, FSR, toma el Roland Rating como criterio para realizar una comparativa entre el salario real y el merecido, lo que hace que los resultados dependan de la bondad de las limitaciones de ese sistema. En cualquier caso, hemos analizado la correlación existente entre los salarios merecidos de 60 jugadores, según FSR, y los reales en la temporada 2005/2006 en la NBA: 0.36 (Pearson) y 0.42 (Spearman), lo que podría ser indicativo de una distribución de salarios reales bastante cuestionable, es decir, una muestra de irracionalidad en las decisiones directivas.

Estadística o experiencia

Llegados a este punto, una cuestión empieza a emerger entre algunos analistas: ¿Está realmente justificada toda esta revolución estadística? Es decir, si todos los esfuerzos realizados en construir esa pléyade de sistemas de valoración están justificados por la ventaja competitiva que ellos proporcionan a los decisores. Planteamos al comienzo de esta trilogía de artículos que esa premisa se daba por hecha entre los analistas, o lo que es lo mismo, la “fiebre estadística” es una fuente de información valiosa para reducir el riesgo en las decisiones en baloncesto. Y así lo atestigua, por ejemplo, algunos de los más importantes equipos en la NBA, donde se tienen en nómina a reconocidos profesionales del análisis estadístico. Y por supuesto, el también comentado fenómeno “Moneyball” es gran parte responsable de estos hechos.

Sin embargo, existe también en este mundo del baloncesto un pensamiento más escéptico en relación a las posibilidades que proporciona tanta complejidad estadística. Es cierto, que esta perspectiva, suele estar más arraigada en personas que desconocen las posibilidades de la estadística, pero también ha recibido atención por parte de académicos.

Lewin y Rosenbaum (2007) se plantean si algunos de los sistemas más utilizados en la NBA (WinsProduced, Eficiencia o Adjusted Plus Minus) tienen más importancia en predecir los salarios de los jugadores en relación a dos métricas mucho más sencillas: los puntos por partido y los minutos por partido. Lewin y Rosenbaum (2007) encuentran que las métricas más sofisticadas no superan en importancia relativa a las más simples en explicar los salarios de los jugadores y las victorias de los equipos, y concluyen que a veces el juicio subjetivo de los expertos puede ser más útil que cualquier método de valoración de jugadores derivado numéricamente. Algunos aspectos de la investigación de Lewin y Rosenbaum (2007) son cuestionables, como por ejemplo, el hecho de que los puntos por partido correlacionen más con los salarios que ninguna otra variable es más bien un síntoma de la irracionalidad de algunos decisores, que no ven más allá de esta estadística para evaluar el rendimiento de los jugadores,

⁶ Los salarios de los jugadores correspondientes a la temporada actual, 2009/2010 pueden consultarse en: <http://hoopshype.com/salaries.htm> Como puede comprobarse, jugadores como L. James o D. Wade, no están entre los 20 primeros de la liga, lo que contrasta con jugadores como por ejemplo J. O’Neill o R. Lewis.

(como bien indican Berri, Schmidt y Brook, 2006), que un reflejo de las carencias de los sistemas más sofisticados. Como explicábamos anteriormente, es simplemente un problema de escoger la variable criterio inadecuada. No obstante, Lewin y Rosenbaum (2007), amparados también por reflexiones de otros reputados analistas, como Dean Oliver, hacen un buen trabajo en mostrar ciertas evidencias que reflejan las carencias de ciertos sistemas estadísticos complejos en relación a los juicios subjetivos de expertos.

Y es que esa comparativa entre la capacidad predictiva de diferentes modelos estadísticos en relación a los juicios de expertos está empezando a analizarse cada vez más en la ciencia del deporte, siendo algunos resultados muy curiosos. Por ejemplo, O'Donoghue y Williams (2004) muestran como diferentes modelos estadísticos, que trabajan con distintas asunciones y algoritmos, no son capaces de mejorar de forma importante el éxito en las predicciones de humanos en la Copa del Mundo de Rugby 2003. Es más, la capacidad predictiva de un grupo de expertos fue incluso superior. Resultados similares también se encontraron para diversas predicciones realizadas en relación a los resultados del Mundial de Fútbol 2002 (O'Donoghue et al., 2004). Si diversos sistemas cuantitativos, que manejan un gran volumen de información (mucho más que la procesable por el ser humano) no son capaces de mejorar las predicciones de expertos analistas, entonces habría que comenzar a cuestionar si realmente la sofisticación estadística es útil de manera práctica.

Obviamente hay muchas evidencias que dicen lo contrario. Al margen del ejemplo que ilustra "Moneyball", hemos comentado cómo muchos de los principales equipos de la NBA tienen gran confianza en su staff estadístico. Un ejemplo claro es el trabajo de Daryl Morey, General Manager de los Houston Rockets desde 2006, y un firme defensor de las decisiones de contratación de jugadores basadas en la sofisticación estadística, llevando a los Rockets a unas cotas de rendimiento sorprendentes, fundamentada en jugadores con "poco nombre" pero excelso desempeño, y alabadas por toda la NBA. Además, en ocasiones la complejidad inherente a las interacciones en el campo de juego lleva a que surjan principios contra-intuitivos, que serían incapaces de percibirse si un modelo matemático explicativo. Así, por ejemplo, Skinner (2010), muestra como la estrategia de juego más eficiente para un equipo puede ser la de perder a su jugador a priori más determinante.

En cualquier caso, y como vimos en la segunda parte de esta trilogía de artículos, la decisión de conceder galardones individuales en las principales competiciones del mundo está dominada por los criterios cualitativos, basados en paneles de expertos. En base a todas estas reflexiones, no creemos prudente cuestionar abiertamente ese sistema. Más bien, podemos recomendar que si ese panel de expertos sigue las directrices de Surowiecki (2004): son independientes en su decisión, tienen características heterogéneas y existe la suficiente diversidad de opinión, es muy posible que las decisiones sobre esos galardones sean bastante "sabias" o "justas". El mismo razonamiento se extiende cuando se incluye la votación popular (caso por ejemplo del MVP de la temporada en la Liga ACB), aunque obviamente esas condiciones ideales son difíciles de cumplir en su totalidad, como el propio Surowiecki (2004) reconoce.

Futuro

Actualmente se sigue trabajando duramente en mejorar los sistemas de valoración de jugadores. En APBRmetrics podemos encontrar propuestas novedosas prácticamente cada mes, y discusiones sobre las existentes casi a diario. El mundo académico sigue también esta tendencia, donde trabajos como las de Cooper, Ruiz y Sirvent (2009), por ejemplo, son muy interesantes. Estos autores utilizan el Análisis Envolvente de Datos como método para analizar la eficiencia de los jugadores, y lo aplican a la Liga ACB. Este método tiene un alto nivel de complejidad matemática, y presenta la ventaja de incorporar información cualitativa a la determinación del rango de valores posible de los pesos de las categorías estadísticas. Es una línea de investigación muy interesante para el futuro, y donde estos autores siguen trabajando.

Para que se puedan seguir mejorando los sistemas de valoración se deben disponer de los datos de *play by play*, los cuales son fundamentales para explotar adecuadamente los números de los partidos. Como hemos visto en el segundo artículo, casi todas las competiciones importantes disponen ya de esos datos, y otras que aún no cuentan con ellos tienen planes inmediatos para implementar esa tecnología. Esto ayudaría a que analistas independientes y a que las propias ligas pudieran ofrecer estadísticas sofisticadas como las que hemos repasado en esta investigación. Para ello se necesita que, como ocurre en la NBA, analistas realicen el tedioso trabajo de exportar esos datos a archivos que puedan ser manejados por el resto de la comunidad del baloncesto, aunque lo ideal sería que las propias ligas pusieran a disposición pública y gratuita esas bases de datos.

Este es uno de los grandes problemas que explica la diferencia entre el desarrollo de las estadísticas en Estados Unidos y en el resto del mundo. No obstante, existen ya ejemplos en España de inteligentes intentos de explotación de datos de *play by play*, aunque carecen de la sistematización y desarrollo de Estados Unidos. Así, Francisco Navarro y su valoración SEDENA⁷, o Javier González y sus estudios realizados para MueveteBasket⁸, son exponentes de las bondades que puede conllevar realizar análisis de *play by play*. Desde luego, son un ejemplo de trabajos brillantes e innovadores en España.

A nivel europeo, se están empezando a gestar proyectos que pretenden acercarse a los realizados en Estados Unidos (ej. Basketball-Reference.com, 82games.com o Basketballvalue.com), con el fin de poner el baloncesto en Europa en las cotas analíticas que merece. Alessandro Mazza, analista italiano de www.basketcentral.it, está trabajando en ello, aunque de momento el proyecto no consigue despegar. Otros analistas europeos, como Steven Houston, han realizado trabajos interesantes con los jugadores de la Euroliga. No obstante, quizá el principal exponente en cuanto a análisis avanzado en el baloncesto europeo es Simon Jatsch, creador de www.in-the-game.org, y uno de los pioneros en el análisis del *play by play*, y de la diseminación de estadísticas complejas en nuestro continente. Otras webs, como www.draftexpress.com, también están realizando un gran trabajo analítico, aportando diferentes

⁷ Actualmente, los datos ofrecidos por Francisco Navarro ya no están disponibles en Internet.

⁸ <http://www.muevetebasket.es/autor.php?a=474>

estadísticas avanzadas (PER; WinScore, estadísticas por posesión, etc.) para diversas competiciones europeas, que no se encuentran en las webs oficiales de esas ligas.

Las competiciones oficiales podrían realizar el esfuerzo de mejorar la información estadística proporcionada, al menos en lo que se refiere a los *box-score*. Para ello, podrían añadir estadísticas más avanzadas, tanto ofensivas como defensivas (utilizando algunos de los índices más simples descritos en la primera parte de esta trilogía), o incluyendo otras estadísticas básicas para realizar análisis posteriores, como la posesión, por ejemplo. De nuevo Javier González, esta vez aplicado a la NBA⁹, muestra con su brillante trabajo que esa mejora de los *box-score* es factible, y que proporciona información muy valiosa.

La paulatina incorporación de matemáticos e investigadores y analistas con alta formación en ciencias está enriqueciendo las propuestas de valoración, y llevándolas hacia un terreno cada vez más sofisticado a nivel analítico. Así, de los clásicos usos de métodos estadísticos más básicos, como la regresión lineal, se están trabajando en la aplicación de métodos como la regresión en cresta, métodos bayesianos, modelos multinivel, programación lineal, y otros métodos de minería de datos. El foro de APBRmetrics es una plataforma increíblemente atractiva para compartir y discutir esos avances. Pero no sólo ese foro “informal” está disponible, sino que la creación de la revista *Journal of Quantitative Analysis in Sports*, o las conferencias periódicas organizadas en lugares emblemáticos de la ciencia como Harvard o el MIT, sobre el análisis cuantitativo en el deporte, son fuentes ineludibles para continuar avanzando.

Otro proyecto actualmente en marcha es el de Martínez y Martínez (2010b), un panel de expertos a nivel mundial que está evaluando varios de los temas controvertidos de esta investigación. Así, estos autores están recogiendo la opinión de diversos estamentos relacionados con el mundo de la canasta (jugadores, aficionados, periodistas, analistas, entrenadores, *bloggers*, investigadores, etc.) en relación a la definición del concepto de valor de un jugador, la adecuación de las valoraciones cualitativas, o de las diferentes métricas existentes. Uno de los objetivos de este proyecto es intentar establecer por consenso algunas definiciones e ideas que podrían ayudar a mejorar los sistemas de valoración en el futuro.

En relación al posible cambio en los sistemas de valoración utilizados en las competiciones oficiales y en los *fantasy games*, creemos que todavía es pronto para que este se produzca. Y más teniendo en cuenta los resultados del análisis de correlación realizado en esta investigación, que demuestra que esas métricas (teóricamente débiles en su fundamento) proporcionan resultados casi idénticos a aquellas más sofisticadas. Además, las limitaciones de sistemas más complejos deben ser consideradas. Hay que recordar también que la facilidad de interpretación de los sistemas utilizados actualmente es otra barrera más para el cambio, ya que tanto las ligas como los *fantasy games* son organizaciones cuyo modelo de negocio se basa en los aficionados, y por ello la sencillez de esas estadísticas es un valor añadido.

Es cierto que muchos equipos profesionales, fuera de la NBA, utilizan de manera privada análisis más sofisticados de los que son visibles en las estadísticas oficiales. Felipe (2005) muestra ejemplos de ello en

⁹ <http://userscripts.org/scripts/show/7817>

la Liga ACB. No obstante, son casos puntuales y no existe un desarrollo similar, ni tan sofisticado, como lo que ocurre en la NBA. Pero más allá del uso que los equipos puedan hacer en la trastienda, está el significado social de esta revolución estadística¹⁰. Es decir, el desarrollo de plataformas y proyectos que faciliten la integración de datos, la explotación de esos datos por comunidades heterogéneas de analistas, aficionados e investigadores, la co-creación de contenidos estadísticos por parte de esas comunidades, y la repercusión de ese movimiento en el baloncesto profesional y en los medios de comunicación. En eso, Estados Unidos todavía está muy por delante.

Conclusión final

En esta trilogía de artículos hemos realizado un análisis profundo del estado del arte referente a los sistemas de valoración de jugadores de baloncesto, una cuestión con una innegable trascendencia económica y social en este deporte.

El gran número de sistemas repasado ofrece un dibujo muy atractivo para la gestión deportiva, a la hora de tomar decisiones directivas con una base más racional. Hemos dado orientaciones para elegir entre esa pléyade de opciones, en aras de escoger aquellos sistemas que tengan un menor número de limitaciones. No obstante, aún queda mucho camino por recorrer, con el fin de encontrar un consenso sobre qué sistema o sistemas son la respuesta definitiva a las necesidades del decisor. Tal vez, uno de los factores que incidan en que aún exista tanta controversia es que el baloncesto es un juego de equipo, donde el objetivo es la victoria (del equipo), y donde el desempeño conjunto emerge de las interacciones individuales. Por ello, intentar individualizar el éxito del equipo trae diversos problemas teóricos y metodológicos, dificultades que los analistas e investigadores tratan de superar con innovadores y sofisticadas propuestas.

Las competiciones oficiales y los *fantasy games* utilizan sistemas de valoración muy sencillos, y los que tienen un menor fundamento teórico. Sin embargo, la incorporación de las soluciones tecnológicas del *play by play*, hace que el abanico de posibilidades para la explotación de datos sea enorme. Aún así, tal vez es comprensible que las competiciones quieran seguir manteniendo sus sistemas actuales que, aunque rudimentarios y deficientes en su mayoría, no intentan cambiar debido a que las alternativas disponibles también están cuestionadas, y a que es posible que prefieran seguir utilizando esos sistemas muy limitados, para no desorientar al gran público al que se dirigen. En su mano tienen las herramientas para cambiar esta tendencia.

Wayne Winston, uno de los analistas más importantes, afirma que el 80% de lo que sucede en un partido de baloncesto no se refleja en el *box-score* (Winston, 2009). Ese es uno de los argumentos que utiliza para defender su sistema Plus Minus ajustado. Sin embargo, las posibilidades de los *box-score* se multiplican si se es capaz de aportar estadísticas avanzadas derivadas del *play by play*. Como actualmente ocurre en

¹⁰ Un claro ejemplo de esta “revolución” social se ha dado en la última edición (2010) del Congreso celebrado en el MIT (MIT Sloan Sports Analytics Conference), donde la expectación ha sido tal, que han asistido en torno a 1000 personas, habiendo una lista de espera de 400 que se han quedado sin plaza (Pelton 2010; Sports Business Daily, 2010).

algunas webs especializadas en Estados Unidos, las competiciones podrían ofrecer versiones simplificadas y avanzadas del *box-score*, lo que ayudaría a obtener una visión más fidedigna de las aportaciones individuales de los jugadores.

Como hemos repetido a lo largo de esta investigación, el debate continúa, por lo que este trabajo no pretende ser un punto de inflexión, sino simplemente una herramienta de ayuda para el decisor en aras de realizar predicciones con un menor riesgo. Además creemos que hemos dado los fundamentos para estimular la reflexión acerca del uso de los sistemas de valoración, y de las posibilidades que las competiciones oficiales y los *fantasy games* tienen para ofrecer a su público objetivo. Finalmente, confiamos en que esta exhaustiva recopilación de la situación de los sistemas de valoración en todo el mundo facilite la integración de la información entre diferentes estamentos, competiciones, analistas y aficionados, en pos de conseguir mejorar el análisis “objetivo” y “racional” del deporte de la canasta.

Agradecimientos

Mi más sincero agradecimiento a: *Francisco Navarro, Javier González, Franco Pinotti, Arkadios Chasiride, Jean-Luc Monscheau, Alesssandro Mazza, Steven Houston, John-Charles Bradbury, Mike Goodman y Dick Mays.*

Referencias

- Apgar, V. (1953). A proposal for a new method of evaluation of the newborn infant. *Current Researches in Anaesthesia & Analgesia*, 32 (4), 260–267
- Berri, D. J. (1999). Who is most valuable? Measuring the player’s production of wins in the National Basketball Association. *Managerial and Decision Economics*, 20 (8), 411-427.
- Berri, D. J. (2008). A simple measure of worker productivity in the National Basketball Association. En Brad Humphreys and Dennis Howard (Eds). *The Business of Sport* (pp. 1-40); Westport, Conn: Praeger.
- Berri, D. J., y Bradbury, J. C. (2010). Working in the land of metricians. *Journal of Sports Economics*, 11 (1), 29-47.
- Berri, D. J., Schmidt, M. B., y Brook, S. L. (2006). *The wages of wins: Taking measure of the many myths in modern sport*. Palo Alto,CA: Stanford University Press.
- Bodvarsson, O. B. y Brastow, R. T. (1998). Do employers pay for consistent performance?: Evidence from the NBA. *Economic Inquiry*, 36 (1), 145-160.
- Colditz, G. A., Atwood, K. A., Emmons, K., Monson, R. R., Willett, W. C., Trichopoulos, D., y Hunter, D. J. (2000). Harvard report on cancer prevention. Volume 4: Harvard Cancer Risk Index. *Cancer Causes and Control*, 11, 477-488.
- Cooper, W. W., Ruiz, J. L., Sirvent, I. (2009). Selecting non-zero weights to evaluate effectiveness of basketball players with DEA. *European Journal of Operational Research*, 195 (2), 563–574.
- Falk, D. (2009) *The bald truth*. Pocket books.
- Felipo, J (2005). *Fórmulas para ganar: La revolución estadística del basket*. Barcelona: Zona 131.

- Hollinger, J. (2005). *Pro Basketball Forecast*. Washington, D.C.: Potomac, Inc.
- Goodman, M. (2010, 11 Enero). eWins, WinShares and Ewa. Mensaje enviado al foro de APBRmetrics. Descargado desde <http://sonicscentral.com/apbrmetrics/viewtopic.php?t=2440&postdays=0&postorder=asc&start=0>
- Kubatko, J., Oliver, D., Pelton, K., y Rosenbaum, D. T. (2007). A starting point for analyzing basketball statistics. *Journal of Quantitative Analysis in Sports*, 3 (3), Article 1.
- Lewin, D., y Rosenbaum, D. T. (2007). The pot calling the kettle black. Are NBA statistical models more irrational than “irrational” decision-makers?. *New England Symposium on Statistics in Sports*. Harvard University Science Center
- Martínez, J. A. (2010a). Una revisión de los sistemas de valoración de jugadores de baloncesto (I). Descripción de los métodos existentes. *Revista Internacional de Derecho y Gestión del Deporte*, 10
- Martínez, J. A. (2010b). Una revisión de los sistemas de valoración de jugadores de baloncesto (II). Competiciones oficiales y ligas de fantasía. *Revista Internacional de Derecho y Gestión del Deporte*, 11.
- Martínez, M., Choi, T., Martínez, J. A. y Martínez, A. R. (2009). ISO 9000/1994, ISO 9001/2000 and TQM: The performance debate revisited. *Journal of Operations Management*, 27 (6), 495-511.
- Martínez, J. A. y Martínez, L. (2009a). La validez discriminante como criterio de validación de escalas en marketing, ¿teoría o estadística? *Universitas Psychologica*, 8 (1), 27-36.
- Martínez, J. A. y Martínez, L. (2009b). El análisis factorial confirmatorio y la validez de escalas en modelos causales. *Anales de Psicología*, 25 (2), 368-374.
- Martínez, J. A., y Martínez, L. (2010a). Un método probabilístico para las clasificaciones estadísticas de jugadores en baloncesto. *Revista Internacional de Ciencias del Deporte*, 18 (6), 13-36.
- Martínez, J. A., y Martínez, L. (2010b). Búsqueda del consenso sobre la valoración de jugadores de baloncesto. *Trabajo en curso*
- O'Donoghue, P.G., Dubitzky, W., Lopes, P., Berrar, D., Lagan, K., Hassan, D., Bairner, A., y Darby, P. (2004). An evaluation of quantitative and qualitative methods of predicting the 2002 FIFA World Cup. *Journal of Sports Sciences*, 22 (6), 513-514.
- O'Donoghue, P.G., y Williams, J. (2004). An evaluation of human and computer-based predictions of the 2003 rugby union world cup. *International Journal of Computer Science in Sport (e)*, 3(1), 5-22
- Oliver, D. (2004). *Basketball on paper. Rules and tools for performance analysis*. Washington, D. C.: Brassey's, INC.
- Pelton, K. (2010, 8 Marzo). Entering the mainstream. Descargado desde <http://www.basketballprospectus.com/article.php?articleid=972>
- Paine, N. (2009). Building an Updated Statistical +/- Model for 2009. Descargado desde <http://www.basketball-reference.com/blog/?p=2191>
- Schaller, J. (2001, 15 Junio). The TPR formula; A lesson in logic. Descargado desde <http://schallertpr.alliancesports.com/content.asp?CID=26263>
- Skinner, B. (2010). The price of anarchy in basketball. MIT Sloan Sports Analytics Conference, Harvard, 6 Marzo.

Sports Business Daily (2010, 8 Marzo). MIT Sloan Sports Analytics Conference draws big names, big crowd. Descargado desde

<https://www.sportsbusinessdaily.com/index.cfm?fuseaction=sbd.main&articleID=137571>

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom Shapes Business, Economies, Societies and Nations* Little, Brown

Winston, W. L. (2009). *Mathletics*. New Jersey: Princeton University Press.