# Confiabilidad del Instrumento de Evaluación Diagnóstica del Ingreso al Bachillerato

## Reliability of the assesment instrument for high school entry

Oscar Luis Ochoa Martínez
*Colegio de Estudios Científicos y Tecnológicos del Estado de Durango*
chokar128@hotmail.com

Edgar Ochoa Martínez
*Colegio de Estudios Científicos y Tecnológicos del Estado de Durango*
jheoom_kirad@live.com.mx

Gloria Rocío Nery León
Escuela Secundaria General México
neryle@hotmail.com

**Resumen**

Al inicio de cada ciclo escolar en el subsistema de Educación Media Superior (EMS) del Colegio de Estudios científicos y Tecnológicos de los Estado de Durango (CECyTED) y en todos los Estados del país, se aplica el Instrumento de evaluación diagnóstica del ingreso al bachillerato en la modalidad de test-retest de forma equivalente, mediando entre ambas aplicaciones un curso de inducción. El propósito de éste proceso consiste en actualizar y homologar el conocimiento de los aspirantes en las áreas de habilidad matemática (HM) y habilidad lectora (HL), dada su importancia se realizó el estudio con el objetivo de medir el grado de confiabilidad del Instrumento, partiendo del hecho de que en la revisión bibliográfica previa no existe evidencia alguna de que haya sido objeto de algún proceso de estandarización. El tipo de estudio es de carácter instrumental, con base en él y para lograr el objetivo, se realizaron dos pruebas psicométricas; en un primer acercamiento se determinó la calidad de cada uno de los reactivos en función de sus índices de dificultad y discriminación, como referencia a la posterior determinación del coeficiente de KR-20. La muestra en el estudio fue de 34 participantes que cumplieron con el proceso de evaluación diagnostica. Con los valores obtenidos del coeficiente KR-20 de HM y HL en test y retest, se concluyó que el Instrumento es confiable pero no alcanza los niveles de un examen estandarizado.

**Palabras claves**

Confiabilidad, instrumento, evaluación diagnóstica, bachillerato.

**Abstract**

At the beginning of each school year in the subsystem of Higher Secondary Education (HSE) of the College of Scientific and Technological Studies of the State of Durango (CECyTED) and in all the States of the country, the evaluation tool of admission to the school is applied in the test-retest modality in an equivalent way, mediating between both applications an induction course. The purpose of this process is to update and standardize the knowledge of the applicants in the areas of mathematical ability (MA) and reading ability (RA). Given its importance, the study was conducted with the objective of measuring the degree of reliability of the Instrument, parting from the fact that

previous bibliographic review shows no evidence of it being the subject of any standardization process. The type of study is instrumental, two psychometric tests were carried out; In a first approach, the quality of each of the questions was determined according to their difficulty and discrimination indexes, as a reference to the subsequent determination of the KR-20 coefficient. The sample in the study was of 34 participants who complied with the diagnostic evaluation process. With the values obtained from the KR-20 coefficient of MA and RA in test and retest, it was concluded that the instrument is reliable but does not achieve the levels of a standardized test.

**Keywords**
Reliability, instrument, diagnostic evaluation, baccalaureate.

## Introducción

### *Planteamiento del problema*

Para contribuir al logro de su propósito a nivel nacional, los Colegios de Estudios llevan a cabo un proceso de evaluación diagnóstica del ingreso al bachillerato, bajo el supuesto de que al final el estudiante podrá realizar un autodiagnóstico con ayuda del material que incluye actividades y ejercicios ubicados en un contexto específico; es decir, no son conocimientos aislados que pretendan la memorización, su propósito es fortalecerlo en la habilidad matemática y lectora para que aproveche sus fortalezas y emprenda acciones para aminorar sus debilidades.

Tomando en cuenta las diversas opiniones acerca del complicado tema de la evaluación y sus diversas conceptualizaciones, éste trabajo adopta como objeto de estudio el "Instrumento de evaluación diagnóstica de ingreso al bachillerato", con el objetivo general de medir su grado de confiabilidad y, con base en sus resultados, contar con indicadores que le permitan emitir un juicio acerca de su contribución al desarrollo de las habilidades para el que fue construido.

### *Objetivo*

Medir el grado confiabilidad del Instrumento de evaluación diagnóstica del ingreso al bachillerato.

### *Justificación*

La SEMS inició desde algunos años el proceso de reforma educativa; una de las estrategias de reforma es diagnosticar a través de un instrumento de evaluación elaborado con reactivos de opción múltiple, el nivel académico de los estudiantes que ingresan al subsistema, tanto en HM, como HL.

Para el estudiante la evaluación diagnóstica del ingreso al bachillerato representa un acercamiento a las habilidades académicas requeridas para ingresar al nivel medio superior; y para la Institución, un diagnóstico que le permitirá establecer Programas de Mejora Continua, ya que al conocer las deficiencias de sus alumnos en las áreas de HM y HL, entonces contarán con información a partir de la cual podrán emprender acciones tendientes a promover y fortalecer aquellas que ellos aún no dominan, lo que supone traerá como resultado, entre otros aspectos, aminorar el índice de reprobación y mejorar la calidad educativa.

## Sustento teórico

Para inicio del ciclo escolar 2010-2011 y aparejado al proceso de reforma, la evaluación diagnóstica retoma un interés especial y, la Secretaría de Educación Pública (2010) coordinó trabajos con la SEMS, organismo que a través de la COSCAC comenzó a normar y reglamentar el proceso de evaluación diagnóstica del ingreso al Bachillerato; para tal efecto, emitió una *"Guía de estudio para el*

*examen de habilidad matemática y habilidad verbal"*, la base del sustento fue:

> (…) la incorporación de estudiantes al nivel medio superior es un proceso que reviste especial importancia porque la recuperación de conocimientos previos y la construcción de aprendizajes elementales constituirá la base que permita, durante su formación, desarrollar las competencias genéricas, disciplinares y profesionales que finalmente conformen el perfil de egreso (p. 2).

Cada modificación de conducta que se produzca en los estudiantes, en los profesores, o en las experiencias de aprendizaje en general, aporta elementos de diagnóstico que servirán de guía para replantear los objetivos o para una nueva selección y organización de las actividades o de los mismos instrumentos de evaluación. "La formulación de objetivos, expresados en relación con lo que el alumno puede llegar a evidenciar a través de comportamientos específicos, constituye una tarea básica, absolutamente necesaria para el planeamiento de las unidades y el control real de los resultados" (Lafourcade, 1969, p. 46).

Para llevar a cabo un proceso adecuado de diagnóstico, es necesario considerar las características del contexto, las interacciones de los actores sociales y la existencia de problemas o situaciones susceptibles de modificación cuyo resultado facilite la toma de decisiones para intervenir, de acuerdo a esto Camacho (2007) afirma lo siguiente:

> Evaluación diagnóstica. A través de ella se pueden detectar los distintos saberes, actitudes y expectativas de un grupo de estudiantes, y permiten al docente-facilitador tener claridad sobre cómo intervenir en el proceso de aprendizaje. También favorece conocer el contexto y las condiciones en que podrá dirigir su actividad educativa. Es sumamente útil para obtener información valiosa respecto a los conocimientos previos de los alumnos y organizar con mayor realidad las actividades de aprendizaje previstas (p. 207).

## Metodología

### *Método de investigación*

El trabajo está enfocado medir el grado de confiabilidad del Instrumento de evaluación objeto de estudio, razón por la que la investigación se considera de carácter instrumental puesto que "…se consideran como pertenecientes a esta categoría todos los estudios encaminados al desarrollo de pruebas y aparatos, incluyendo tanto el diseño (o adaptación) como el estudio de las propiedades psicométricas de los mismos." (Montero y León, 2005, citado por Barraza, 2010).

Uno de los primeros acercamientos para determinar el grado de confiabilidad del Instrumento, fue el cálculo del índice de dificultad e índice de discriminación de cada uno de los reactivos del HM y HL de *test-retest*; esta medida se realizó mediante el siguiente procedimiento:

> (…) un grupo superior que consta del 27% de que obtiene las calificaciones más altas, un grupo inferior del 27% de que obtiene las calificaciones más bajas y el 46% restante en el grupo intermedio. Cuando la cantidad de sujetos es reducida, pueden emplearse grupos superiores e inferiores del 50 por ciento de las calificaciones totales de la prueba (Aiken, 1996, p. 65).

Para determinar el grado de fiabilidad del Instrumento, propiedad psicométrica que hace referencia a la ausencia de errores de medida, se utilizó la prueba de formas alternativas, éste modelo permite el cálculo del coeficiente de correlación de Pearson o coeficiente de fiabilidad, "(…) su valor es indicativo de la consistencia y estabilidad de las puntuaciones generadas en el proceso de aplicación de un *test y retest* o, del grado de

equivalencia entre la aplicación de sus dos formas paralelas. (Martínez, 1996, p. 82)

***Selección de la muestra***

Los datos recolectados fueron las respuestas a las preguntas correspondientes al *test* de HM y HL que realizaron 34 estudiantes, la selección de los participantes fue de forma determinística observando que fue el máximo marco poblacional que cubrió con los requisitos adecuados para el objetivo de la investigación; en este sentido Babbie (2000) afirma que "(…) en muchas situaciones de investigación el muestreo probabilístico es imposible o inadecuado, y lo más convenientes las técnicas de muestreo no probabilístico" (p. 173).

En aplicación del *test*, el Instrumento estuvo integrado por 37 reactivos en la disciplina de HM (el ítem 15 fue descartado por opciones incorrectas), y el de HL por 39 reactivos; en *retest* se computaron 37 reactivos HM y 39 reactivos para HL.

***Procedimientos para el análisis de la información***

La preparación de datos y las pruebas correspondientes al estudio se realizaron con el uso del programa de cálculo Excel y el programa estadístico SPSS en su versión 22. En el cuadro número 1 (ver anexo 1) se encuentra el código de las variables de estudio.

**Prueba de confiabilidad de la HM y HL de test y retest**

El concepto de confiabilidad o reproducibilidad implica la cantidad de error que se comete al realizar cualquier medida; en la práctica educativa es común dudar acerca de la confiabilidad o repetibilidad de una prueba; si un resultado no es reproducible, el valor y la utilidad de la prueba son pobres. Sobre éste concepto Dueñas, (1998) opina lo siguiente:

> Esta fiabilidad puede ser estimada en diferentes maneras (…) La equivalencia es un procedimiento largo y costoso. Exige construir dos test "paralelos", de naturaleza y dificultades análogas. Es necesario calcular la correlación existente entre las respuestas dadas por los mismos sujetos a ambas pruebas (pp. 26, 27).

Para este caso, se estimó la confiabilidad de consistencia interna de las pruebas de HM y HL en *test y retest*, utilizando para tal efecto el método de Kuder y Richardson; Cohen y Swerdlink (2006) amplían la información al describir "(..) que las pruebas son homogéneas si contienen reactivos que midan un solo rasgo (p.137).

En cuanto a la valoración del coeficiente de confiabilidad se tomó como base los rangos expuestos por Cohen y Swerdlink (2006), bajo la consideración de que puede haber flexibilidad de acuerdo a los propósitos que persiga la aplicación de un examen.

**Resultados**

***Índices de dificultad y discriminación***

Los índices de dificultad y discriminación de un reactivo proporcionan información acerca de su calidad y, en su conjunto, de la calidad del examen que integra determinado set de reactivos; respecto a esta situación Aiken (1996) designa el término "discriminación" con una letra "D" y afirma que:

> Por lo general, el término de validez de los reactivos se refiere a la relación de un reactivo con un criterio externo (…) Para elaborar una prueba que produce calificaciones con alta correlación con el criterio externo, debemos seleccionar reactivos que tengan correlaciones bajas entre sí pero altas con el criterio (p. 67).

La calidad de un reactivo de opción múltiple está en función del valor de su índice de

dificultad e índice de discriminación; para determinar el cumplimiento de esta propiedad psicométrica de cada uno de los reactivos de *test y retest* de HM y HL, se consideró como requisito permanecer dentro del siguiente intervalo de valores:

$$0.2 \leq \textit{Índice dificultad} \leq 0.8$$
$$0.2 \leq \textit{Índice discriminación} \leq 1$$

***Resultado de los índices de dificultad y discriminación en test y retest***

En el anexo 1, se encuentra el valor calculado del índice de dificultad y discriminación para cada uno de los reactivos de *test y retest* de HM y HL; en ella se puede observar que las cantidades mayores negativas son para los índices de discriminación del *test y retest* de HL.

Con base en el rango de valores del intervalo propuesto y en atención a la tabla se observa que los valores del índice de dificultad del *test* de HM (IDIFTHM) e índice de discriminación del *test* de HM (IDISTHM), 24 reactivos se consideran de buena calidad lo que equivale al 61.1%; de los 14 descartados.

Con respecto a los valores del índice de dificultad del *test* de HL (IDIFTHL) e índice de discriminación del *test* de HL (IDISTHL), 21 reactivos se consideran de buena calidad lo que equivale al 53.84%; de los 18 descartados.

En los valores del índice de dificultad del *retest* de HM (IDIFRTHM) e índice de discriminación de HM (IDISRTHM), 22 reactivos se consideran de buena calidad lo que equivale al 59.45%; de los 15 descartados.

En los valores del índice de dificultad del *restest* de HL (IDIFRTHL) e índice de discriminación del *retest* de HL (IDISRTHL), 23 reactivos se consideran de buena calidad lo que equivale al 58.97%; de los 16 descartados.

De la misma tabla se obtuvo el valor de los siguientes estadísticos:

- En el *test* de HM, el índice de dificultad es muy bueno con un valor en su media de 0.43 y con un valor del índice de discriminación aceptable con una media igual a 0.25.
- En el *test* de HL, el índice de dificultad es bueno con un valor en su media de 0.34 y con un valor del índice de discriminación no aceptable con una media igual a 0.16, por abajo del límite inferior.
- En el *retest* de HM, el índice de dificultad es bueno con un valor en su media de 0.38 y con un valor del índice de discriminación aceptable con una media igual a 0.24.
- En el *retest* de HL, el índice de dificultad es muy bueno con un valor en su media de 0.41 y con un valor del índice de discriminación aceptable con una media igual a 0.24.

***Concentrado de puntuaciones***

En el anexo 2, se puede apreciar el número de casos procesados de las puntuaciones PTHM, PTHL, PRTHM y PRTHL; todas ellas corresponden al número de participantes en el estudio.

***Resultado de las pruebas de fiabilidad de test y retest de HM y HL***

En el anexo 3, se aprecia el concentrado de los valores de los estadísticos de confiabilidad de la HM y HL del *test y retest*.

**Interpretación de resultados**

***Dificultad y discriminación del test y retest de HM y HL***

Los valores de los estadísticos de los índices de dificultad y discriminación de los set de reactivos del *test* de HM son los siguientes: el índice de dificultad tiene un valor en su media de 0.43 que se considera muy bueno, mientras que la media del índice de discriminación es aceptable con un valor de 0.25, cercano al límite inferior de la escala seleccionada; en la misma prueba para HL, se obtuvo un valor en su índice de dificultad de 0.34 que se

considera bueno, no así el índice de discriminación que tuvo una media de 0.16, fuera del límite inferior.

En el *retest* de HM, el índice de dificultad tuvo un ligero descenso con un valor de 0.38 mientras que el índice de discriminación se mantuvo prácticamente igual con un valor de 0.24; en la misma prueba para HL, el índice de dificultad también es bueno con un valor de 0.41, mientras que el índice de discriminación si logró situarse dentro de la escala con un valor de 0.24.

En términos generales las medias del índice de dificultad del *test y retest* para HM es muy bueno pues promedian un valor de 0.405 cantidad cercana al valor óptimo del 0.5, mientras que las medias de su índice de discriminación promedian un valor de 0.245, nivel bajo pero aceptable; en cuanto a la media del índice de dificultad del *test y retest* para HL también se puede considerar bastante aceptable con un valor de 0.375 mientras que las medias de su índice de discriminación promedian un valor de 0.2, apenas al límite inferior de la escala.

**Pruebas de confiabilidad de test y retest de HM y HL**

En las pruebas de fiabilidad de *test* y *retest* de HM y HL, se obtuvieron los siguientes resultados: en el *test* de HM se obtuvo un valor alfa de 0.787 y en el *retest* un valor de 0.810; para HL en el *test* se obtuvo un valor alfa de 0.562, incrementando su cantidad en el retest con un valor de 0.787; los estadísticos de fiabilidad de HM son bastante aceptables mientras que el del *test* de HL alcanzó lo suficiente; en general los valores de los estadísticos no alcanzan el nivel requerido para sustentar la estandarización del Instrumento.

**Conclusiones**

Obtener la medida del índice de dificultad e índice de discriminación de los reactivos que integran el Instrumento de evaluación diagnóstica fue fundamental en cuanto a la toma de decisión para diferenciar el estudio entre HM y HL, la razón fue porque desde el momento se observó que la media del índice de discriminación del test de HL no alcanzó el nivel mínimo en la escala seleccionada, se tuvo la percepción de la posible existencia de irregularidades en esta habilidad, a lo anterior se suma que aún y cuando en promedio los índices de dificultan de *test y retest* son bastante aceptables, en promedio los índices de discriminación son bajos pues se sitúan cerca del límite inferior y el poder discriminativo de un reactivo marca diferencia en cuanto a su validez y, por tanto, de la de un set de reactivos.

En las pruebas de fiabilidad que se realizaron por el método de consistencia interna a través del coeficiente de KR-20 se encontró que las puntuaciones de HM tuvieron valores bastante aceptables logrando en *test* un $r = 0.787$ y en el *retest* un $r = 0.81$; para las puntuaciones de HL, los coeficientes resultaron más pobres, en el *test* se obtuvo un $r = 0.562$ alcanzando apenas la suficiencia y en el *retest* aumentó con un $r = 0.787$. Con los valores del coeficiente KR, se concluye que el Instrumento es confiable pero no alcanza los niveles de un examen estandarizado.

**Recomendaciones**

Trabajar en la depuración del banco de reactivos; ya sea eliminando los peores y/o modificarlos con base en un sustento teórico para luego pilotearlos y tomar decisiones en cuanto a su permanencia; de esta manera con trabajo y a través del tiempo se podrá contar con items para elaborar un Instrumento estandarizado.

**Referencias**

Aiken, L. (1996). *Tests psicológicos y evaluación* (8a ed.). Distrito Federal,

México: Prentice Hall Hispanoamericana, S. A.

Babbie, E. (2000). *Fundamentos de la investigación social.* Distrito Federal, México: Thompson Editores, S. A. de C. V.

Barraza, A. (2010). Validación del inventario de expectativas de autoeficacia académica en tres muestras secuenciales. *CPUe, Revista de Investigación Educativa*, 1-30.

Barraza, A., Carrasco, R., y Arreola, M. (2009). *Bournout estudiantil: un estudio exploratorio*. Recuperado el 7 de Marzo de 2015, de http://www.comie.org.mx/congreso/memoriaelectronica/v10/pdf/area_tematica_16/ponencias/0614-F.pdf

Bloom, M., Hastings, J., y Madaus, G. (1975). *Evaluación del aprendizaje.* Buenos Aires., Argentina: Editorial Troquel S. A.

Camacho, R. (2007). *¡Manos arriba! El proceso de enseñanza aprendizaje.* Distrito Federal, México: ST Editorial.

Cohen, R., y Swerdlink, M. (2006). *Pruebas y evaluación psicológicas.* (6a ed.). Distrito Federal, México: McGraw Hill Interamericana Editores, S. A. de C. V.

Dueñas, M. et al. (1998). *El libro de los tests.* Distrito Federal, México: Planeta Mexicana, S. A. de C. V.

Lafourcade, P. (1969). *Evaluación de os aprendizajes.* Argentina: Kapelusz.

Martínez, R. (1996). *Psicometría: Teoría de los tests psicológicos y educativos.* Madrid, España: Editorial Síntesis, S. A.

Secretaría de Educación Pública. (2010). *Guía para el examen de habilidad matemática y habilidad verbal.* Distrito Federal, México: SEP. SEMS. COSDAC.

## Anexos

**Anexo 1**

*Valor de los índices de dificultad y discriminación de test y retest*

| Resúmenes de casos[a] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **IDIF THM** | **IDIS THM** | **IDIF THL** | **IDIS THL** | **IDIF RTHM** | **IDIS RTHM** | **IDIF RTHL** | **IDIS RTHL** |
| 1 | .79 | .18 | .21 | .29 | .75 | 0.00 | .09 | -.18 |
| 2 | .76 | .12 | .35 | .35 | .19 | 0.00 | .53 | .47 |
| 3 | .44 | .29 | .41 | .12 | .25 | .17 | .06 | -.12 |
| 4 | .15 | -.06 | .41 | .35 | .28 | .44 | .06 | 0.00 |
| 5 | .21 | .06 | .59 | .24 | .44 | .22 | .47 | .59 |
| 6 | .71 | .35 | .32 | .18 | .66 | .22 | .44 | .53 |
| 7 | .41 | .35 | .53 | 0.00 | .66 | .17 | .12 | .12 |
| 8 | .38 | .41 | .26 | .06 | .34 | .50 | .24 | 0.00 |
| 9 | .62 | .18 | .26 | .41 | .78 | .28 | .59 | .47 |
| 10 | .26 | -.06 | .24 | .12 | .13 | .06 | .41 | .24 |
| 11 | .41 | .59 | .21 | .18 | .56 | .11 | .50 | .29 |
| 12 | .38 | .41 | .12 | 0.00 | .03 | .06 | .29 | 0.00 |
| 13 | .41 | .24 | .50 | .06 | .44 | .22 | .50 | .18 |
| 14 | .47 | .35 | .59 | .47 | .44 | 0.00 | .60 | .06 |
| 15 | .32 | .29 | .41 | .24 | .38 | .56 | .53 | .24 |
| 16 | .65 | .47 | .47 | .12 | .31 | .44 | .53 | .35 |
| 17 | .76 | .35 | .35 | .12 | .31 | .44 | .26 | -.29 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 18 | .38 | .18 | .32 | .41 | .56 | .28 | .29 | .12 |
| 19 | .32 | .06 | .56 | .18 | .03 | .06 | .29 | .35 |
| 20 | .41 | .24 | .29 | 0.00 | .44 | .44 | .65 | .47 |
| 21 | .21 | .29 | .59 | -.24 | .28 | .28 | .68 | .29 |
| 22 | .44 | .29 | .44 | .18 | .31 | .22 | .12 | -.12 |
| 23 | .15 | .06 | .38 | .18 | .53 | .17 | .47 | .24 |
| 24 | .12 | 0.00 | .35 | .12 | .25 | .33 | .62 | .18 |
| 25 | .44 | .06 | .09 | .06 | .09 | -.17 | .38 | .29 |
| 26 | .41 | .35 | .26 | .29 | .06 | -.11 | .62 | .53 |
| 27 | .56 | .41 | .26 | .06 | .66 | .17 | .32 | -.06 |
| 28 | .47 | .35 | .32 | .41 | .47 | .28 | .59 | .24 |
| 29 | .53 | .3 | .21 | -.06 | .69 | .56 | .38 | .41 |
| 30 | .59 | .71 | .65 | .59 | .31 | -.11 | .41 | .35 |
| 31 | .44 | .41 | .44 | .18 | .19 | .17 | .41 | .47 |
| 32 | .47 | .35 | .26 | .18 | .50 | .56 | .59 | .47 |
| 33 | .65 | .12 | .35 | .35 | .38 | .22 | .50 | .41 |
| 34 | .35 | .24 | .24 | .24 | .50 | .67 | .47 | .71 |
| 35 | .35 | .12 | .38 | .06 | .28 | .50 | .38 | .06 |
| 36 | .21 | .06 | .35 | .12 | .22 | .28 | .47 | .59 |
| 37 | | | .09 | -.06 | .44 | .22 | .35 | 0.00 |
| 38 | | | .21 | -.06 | | | .50 | .29 |
| 39 | | | .21 | -.06 | | | .21 | .18 |
| T N | 36 | 36 | 39 | 39 | 37 | 37 | 39 | 39 |

a. Limitado a los primeros 39 casos.

**Anexo 2**

*Concentrado de puntuaciones*

| Resúmenes de casos[a] | | | | | | |
|---|---|---|---|---|---|---|
| | | **PTHM** | **PTHL** | **PRTHM** | **PRTHL** | |
| 1 | | 15 | 14 | 14 | 20 | |
| 2 | | 15 | 15 | 14 | 20 | |
| 3 | | 14 | 13 | 14 | 21 | |
| 4 | | 11 | 15 | 13 | 20 | |
| 5 | | 14 | 16 | 17 | 11 | |
| 6 | | 19 | 18 | 20 | 23 | |
| 7 | | 22 | 15 | 22 | 20 | |
| 8 | | 17 | 12 | 17 | 13 | |
| 9 | | 17 | 11 | 16 | 12 | |
| 10 | | 21 | 6 | 18 | 16 | |
| 11 | | 11 | 11 | 2 | 0 | |
| 12 | | 27 | 22 | 22 | 27 | |
| 13 | | 10 | 11 | 10 | 8 | |
| 14 | | 26 | 12 | 28 | 21 | |
| 15 | | 23 | 15 | 20 | 16 | |
| 16 | | 15 | 7 | 11 | 13 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| 17 | | 12 | 12 | 12 | 15 | |
| 18 | | 10 | 14 | 10 | 14 | |
| 19 | | 15 | 5 | 20 | 17 | |
| 20 | | 14 | 16 | 12 | 15 | |
| 21 | | 3 | 10 | 9 | 15 | |
| 22 | | 21 | 14 | 24 | 23 | |
| 23 | | 13 | 9 | 5 | 8 | |
| 24 | | 8 | 15 | 9 | 11 | |
| 25 | | 18 | 4 | 11 | 16 | |
| 26 | | 9 | 14 | 9 | 15 | |
| 27 | | 19 | 16 | 13 | 20 | |
| 28 | | 8 | 15 | 13 | 7 | |
| 29 | | 17 | 23 | 14 | 18 | |
| 30 | | 16 | 13 | 10 | 10 | |
| 31 | | 26 | 17 | 28 | 26 | |
| 32 | | 6 | 13 | 11 | 17 | |
| 33 | | 21 | 15 | 21 | 8 | |
| 34 | | 19 | 21 | 19 | 26 | |
| Total N | | 34 | 34 | 34 | 34 | |

a. Limitado a los primeros 39 casos.

**Anexo 3**

*Valor de los estadísticos de confiabilidad de la HM y HL del test y retest*

| Test | Estadístico KR-20 | No. de elementos | Resultado |
|---|---|---|---|
| THM | .787 | 36 | Bueno |
| THL | .562 | 39 | Suficiente |
| RTHM | .810 | 37 | muy bueno |
| RTHL | .787 | 39 | Bueno |