

“Confiabilidad inter-evaluador de la medición de presión inspiratoria máxima: Influencia de la experiencia del evaluador”

Inter-rater reliability in maximal expiratory pressure measurement:
Influence of evaluator's experience

Mario Gutiérrez Punoy¹, Mario Órdenes Salinas², Karimé González Gajardo²

¹ Estudiante, Escuela de Kinesiología, Facultad de Ciencias de la Salud, Universidad Católica del Maule, Talca, Chile.

² Estudiante Magister en Kinesiología Universidad Católica del Maule, Talca.

Título Abreviado
Confiabilidad inter-evaluador en la medición de PIM

Información del Artículo
Recepción: 1 de Agosto 2016
Aceptación: 10 de Noviembre 2017

RESUMEN

Objetivo: Determinar el nivel de acuerdo en la medición de presión inspiratoria máxima (PIM) entre un evaluador experto y un estudiante de kinesiología. **Metodología:** Se midió PIM con pletismógrafo corporal por dos evaluadores (Evaluador A, estudiante; Evaluador B, experto), en un grupo de sujetos sanos de la Región del Maule. La confiabilidad inter-evaluador fue analizada con el coeficiente de correlación intraclase (CCI) y se utilizó el gráfico de Bland-Altman para diferencias individuales. **Resultados:** El análisis con CCI mostró una excelente correlación entre ambos evaluadores ($ICC1er:0.914$; $ICC2do=0.915$; $ICC3ro=0.925$) para la medición de PIM. Sin embargo, el gráfico de Bland-Altman demostró una tendencia hacia un mejor rendimiento con el evaluador experimentado y diferencias en promedio de 2,69 cmH₂O. **Discusión:** Los resultados sugieren que el estudiante adquirió un óptimo dominio en la medición de PIM, por lo que las diferencias observadas en el rendimiento podrían deberse a habilidades que no se relacionan con el cumplimiento del protocolo, sino con el desarrollo de competencias transversales, como la optimización de la relación evaluador-paciente o la capacidad de distinguir un gesto motor apropiado durante la ejecución del test. **Conclusión:** Un corto periodo de inducción puede entregar un excelente nivel de confiabilidad en la aplicación de la medición de PIM en un estudiante en formación.

Palabras clave: musculatura ventilatoria, confiabilidad, pruebas de función pulmonar.

ABSTRACT

Objective: To determine the level of agreement on measurement of maximal inspiratory pressure (MIP) between an expert evaluator and a student of kinesiology. **Methods:** MIP was measured with a body plethysmograph by two evaluators (rater A, student; rater B, expert) in a group of healthy subjects of the Maule Region. The inter-rater reliability was analyzed with intraclass correlation coefficient (ICC) and the Bland-Altman plot was used for individual differences. **Results:** ICC analysis showed excellent correlation between the two evaluators ($ICC1er: 0.914$; $ICC2do = 0.915$; $ICC3ro = 0.925$). However, the Bland-Altman showed a trend toward better performance with experienced assessor and differences in average 2,69 cmH₂O. **Discussion:** The results suggest that the student has acquired an excellent mastery in measuring P_{Imax}, so the observed differences in performance could be due to skills that are not related to compliance with the protocol, but with the development of transversal competences, as optimizing the evaluator-patient relationship or the ability to distinguish an appropriate gesture engine during the test run. **Conclusion:** A short practice period is useful for obtaining high levels of reliability in assessment of MIP in students at professional practice.

Key words: respiratory muscles, reliability, pulmonary function tests.

Introducción

La presión inspiratoria máxima (PIM) se define como la mayor presión generada durante una maniobra inspiratoria forzada contra una vía aérea ocluida¹. Es un procedimiento clínico simple, no invasivo y ampliamente utilizado que refleja de forma indirecta la capacidad de generar fuerza de la musculatura inspiratoria durante una contracción breve, cuasi-estática. Sin embargo, una de las desventajas de esta medición radica en que el esfuerzo máximo se encuentra altamente condicionado por el compromiso que el paciente demuestre con la prueba, lo que explica la variabilidad observada en los valores normales, respecto a otras pruebas de función ventilatoria^{2,3}. La sinceridad del esfuerzo se refiere a la motivación consciente del paciente por rendir óptimamente durante una evaluación⁴. Si esto no se cumple, existe riesgo de clasificar erróneamente la severidad de la disfunción lo que podría prolongar el periodo de recuperación o incrementar los gastos en salud.

Otro factor que contribuye a la variabilidad del rendimiento en la medición de PIM corresponde al efecto-aprendizaje, una importante fuente de error sistemático que incrementa la posibilidad de, por ejemplo, reportar mejoras en la fuerza que en realidad no existen^{5,6}.

Todo lo anterior refuerza que el personal a cargo de las pruebas de función ventilatoria debe demostrar un óptimo dominio del equipamiento, además de obtener un rendimiento óptimo, reproducible y sincero del paciente⁷. Si bien no se ha profundizado en torno a cómo impacta la experiencia del evaluador en el rendimiento que presenten los pacientes en la medición de PIM ni cuánto es el tiempo necesario para conseguirlo, existe acuerdo en que los evaluadores experimentados proveen instrucciones claras y precisas, proyectan una imagen segura hacia el paciente y optimizan el tiempo de operación de las mediciones^{7,8}. El proceso de aprendizaje de una técnica es continuo e involucra un modelo clínico de educación del tipo uno a uno, con el fin de que asegurar un traspaso adecuado de conocimientos y la correcta aplicación de los procedimientos empleados en el laboratorio⁹.

Así, se hace necesario constatar si existe concordancia entre las mediciones ejecutadas por diferentes evaluadores, garantizando una correcta decisión clínica. El objetivo de este estudio consiste en determinar si es posible la adquisición de confiabilidad en la medición de PIM de un estudiante en formación durante su prác-

tica profesional en el Laboratorio de Función Ventilatoria de Universidad Católica del Maule.

Método

Estudio observacional, de corte transversal desarrollado en el Laboratorio de Función Ventilatoria de la Universidad Católica del Maule durante un período de 8 semanas (320 horas presenciales), correspondientes a la práctica profesional de un estudiante de Kinesología.

• Participantes

Se reclutó una muestra por conveniencia de sujetos de la región del Maule. Los criterios de inclusión fueron sujetos mayores de 25 años de edad, sin presencia de sintomatología respiratoria aguda y con valores espirométricos normales. Los sujetos con hábito tabáquico, alteraciones en tórax o en columna vertebral o Índice de masa corporal (IMC) > 30 kg/m², fueron excluidos. Todos los participantes firmaron un consentimiento informado. Durante la primera semana el estudiante adquirió un manejo básico en el protocolo de PIM y aplicó la medición a 10 sujetos sanos, bajo la supervisión de un guía con experiencia en la técnica.

• Protocolos

Primero, los sujetos fueron registrados en una base de datos y se realizó la medición de espirometría para determinar el cumplimiento de los criterios de inclusión. La presión inspiratoria máxima se midió por ambos evaluadores: Evaluador A (estudiante) y Evaluador B (experto). Cada paciente fue asignado a uno de los dos evaluadores para comenzar, quienes aplicaron el protocolo de PIM, respetando las indicaciones de la Sociedad Americana del Tórax¹. Se les indicó a los participantes usar clip nasal e insertar una pieza bucal entre los dientes. Luego de 5 ciclos a volumen corriente, se les

Tabla 1: Características de la muestra

Variable	Promedio ± DE
Edad (años)	54,4 ± 11,8
Peso (Kg)	65,9 ± 12,7
Talla (m)	1,52 ± 26,9
VEF1 (l/seg)	2,74 ± 0,74
CVF (l)	3,40 ± 0,98

DE: desviación estándar; VEF1: volumen espiratorio forzado al primer segundo; CVF: capacidad vital forzada.

solicitó exhalar completamente para luego inhalar con tanta fuerza como les fuera posible contra la vía aérea ocluida. Los tres resultados mayores que tuvieran una variabilidad máxima de un 5% fueron incluidos en el análisis. Se respetó un tiempo mínimo de 30 segundos de descanso entre cada medición, o hasta que el paciente indicara que estaba listo para realizar la siguiente maniobra. Todas las mediciones fueron tomadas en un ambiente tranquilo y sin distracciones, en posición sedente y utilizando un pletismógrafo corporal (MED-Graphics Platinum Elite Series).

- Análisis estadístico

La estadística descriptiva se presenta como media y desviación estándar. La normalidad de los datos obtenidos fue analizada con la prueba de Shapiro-Wilk. La confiabilidad se evaluó a través del Coeficiente de Correlación Intraclass (ICC), con un intervalo de confianza del 95%. El análisis de acuerdo se obtuvo del coeficiente de variación y el error estándar de la medición. Se fijó un nivel de significancia de $p \leq 0,05$. Las diferencias individuales fueron exploradas mediante el método gráfico de Bland y Altman. Se usó el programa SPSS v15.0 para el análisis estadístico y los gráficos se generaron utilizando el programa Graph Pad Prism v6.0.

Resultados

La muestra consistió en 32 adultos entre 30 y 75 años de edad. En la Tabla 1 se presenta una descripción de la muestra. Todos los individuos fueron evaluados por ambos evaluadores. La Tabla 2 resume los valores de confiabilidad inter-evaluador para todos los intentos de medición de PIM (Evaluador A y Evaluador B). El análisis con el coeficiente de correlación intra-clase muestra un excelente nivel de confiabilidad inter-evaluador en la

medición de PIM en todos los intentos ($CCI_{1er} = 0.924$, $CCI_{2do} = 0.928$, $CCI_{3ro} = 0.891$). Sin embargo, el método gráfico de Bland-Altman, que considera las diferencias individuales entre el evaluador A y B en relación al rendimiento promedio de los 32 sujetos, muestra que existe una tendencia hacia un mejor rendimiento con el evaluador B (Figura 1). Más aún, rendían en promedio 2,69 cmH₂O más cuando eran evaluados por el evaluador experto. Estos resultados sugieren que, si bien el estudiante adquirió habilidades técnicas y discriminatorias que le permitían identificar variaciones interindividuales en los resultados de los pacientes, existe un sesgo de medición sistemático que no se explica por diferencias en el protocolo de medición.

Discusión

La PIM es el único índice de fuerza muscular inspiratoria disponible de uso masivo, tanto en mediciones clínicas de función pulmonar como en investigación. Sin embargo su utilidad depende de que el paciente alcance un esfuerzo máximo verdadero²⁻⁴. Los motivos de la amplia variabilidad de valores normativos de PIM en distintas poblaciones pueden deberse a este u otros factores, entre los que figuran la experiencia del evaluador¹⁰. Así, el objetivo de este estudio fue valorar la adquisición de confiabilidad en la medición de PIM de un estudiante en formación profesional, mediante la comparación de sus resultados obtenidos con los de un evaluador experto.

El primer hallazgo de este estudio consiste en el excelente grado de acuerdo que existe entre ambos evaluadores. Nuestros valores son similares a los reportados por estudios previos de medición de función ventilatoria entre estudiantes y profesores¹¹. La confiabilidad se define como el grado en el cual la técnica de medición

Tabla 2: Confiabilidad Inter-evaluador.

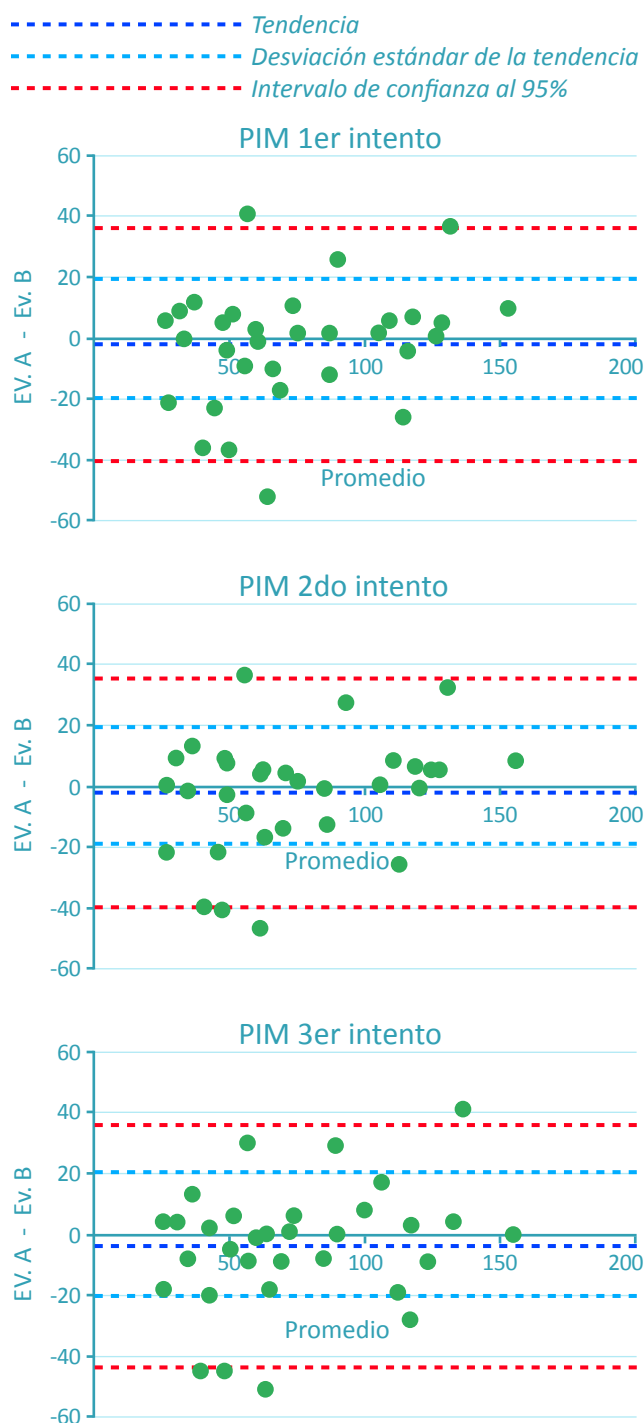
Variable	Evaluador A (promedio±DE)	Evaluador B (promedio±DE)	CCI	IC 95%
1 ^{er} PIM (cmH ₂ O)	130,7 ± 30,1	140,7 ± 31,7	0,924	0,884 - 0,963
%p/c	119,2 ± 20,1	128,5 ± 22,6	0,887	0,768 - 0,945
2 ^{do} PIM (cmH ₂ O)	130,9 ± 30,7	139,7 ± 31,5	0,928	0,853 - 0,965
%p/c	119,3 ± 20,3	127,8 ± 22,6	0,886	0,766 - 0,944
3 ^{ro} PIM (cmH ₂ O)	131,5 ± 30,8	140,4 ± 31,8	0,891	0,778 - 0,946
%p/c	119,9 ± 20,6	127,7 ± 23,2	0,860	0,715 - 0,931

DE: Desviación Estándar; CVF: Capacidad Vital Forzada; CI: Capacidad Inspiratoria; PIM: Presión Inspiratoria Máxima; % p/c: Porcentaje del valor predicho.

puede asegurar resultados consistentes luego de mediciones repetidas sobre el mismo sujeto¹². Por lo tanto, es útil para minimizar la variación causada por factores técnicos, permitiendo que las variaciones biológicas propias de la evolución de la enfermedad puedan ser apropiadamente ponderadas¹³. Ha sido utilizada en otros estudios para comparar el nivel de acuerdo entre evaluadores expertos y novatos en distintas herramientas de medición en terapia física^{14,15}. Por otra parte, al comparar el rendimiento obtenido por los sujetos al ser evaluados por el estudiante y por el evaluador experto, existe un comportamiento diferenciado. Al comparar la diferencia entre los resultados de ambos evaluadores en relación su promedio con la gráfica de Bland-Altman, se evidencia una marcada tendencia hacia un mejor rendimiento con el evaluador experto, de 2,69 cmH₂O (Figura 1). Estos resultados sugieren que el estudiante efectivamente adquirió un dominio técnico del procedimiento de medición de PIM y en su interpretación, ya que fue capaz de distinguir comportamientos diferenciados entre los paciente en concordancia con el juicio del evaluador experimentado.

Ya que todos los participantes recibieron similares instrucciones para la realización de la prueba y fueron medidas por el mismo instrumento, es poco probable que las variaciones en los resultados se deban a diferencias relacionadas con la técnica de medición. Una revisión que aborda el proceso de reflexión sobre la acción entre terapeutas físicos experimentados versus novatos señala que las experiencias previas vividas por el evaluador constituyen un marco de apoyo para la toma de decisiones, al permitirle establecer comparaciones entre los diferentes rendimientos y disminuir la impresión que puede traer un rendimiento inesperado o novedoso¹⁷. En ese sentido, el desarrollo de la práctica reflexiva de un evaluador que recién se interioriza en una técnica debe ser complementado con estrategias de aprendizaje que incorporen dimensiones tanto iterativas como verticales, apropiadas para el logro de las competencias transversales que cobran especial relevancia en las mediciones de fuerza muscular^{5,6}. Los evaluadores experimentados son capaces de proveer instrucciones más claras y efectivas a los pacientes, de tal manera que facilitan la comprensión del test como una forma de potenciar la fiabilidad de los resultados obtenidos. Por otra parte, existe evidencia reciente que destaca la alianza terapéutica como un factor relevante en el comportamiento de los pacientes y en la disposición que presentan en la sesión, con un fuerte impacto

Figura 1: Gráficos de Bland-Altman para los 3 intentos.



sobre en los resultados de la intervención¹⁸.

Si bien los sujetos fueron evaluados en distinto orden por cada evaluador, no se controló apropiadamente el efecto de aprendizaje con una aleatorización del orden de evaluación de ambos evaluadores, lo que constituye una limitación de este estudio cuyo impacto no se puede cuantificar. Otra limitación importante consiste en que no fue posible generar un doble ciego, ya que los participantes podían identificar al estudiante y al profesor y eventualmente modificar su conducta ante cada

uno de ellos. Este hecho pudo pasar desapercibido para el estudiante, al no poseer la habilidad de discriminar entre un gesto motor propio de un rendimiento máximo, de uno sub-máximo igualmente reproducible².

En este punto, es importante cuestionarse la necesidad de complementar las estrategias de enseñanza para la pesquisa efectiva de datos, con énfasis en la adquisición de competencias que potencien la relevancia de la relación kinesiólogo-paciente y desarrollar la capacidad de determinar un gesto motor óptimo, reproducible y confiable¹⁹. Futuras investigaciones deberían considerar las limitaciones de este estudio para aislar apropiadamente el impacto que puede tener la experiencia del evaluador sobre la toma de decisiones clínicas en pruebas que dependen del compromiso, motivación y comprensión del paciente.

Conclusión

De los resultados obtenidos se puede concluir que un corto período, como el de práctica profesional, entrega un excelente nivel de confiabilidad en la aplicación de la medición de PIM en un estudiante en formación.

Conflicto de interés

Los autores declaran no tener conflicto de interés.

Bibliografía

1. American Thoracic Society/European Respiratory Society. ATS/ERS Statement on respiratory muscle testing. *Am J Respir Crit Care Med* 2002; 166:518.
2. Aldrich T, Spiro P. Maximal inspiratory pressure: does reproducibility indicate full effort?. *Thorax* 1995;50:40-43
3. Sachs M., Enright P., Hinckley K., Jiang R., Barr R.; Multi-Ethnic Study of Atherosclerosis Lung Study (2009) Performance of maximum inspiratory pressure tests and maximum inspiratory pressure reference equations for 4 race/ethnic groups. *Respir Care*;54(10):1321-8.
4. Lechner, D, Bradbury, S, & Bradley, L. Detecting sincerity of effort: a summary of methods and appro-

aches. *Phys. Ther.* 1998; 75(8), 867-888.

5. Wen AS, Woo MS, Keens TG. How many manoeuvres are required to measure maximal inspiratory pressure accurately? *Chest* 1997; 111: 802-07.)
6. Assessment of maximum inspiratory pressure (PImax): prior submaximal respiratory muscle activity ('warm-up') enhances PImax and attenuates the learning effect of repeated measurement. S. Volianitis, A.K. McConnell, D.A. Jones
7. Laghi F, Tobin MJ. Disorders of the respiratory muscles. *Am J Respir Crit Care Med.* 2003;168(1):10-48.
8. Puente Maestú L., García de Pedro J. Las pruebas funcionales respiratorias en las decisiones clínicas. *Arch Bronconeumol.* 2012;48(5):161-169.
9. Ferreira EV. Respiratory muscles: myths and secrets. *J Bras Pneumol.* 2015 Mar-Apr;41(2):107-9.
10. American Thoracic Society. Lung Function Testing: Selection of Reference Values and Interpretative Strategies. *American Review of Respiratory Disease*, Vol. 144, No. 5 (1991), pp. 1202-1218.
11. Garrido, F. y Muñoz, R. (2015). Estudio transversal de confiabilidad inter-evaluador para la evaluación de peak del flujo expiratorio, capacidad inspiratoria y presión inspiratoria máxima, *Reem* 2(2), 25-32.
12. López, A. Confiabilidad: teoría y aplicaciones. *Revista oficial del colegio de kinesiólogos de Chile* 2003; 70: 5-8
13. Dimitriadis Z, Kapreli E, Konstantinidou I, Oldham J, Strimpakos N. Test/retest reliability of maximum mouth pressure measurement with the MicroRPM in healthy volunteers. *Respir Care.* 2011;56:776-82
14. Cipriany-Dacko LM, Innerst D, Johannsen J, Rude V. Interrater reliability of the Tinetti Balance Scores in novice and experienced physical therapy clinicians. *Arch Phys Med Rehabil.* 1997;78:1160-1164.
15. Figueiredo K, de Lima K, Cavalcanti Maciel A, Guerra RO. Interobserver reproducibility of the Berg Balance Scale by novice and experienced physiotherapists. *Physiother Theory Pract.* 2009 Jan-Feb;25(1):30-6.
16. Sclausser P, Franco P, Fregonezi G, Sheel A, Chung

F, and Reid W. Reference values for maximal inspiratory pressure: a systematic review. *Canadian Respiratory Journal* 2013; 21, 43-50.

17. Wainwright S, Shepard K, Harman L, Stephens J. Novice and experienced physical therapist clinicians: a comparison of how reflection is used to inform the clinical decision-making process. *Phys Ther.* 2010;90(1):75-88.

18. Fuentes J, Armijo-Olivo S, Funabashi M, et al. Enhanced therapeutic alliance modulates pain intensity and muscle pain sensitivity in patients with chronic low back pain: an experimental controlled study. *Phys Ther.* 2013;94:477-489.

19. Mann, K., Gordon, J., MacLeod, A. (2009). Reflection and reflective practice in health professions education: a systematic review. *Adv in Health Sci Educ* 14:595-621

Correspondencia

Karimé González Gajardo
E-mail: kgonzalezgajardo@gmail.com
Departamento de Kinesiología, Universidad Católica del Maule, Avda. San Miguel 3605. PO Box 617, Talca, Chile.