

# CONFIABILIDAD ÍTER-OBSERVADOR DEL MÉTODO DE EVALUACIÓN DE RIESGO INDIVIDUAL

Yordán Rodríguez Ruiz\*  
Jobany J. Heredia Rico\*\*

Recibido en octubre 22 de 2012, aceptado en abril 2 de 2013

## Resumen

**Objetivo:** El objetivo del trabajo es evaluar el nivel de confiabilidad íter-observador de la herramienta ergonómica: Evaluación de Riesgo Individual. La confiabilidad íter-observador de un método de evaluación, valora la capacidad del mismo para proporcionar evaluaciones similares independientemente del observador que lo emplee. Con el método de Evaluación de Riesgo Individual se evalúan diferentes variables las cuales se asocian a la postura del tronco, del brazo, de la muñeca y del cuello; a la interacción de estas con su frecuencia de movimiento, además otras vinculadas con el ritmo de trabajo y el esfuerzo realizado. **Materiales y Métodos:** El experimento se realizó entrenando 42 observadores, la mitad recibió tres horas de entrenamiento y el resto seis, los cuales evaluaron siete tareas representativas de la industria cubana. El entrenamiento y la evaluación de los videos se realizaron en el periodo comprendido entre enero y marzo de 2011. Desde el punto de vista estadístico, la evaluación de confiabilidad se basó fundamentalmente en el cálculo del coeficiente de correlación intra-clase. **Resultados:** Se evidenció que al emplear el método de Evaluación de Riesgo Individual, la confiabilidad íter-observador es aceptable, existiendo un mayor acuerdo en la valoración dada a las variables dicotómicas que a las ordinales. **Conclusión:** Se comprobó la necesidad de mejorar la capacidad de los evaluadores para valorar zonas del cuerpo pequeñas y para determinar la frecuencia con que ocurren determinadas tareas. Son de esperarse mayores niveles de acuerdo entre los evaluadores que utilicen esta herramienta en la práctica.

## Palabras clave

Enfermedades músculo-esqueléticas, evaluación, ingeniería humana, métodos y procedimientos estadísticos, reproducibilidad de resultados (*fuentes: DeCS, BIREME*).

---

\* Ingeniero Industrial. Ph.D en Ciencias Técnicas. Instituto Superior Politécnico "José A. Echeverría", Facultad de Ingeniería Industrial. La Habana, Cuba.

Autor de correspondencia: Correo electrónico: yordanr@ind.cujae.edu.cu

\*\* Ingeniero Industrial. Máster en Tecnologías en Apoyo a la Decisión. Instituto Superior Politécnico "José A. Echeverría", Facultad de Ingeniería Industrial. La Habana, Cuba.

## INDIVIDUAL RISK ASSESSMENT METHOD INTER-RATER RELIABILITY

### Abstract

**Objective:** The objective of this work is the evaluation of the inter-rater reliability level of the ergonomic tool: Individual Risk Assessment. The inter-rater reliability of any assessment method values its capacity for providing similar assessment independently from the rater who uses it. Using the Individual Risk Assessment method different variables associated with: the back, arm, wrist and neck postures, and their interaction with the frequency of movement; their interaction with movement frequency; work rhythm and effort carried out, are evaluated. **Materials and Methods:** Forty-two raters were trained to conduct the experiment, one half took a three-hours training and the rest a six-hours training. All of them assessed seven tasks representative of the Cuban industry. The training and the assessment of the videos were carried out between January and March 2011. From the statistical viewpoint, the reliability assessment was mainly based on the intra-class correlation coefficient calculation. **Results:** It is shown that while using the Individual Risk Assessment method the inter-rater reliability is acceptable, existing a higher agreement on the valuation given to dichotomy variables than the ones given to the ordinal variables. **Conclusion:** The need to improve the rater's skills to assessment small body areas and to determine the frequency in which some tasks are carried out was confirmed. Higher levels of agreement between raters using this tool on practical environments are expected.

### Key words

Musculoskeletal diseases, evaluation, human engineering, statistical methods and procedures, reproducibility of results (*source: DeCS, BIREME*).

## CONFIABILIDADE INTER OBSERVADOR DO METODO DE AVALIAÇÃO DE RISCO INDIVIDUAL

### Resumo

**Objetivo:** O objetivo do trabalho é avaliar o nível de confiabilidade inter - observador da ferramenta ergonômica: Avaliação de Risco individual. A confiabilidade inter observador dum método de avaliação, valora a capacidade do mesmo para proporcionar avaliações similares independentemente do observador que o empregue. Com o método de Avaliação de Risco Individual se avaliam diferentes variáveis as quais se associam à postura do tronco, do braço do pulso e do colo; à interação destas com sua frequência de movimento, além outras vinculadas com ritmo de trabalho e o esforço realizado. **Materiais e Métodos:** O experimento se realizou treinando 42 observadores, a metade recebeu três horas de treinamento e o resto seis, os quais avaliaram sete tarefas representativas da indústria cubana. O treinamento e a avaliação dos vídeos se realizaram no período compreendido entre janeiro e março de 2011. Desde o ponto de vista estatístico, a avaliação de conformidade, se baseou fundamentalmente no cálculo do coeficiente de correlação intraclasse. **Resultados:** Se evidenciou que ao empregar o método de Avaliação de Risco individual, a confiabilidade inter observador é aceitável, existindo um maior acordo na valoração dada as variáveis dicotômicas que às ordinais. **Conclusões:** Se comprovou a necessidade de melhorara capacidade dos avaliadores para valorar zonas do corpo pequenas e para determinar a frequência com que ocorrem determinadas tarefas. São de esperar se maiores níveis de acordo entre os avaliadores que utilizem esta ferramenta na prática.

### Palavras chave

Doenças músculo - esqueléticas, avaliação, engenharia humana, métodos e procedimentos estatísticos, reprodutibilidade de resultados (*fonte: DeCS, BIREME*).

## INTRODUCCIÓN

Desde hace varios años las estadísticas internacionales muestran un crecimiento en la incidencia y prevalencia de los Desórdenes Músculo-Esqueléticos (DMEs) de origen laboral (1). Por lo que se hace impostergable incrementar las acciones dirigidas a la prevención de estas enfermedades ocupacionales. En este sentido, es necesaria la creación de métodos que permitan a personal no experto contar con herramientas para evaluar trabajadores expuestos a factores de riesgo relacionados con los DMEs, identificarlos y cuantificar su magnitud, sirviendo de guía para emprender acciones y además que permitan evaluar el impacto de las intervenciones realizadas a nivel de puesto de trabajo (2). Esto contribuye a la evaluación masiva de puestos de trabajo, especialmente en los países del Tercer Mundo, donde el personal dedicado a esta actividad no posee niveles elevados de conocimiento y entrenamiento para usar los métodos disponibles, pues en su mayoría fueron concebidos para personal especializado (3, 4).

Actualmente existen varios métodos o herramientas disponibles para evaluar la exposición de factores de riesgo relacionados con los DMEs, o identificar trabajos potencialmente riesgosos o factores de riesgo en el trabajo. Estos se pueden agrupar en tres categorías fundamentales: medición directa, auto-reporte y observacionales.

Los de medición directa, tales como electrogoniómetros, inclinómetros, potenciómetros y monitor lumbar, ofrecen gran precisión en las mediciones si son utilizados adecuadamente (5, 6). Son empleados para el estudio de tareas simuladas y permiten registrar las coordenadas de todo el cuerpo en tiempo real. Como principales desventajas se pueden mencionar que los sensores colocados sobre el sujeto pueden causar molestias y provocar cambios en su comportamiento, se requiere de tiempo para el análisis e interpretación de los datos, calibración de los instrumentos, y

de grupos de trabajo con elevada calificación y experiencia. Debido a las limitaciones tecnológicas, por ejemplo el número de canales de entrada disponibles de un instrumento y el espacio para almacenar información, estos métodos generalmente están limitados a un pequeño número de partes del cuerpo y participantes (5). Los costos asociados con los métodos de medición directa son usualmente elevados, lo que los hace imprácticos en grandes estudios epidemiológicos sobre DMEs o para la mayoría de los practicantes de la ergonomía (6).

Los de auto-reporte son fáciles de usar, aplicables a un amplio espectro de situaciones de trabajo y permiten estudiar muestras grandes a costos relativamente bajos (6). Como principales desventajas se pueden mencionar la necesidad de grandes tamaños de muestras para lograr representatividad, y que la información obtenida a partir de la percepción del trabajador es generalmente imprecisa y no confiable (6). Los practicantes de la ergonomía con frecuencia usan este tipo de método en las evaluaciones de puestos de trabajo.

Los observacionales son de uso práctico en una amplia gama de puestos de trabajo, permiten estudiar muestras grandes a costos relativamente bajos y en su empleo no es necesario irrumpir en el trabajo (6). Como principales desventajas se pueden mencionar que el sistema de puntuación es principalmente hipotético, se requiere de conocimientos y experiencia por parte del observador y el sesgo de este cuando realiza la evaluación (6). Los métodos observacionales han sido considerados una herramienta práctica y de confiabilidad razonable en las investigaciones epidemiológicas acerca de los DMEs; por lo que han ganado popularidad entre los practicantes de la seguridad y salud en el trabajo u otras profesiones afines, ya que en alguna medida responden a las necesidades de estos, que por lo general tienen tiempo y recursos limitados, y están necesitados de guías y criterios para establecer prioridades en las intervenciones (3, 6).

Con el objetivo de proveer a los practicantes de la seguridad y salud en el trabajo de un método de fácil aprendizaje y aplicación, se desarrolla el método observacional Evaluación de Riesgo Individual (ERIN) (7, 8). Este permite a personal no experto y con relativo poco entrenamiento, evaluar puestos de trabajo, identificar factores de riesgos y conocer, producto del análisis, qué acciones deben tomarse para reducir el riesgo. Su aplicación solo requiere de hoja y lápiz, aunque su efectividad estará influenciada por el entrenamiento y los conocimientos previos del observador. Fue concebido como una herramienta de riesgo preventivo más que una respuesta al reporte de enfermedades músculo-esqueléticas.

Cuando se desarrolla un método ergonómico de evaluación, es importante realizar estudios para evaluar su confiabilidad y validez. Estos deben realizarse en el contexto y con los usuarios para los cuales fue diseñado. Un método es considerado válido cuando mide lo que intenta medir. El concepto de validación es simple, sin embargo la evaluación de la validación de una medida, es escurridiza. Un instrumento puede ser confiable pero no válido, pero para que un instrumento sea válido, este debe ser confiable. Por lo tanto una condición necesaria, aunque no suficiente, para lograr la validez, debe ser la confiabilidad.

En los métodos observacionales por lo general se estudia la confiabilidad inter-observador. Esta se refiere a la capacidad de la herramienta para producir los mismos resultados independientemente de quién la utilice (5, 9). Por lo tanto, una alta confiabilidad inter-observador implicará que los evaluadores tengan un acuerdo elevado en las evaluaciones que realizan. Estadísticamente esto significa que la variabilidad en las evaluaciones realizadas atribuida a las diferencias entre los evaluadores,

sea pequeña (9, 10). El estudio de esta importante propiedad en los métodos observacionales existentes, es reportado en una revisión extensa y reciente realizada por Takala (2010). De los 30 métodos incluidos en esta revisión, solamente 17 métodos reportaron haber realizado estudios de confiabilidad inter-observador (3).

El objetivo del presente trabajo es evaluar la confiabilidad inter-observador del método ERIN al ser usado por personal no experto.

## MÉTODOS

### Evaluación del Riesgo Individual (ERIN)

ERIN es desarrollado para que personal no experto con un mínimo de entrenamiento, realice la evaluación masiva de puestos de trabajo y mida el impacto de las intervenciones ergonómicas, comparando el riesgo global antes y después, esperando una disminución. Con ERIN se evalúa la postura de las cuatro regiones corporales (tronco, brazo, muñeca y cuello) y la interacción de estas con su frecuencia de movimiento. Se evalúa el ritmo de trabajo, que está dado por la interacción entre la velocidad de trabajo y la duración efectiva de la tarea; el esfuerzo, resultado de la interacción del esfuerzo percibido por el evaluador y su frecuencia, y la autovaloración en la cual se le pregunta al sujeto su percepción del estrés ocasionado por el trabajo. Como resultado final ERIN ofrece el nivel de riesgo de padecer un DMEs, a partir del nivel de riesgo global calculado por la suma del riesgo de las siete variables incluidas, recomendando diferentes niveles de acción ergonómica. Estos niveles de riesgo son mostrados en la Tabla 1.

**Tabla 1. Niveles de riesgo y acción ergonómica recomendada según el riesgo global en ERIN**

Riesgo global	Nivel de riesgo	Acción ergonómica
7-14	Bajo	No son necesarios cambios
15-23	Medio	Se requiere investigar a fondo, es posible realizar cambios
24-35	Alto	Se requiere realizar cambios en breve periodo de tiempo
≥ 36	Muy alto	Se requiere de cambios inmediatos

Fuente: *Elaboración propia.*

Para la evaluación de ERIN es necesario observar las tareas durante varios ciclos de trabajo, para identificar la postura crítica para cada parte del cuerpo evaluada por separado. Esto significa que no necesariamente exista una postura crítica para todas las regiones a la vez. Por ejemplo, el momento en el que se coloca una carga en un estante situado en un nivel superior a la altura de los hombros, pudiera ser el momento donde ocurre la postura crítica para los brazos y no para el tronco.

El observador tiene que seleccionar qué parte del cuerpo evaluar, la derecha o la izquierda. Para ello debe valorar qué parte se encuentra sometida a mayor carga. En caso de dudas, se recomienda evaluar las dos partes y asumir el riesgo mayor. En las Figuras 1 y 2 se muestra la hoja de campo ERIN utilizada por los observadores para evaluar los puestos de trabajo<sup>1</sup>.

### Descripción general del estudio

Con el objetivo de evaluar la confiabilidad inter-observador de la herramienta ergonómica ERIN se diseñó un grupo de experimentos. Estos consisten en evaluar siete tareas representativas de la industria cubana empleando la hoja de campo ERIN. Estas tareas habían sido filmadas anteriormente con otros propósitos y fueron proyectadas con un proyector digital. Para ello, los evaluadores recibieron un entrenamiento previo que les permitiera realizar las evaluaciones. Antes de comenzar la evaluación de los puestos de trabajo grabados, se aclaró a los evaluadores aspectos relacionados al uso del método ERIN y de las actividades planificadas en el estudio.

<sup>1</sup> Aunque la hoja de campo se muestra dividida en dos partes con el fin de conseguir una mejor calidad en la representación de la misma, en la práctica ambas partes conforman una única plantilla de evaluación.

Figura 1. Hoja de campo de ERIN

### ERIN: Evaluación del Riesgo Individual

Considere los pasos 1, 2 y 3 para las variables Tronco, Brazo, Muñeca y Cuello ; para las variables Ritmo, Esfuerzo y Autovaloración el paso 4.

**Pasos:**

1. Observe al trabajador y seleccione la postura crítica para la región del cuerpo evaluada. (Auxiliarse con las figuras y el texto).
2. Adicione el ajuste en caso que corresponda para obtener la Carga postural.
3. Determine el riesgo por variable dado por la interacción entre la Carga postural y el movimiento de la región del cuerpo; anótelo en la casilla correspondiente.
4. Determine el valor de riesgo para las variables Ritmo, Esfuerzo y Autovaloración según se indica en cada tabla; anótelos en la casilla correspondiente.
5. Sume los valores de riesgo para obtener el **Riesgo Total**.
6. Determine el **Nivel de Riesgo** correspondiente.

**Tronco**

Nivel de riesgo		1	2	3
Carga postural	Estático más de un minuto	1	1	1
	Poco frecuente < 5 veces/min	1	2	2
	Frecuente 6-10 veces/min	2	3	3
	Muy frecuente >10 veces/min	3	4	4

Ajuste: +1 si el Tronco está girado y/o doblado

**Brazo**

Nivel de riesgo		1	2	3
Carga postural	Estático más de un minuto	1	1	1
	Poco frecuente	1	2	2
	Frecuente	2	3	3
	Muy frecuente	3	4	4

Ajuste: +1 si existe abducción -1 si el peso del Brazo está apoyado

**Muñeca**

Nivel de riesgo		1	2	Ajuste
Carga postural	Poco frecuente < 10 veces/min	1	1	1
	Frecuente 11-20 veces/min	2	2	2
	Muy frecuente >20 veces/min	3	3	3

Ajuste: +1 si la Muñeca está desviada o girada

**Cuello**

Nivel de riesgo		1	2
Carga postural	Estático más de un minuto	1	1
	Algunas Veces	2	2
	Constantemente	3	3

Ajuste: +1 si el Cuello está girado y/o doblado

Niveles de Riesgo		
Riesgo Total	Nivel de riesgo	Acción recomendada
●	7-14	Bajo No son necesarios cambios
●	15-23	Medio Se requiere investigar a fondo, es posible realizar cambios
●	24-35	Alto Se requiere realizar cambios en un breve periodo de tiempo
●	+36	Muy Alto Se requiere de cambios inmediatos

**Ritmo**

Duración efectiva de la tarea en (horas)	Velocidad de trabajo				
	Muy lento (Ritmo muy relajado)	Lento (Tomándose su tiempo)	Normal (Velocidad normal de movimiento)	Rápido (Posible de soportar)	Muy Rápido (Difícil o imposible de soportar)
< 2 h	1	1	3	4	5
2-4 h	1	2	3	5	6
4-8 h	2	3	4	6	7
> 8 h	2	4	5	7	7

**Esfuerzo**

Clasificación	Esfuerzo percibido	Frecuencia		
		< 5 por minuto	5-10 por minuto	>10 por minuto
Liviano	Relajado (Esfuerzo poco notorio)	1	2	6
Algo Pesado	Esfuerzo claro-Perceptible	1	2	6
Pesado	Esfuerzo evidente-expresión facial sin cambios	3	7	8
Muy Pesado	Esfuerzo sustancial-cambios en la expresión facial	6	8	9
Casi Máximo	Uso de hombros y tronco para hacer esfuerzos	7	8	9

**Autovaloración**

Descripción	Riesgo
Nada estresante	1
Un poco estresante	2
Estresante	3
Muy estresante	4
Excesivamente estresante	5

Empresa: \_\_\_\_\_

Puesto de trabajo: \_\_\_\_\_

Trabajador: \_\_\_\_\_

Fecha: \_\_\_\_\_

**Riesgo Total**

=

## Observadores

Participaron voluntariamente 42 estudiantes de la carrera de Ingeniería Industrial de tercer y cuarto año. Estos son denominados “observadores” y evaluaron siete tareas empleando la hoja de campo ERIN. La edad promedio de los evaluadores fue de 20 años. Ninguno de los observadores había realizado evaluaciones de riesgo de DMEs en puestos de trabajo con anterioridad, por lo que se pueden clasificar como personal no experto y sin experiencia en esta actividad, coincidiendo con el perfil de los usuarios para los que fue concebido ERIN.

## Entrenamiento de los observadores

Días antes de la realización del experimento, fueron organizadas varias sesiones de entrenamiento de 1 hora y 30 minutos de duración. Estas estuvieron dirigidas a que los futuros observadores conocieran el impacto económico de los DMEs ocupacionales y los factores de riesgo que los ocasionan. Además, a brindarles una descripción y argumentación sobre las variables incluidas en ERIN, sobre la necesidad de la creación de esta herramienta, así como una instrucción en el uso de la hoja de campo ERIN. También, se ejecutó la resolución de casos prácticos a partir de la proyección de videos y se discutieron los resultados. Ninguno de los videos seleccionados para el entrenamiento fue utilizado en el estudio. Los observadores debían asistir al menos a dos sesiones para poder participar en el experimento. Para conocer la influencia del tiempo de entrenamiento en la confiabilidad, la mitad de los observadores recibió tres horas de entrenamiento y el resto seis horas.

## Selección de los videos

Siete videos fueron tomados de investigaciones anteriores y para su selección fueron valorados los cinco aspectos siguientes: variedad de las tareas y representatividad de diferentes ramas de la industria, calidad de la película filmica, visibilidad de los ángulos corporales desde diferentes vistas,

trabajos de pie y sentado y representación de ciclos de trabajos completos. De las siete tareas, tres fueron en posición de sentado (colocación de bandejas de comida en carro de distribución, destripe de hojas de tabaco y estirado de hojas de tabaco) y cuatro en posición de pie (forrado de cajas de tabaco, fregado de cajas, limpieza de clóset y pesaje de alimentos). La proyección de los videos fue aleatoria y se repitieron hasta que todos los observadores terminarán de realizar la evaluación. Para establecer el orden aleatorio de los videos se utilizó la función de números aleatorios del programa Microsoft Excel.

## Ejecución del experimento

A medida que cada evaluador terminaba la evaluación de cada tarea se recogía la hoja de campo ERIN, y eran revisadas preliminarmente por personal capacitado para comprobar que habían sido completados todos los aspectos. En caso de detectar alguna irregularidad, la hoja de campo era devuelta y corregida por el observador en el momento. En cada hoja se registraba el puesto de trabajo evaluado, el nombre del observador y el grupo en que participó. La muestra quedó conformada por 42 observadores.

## Métodos estadísticos

Existen diferentes técnicas estadísticas que se pueden utilizar para evaluar la confiabilidad inter-observador; los métodos más comunes incluyen el estadístico kappa, el porcentaje de acuerdo entre evaluadores y el coeficiente de correlación intra-clase (ICC) (5, 9, 11). No obstante, es importante comprender las diferencias entre las diferentes técnicas así como la validez y utilidad de la aplicación de cada una en un análisis particular.

La forma más fácil de evaluar el acuerdo entre evaluadores es calcular el porcentaje de acuerdo entre estos. Este se calcula dividiendo el número de veces que los evaluadores coinciden en la evaluación otorgada a un mismo objeto, entre

el número total de objetos (puestos de trabajo) evaluados. El porcentaje de acuerdo ha sido considerado una medida inadecuada debido a que no tiene en cuenta el acuerdo atribuible al azar, por lo que tiende a dar resultados de confiabilidad optimistas (5, 12).

El estadístico kappa es una medida que sí considera el acuerdo atribuible al azar, ya que compara la medida de acuerdo observada con el grado de concordancia debida solamente al azar (10, 12). Uno de los problemas que presenta este estadístico es que solo puede ser utilizado en datos con escala nominal (10, 12). Algunos autores plantean que este estadístico puede no ser apropiado como un índice de confiabilidad inter-observador, y por tanto recomiendan su uso si el objetivo del estudio se limita a verificar que el acuerdo entre los observadores es mayor que el acuerdo atribuible al azar (9, 10). Para datos con escala ordinal puede ser utilizado el estadístico kappa ponderado, el cual asigna un peso a las diferencias entre las categorías, otorgándole mayor peso a las diferencias entre categorías distantes (10, 12). La elección de los pesos asignados puede afectar su interpretación, ya que diferentes investigadores pueden establecer diferentes pesos (10, 12). Este problema puede ser evitado si son empleadas ponderaciones estándares (12). En caso de utilizarse la ponderación cuadrática, el kappa ponderado es equivalente al coeficiente de correlación intra-clase (ICC) (12). Existen varios tipos de coeficientes de correlación intra-clase (ICCs) y cada tipo puede ser potencialmente aplicado a diferentes situaciones (10).

Los ICCs son ampliamente empleados para evaluar la confiabilidad inter-observador (9-11). Su cálculo se basa en separar en dos partes la variabilidad total de las evaluaciones realizadas. Por una parte, la variación debida a la diferencia entre los elementos a evaluar (en este caso los puestos de trabajo evaluados); y por otra, la variabilidad atribuible a la diferencia entre las percepciones de los observadores (13, 14). El ICC se define como la proporción de la varianza total

que es debida a las diferencias entre los puestos de trabajo evaluados (13, 14). Esto implica que mientras mayor valor tenga este coeficiente, menor variabilidad es atribuible a la diferencia entre las percepciones de los observadores, es decir, mayor acuerdo existe entre estos. Para la evaluación de la confiabilidad inter-observador los ICCs han demostrado ser los más precisos y flexibles (10). Incluso, se plantea que es válido el uso de algunos ICCs para valorar la confiabilidad en variables numéricas (11, 13).

El ICC (2.1) es ampliamente utilizado en los casos en que se cumplan las siguientes condiciones (10, 13):

1. El objetivo del estudio es evaluar el acuerdo entre los evaluadores y no la consistencia entre sus evaluaciones.
2. Se toma una muestra aleatoria de los objetos a evaluar (puestos de trabajo) y de los evaluadores, de forma que los resultados sean independientes de la muestra seleccionada y puedan generalizarse a otros elementos de ambas poblaciones (puestos de trabajo y evaluadores).
3. Cada evaluador, evalúa una sola vez las variables asociados a un determinado objeto, y por lo tanto se trabaja con evaluaciones simples y no con promedios.

El ICC (2.1) es empleado generalmente para evaluaciones de confiabilidad con variables de escala ordinal, aunque también se reporta su uso en variables numéricas (10). Otro ICC es el ICC-Kuder-Richardson-20 (KR-20), cuyo uso se recomienda sobre los otros ICCs en el caso de variables dicotómicas (10). Otro estadístico que ha sido utilizado para medir acuerdo entre evaluadores para variables ordinales, es el coeficiente de concordancia de Kendall (W) (15, 16). Este coeficiente es un índice de la divergencia entre el acuerdo efectivo mostrado en los datos y el máximo acuerdo posible. Su valor varía desde 0 (no existencia de acuerdo) hasta 1 (acuerdo perfecto) (15).



### Elección de los métodos estadísticos para evaluar la confiabilidad inter-observador

De las variables evaluadas con ERIN, las variables ajuste del tronco, ajuste del brazo, postura muñeca, ajuste muñeca, postura cuello y ajuste cuello son dicotómicas pues solo toman dos posibles valores. Para estas variables se utiliza como métrica de confiabilidad el estadístico ICC KR-20. Para las restantes variables, cuya escala es ordinal pues toman al menos 3 valores a los cuales se les puede asociar un orden, se utiliza el estadístico ICC (2.1). La utilización de este método estadístico es válida, ya que fue seleccionada aleatoriamente una muestra de evaluadores y puestos de trabajo de poblaciones consideradas infinitas. Además, los puestos de trabajo son evaluados solo una vez por cada evaluador. Para evaluar el valor obtenido al calcular un ICC, se recomienda utilizar las siguientes categorías (13): si el ICC < 0,40, la confiabilidad es pobre; si el ICC está entre 0,40 y 0,75, la confiabilidad es de moderada a buena

(generalmente valores mayores o iguales que 0,7 indican que la confiabilidad es buena); y si el ICC > 0,75, la confiabilidad es excelente. Para enriquecer el análisis en el caso de las variables con escala ordinal, también se utiliza como medida de acuerdo el coeficiente de concordancia de Kendall. Los niveles de acuerdo medidos con W son interpretados de la siguiente forma: acuerdo pobre,  $0,0 < W < 0,6$ ; acuerdo moderado,  $0,6 \leq W < 0,7$ ; buen acuerdo,  $0,7 \leq W < 0,8$  y, acuerdo fuerte,  $0,8 \leq W \leq 1,0$  (15).

### RESULTADOS

Los resultados de la confiabilidad inter-observador son mostrados en la Tabla 2. La clasificación mostrada en la última columna de esta Tabla, se realiza según el valor obtenido para el ICC pues, como se ha comentado, este índice es reconocido como el más preciso.

**Tabla 2. Resultados de confiabilidad**

VARIABLES ERIN	ICC(2.1)	ICC RK-20	Kendall	Clasificación
Postura Tronco	0,768	*	0,807	Excelente
Ajuste Tronco	**	0,964	*	Excelente
Movimiento Tronco	0,574	*	0,667	Moderada
Postura Brazo	0,594	*	0,638	Moderada
Ajuste Brazo	**	0,754	*	Excelente
Movimiento Brazo	0,295	*	0,339	Pobre
Postura Muñeca	**	0,836	*	Excelente
Ajuste Muñeca	**	0,081	*	Pobre
Movimiento Muñeca	0,088	*	0,134	Pobre
Postura Cuello	**	0,951	*	Excelente
Ajuste Cuello	**	0,896	*	Excelente
Movimiento Cuello	0,298	*	0,356	Pobre
Velocidad de Trabajo	0,271	*	0,329	Pobre
Esfuerzo	0,728	*	0,725	Buena
Frecuencia de Esfuerzo	0,278	*	0,359	Pobre
Riesgo Total	0,465	*	**	Moderada
Niveles de Riesgo	0,375	*	0,471	Pobre
Niveles de Riesgo Modificado	**	0,944	*	Excelente

\* El estadístico no es aplicable dada la escala de medición de la variable.

\*\* No se calcula pues no es el estadístico recomendado según la escala de la variable en cuestión.

Fuente: Elaboración propia.

Los resultados de la confiabilidad inter-observador entrenamiento entre los observadores, son considerando las diferencias en las horas de mostrados en la Tabla 3.

**Tabla 3. Resultados de confiabilidad considerando el tiempo de entrenamiento**

Variables ERIN	n = 21 observadores 3 horas de entrenamiento			n = 21 observadores 6 horas de entrenamiento		
	ICC(2.1)	RK-20	Kendall	ICC(2.1)	RK-20	Kendall
Postura Tronco	0,75	*	0,804	0,792	*	0,827
Ajuste Tronco	**	0,896	*	**	0,952	*
Movimiento Tronco	0,532	*	0,65	0,604	*	0,692
Postura Brazo	0,625	*	0,679	0,577	*	0,619
Ajuste Brazo	**	0,792	*	**	0,053	*
Movimiento Brazo	0,246	*	0,315	0,328	*	0,386
Postura Muñeca	**	0,619	*	**	0,74	*
Ajuste Muñeca	**	0	*	**	0,267	*
Movimiento Muñeca	0,077	*	0,13	0,08	*	0,164
Postura Cuello	**	0,863	*	**	0,939	*
Ajuste Cuello	**	0,682	*	**	0,859	*
Movimiento Cuello	0,331	*	0,405	0,271	*	0,342
Velocidad de Trabajo	0,212	*	0,32	0,352	*	0,382
Esfuerzo	0,73	*	0,729	0,72	*	0,736
Frecuencia de Esfuerzo	0,351	*	0,432	0,193	*	0,311
Riesgo Total	0,503	*	**	0,461	*	**
Niveles de Riesgo	0,384	*	0,478	0,386	*	0,509
Niveles de Riesgo Modificado	**	0,835	*	**	0,863	*

Fuente: Elaboración propia.

\* El estadístico no es aplicable dada la escala de medición de la variable.

\*\* No se calcula pues no es el estadístico recomendado según la escala de la variable en cuestión.

## DISCUSIÓN

Todas las variables dicotómicas (ajuste tronco, ajuste brazo, postura muñeca, ajuste de cuello, postura de cuello), excepto ajuste de muñeca (0,081) presentan valores de ICC KR-20 mayores de 0,75, de las cuales cuatro (ajuste del tronco, postura muñeca, postura del cuello y ajuste del cuello) presentan un valor mayor que 0,83. Esto indica que el acuerdo de los observadores al evaluar estas cinco variables, es excelente. Solamente no hubo acuerdo entre los observadores

en la detección de la posible desviación o giro en la muñeca de los trabajadores que realizaban las tareas (ICC KR-20 = 0,081). Como ha sido evidenciado en otros estudios, las evaluaciones de zonas del cuerpo más pequeñas, como es el caso de la muñeca en este estudio, tienden a mostrar un menor grado de acuerdo entre los observadores (3, 9, 10, 12, 16, 17). En las 10 variables ordinales (postura y movimiento del tronco, postura y movimiento del brazo, movimiento muñeca, movimiento cuello, velocidad de trabajo, esfuerzo, frecuencia del esfuerzo y niveles de riesgo) los

valores de acuerdo variaron en mayor medida que en las dicotómicas. Solamente en la evaluación de la postura del tronco puede considerarse que el acuerdo es excelente, pues solo en esta variable el ICC (2.1) y el coeficiente de Kendall excedieron los valores que así lo evidencian (0,768 y 0,807, respectivamente). En las evaluaciones de las variables: movimiento tronco, postura brazo y esfuerzo, el nivel de acuerdo es de moderado a bueno ya que los valores del ICC (2.1) varían entre 0,574 y 0,728, y los de Kendall entre 0,638 y 0,755.

En la evaluación de: movimiento del brazo, movimiento de la muñeca, movimiento del cuello, velocidad de trabajo y frecuencia del esfuerzo se obtuvo un grado de acuerdo pobre. Los valores del ICC (2.1) variaron desde 0,088 a 0,298, y los del W desde 0,134 a 0,359. Generalmente la observación de tareas dinámicas es más complicada que las estáticas (16, 17). Esto indica que la clasificación de acciones dinámicas mediante la observación en tiempo real, es un aspecto en el que se debe hacer énfasis para mejorar el acuerdo en personal no experto. La estimación de estas variables hubiera tenido un mejor comportamiento, si los evaluadores hubieran tenido la oportunidad de disminuir la velocidad de movimiento en el video o detenerlo para contabilizar las acciones (11, 12). En este caso, como fue explicado, no ocurrió de esa manera. No obstante, en los estudios de campo los evaluadores pueden filmar y estimar con mayor precisión estos aspectos, lo que supone mayores niveles de acuerdo a los reportados en este estudio.

En la variable nivel de riesgo se obtuvo un bajo acuerdo entre los evaluadores pues el ICC (2.1) = 0,375, y W = 0,471, aunque estos valores están próximos a 0,4 y 0,6, límites que indican un acuerdo aceptable para el ICC (2.1) y W, respectivamente. Otros estudios han analizado el nivel de riesgo considerando solo dos categorías: riesgo y no riesgo (10). Al definir una nueva variable (nivel de riesgo modificado) con esta característica, es decir, estableciendo dos categorías para evaluar el

nivel de riesgo: la primera resultado de agrupar los niveles de riesgo “bajo” y “medio” en la categoría “poco riesgo”, y la segunda resultado de agrupar los niveles de riesgo “alto” y “muy alto” en la categoría “riesgo considerable”, se obtiene que el ICC KR-20 es 0,944, indicando en este caso una excelente confiabilidad y por lo tanto que el método es adecuado con este fin. En la variable riesgo total que se obtiene al sumar el valor de riesgo por variable, se obtuvo un valor de ICC (2.1) de 0,465 indicando una confiabilidad de moderada.

En la observación visual de puestos de trabajos, se considera que el acuerdo entre los observadores varía dependiendo del método empleado, los riesgos observados, el conocimiento y entrenamiento de los observadores, y quizás de las características del puesto (11, 17). El acuerdo en la observación de posturas es un aspecto complicado. Muchos factores pueden influir en el acuerdo. Estos pueden incluir, aunque no limitado a esto: a) parámetros de postura específicos, b) variación de la postura de una población de estudio o en un puesto de trabajo, c) distribución de la postura entre las diferentes categorías angulares, d) estrategia de categorización de posturas, e) entrenamiento y experiencia del evaluador, f) posición del observador relativo a la persona estudiada o calidad de la imagen de video, g) tipos de puestos de trabajo, h) definición de la postura error de estimación del evaluador, i) las herramientas usadas, tales como el diseño de interfaces e instrucciones (5). Algunos parámetros posturales fueron más fáciles de observar que otros. En general, las posturas de las partes del cuerpo más grandes, tales como el tronco y el brazo, fueron más fáciles de observar que las partes del cuerpo pequeñas como la muñeca. Esto se ve reflejado en mayores valores de ICC y Kendall en la Tabla 2.

En la Tabla 4 se resumen los resultados obtenidos para las variables dicotómicas, separados según las horas de entrenamiento de los evaluadores.

**Tabla 4. Resumen de los resultados para variables dicotómicas**

VARIABLES ERIN	RK-20 (3 horas)	Clasificación	RK-20 (6 horas)	Clasificación
Ajuste Tronco	0,896	Excelente	0,952	Excelente
Ajuste Brazo	0,792	Excelente	0,053	Pobre
Postura Muñeca	0,619	Moderada	0,74	Moderada
Ajuste Muñeca	0	Pobre	0,267	Pobre
Postura Cuello	0,863	Excelente	0,939	Excelente
Ajuste Cuello	0,682	Moderada	0,859	Excelente
Nivel de Riesgo Modificado	0,835	Excelente	0,863	Excelente

Fuente: Elaboración propia.

Los índices de confiabilidad obtenidos para las variables dicotómicas, generalmente son mayores para el grupo que recibió mayor tiempo de entrenamiento. No obstante a estas diferencias en los índices, en las variables ajuste tronco, postura del cuello, postura muñeca, ajuste muñeca y nivel de riesgo modificado, la clasificación de confiabilidad fue la misma en ambos grupos de acuerdo a los rangos establecidos para estos índices. Para ambos grupos la clasificación fue de excelente en el ajuste del tronco, la postura del cuello y el nivel de riesgo modificado; de moderada para la postura de la muñeca y pobre en el ajuste de la muñeca. En la variable ajuste

de cuello existió diferencia en la clasificación de confiabilidad debido al tiempo de entrenamiento, siendo de moderada para el grupo de tres horas y excelente para el grupo de 6 horas. La mayor diferencia, en este caso inexplicable, ocurrió en la variable ajuste brazo, ya que los evaluadores menos entrenados presentaron una confiabilidad excelente, mientras que para los de más tiempo de entrenamiento la confiabilidad fue pobre.

En la Tabla 5 se resumen los resultados obtenidos para las variables ordinales, separados según las horas de entrenamiento de los evaluadores.

**Tabla 5. Resumen de los resultados para variables ordinales**

<b>Variabes ERIN</b>	<b>ICC(2.1) (3 h)</b>	<b>Kendall (3 h)</b>	<b>Clasificación</b>	<b>ICC(2.1) (6 h)</b>	<b>Kendall (6 h)</b>	<b>Clasificación</b>
Postura Tronco	0,75	0,804	Excelente	0,792	0,827	Excelente
Movimiento Tronco	0,532	0,65	Moderada	0,604	0,692	Moderada
Postura Brazo	0,625	0,679	Moderada	0,577	0,619	Moderada
Movimiento Brazo	0,246	0,315	Pobre	0,328	0,386	Pobre
Movimiento Muñeca	0,077	0,13	Pobre	0,08	0,164	Pobre
Movimiento Cuello	0,331	0,405	Pobre	0,271	0,342	Pobre
Velocidad de Trabajo	0,212	0,32	Pobre	0,352	0,382	Pobre
Esfuerzo	0,73	0,729	Buena	0,72	0,736	Buena
Frecuencia de Esfuerzo	0,351	0,432	Pobre	0,193	0,311	Pobre
Nivel de Riesgo	0,384	0,478	Pobre	0,386	0,509	Pobre

Fuente: Elaboración propia.

En la evaluación de las variables ordinales, aunque generalmente los valores de los índices del ICC (2.1) y del coeficiente de Kendall fueron mayores en el grupo que recibió mayor tiempo de entrenamiento, la clasificación basada en las categorías de confiabilidad establecidas en la sección 2.7.1, fue similar para todas las variables en los dos grupos.

En la variable riesgo total, de escala numérica, no se encontró diferencia en la clasificación de la confiabilidad (de moderada a buena), pues los valores del ICC (2.1) fueron de 0,503 y 0,461 en el grupo de tres y seis horas, respectivamente (ver Tabla 3).

En general, los resultados de confiabilidad obtenidos pueden valorarse como adecuados ya que con 10 de las 17 variables que se registran con ERIN (59%), la confiabilidad inter-observador

evaluada según los índices calculados fue valorada de al menos moderada (ver Tabla 2). Específicamente, seis variables mostraron índices que indican confiabilidad excelente (35%), una variable mostró buena confiabilidad en sus evaluaciones (6%), y tres revelaron índices indicativos de confiabilidad moderada (18%). De las siete variables cuyos resultados de confiabilidad es pobre (41%), seis (35%) están asociadas, ya sea a la evaluación de zonas del cuerpo de difícil observación como la muñeca, o a la determinación de la frecuencia o la velocidad con que ocurren determinadas tareas, indicadores que también han mostrado bajos valores de confiabilidad en otros estudios (6). La otra variable con resultados bajos de confiabilidad inter-observador fue el nivel de riesgo, la cual, al modificarse para convertirla en dicotómica (existencia o no de riesgo) tal como se reporta en otros estudios (10), aumenta el índice de confiabilidad de tal forma que esta se clasifica

como excelente. Otro elemento que justifica la valoración positiva dada a los resultados obtenidos, es que las horas de entrenamiento fueron pocas si se comparan con el tiempo dedicado a esta tarea en otros estudios (10, 18).

### **Limitaciones del estudio realizado**

Un primer aspecto es que los participantes solo realizaron una sola evaluación por tarea, lo que impide el uso del promedio de las mediciones y aumenta el impacto negativo de los errores de los evaluadores en los índices de confiabilidad a partir del ICC pues, como fue expresado, el uso del promedio de las mediciones proporciona mayores índices de confiabilidad que los obtenidos con las mediciones individuales. Por lo que en un estudio donde se utilice el promedio de las mediciones, pudiera reportar índices de ICCs mayores que en el presente.

También el hecho de que la diferencia en el tiempo de entrenamiento entre los evaluadores fue solamente de 3 horas, lo que no permite arribar a conclusiones definitivas sobre la influencia de este factor en la confiabilidad del método ERIN. Aunque ERIN fue desarrollado para que personal no experto con un mínimo de entrenamiento pueda utilizarlo, debiera valorarse la posibilidad de incrementar las horas de entrenamiento, puesto que en otros estudios dedicaron mayor tiempo de entrenamiento a pesar de que el personal era experto y con experiencia (10, 17).

### **Fortalezas del estudio realizado**

Una de las fortalezas de este estudio es el rigor del análisis estadístico realizado, pues los métodos seleccionados fueron apropiados y los resultados fueron reportados de manera que se posibilita una interpretación adecuada de estos. Otro aspecto positivo es el número de evaluadores utilizados, ya que se plantea que al menos 30 individuos deben ser involucrados para realizar mediciones o pruebas de confiabilidad (10), y en

este caso participaron 42 evaluadores, superando lo recomendado por estos autores. Otra fortaleza es que los observadores que participaron en este estudio, representan adecuadamente el espectro de personas que pueden utilizar el método ERIN en la evaluación ergonómica de puestos de trabajo en la industria cubana.

También el hecho de que en este estudio participaron observadores con diferentes niveles de entrenamiento. Esta situación es semejante a la realidad, donde no todo el personal que empleará la herramienta tiene la misma formación y entrenamiento. En la mayoría de los métodos existentes similares a ERIN, la literatura no menciona la duración del tiempo de entrenamiento que es necesario para alcanzar la habilidad adecuada en el uso de la herramienta (3).

## **CONCLUSIONES**

Con el estudio realizado se comprobó que la confiabilidad inter-observador del método ERIN es adecuada, y por lo tanto el uso de esta herramienta produce en general resultados similares independientemente de quién la utilice. A pesar de que algunos valores de ICC en determinadas variables fueron bajos, son considerados como aceptables pues a diferencia de otros estudios donde se reportan resultados con personal experto, los participantes en este son no expertos.

Se obtuvieron valores diferentes de confiabilidad para cada uno de los dos grupos con diferente entrenamiento, pero estos valores están asociados a un mismo grado de confiabilidad, es decir, según las escalas establecidas para valorar la confiabilidad, pertenecen a una misma categoría. Solo para las variables ajuste cuello y ajuste brazo, se obtuvo una diferencia notable en la confiabilidad debida al tiempo de entrenamiento. Por lo que los resultados obtenidos muestran que la diferencia de tres horas en el tiempo de entrenamiento en el uso del método ERIN, no produce mejoras sustanciales

en la confiabilidad inter-observador. A pesar de esto, en 12 de las 17 variables (70%) el valor de los índices de confiabilidad fue mayor en el grupo de más tiempo de entrenamiento, evidenciando la tendencia al incremento del acuerdo dentro de este grupo.

En las variables con escala ordinal, aquellas que tienen al menos tres categorías de respuesta a las cuales se puede asociar un orden, se obtuvo en general menor acuerdo, por ende la confiabilidad inter-observador del método es mayor para las variables dicotómicas, lo que significa que con estas variables se puede estar más seguro de que la evaluación que haga un individuo cualquiera, no diferirá considerablemente de las posibles evaluaciones que puedan hacer otros individuos.

A pesar de que los resultados de confiabilidad obtenidos son positivos, debe mejorarse la capacidad de los evaluadores al emitir sus valoraciones en cuanto a las zonas del cuerpo más pequeñas, y en la determinación de la frecuencia o la velocidad con que ocurren determinadas tareas.

Para lograr esto, puede valorarse la posibilidad de perfeccionar el entrenamiento que se brinda a los evaluadores, lo cual puede implicar, aumentar el tiempo de entrenamiento y/o mejorar las habilidades relacionadas con la identificación de la frecuencia de movimiento de diferentes partes del cuerpo, así como de elementos de interés en zonas pequeñas de este.

Son de esperarse mayores niveles de acuerdo entre los evaluadores que utilicen esta herramienta en la práctica, por cuanto en este caso estos tendrían la oportunidad de filmar las tareas para después disminuir la velocidad de movimiento en el video o detenerlo para contabilizar las acciones, y así estimar con mayor precisión estos aspectos que son más difíciles de discernir.

Para concluir finalmente sobre la efectividad del método ERIN para identificar factores de riesgos asociados a desórdenes músculo-esqueléticos, debe complementarse el estudio realizado con otros enfocados en evaluar su validez y su confiabilidad intra-observador.

## REFERENCIAS BIBLIOGRÁFICAS

1. USA. National Institute for Occupational Safety and Health. *Musculoskeletal Disorders and Workplace Factors: A Critical Review of Epidemiologic Evidence for Work-Related Musculoskeletal Disorders of the Neck, Upper Extremity, and Low Back*. Cincinnati; 1997.
2. European Agency for Safety and Health at Work. *Work-related musculoskeletal disorders in the EU – Facts and figures*. Luxembourg: Publications Office of the European Union; 2010.
3. Takala EP. Systematic evaluation of observational methods assessing biomechanical exposures at work. *Scandinavian Journal of Work, Environmental and Health*, 2010;36(1):3-24.
4. Li G, Buckle P. Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods. *Ergonomics*. 1999;42(5):674-95.
5. Bao S. Two posture analysis approaches and their application in a modified Rapid Upper Limb Assessment evaluation. *Ergonomics*. 2007;50(12):2118-36.
6. David G. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occupational Medicine*. 2005;55(3):190-99.
7. Rodríguez Y, Guevara-Velasco C. Empleo de los métodos ERIN y RULA en la evaluación ergonómica de puestos de trabajo. *Ingeniería Industrial*. 2011;32(1):19-27.
8. Rodríguez Y, Viña S, Montero R. ERIN: un método práctico de evaluación de riesgo de desórdenes músculo-esqueléticos de origen laboral. Paper presented at: III Congreso Latinoamericano de Ergonomía; 2010 Agosto 2-6; Río de Janeiro, Brasil.
9. Bao S, Howard N, Spielholz P, Silverstein B, Polissar N. Inter-rater Reliability of Posture Observations Human Factors. *The Journal of the Human Factors and Ergonomics Society*. 2009;51:292-309.
10. Stevens EM Jr, Vos GA, Stephens JP, Moore JS. Inter-rater reliability of the strain index. *Journal of Occupational and Environmental Hygiene*. 2004;11(1):745-51.
11. Xu X, Chang CC, Faber GS, Kingma I, Dennerlein JT. The validity and interrater reliability of video-based posture observation during asymmetric lifting tasks. *Human Factors*. 2011;53(4):371-82.
12. Burt S, Punnett L. Evaluation of interrater reliability for posture observations in a field study. *Applied Ergonomics*. 1999;30(2):121-35.
13. Shrout P, Fleiss J. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*. 1979;86(2):420-28.
14. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*. 1996;1(1):30-46.
15. Siegel S, Castellan NJ. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill; 1988.
16. David G, Woods V, Li G, Buckle P. The development of the Quick Exposure Check (QEC) for assessing exposure to risk factors for work-related musculoskeletal disorders. *Applied ergonomics*. 2008;39(1):57-69.
17. Parkab JK, Boyerb J, Tesslerb J, Caseybc J, Schemmb L, Goreb R, et al. Inter-rater reliability of PATH observations for assessment of ergonomic risk factors in hospital work. *Ergonomics*. 2009;52(7):820-29.
18. Toivonen R, Viikari-Juntura E, Ketola R. Interobserver repeatability and validity of an observation method to assess physical loads imposed on the upper extremities. *Ergonomics*. 2001;44(2):119-31.