



# Cómo resolver el problema de pérdida de información conductual en el diseño de *Behaviorally Anchored Rating Scales-BARS*. El caso de la medición de la eficiencia docente en el contexto universitario

**How to deal with the problem of loss of behavioral information during the construction of Behaviorally Anchored Rating Scales (BARS). The teaching efficiency assessment case in the university context**

MATOSAS-López, Luis [1](#); LEGUEY-Galán, Santiago [2](#) y LEGUEY-Galán, Sonsoles [3](#)

Recibido: 12/02/2019 • Aprobado: 22/05/2019 • Publicado 10/06/2019

## Contenido

- [1. Introducción](#)
  - [2. Metodología](#)
  - [3. Resultados](#)
  - [4. Discusión y conclusiones](#)
- [Referencias bibliográficas](#)

### RESUMEN:

Uno de los problemas identificados en la literatura sobre Behaviorally Anchored Rating Scales (BARS) es el de la pérdida de información conductual durante el diseño del cuestionario. Este trabajo detalla el proceso de diseño de un instrumento con BARS, en el que participan 1.198 estudiantes y 15 docentes. El diseño postulado por los autores añade, con respecto al diseño tradicional, una etapa adicional que permite reducir la pérdida de información conductual sin trastocar la validez y fiabilidad del instrumento resultante.

**Palabras clave:** Eficiencia docente, evaluación del profesorado, ejemplos de comportamiento, universidad

### ABSTRACT:

One of the problems identified on Behaviorally Anchored Rating Scales (BARS) literature is the loss of behavioral information during the design of the questionnaire. This paper describes the process of designing an instrument with BARS, in which 1.198 students and 15 teachers participate. The design postulated by the authors adds, in comparison with the traditional design, an additional stage that allows reducing the loss of behavioral information without affecting the validity and reliability of the resulting instrument.

**Keywords:** Teaching efficiency, teacher evaluation, behavioral examples, university

## 1. Introducción

Las escalas de valoración ancladas conductualmente, más conocidas por su acrónimo en inglés BARS (Behaviorally Anchored Rating Scales), aparecen a principios de los años sesenta, con la

intención de reducir la influencia de las variables de sesgo en los procesos de evaluación de la eficiencia en diferentes contextos laborales (Bernardin & Smith, 1981). La manera en la que se mide el desempeño profesional puede resultar a menudo ambigua y subjetiva. De igual forma, es habitual que en estos procesos no se ofrezcan definiciones explícitas de cuáles son las categorías a considerar o de cuál es el desempeño esperado para cada nivel de valoración. En consecuencia, estas mediciones quedan supeditadas a la interpretación que los evaluadores hacen de las definiciones y niveles de desempeño. Llegando esta situación a ocasionar diferentes calificaciones, incluso cuando existen comportamientos idénticos.

El objetivo de las BARS es atenuar el impacto de las interpretaciones realizadas por parte del evaluador, definiendo el desempeño en términos conductuales y ofreciendo ejemplos concretos de acciones que ilustran los distintos niveles de eficiencia considerados en esa labor (Smith & Kendall, 1963). El valor diferencial de estas escalas con respecto a otros sistemas de medición radica, por tanto, en el empleo de ejemplos de comportamiento para la representación de cada uno de los puntos de anclaje que constituyen la escala de cada una de las categorías a evaluar. El uso de ejemplos de comportamiento en los puntos de anclaje de la escala ayuda a garantizar, por parte del evaluador, una comprensión más estandarizada y uniforme del rendimiento en esa categoría profesional, y esto permite realizar interpretaciones más consistentes, precisas y objetivas (Bernardin & Beatty, 1984).

Aunque la primera propuesta de BARS corresponde a Smith y Kendall (1963), el antecedente natural de este tipo de mediciones es la técnica de incidentes críticos presentada por Flanagan (1954). Esta técnica postula la recolección estructurada y planificada de ejemplos de conductas propias del trabajo en cuestión a través de entrevistas personales, grupales, e incluso encuestas, y siempre con la participación de individuos experimentados en esa actividad. La cercanía de la relación entre la técnica de incidentes críticos y las BARS es tal que Campbell, Dunnette, Arvey y Hellervik (1973) llegan a definir las BARS como escalas basadas en incidentes críticos.

Aun cuando la metodología original de Smith y Kendall (1963) ha sido objeto de variaciones por parte de diferentes autores, los pasos fundamentales son los siguientes: a) se definen de manera detallada las categorías más destacables de la actividad profesional a evaluar; b) un grupo de sujetos con vinculación directa con la actividad proporcionan incidentes críticos eficientes e ineficientes para cada una de las categorías consideradas en el trabajo en cuestión; c) los incidentes críticos son depurados para eliminar aquellos que aparecen duplicados o que resultan ambiguos; d) a continuación, otro grupo de sujetos pasa a reubicar los incidentes críticos en la categoría para la que fueron formulados, eliminándose aquellos que no son correctamente reasignados por la mayor parte de participantes; e) los incidentes críticos que superan la reubicación anterior vuelven a ser valorados por los participantes en una escala de tipo ordinal; f) y por último, los investigadores proceden a seleccionar los incidentes críticos que servirán de punto de anclaje para cada uno de los niveles de desempeño en la escala final.

## 1.1. Ámbitos de aplicación

Desde que las BARS se desarrollaron por primera vez para evaluar el desempeño de profesionales del ámbito de la sanidad (Smith & Kendall, 1963), estas escalas han sido empleadas para medir la forma en que las personas desempeñan diversos roles en multitud de contextos laborales (ver Tabla 1).

**Tabla 1**  
Ámbitos de aplicación de las BARS

Autor	Contexto
Smith y Kendall (1963)	Ámbito sanitario
Landy y Guion (1970)	Ámbito de la ingeniería
Fogli, Hulin y Blood (1971)	Ámbito comercial
Williams y Seiler (1973)	Ámbito de la ingeniería
Arvey y Hoyle (1974)	Ámbito informático

Bearden, Wagner y Simon (1988)	Ámbito de defensa
Woods, Sciarini y Breiter (1998)	Ámbito turístico
Catano (2007)	Ámbito de la seguridad civil
MacDonald y Sulsky (2009)	Ámbito de gestión
Levashina, Hartwell, Morgeson y Campion (2014)	Ámbito de recursos humanos

Fuente: Elaboración propia.

Smith y Kendall (1963) utilizan escalas de comportamiento para valorar el rendimiento de asistentes de enfermería, Landy y Guion (1970) analizan la motivación en el desempeño laboral de ingenieros, Fogli et al. (1971) desarrollan BARS para estudiar el rendimiento de los vendedores de una cadena minorista de alimentación, Williams y Seiler (1973) exploran el esfuerzo y la productividad de profesionales de la ingeniería en el sector industrial, Arvey y Hoyle (1974) aplican estas escalas para medir la labor de programadores-analistas de sistemas, Bearden et al. (1988) emplean BARS para evaluar el trabajo de operadores mecánicos de la marina, Woods et al. (1998) postulan la utilización de escalas de comportamiento para obtener información sobre la actividad de profesionales en el sector hotelero, Catano (2007) utiliza BARS para recabar información destinada a la toma de decisiones de promoción de oficiales del cuerpo de policía, MacDonald y Sulsky (2009) emplean este sistema para comparar la eficiencia de modelos de gestión en culturas empresariales orientales y occidentales, y Levashina et al. (2014) usan estas escalas para analizar el comportamiento y la adecuación de candidatos a procesos de selección de personal.

Asimismo, desde los años setenta, las BARS han sido empleadas intensamente también para evaluar la actividad en el contexto educativo llegando a alcanzar una buena acogida. Bernardin (1977) usa escalas con ejemplos de comportamiento para medir la eficiencia docente de profesores universitarios, Kavanagh y Duffy (1978) emplean este tipo de escalas para evaluar un programa de educación a distancia orientado a mejorar las competencias lectoras, Hom, DeNisi, Kinicki y Bannister (1982) utilizan este sistema para valorar a los docentes de un programa de cursos de verano, Ohland et al. (2012) emplean BARS para medir el desempeño de los miembros de equipos de trabajo en el contexto universitario en estudios de empresa, Fernández Millán y Fernández Navas (2013) usan estas escalas para evaluar la eficiencia de educadores sociales en centros de protección de menores, y Martin-Raugh, Tannenbaum, Tocci y Reese (2016) aplican BARS para valorar la práctica docente en el desarrollo de debates grupales y de la producción de materiales en profesores de Inglés y Matemáticas en educación primaria.

## 1.2. Ventajas de las BARS

No es una casualidad que en los últimos cincuenta años las escalas con ejemplos de comportamiento hayan mantenido el interés de la comunidad investigadora de muy distintos ámbitos. Parte del éxito de las BARS radica en las ventajas psicométricas que este tipo de instrumento plantea con respecto a otros sistemas de medición.

En este capítulo destaca su capacidad para reducir tanto el efecto halo como el error de lenidad. El efecto halo se define como la tendencia del evaluador a extrapolar la calificación otorgada en una determinada pregunta al resto de ítems de la encuesta (Bernardin, 1977). El error de lenidad, por su parte, es la propensión del evaluador a valorar al sujeto de manera demasiado alta o demasiado baja en todos los ítems de la encuesta (Sharon & Bartlett, 1969). Son muchas las investigaciones que demuestran como las BARS tienden a producir menor efecto halo y error de lenidad que otro tipo de escalas como puedan ser las escalas de Likert (Bernardin, Alvares, & Cranny, 1976; Borman & Dunnette, 1975; Campbell et al., 1973). Smith y Kendall (1963) afirman que la reubicación y retención durante el proceso de construcción de tan solo aquellos incidentes críticos que alcanzan un sustancial nivel de acuerdo entre los participantes resulta clave en estas reducciones.

Otros beneficios de las BARS son las mejoras en la validez de la medición y las reducciones en la influencia de variables de sesgo durante la evaluación (Debnath et al., 2015; Martin-Raugh et al., 2016; Ohland et al., 2012). Murphy y Pardaffy (1989) apuntan que este hecho puede derivarse de

la disgregación total entre las escalas, es decir, el desarrollo de escalas independientes para cada categoría de desempeño aísla a cada una de ellas de los sesgos que pudieran derivarse de otras. En la misma línea, también existen estudios que indican que la utilización de BARS redundaría en mejoras de la fiabilidad entre evaluadores. Bernardin (1977) define este concepto como el grado de acuerdo alcanzado entre los evaluadores en varias categorías para cada sujeto calificado. Aunque no existe un consenso total en torno a este punto (Campbell et al., 1973; Schwab, Heneman, & DeCotiis, 1975), la mayoría de estudios apuntan que las escalas con ejemplos de comportamiento proporcionan indicadores de fiabilidad entre evaluadores moderadamente superiores a los encontrados en otros sistemas de medición (Bearden et al., 1988; W. C. Borman & Vallon, 1974; Debnath, Lee, & Tandon, 2015; Williams & Seiler, 1973; Zedeck & Baker, 1972). De acuerdo con Bernardin y Smith (1981) la utilización de ejemplos de comportamiento para la representación de cada punto de anclaje de la escala, unido al uso de categorías claramente separadas entre sí, contribuye a esta mejora en la fiabilidad.

Los beneficios de las escalas con episodios de comportamiento son tales que algunos autores llegan a afirmar de manera categórica que las BARS son superiores técnica o psicométricamente a cualquier otro sistema de medición (Borman & Dunnette, 1975; Goodale & Burke, 1975). Esta superioridad en términos psicométricos es, a menudo, atribuida al rigor con el que se desarrolla el instrumento de medición (Borman, 1991), a la implicación directa de sujetos vinculados con la actividad a evaluar en el diseño de la escala (Bernardin & Beatty, 1984), e incluso a los beneficios derivados del uso de una terminología familiar para el evaluador en el cuestionario final (Jacobs, Kafry, & Zedeck, 1980).

### 1.3. Inconvenientes de las BARS

Aunque las escalas con ejemplos de comportamiento han demostrado, con el transcurso de los años, brindar importantes beneficios en la evaluación del rendimiento profesional, estos instrumentos no están exentos de limitaciones. La literatura sobre BARS destaca tres problemas a este respecto: a) la elevada inversión de tiempo y esfuerzo requerida para el diseño y construcción del instrumento; b) la dificultad para obtener ejemplos de comportamiento representativos de los puntos medios de la escala; y c) la pérdida de información conductual sufrida a lo largo del proceso de construcción.

Aunque parte del potencial de las BARS reside en su minucioso diseño, este es también uno de los inconvenientes asociados a su utilización. La gran cantidad de tiempo y esfuerzo necesario para recopilar y depurar los incidentes críticos a lo largo de las sucesivas etapas que conlleva el diseño de la escala puede suponer una barrera para aquellos interesados en este tipo de cuestionarios (Goodale & Burke, 1975; Stoskopf, Glik, Baker, Ciesla & Cover, 1992).

Otro de los problemas identificados en la literatura es la dificultad para alcanzar grados de acuerdo sustanciales entorno a los incidentes críticos que ilustran los puntos de anclaje intermedios en la escala. Mientras que los niveles de acuerdo en torno a los puntos extremos - los de menor y mayor eficiencia - suelen presentar un quorum relativamente amplio, la gestión de los valores intermedios supone un reto para el investigador. (Debnath et al., 2015; Hauenstein, Brown & Sinclair, 2010).

No obstante, de entre los tres problemas señalados, el de la pérdida de información conductual durante el diseño del instrumento es el asunto tratado de forma más recurrente en el acervo científico especializado en esta área de investigación. La revisión de la literatura sobre BARS realizada por Schwab et al. (1975) pone de manifiesto la constante y sustancial pérdida de información en la construcción de este tipo de escalas. El propio autor define esta pérdida de información como consecuencia directa de la eliminación de ejemplos de comportamiento a lo largo de los cribados realizados durante el proceso de construcción del cuestionario. Este problema de pérdida de información en forma de episodios de comportamiento ha quedado constatado en múltiples estudios en el transcurso de los años.

**Tabla 2**  
Pérdidas de información en el diseño de BARS en trabajos previos.

Autor	Ejemplos de comportamiento iniciales	Ejemplos de comportamiento en el instrumento final	% Ejemplos de comportamiento perdidos (Pérdida de información)

Harari y Zedeck (1973)	310	78	74.84%
Zedeck, Imperato, Krausz y Oleno (1974)	420	177	57.86%
Goodale y Burke (1975)	360	70	80.56%
Borman y Dunnette (1975)	400	70	82.50%
Bernardin et al. (1976)	497	63	87.32%
Dickinson y Zellinger (1980)	731	65	91.11%
Carretta y Walters (1991)	261	37	85.82%
Pounder (2000)	592	40	93.24%
Kell et al. (2017)	681	128	81.20%
Klieger et al. (2018)	430	48	88.84%

Fuente: Elaboración propia

A pesar de los esfuerzos del investigador, al comenzar el diseño de la escala, por recabar incidentes críticos capaces de ilustrar los diferentes niveles de desempeño en la actividad, la mayor parte de estos resultan desechados a lo largo del proceso. Este hecho, ocasiona la pérdida de una importante cantidad de información conductual, información que, por otro lado, constituye la piedra angular en este tipo de escala. La merma de ejemplos de comportamiento desvirtúa, por tanto, el potencial del propio instrumento. Así, la pérdida de información ocasionada por el depurado de episodios de comportamiento, durante el desarrollo de la escala, se revela como uno de los problemas trascendentales para en la utilización de BARS.

## 1.4. Objetivo

El presente trabajo tiene el propósito de dar respuesta al problema de pérdida de información conductual identificado en la literatura previa sobre BARS. Para ello los autores postulan un proceso de diseño de BARS que presenta una etapa adicional (etapa 5) sobre el procedimiento de construcción habitual. Asimismo, los investigadores abordan también el análisis de la calidad psicométrica de la escala resultante mediante la exploración de la validez y la fiabilidad del instrumento obtenido.

Considerando la intensidad con la que las BARS han sido empleadas en el ámbito de la educación, los autores abordan la presente investigación en el contexto de la evaluación de la eficiencia del profesorado dentro del ámbito universitario.

## 2. Metodología

La investigación, enmarcada dentro del ámbito educativo, se contextualiza en el sistema universitario en España. En el estudio toman parte 15 profesores y un colectivo de 1.198 estudiantes, todos ellos pertenecientes a la Universidad Rey Juan Carlos (en adelante URJC). La muestra de estudiantes se toma por muestreo incidental (Mayorga Fernández & Ruiz Baeza, 2002) contando con la participación de alumnos de distintos grados y cohortes. La muestra de 1.198 participantes queda distribuida en dos grandes grupos. Un primer colectivo de 974 estudiantes que participa en el diseño del instrumento de medición, y un segundo colectivo de 224 alumnos que toma parte en los análisis de validez y fiabilidad del cuestionario obtenido.

### 2.1. Instrumento BARS para reducir la pérdida de información

El procedimiento empleado sigue las pautas de la metodología diseñada originalmente por Smith y Kendall (1963). Sobre esta base los autores introducen una etapa adicional (etapa 5) con respecto

al procedimiento tradicional. En esta etapa los autores agrupan los incidentes críticos en Aspectos de Comportamiento Básico o ACB (Matosas, 2018), en un paso que permite reducir la pérdida de información conductual. El proceso de construcción comprende siete etapas sucesivas. La construcción del instrumento BARS implica al colectivo de 15 docentes y al primer grupo de estudiantes considerados en la muestra ( $n = 974$ ).

### **Etapas 1. Descripción de las categorías**

Un panel de profesores ( $n = 5$ ) efectúa una descripción pormenorizada de las diez categorías docentes consideradas. Las categorías son: Introducción a la asignatura, Descripción del sistema de evaluación, Gestión del tiempo, Disponibilidad general, Coherencia organizativa, Implementación del sistema de evaluación, Resolución de dudas, Capacidad explicativa, Facilidad de seguimiento y Satisfacción general.

### **Etapas 2. Ejemplos de comportamiento**

Un grupo de estudiantes ya egresados ( $n = 25$ ) proporciona ejemplos de comportamientos eficientes e ineficientes en la labor docente para cada una de las categorías contempladas. En consonancia con Flanagan (1954), los ejemplos de comportamiento o incidentes críticos son recogidos mediante la utilización de entrevistas grupales. Los participantes en esta etapa proporcionan 321 incidentes críticos.

### **Etapas 3. Selección de ejemplos de comportamiento**

Un segundo panel de profesores ( $n = 5$ ) revisa los ejemplos de comportamiento recopilados durante la segunda etapa para eliminar redundancias o ejemplos ambiguos. El número de incidentes críticos se reduce aquí de 321 a 278 elementos en base a las recomendaciones realizadas por los docentes del panel.

### **Etapas 4. Reubicación**

Un colectivo de estudiantes ( $n = 70$ ) aborda la etapa de reubicación. Este paso implica la reclasificación de los 278 incidentes críticos, que superaron el paso anterior, en la categoría docente para la que fueron formulados. Esta reubicación es realizada mediante un cuestionario *online*. En línea con Bernardin (1977) se mantienen solo aquellos episodios de comportamiento que son correctamente reasignados en la categoría a la que pertenecen, por al menos un 80% de los participantes. En esta etapa el número de elementos pasa de 278 a 229.

### **Etapas 5. Agrupación en Aspectos Básicos del Comportamiento (ACB)**

Un panel de profesores ( $n = 5$ ) revisa minuciosamente los 229 incidentes críticos que sobreviven a la etapa de reubicación, con el propósito de identificar los aspectos a los que aluden de forma recurrente los episodios de cada categoría. Identificados estos aspectos, en línea con Matosas (2018), los incidentes críticos clasificados en cada categoría docente son agrupados en subcategorías de episodios sintetizados, que pasan a ser denominadas Aspectos de Comportamiento Básico (ACB). El panel de profesores concluye que cada dimensión puede redefinirse utilizando cuatro ACB diferenciados e inequívocos. De esta forma se obtienen un total de cuarenta ACB considerando las diez categorías de interés. De acuerdo con las sugerencias de Flanagan (1954), para la definición de incidentes críticos, los cuarenta ACB son adaptados para proporcionar afirmaciones concisas y descriptivas, manteniendo el vocabulario original de los estudiantes.

Un total de 215 incidentes críticos quedan agrupados y sintetizados en una de las cuarenta subcategorías de ACB creadas. Tan solo catorce incidentes críticos no pueden ser reagrupados en ninguna de las cuarenta categorías por no guardar relación alguna con ningún otro ejemplo de comportamiento. En esta etapa se pasa, por tanto, de 229 incidentes críticos a 215 que resultan agrupados adoptando la forma de 40 ACB.

### **Etapas 6. Evaluación dual de episodios de comportamiento**

Un colectivo de estudiantes de grado ( $n = 879$ ) realiza el trabajo de evaluación dual de esta etapa. El objetivo de esta fase es ordenar los ACB desde la perspectiva del estudiante, considerando la importancia que cada ACB tiene para el alumnado.

Los participantes realizan este trabajo mediante el empleo de un cuestionario *online* en el que se abordan dos tareas diferenciadas. La primera es la evaluación de los ACB de cada categoría, la segunda es la puntuación de una afirmación representativa de cada categoría usando una escala tipo Likert. Para la ejecución de ambas tareas, se pide a los estudiantes que tengan en cuenta el rendimiento de uno de sus profesores durante el cuatrimestre anterior.

**a) Evaluación de ACB**

En primer lugar, los estudiantes evalúan los cuatro ACB incluidos en cada categoría utilizando un cuestionario que emplea un método de evaluación dicotómico en términos de "Cumplido" o "No cumplido".

**Figura 1**  
Evaluación de ACB en la categoría Resolución de dudas.

		Cumplido	No cumplido
<b>ACB1</b>	El profesor/a genera un ambiente adecuado que anima a los estudiantes a formular dudas o compartir opiniones	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<b>ACB2</b>	El profesor/a establece pausas específicas durante las clases para formular dudas	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<b>ACB3</b>	El profesor/a es capaz de hacer frente a las dudas de los estudiantes presentando un solo concepto de varias maneras	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<b>ACB4</b>	El profesor/a resuelve las dudas de los estudiantes usando ejemplos prácticos / materiales de apoyo para afianzar la idea	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Fuente: Elaboración propia

En este punto, en lugar de dar una valoración ordinal a incidentes críticos aislados, como se haría en la metodología tradicional, los sujetos evalúan ACB. Considerando que cada grupo de cuatro ACB es tratado de manera independiente dentro de cada categoría, la evaluación dicotómica ("Cumplido" o "No cumplido") de los ACB de cada dimensión genera dieciséis potenciales escenarios por categoría y un total de 160 combinaciones de ACB. Así, por ejemplo, una situación como la reflejada en la figura 1 representa un escenario en el que el docente satisface o cumple con los episodios sintetizados en los ACB1, ACB2, ACB3 y ACB4.

**b) Evaluación de la afirmación representativa de la categoría con escala Likert**

En segundo lugar, el mismo grupo de estudiantes, manteniendo como referencia el rendimiento del profesor previamente seleccionado, y de nuevo mediante el mismo cuestionario *online*, puntúa una afirmación representativa de cada categoría empleando una escala tipo Likert de cinco niveles (1=Muy en desacuerdo, 5=Muy de acuerdo).

**Figura 2**  
Evaluación de una afirmación representativa de la categoría Resolución de dudas.

	1	2	3	4	5
El profesor/a aclara adecuadamente las dudas de las distintas actividades propuestas en la asignatura	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>1=Muy en desacuerdo / 5=Muy de acuerdo</i>					

Fuente: Elaboración propia.

La primera valoración en este sistema de evaluación dual genera 160 combinaciones de ACB. La segunda indica la puntuación asociada a cada combinación de ACB, desde la perspectiva del estudiante. Este proceso de evaluación dual ordena los ACB en términos de importancia para los participantes. Esto permite a los investigadores determinar la combinación de ACB esperada por el estudiante para cada nivel de desempeño.

Por ejemplo, a la vista de las figuras 1 y 2 indicaremos que, desde la perspectiva del estudiante, en la categoría Resolución de dudas el escenario en el que el profesor satisface los ACB1, ACB2, ACB3 y ACB4 (ver Figura 1), correspondería al nivel máximo de eficiencia (ver Figura 2).

## Etapa 7. Generación de la escala final

Completadas las dos tareas anteriores, se calculan los promedios y las desviaciones típicas para cada una de las 160 combinaciones de ACB generadas. El promedio identifica el punto de la escala que esa combinación de ACB podría ocupar en el cuestionario final; mientras que la desviación típica describe el grado de acuerdo entre los evaluadores a la hora de ubicar esa combinación de ACB en el punto indicado por el promedio. En línea con Dickinson y Zellinger (1980), los autores conservan para su posible inclusión en el instrumento final solo aquellas combinaciones de ACB con desviaciones típicas iguales o inferiores a 1.25.

Empleando estas combinaciones de ACB supervivientes como puntos anclaje los autores proceden a generar la escala final. Para ello se definen cuatro intervalos de clase (de 1 a 1.80 / de 1.81 a 2.60 / de 2.61 a 3.40 / de 3.41 a 4.20 / de 4.21 a 5), asignándose las combinaciones de ACB a cada intervalo de acuerdo con el valor promedio obtenido en la evaluación dual anterior. En consecuencia, cada combinación de ACB cae en un intervalo de clase. Cuando más de una combinación de ACB cae en un determinado intervalo, la elección se realiza tomando la combinación de ACB que muestra, en términos de desviación típica, el mayor grado de acuerdo entre los estudiantes.

**Figura 3**  
Escala final para la categoría Resolución de dudas.

1	El profesor/a <b>NO</b> genera un ambiente adecuado que anime a los estudiantes a formular dudas o compartir opiniones; <b>NO</b> establece pausas específicas durante las clases para formular dudas; <b>NO</b> es capaz de hacer frente a las dudas de los estudiantes presentando un solo concepto de varias maneras y <b>NO</b> resuelve las dudas de los estudiantes usando ejemplos prácticos / materiales de apoyo para afianzar la idea
2	El profesor/a establece pausas específicas durante las clases para formular dudas
3	El profesor/a genera un ambiente adecuado que anima a los estudiantes a formular dudas o compartir opiniones y establece pausas específicas durante las clases para formular dudas
4	El profesor/a genera un ambiente adecuado que anima a los estudiantes a formular dudas o compartir opiniones; es capaz de hacer frente a las dudas de los estudiantes presentando un solo concepto de varias maneras y resuelve las dudas de los estudiantes usando ejemplos prácticos / materiales de apoyo para afianzar la idea
5	El profesor/a genera un ambiente adecuado que anima a los estudiantes a formular dudas o compartir opiniones; establece pausas específicas durante las clases para formular dudas; es capaz de hacer frente a las dudas de los estudiantes presentando un solo concepto de varias maneras y resuelve las dudas de los estudiantes usando ejemplos prácticos / materiales de apoyo para afianzar la idea

Fuente: Elaboración propia

La figura 3 representa la escala final resultante para la categoría docente Resolución de dudas. El primer punto de anclaje muestra una situación de incumplimiento en los ACB1, ACB2, ACB3 y ACB4; el segundo punto de anclaje coincide con el cumplimiento del ACB2; el tercer punto de anclaje representa el cumplimiento de ACB1 y ACB2 de forma simultánea, el cuarto punto de anclaje refleja en el logro de ACB1, ACB3 y ACB4, y el punto más alto se alcanza cuando el profesor cubre las expectativas de los estudiantes en los cuatro ACB (ver Figura 1 para identificar ACB).

## 2.2. Validez y fiabilidad del instrumento BARS

Una vez construido el instrumento BARS se procede a analizar la calidad psicométrica de este, explorando su validez y fiabilidad. Para tal tarea se administra el cuestionario BARS al segundo grupo de estudiantes considerados en la muestra (n = 224), con el objetivo de que estos realicen la evaluación de la labor docente de un profesor de referencia tras la finalización de uno de sus cursos. Una vez recabados los datos, se estudia la validez de constructo y la fiabilidad del instrumento utilizando los resultados obtenidos.

La validez de constructo, en línea con otros estudios de validación de cuestionarios en el contexto universitario (Gutiérrez-Castillo, Juan-Jesús Cabero-Almenara & Estrada-Vidal, 2017; Luna



Serrano, 2015), se explora empleando un análisis factorial exploratorio (AFE). Por su parte, el análisis de fiabilidad se aborda considerando el coeficiente Alfa de Cronbach (Fernández Millán & Fernández Navas, 2013; Lizasoain-Hernández, Etxeberria-Murgiondo & Lukas-Mujika, 2017). Todos los análisis son desarrollados empleando el software de análisis estadístico IBM SPSS en su versión 25.

---

## 3. Resultados

### 3.1. Instrumento BARS para reducir la pérdida de información

El proceso de construcción descrito por los autores refleja un diseño en el que se parte de 321 incidentes críticos. De estos, 215 quedan representados en el instrumento final adoptando la forma de 40 ACB. Esto supone, por tanto, la pérdida de la información de comportamiento contenida en 106 incidentes críticos. Por ende, los ejemplos de conductas representativas de la labor docente recogidos en estos 106 elementos, ya sea por problemas de duplicidad, ambigüedad o por haber sido desechados en alguna de las fases de depuración durante el diseño, no quedan representados de manera alguna en el cuestionario final. En consecuencia, la información contenida en estos 106 episodios de comportamiento refleja la pérdida de información sufrida durante la construcción de la escala, pérdida que representa el 32.02% de los ejemplos de comportamiento iniciales.

Asimismo, el análisis de los datos de pérdida de información conductual sufrida con posterioridad a la etapa de reubicación (fase clave en el diseño de BARS) muestra como de los 229 incidentes críticos restantes tras esta fase, solo catorce son descartados antes de conformar las 40 subcategorías de ACB. Este paso de 229 ejemplos de comportamiento a 215 representa la pérdida de información experimentada tras la etapa de reubicación, pérdida que queda, por tanto, cuantificada en un 4.36%.

### 3.2. Validez y fiabilidad del instrumento BARS

Antes de proceder con el AFE requerido para analizar la validez de constructo del cuestionario final se calcula la medida de adecuación muestral Kaiser-Meyer-Olkin y la prueba de esfericidad de Bartlett. El propósito de la extracción de ambos estadísticos es el de evaluar la adecuación de los datos a la técnica de AFE. El valor de Kaiser-Meyer-Olkin obtenido es de .942, superando el valor recomendado de .60 (Kaiser, 1974). Por su parte, la prueba de esfericidad de Bartlett logra una significación del .000 (Bartlett, 1954). Ambos resultados corroboran la existencia de correlaciones suficientes, ratificando así la idoneidad del análisis.

El AFE se desarrolla empleando el método de extracción de componentes principales, con rotación Varimax, y extracción de factores siguiendo el criterio de auto valores superiores a 1. Las cargas observadas en la matriz de componentes rotados muestran la estructura factorial del instrumento, revelando la existencia de dos constructos subyacentes. Estos constructos son capaces de explicar un 79.095% de la varianza total del instrumento, siendo el detalle de la composición de cada uno de ellos el descrito a continuación.

- Factor 1. Constructo capaz de explicar el 43.134% de la varianza y que recoge siete categorías o preguntas. Estas son: Gestión del tiempo, Resolución de dudas, Satisfacción general, Capacidad explicativa, Disponibilidad general, Introducción a la asignatura y Facilidad de seguimiento. Este factor queda identificado con la denominación "Actitud y aptitud docente".
- Factor 2. Constructo capaz de explicar el 35.960% de la varianza y que recoge tres categorías o preguntas. Estas son: Coherencia organizativa, Descripción del sistema de evaluación e Implementación del sistema de evaluación. Este factor queda identificado con la denominación "Organización y evaluación del curso".

Por otro lado, la fiabilidad del cuestionario se examina empleando el coeficiente Alfa de Cronbach. La consistencia interna de los elementos que conforman el primer constructo muestra un coeficiente Alfa de Cronbach de .842. Por su parte el Alfa de Cronbach para el segundo constructo arroja un valor de .786. En línea de con Kerlinger, Lee, Pineda y Mora Magaña (2002), los valores identificados para cada uno de los dos factores del instrumento final son considerados como óptimos. Asimismo, la fiabilidad del instrumento en su conjunto presenta un coeficiente Alfa de Cronbach de .871.

---

## 4. Discusión y conclusiones

Aunque las BARS posibilitan mejoras psicométricas con respecto a otros sistemas de medición, estas escalas también cuentan con limitaciones. La pérdida de información conductual durante el diseño del instrumento es uno de los problemas más delicados. Dado que parte del potencial de estos cuestionarios radica en el uso de ejemplos de comportamiento en la construcción de la escala, una excesiva merma de información puede llegar a desvirtuar el valor del instrumento.

El presente trabajo postula un proceso de construcción capaz de reducir sustancialmente la pérdida de información conductual durante el diseño de la escala. La propuesta de los autores incorpora una etapa adicional (etapa 5) con respecto al procedimiento tradicional. Esta etapa tiene el objetivo de agrupar los incidentes críticos de cada categoría en subcategorías de episodios sintetizados o ACB. La nueva fase permite al investigador retener la información de comportamiento de varios incidentes críticos, de manera simultánea, en un único ACB. Este ACB, redactado y adaptado de la manera oportuna por el investigador, sirve en última instancia de punto de anclaje para la escala en el instrumento final. En consecuencia, la agrupación de incidentes críticos en ACB introducida por los autores (etapa 5), unida a la posterior evaluación dual de la información, reduce notablemente la pérdida de información durante el proceso de construcción.

La comparación de la pérdida de información, entendida como minoración de incidentes críticos, experimentada en términos porcentuales en la presente metodología con la sufrida en estudios previos que aplican el procedimiento tradicional, revela reducciones significativas. En el trabajo desarrollado por Zedeck et al. (1974), que destaca por ser uno de los que refleja menor merma de incidentes críticos en la literatura, se pierde el 57.86% de la información conductual. Por su parte, en el trabajo de Pounder (2000), estudio que recoge una de las pérdidas de información más cuantiosas identificadas por la literatura, se desecha el 93.24% de la información de comportamiento considerada inicialmente. La pérdida de información conductual, del 32.02%, experimentada en el presente trabajo, deja diferencias de 25.84 puntos porcentuales con respecto al primer ejemplo y 61.22 con respecto al segundo.

Cabe destacar también el dato de pérdida de información, en la metodología descrita, con respecto a estudios en los que se aplican estándares de reubicación menos exigentes. El estándar de reubicación del 80%, empleado en el presente trabajo, establece un elevado umbral de validación durante esta etapa; esto permite contar con incidentes críticos de calidad y bien definidos, pero también dificulta que un ejemplo de comportamiento se mantenga y favorece la eliminación de información conductual durante el diseño. A pesar de ello, el presente estudio, aplicando un estándar de reubicación del 80%, refleja una pérdida de información de comportamiento (32.02%) sensiblemente inferior a la experimentada en otras investigaciones. Por ejemplo, el trabajo de Harari y Zedeck (1973) - trabajo que con un estándar de reubicación del 60% refleja uno de los estándares menos exigentes de la literatura - muestra como de los 310 incidentes críticos iniciales se pasa a 78 en el instrumento final, revelando la pérdida del 74.84% de la información conductual. Por tanto, la agrupación de ejemplos de comportamiento en ACB propuesta por los autores en el paso quinto, es capaz de moderar la pérdida de información con independencia del estándar de reubicación empleado.

Asimismo, los autores resaltan también el dato de pérdida de información de comportamiento experimentado con posterioridad a la mencionada etapa de reubicación. Mientras en la presente metodología se observa una pérdida del 4.36%, otros trabajos sufren mermas sustancialmente mayores. Así, por ejemplo, el estudio de Dickinson y Zellinger (1980) muestra como los 123 episodios de comportamiento que superan la fase de reubicación quedan reducidos a 65 en el cuestionario definitivo, mostrando una merma del 47.15%. Mientras la metodología tradicional elimina numerosos episodios de comportamiento en las etapas finales del diseño, el paso de agrupación de incidentes críticos en ACB permite al investigador conservar la mayor parte de las evidencias de conducta proporcionadas por los participantes y usar las mismas en la generación del cuestionario final.

Por último, el presente estudio demuestra que la variación metodológica introducida por los autores en el proceso de construcción con el propósito de minorar la pérdida de información no ocasiona perjuicio alguno en la calidad del instrumento en términos de validez y fiabilidad. La aplicación del cuestionario BARS creado por los autores en la segunda muestra de estudiantes participantes en la investigación revela la existencia de dos constructos claramente diferenciados ("Actitud y aptitud docente" y "Organización y evaluación del curso") capaces de explicar gran parte de la varianza del fenómeno de interés. De la misma forma, los resultados alcanzados durante el análisis de fiabilidad constatan también la robustez del instrumento final, tanto a nivel individual para cada uno de los dos constructos observados, como de manera global para el conjunto del cuestionario.

A la vista de los resultados, los autores concluyen que la etapa de agrupación de episodios de comportamiento en ACB y la posterior evaluación dual desarrollada durante el proceso contribuye a resolver el problema de pérdida de información conductual detectado en la literatura previa sobre BARS; todo ello sin trastocar la validez y fiabilidad del instrumento resultante.

---

## Referencias bibliográficas

- Arvey, R. D., & Hoyle, J. C. (1974). A Guttman approach to the development of behaviorally based rating scales for systems analysts and programmer/analysts. *Journal of Applied Psychology, 59*(1), 61–68. <http://doi.org/10.1037/h0035830>
- Barlett, M. S. (1954). A note on multiplying factors for various chi-squared approximations. *Journal of the Royal Statistical Society, Series B 1*(16), 296–298.
- Bearden, R. M., Wagner, M., & Simon, R. (1988). *Developing Behaviorally Anchored Rating Scales for the Machinist's Mate Rating*. San Diego. Recuperado de <https://apps.dtic.mil/dtic/tr/fulltext/u2/a195403.pdf>
- Bernardin, H. J. (1977). Behavioural expectation scales versus summated scales. *Journal of Applied Psychology, 62*(4), 422–427. Recuperado de <http://psycnet.apa.org/record/1978-09104-001>
- Bernardin, H. J., Alvares, K. M., & Cranny, C. J. (1976). A recomparision of behavioral expectation scales to summated scales. *Journal of Applied Psychology, 61*(5), 564–570. <http://doi.org/10.1037/0021-9010.61.5.564>
- Bernardin, H. J., & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work*. (PWS, Ed.). Boston: Kent Pub. Co.
- Bernardin, H. J., & Smith, P. C. (1981). A clarification of some issues regarding the development and use of behaviorally anchored ratings scales (BARS). *Journal of Applied Psychology, 66*(4), 458–463. <http://doi.org/10.1037/0021-9010.66.4.458>
- Borman, W. (1991). Job Behavior, Performance, and Effectiveness. In M. D. Dunnette & L. M. Hough (Ed.), *Handbook of Industrial and Organizational Psychology* (pp. 271–326).
- Borman, W. C., & Dunnette, M. D. (1975). Behavior-based versus trait-oriented performance ratings: An empirical study. *Journal of Applied Psychology, 60*(5), 561–565. <http://doi.org/10.1037/0021-9010.60.5.561>
- Borman, W. C., & Vallon, W. R. (1974). A view of what can happen when Behavioral Expectation Scales are developed in one setting and used in another. *Journal of Applied Psychology, 59*(2), 197–201. <http://doi.org/10.1037/h0036312>
- Campbell, J. P., Dunnette, M. D., Arvey, R. D., & Hellervik, L. V. (1973). The development and evaluation of behaviorally based rating scales. *Journal of Applied Psychology, 57*(1), 15–22. <http://doi.org/10.1037/h0034185>
- Carretta, T. R., & Walters, L. C. (1991). *The Development of Behaviorally Anchored Rating Scales (BARS) for Evaluating USAF Pilot Training Performance*. Texas.
- Catano, V. M. (2007). Performance Appraisal of Behavior-Based Competencies: a Reliable and Valid Procedure. *Personnel Psychology, 60*, 201–230. <http://doi.org/10.1111/j.1744-6570.2007.00070.x>
- Debnath, S. C., Lee, B., & Tandon, S. (2015). Fifty years and going strong: What makes Behaviorally Anchored Rating Scales so perennial as an appraisal method? *International Journal of Business and Social Science, 6*(2), 16–25.
- Dickinson, T. L., & Zellinger, P. M. (1980). A comparison of the behaviorally anchored rating and mixed standard scale formats. *Journal of Applied Psychology, 65*(2), 147–154. <http://doi.org/10.1037//0021-9010.65.2.147>
- Fernández Millán, J. M., & Fernández Navas, M. (2013). Elaboración de una escala de evaluación de desempeño para educadores sociales en centros de protección de menores. *Intangible Capital, 9*(3), 571–589. <http://doi.org/10.3926/ic.410>
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4), 327–358. <http://doi.org/10.1037/h0061470>
- Fogli, L., Hulin, C. L., & Blood, M. R. (1971). Development of first-level behavioral job criteria. *Journal of Applied Psychology, 55*(1), 3–8. <http://doi.org/10.1037/h0030631>

- Goodale, J. G., & Burke, R. J. (1975). Behaviorally based rating scales need not be job specific. *Journal of Applied Psychology, 60*(3), 389–391. <http://doi.org/10.1037/h0076629>
- GUTIÉRREZ-Castillo, Juan-Jesús CABERO-Almenara, J., & ESTRADA-Vidal, L. I. (2017). Diseño y validación de un instrumento de evaluación de competencia digital docente. *Revista Espacios, 38*(10), 16. Recuperado de <http://www.revistaespacios.com/a17v38n10/17381018.html>
- Harari, O., & Zedeck, S. (1973). Development of Behaviorally Anchored Scales for the Evaluation of Faculty Teaching. *Journal of Applied Psychology, 58*(2), 261–265. <http://doi.org/10.1037/h0035633>
- Hauenstein, N. M. A., Brown, R. D., & Sinclair, A. L. (2010). BARS and those mysterious, missing middle anchors. *Journal of Business and Psychology, 25*(4), 663–672. <http://doi.org/10.1007/s10869-010-9180-7>
- Hom, P. W., DeNisi, A. S., Kinicki, A. J., & Bannister, B. D. (1982). Effectiveness of performance feedback from behaviorally anchored rating scales. *Journal of Applied Psychology, 67*(5), 568–576. <http://doi.org/10.1037/0021-9010.67.5.568>
- Jacobs, R., Kafry, D., & Zedeck, S. (1980). Expectations of behaviorally anchored rating scales. *Personnel Psychology, 33*(3), 595–640. <http://doi.org/10.1111/j.1744-6570.1980.tb00486.x>
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36. <http://doi.org/10.1007/BF02291575>
- Kavanagh, M. J., & Duffy, J. F. (1978). An extension and field test of the retranslation method for developing rating scales. *Personnel Psychology, 31*(3), 461–470. <http://doi.org/10.1111/j.1744-6570.1978.tb00455.x>
- Kell, H. J., Martin-Raugh, M. P., Carney, L. M., Inglese, P. A., Chen, L., & Feng, G. (2017). *Exploring Methods for Developing Behaviorally Anchored Rating Scales for Evaluating Structured Interview Performance*. Princeton. Recuperado de <https://files.eric.ed.gov/fulltext/EJ1168380.pdf>
- Kerlinger, F. N., Lee, H. B., Pineda, L. E., & Mora Magaña, I. (2002). *Investigación del comportamiento. Métodos de investigación en ciencias sociales* (4a). México D.F.: McGraw-Hill. Recuperado de <https://www.casadellibro.com/libro-investigacion-del-comportamiento-metodos-de-investigacion-en-ciencias-sociales-4-ed/9789701030707/814762>
- Klieger, D. M., Kell, H. J., Rikoon, S., Burkander, K. N., Bochenek, J. L., & Shore, J. R. (2018). Development of the Behaviorally Anchored Rating Scales for the Skills Demonstration and Progression Guide. *ETS Research Report Series*, (July). <http://doi.org/10.1002/ets2.12210>
- Landy, F. J., & Guion, R. M. (1970). Development of scales for the measurement of work motivation. *Organizational Behavior and Human Performance, 5*(1), 93–103. [http://doi.org/10.1016/0030-5073\(70\)90007-3](http://doi.org/10.1016/0030-5073(70)90007-3)
- Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The Structured Employment Interview: Narrative and Quantitative Review of the Research Literature. *Personnel Psychology, 67*(1), 241–293. <http://doi.org/10.1111/peps.12052>
- Lizasoain-Hernández, L., Etxeberria-Murgiondo, J., & Lukas-Mujika, J. F. (2017). Propuesta de un nuevo cuestionario de evaluación de los profesores de la Universidad del País Vasco. Estudio psicométrico, dimensional y diferencial. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa, 23*(1). <http://doi.org/10.7203/relieve.23.2.10436>
- Luna Serrano, E. (2015). Validación de constructo de un cuestionario de evaluación de la competencia docente. *Revista Electronica de Investigación Educativa, 17*.
- MacDonald, H. A., & Sulsky, L. M. (2009). Rating formats and rater training redux: A context-specific approach for enhancing the effectiveness of performance management. *Canadian Journal of Behavioural Science, 41*(4), 227–240. <http://doi.org/10.1037/a0015165>
- Martin-Raugh, M., Tannenbaum, R. J., Tocci, C. M., & Reese, C. (2016). Behaviourally Anchored Rating Scales: An application for evaluating teaching practice. *Teaching and Teacher Education, 59*, 414–419. <http://doi.org/10.1016/j.tate.2016.07.026>
- MATOSAS, L. (2018). Aspectos de comportamiento básico del profesor universitario en los procesos de valoración docente para modalidades blended learning. *Revista Espacios, 39*(10), 10. Recuperado de <http://www.revistaespacios.com/a18v39n17/18391713.html>
- Mayorga Fernández, M. J., & Ruiz Baeza, V. M. (2002). Muestras utilizadas en investigación educativa en España. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa, 8*(2). Recuperado de [https://www.uv.es/RELIEVE/v8n2/RELIEVEv8n2\\_2.htm](https://www.uv.es/RELIEVE/v8n2/RELIEVEv8n2_2.htm)

- Murphy, K. R., & Pardaffy, V. A. (1989). Bias in Behaviorally Anchored Rating Scales: Global or scale-specific? *Journal of Applied Psychology*, 74(2), 343–346. <http://doi.org/10.1037/0021-9010.74.2.343>
- Ohland, M. W., Loughry, M. L., Woehr, D. J., Bullard, L. G., Felder, R. M., Finelli, C. J., ... Schmucker, D. G. (2012). The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self- and peer evaluation. *Academy of Management Learning and Education*, 11(4), 609–630. <http://doi.org/10.5465/amle.2010.0177>
- Pounder, J. S. (2000). A Behaviourally Anchored Rating Scales Approach to Institutional Self-assessment in Higher Education. *Assessment & Evaluation in Higher Education*, 25(2), 171–182. <http://doi.org/10.1080/713611422>
- Schwab, D. P., Heneman, I. I. I., & DeCotiis, T. A. (1975). Behaviorally anchored rating scales: A review of the literature. *Personnel Psychology*, 28(4), 549–562. <http://doi.org/Article>
- Sharon, A. T., & Bartlett, C. J. (1969). Effect of Instructional Conditions in Producing Leniency on Two Types of Rating Scales. *Personnel Psychology*, 22(3), 251–263. <http://doi.org/10.1111/j.1744-6570.1969.tb00330.x>
- Smith, P. C., & Kendall, L. M. (1963). Retranslation of Expectations : an approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 47(2), 149–155. <http://doi.org/10.1037/h0047060>
- Stoskopf, C. H., Glik, D. C., Baker, S. L., Ciesla, J. R., & Cover, C. M. (1992). The reliability and construct validity of a Behaviorally Anchored Rating Scale used to measure nursing assistant performance. *Evaluational Review*, 16(3), 333–345.
- Williams, W. E., & Seiler, D. A. (1973). Relationship between measures of effort and job performance. *Journal of Applied Psychology*, 57(1), 49–54. <http://doi.org/10.1037/h0034201>
- Woods, R. H., Sciarini, M., & Breiter, D. (1998). Performance appraisals in hotels: Widespread and valuable. *The Cornell Hotel and Restaurant Administration Quarterly*, 39(2), 25–29. [http://doi.org/10.1016/S0010-8804\(98\)80008-6](http://doi.org/10.1016/S0010-8804(98)80008-6)
- Zedeck, S., & Baker, H. T. (1972). Nursing performance as measured by behavioral expectation scales: A multitrait-multirater analysis. *Organizational Behavior and Human Performance*, 7(3), 457–466. [http://doi.org/10.1016/0030-5073\(72\)90029-3](http://doi.org/10.1016/0030-5073(72)90029-3)
- Zedeck, S., Imparato, N., Krausz, M., & Oleno, T. (1974). Development of behaviorally anchored rating scales as a function of organizational level. *Journal of Applied Psychology*, 59(2), 249–252. <http://doi.org/10.1037/h0036521>

- 
1. Departamento de Economía Financiera, Contabilidad e Idioma Moderno. UNIVERSIDAD REY JUAN CARLOS (Madrid, España). Miembro del grupo de investigación Ciberimaginario de la Universidad Rey Juan Carlos y colaborador del grupo de investigación DIMTE de la Universidad Autónoma de Madrid. E-mail: [luis.matosas@urjc.es](mailto:luis.matosas@urjc.es)
  2. Departamento de Economía Financiera, Contabilidad e Idioma Moderno. UNIVERSIDAD REY JUAN CARLOS (Madrid, España). Director del Centro Universitario de Estudios Sociales y Aplicados (CUESA). E-mail: [santiago.leguey@urjc.es](mailto:santiago.leguey@urjc.es)
  3. Departamento de Economía Financiera, Contabilidad e Idioma Moderno. UNIVERSIDAD REY JUAN CARLOS (Madrid, España). E-mail: [sonsoles.leguey@urjc.es](mailto:sonsoles.leguey@urjc.es)
-