

Evaluación de

Escalas Bayley de Desarrollo Infantil - III

 **BAYLEY-III** 

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	Escalas Bayley de desarrollo infantil - III
Autor	Nancy Bayley
Autor de la adaptación española	Departamento de I+D de Pearson Clinical and Talent Assessment en colaboración con el Centro de Desarrollo
Variable(s)	Escala de desarrollo
Áreas de aplicación	Psicología educativa, Neuropsicología
Soporte	Papel y lápiz

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Excelente	5
Fundamentación teórica	Excelente	4,5
Adaptación	Adecuada	3
Análisis de ítems	Adecuado con carencias	2
Validez: contenido	Buena	4
Validez: relación con otras variables	Adecuada	3
Validez: estructura interna	Buena	3,5
Validez: análisis del DIF	Adecuada con carencias	2
Fiabilidad: equivalencia	No aplicable	--
Fiabilidad: consistencia interna	Excelente	5
Fiabilidad: estabilidad	Adecuada	3
Fiabilidad: TRI	No aplicable	--
Fiabilidad inter-jueces	No aplicable	--
Baremos e interpretación de puntuaciones	Adecuados con carencias	2,5

Comentarios generales

Las Escalas Bayley de desarrollo infantil-III evaluadas, son la adaptación de la Bayley Scales of infant and toddler development – Third Edition elaborada por Nancy Bayley. Estas escalas, basándose en una excelente fundamentación teórica, tienen como objetivo la evaluación de los bebés y niños en tres dominios diferentes: cognitivo, lenguaje y motor. Los dominios pueden ser evaluados de manera independiente, lo cual puede ser de gran utilidad para los equipos de evaluación multidisciplinarios.

Es destacable el hecho de que el Bayley-III permite realizar la evaluación del desarrollo desde el primer mes de vida de niño. Dado que se puede aplicar desde los 1 hasta los 42 meses de edad, las escalas son, lógicamente, heteroadministradas

El Bayley-III, a nivel formal, destaca por la calidad de los materiales, de la documentación aportada en el manual, por la calidad de las instrucciones (tanto para el participante como para la persona encargada de realizar la administración, puntuación e interpretación), así como por la facilidad para registrar las respuestas.

Se ha entendido la fiabilidad tanto como consistencia interna (evaluándola mediante el coeficiente alfa) como estabilidad en la medida. Desde ambas perspectivas, los valores obtenidos muestran una excelente fiabilidad de las distintas escalas y grupos de edad.

El apartado de Validez está estructurado siguiendo las recomendaciones propuestas por los Standards. Se aportan diferentes tipos de evidencias de validez como son las evidencias sobre el contenido, sobre los procesos de respuesta, en relación con otras variables, y de estructura interna. Los resultados obtenidos muestran adecuadas propiedades para la evaluación del desarrollo de los niños y niñas.

El apartado de baremos es el más deficitario, ya que los baremos empleados son los de la versión americana de la prueba que están contruidos sobre la muestra original. Aunque se ofrecen argumentos sobre la aplicabilidad de estos baremos en la muestra española, éstos no están apoyados por estudios empíricos que demuestren la equivalencia métrica de las puntuaciones ni la equivalencia de las distribuciones de las poblaciones.

Propuestas de mejora en futuras revisiones

Análisis de los ítems: Se recomienda incrementar la información proporcionada en el manual sobre el estudio del funcionamiento diferencial de los ítems

Fiabilidad: La muestra empleada para el cálculo de la fiabilidad como estabilidad es de origen estadounidense. Es necesario replicar el análisis con una muestra española.

Validez: En algunos casos las muestras empleadas para encontrar evidencias de validez de relación de las puntuaciones del test con otras variables son estadounidenses. Sería necesario replicar estos análisis con muestras españolas.

Baremos: Los baremos deben emplear muestras locales que garanticen su aplicabilidad

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

Escalas Bayley de desarrollo infantil-III (Bayley-III)

1.2. Nombre del test en su versión original:

Bayley Scales of Infant and Toddler Development-Third Edition

1.3. Autor/es del test original:

Nancy Bayley

1.4. Autor/es de la adaptación española:

Departamento de I+D de Pearson Clinical and Talent Assessment en colaboración con el Centro de Desarrollo Infantil y Atención Precoz (CDIAP) Parc Taulí del Servicio de Medicina Pediátrica de la Corporació Sanitària Parc Taulí y el Grupo de Investigación en Atención Temprana (GIAT) de la Universidad de Murcia

1.5. Editor del test en su versión original:

NCS Pearson Inc.

1.6. Editor de la adaptación española:

Pearson Educación

1.7. Fecha de publicación del test original:

2006

1.8. Fecha de publicación del test en su adaptación española:

2015

1.9. Fecha de la última revisión del test:

2015

1.10. Área general de la/s variable/s que pretende medir el test:

Escala de desarrollo, Neuropsicología

1.11. Breve descripción de la/s variable/s que pretende medir el test:

Las escalas Bayley de desarrollo infantil (BSID; Bayley, 1969, 1977) tienen su origen en varias escalas de desarrollo del niño y están apoyadas en una amplia investigación en el desarrollo infantil.

La escala Bayley de desarrollo infantil-III (Bayley-III) es una revisión de la Bayleys Scales of Infant Development-Second Edition (BSID-II; Bayley, 1993), efectuada con el objetivo de mejorar la calidad del instrumento mediante las siguientes actuaciones: actualización de los baremos, mejoras en las propiedades psicométricas, incremento de la utilidad clínica, simplificación de los ítems, actualización de la aplicación de los ítems y actualización de los materiales. No obstante, la escala Bayley-III conserva la naturaleza y el objetivo general de las escalas Bayley, tal y como las concibió su autora, Nancy Bayley.

La Bayley-III evalúa el desarrollo de bebés y niños pequeños en tres dominios: cognitivo, lenguaje y motor, mediante la aplicación de los ítems al niño. Asimismo, también evalúa el comportamiento del niño durante la evaluación y en el hogar; para ello, en la última página del cuadernillo de anotación se proporciona el "Inventario de observación conductual", que cumplimentan tanto el examinador como los padres o el cuidador y que sirve de ayuda para la interpretación.

Los tres dominios evaluados por la Bayley-III se corresponden con la escala Cognitiva; la escala Lenguaje, formada por dos pruebas: Comunicación receptiva y Comunicación expresiva; y la escala Motora, asimismo formada por dos pruebas: Motricidad fina y Motricidad gruesa.

La escala Cognitiva incluye los ítems que evalúan el desarrollo sensoriomotor, la exploración y la manipulación, la relación entre objetos, la formación de conceptos, la memoria y otros aspectos del procesamiento cognitivo.

La escala Lenguaje incluye ítems de comunicación receptiva e ítems de comunicación expresiva, que componen las pruebas de esta escala: Comunicación receptiva y Comunicación expresiva, respectivamente. La prueba Comunicación receptiva está formada por ítems que evalúan las conductas preverbales; el desarrollo del vocabulario; el vocabulario relacionado con el desarrollo morfológico; y la comprensión de marcadores morfológicos. También incluye ítems que miden la referencia social del niño y la comprensión verbal.

La escala Motora está formada por dos pruebas: Motricidad fina y Motricidad gruesa. Las habilidades motoras asociadas con la prensión, la integración perceptivo-motora, la planificación y velocidad motoras se evalúan en la prueba de Motricidad fina. Los ítems miden las habilidades de los niños más pequeños relacionadas con el seguimiento visual, el alcance, la manipulación de objetos y el agarre; así como las habilidades manuales funcionales y las respuestas a informaciones táctiles. La prueba Motricidad gruesa mide principalmente el movimiento de las extremidades y el torso. Los ítems evalúan el posicionamiento estático; el movimiento dinámico, incluidas la locomoción y la coordinación; el equilibrio y la planificación motora.

Estas tres escalas se pueden aplicar de forma independiente, por lo que puede estar especialmente indicada para equipos de evaluación multidisciplinares. Asimismo, permite realizar un seguimiento del rendimiento del evaluado tras iniciar algún programa de intervención.

Respecto a la interpretación de las puntuaciones, a partir de las puntuaciones directas de la escala, es posible obtener puntuaciones escalares, compuestas, percentiles, puntuaciones de edad equivalente y puntuaciones de desarrollo obtenidas a partir del modelo logístico de un parámetro de la TRI. Asimismo, ofrece una clasificación cualitativa para encuadrar el nivel de rendimiento del niño en hasta 7 categorías (desde muy bajo, hasta muy superior). Para cada evaluado, también permite el establecimiento de perfiles desde una perspectiva intra e interindividual.

1.12. Áreas de aplicación:

Psicología educativa, Psicología clínica, Neuropsicología

1.13. Formato de los ítems:

Respuesta dicotómica, si bien el instrumento consta de muy diversos tipos de ítems. Por ejemplo, en la Escala Cognitiva, responder al entorno, observar objetos, explorar objetos, jugar con determinados objetos, manipular objetos, encontrar objetos escondidos, hacer puzles... En la prueba de Comunicación receptiva, reconocer palabras familiares, identificar objetos, entender órdenes, seguir indicaciones, comprender conceptos... En la prueba de Comunicación expresiva, emitir vocalizaciones, realizar inflexiones vocálicas, combinar palabras y gestos, utilizar frases... En la prueba de Motricidad fina, seguir visualmente objetos o personas, agarrar objetos, pasar un objeto de una mano a otra, hacer garabatos, imitar trazos, discriminar formas a través del tacto, construir, dibujar... Finalmente, en la prueba de Motricidad gruesa, agitar las piernas y los brazos, hacer movimientos de gateo, incorporarse, coger el pie con las manos, saltar, caminar, subir escaleras, chutar la pelota, mantener el equilibrio sobre un pie...

1.14. Número de ítems:

Escala Cognitiva (91).

Escala de Lenguaje (97): Comunicación receptiva: 49; y Comunicación expresiva: 48.

Escala Motora (138): Motricidad fina: 66; y Motricidad gruesa: 72.

TOTAL ESCALA (326).

1.15. Soporte:

Administración oral, papel y lápiz, manipulativo

1.16. Cualificación requerida para el uso del test de acuerdo con la documentación aportada:

Nivel C

1.17. Descripción de las poblaciones a las que el test es aplicable:

Niños y niñas de entre 1 y 42 meses de edad. En el manual de aplicación se proporcionan sugerencias generales para determinar las adaptaciones adecuadas para niños y niñas con discapacidad visual, auditiva, en el movimiento y control motor.

1.18. Existencia o no de diferentes formas del test y sus características:

No existen diferentes formas del test. Sin embargo, dependiendo de la edad cronológica de los niños evaluados, se definen los puntos de comienzo de la aplicación. Todas las pruebas disponen de regla de retorno y regla de terminación para garantizar que se aplican los ítems más adecuados para el niño.

1.19. Procedimiento de corrección:

Manual y Automatizado por ordenador

1.20. Puntuaciones:

La escala proporciona puntuaciones directas, escalares, compuestas, percentiles, puntuaciones de edad equivalente y puntuaciones de desarrollo.

Las puntuaciones directas para cada prueba se obtienen a partir de la suma de los puntos obtenidos en cada prueba. Cada ítem respondido correctamente se puntuó con un 1.

Las puntuaciones escalares se calculan a partir de las puntuaciones directas de cada prueba. Tienen de media 10 y de desviación típica 3. Representan el rendimiento del niño con relación a los niños de su misma edad.

Las puntuaciones compuestas son puntuaciones típicas obtenidas a partir de las escalares. Para las escalas de Lenguaje y Motora, se reescalan con media 100 y desviación típica 15.

Las puntuaciones de edad equivalente suponen la edad en la que una puntuación directa es igual a la media para un determinado grupo de edad.

Las puntuaciones de desarrollo se calculan a partir del modelo logístico de un parámetro de la TRI. A continuación se reescalan con media 500 y desviación típica 100.

1.21. Escalas utilizadas:

Puntuaciones basadas en percentiles (centiles), puntuaciones estandarizadas (típicas), puntuaciones escalares (M=10, DT=3).

1.22. Posibilidad de obtener informes automatizados:

Sí

1.23. Tiempo estimado para la aplicación del test:

Entre 30 y 90 minutos (dependiendo de la edad).

1.24. Documentación aportada por el editor:

Manual

1.25. Precio de un juego completo de la prueba:

1045€ (20/09/2020)

1.26. Precio y número de ejemplares del paquete de cuadernillos:

25 cuadernillos de anotación 88,95€ (20/09/2020)

1.27. Precio y número de ejemplares del paquete de hojas de respuesta:

1.28. Precio de la administración y/o corrección, y/o elaboración de informes por parte del editor:

Recarga de 25 perfiles on-line 54,95€ (20/09/2020)

2. VALORACIÓN DE LAS CARACTERÍSTICAS DEL TEST

2.1. Aspectos generales:

Contenido	Valoración	Puntuación
2.1. Calidad de los materiales del test	Excelente	5
2.2. Calidad de la documentación aportada	Excelente	4,5
2.3. Fundamentación teórica	Excelente	4,5
2.4. Adaptación del test	Adecuada	3
2.5. Desarrollo de los ítems del test	Buena	4
2.6. Calidad de las instrucciones para el participante	Excelente	5
2.7. Calidad de las instrucciones (administración, puntuación, interpretación)	Excelente	5
2.8. Facilidad para registrar las respuestas	Excelente	5
2.9. Bibliografía del manual	Buena	4
2.10. Datos sobre el análisis de los ítems	Adecuada, pero con alguna carencia	2

2.11. Validez:

2.11.1. Evidencias de validez de contenido:

Contenido	Valoración	Puntuación
2.11.1.1. Calidad de la representación del contenido o dominio	Adecuada	3
2.11.1.2. Consultas a expertos	Excelente	5

2.11.2. Evidencias de validez basadas en la relación entre las puntuaciones del test y otras variables:

2.11.2.1. Evidencias de validez basadas en la relación entre las puntuaciones del test y otras variables:

Contenido	Valoración	Puntuación
2.11.2.1.1. Diseños empleados	Correlaciones con otros tests	
2.11.2.1.2. Tamaño de las muestras	Adecuada	2
2.11.2.1.3. Procedimiento de selección de las muestras	Incidental	
2.11.2.1.4. Calidad de los tests empleados como criterio o marcador	Adecuada	3
2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos similares	Buena	4
2.11.2.1.6. Promedio de las correlaciones con otros tests que miden constructos no relacionados	No se aporta	--
2.11.2.1.7. Resultados de la matriz multirrasgo-multimétodo	No se aporta	--
2.11.2.1.8. Resultados de las diferencias intergrupo	Excelente	5

2.11.2.2. Evidencias de validez basadas en la relación entre las puntuaciones del test y un criterio:

Contenido	Valoración	Puntuación
2.11.2.2.1. Criterios empleados	MSCA y K-ABC	
2.11.2.2.2. Calidad de los criterios empleados	Excelente	4,5
2.11.2.2.3. Relación temporal entre test y criterio	Concurrente	
2.11.2.2.4. Tamaño de las muestras	Adecuada	2
2.11.2.2.5. Procedimiento de selección de las muestras	Incidental	
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	Excelente	5

2.11.3. Evidencias de validez basadas en la estructura interna:

Contenido	Valoración	Puntuación
2.11.3.1. Resultados del análisis factorial	Excelente	5
2.11.3.2. Funcionamiento diferencial de los ítems	Adecuado con carencias	2

2.11.4. Acomodaciones en la administración del test:

Contenido	Valoración	Puntuación
2.11.4. El manual del test informa sobre las acomodaciones en la administración del test	Sí	

2.11.5. Comentarios generales sobre evidencias de validez:

El apartado de validez del test está estructurado en consonancia con las diversas fuentes de evidencia: evidencia basada en el contenido, evidencia basada en la estructura interna y evidencias basadas en las relaciones con otras variables. También proporciona evidencia basada en los procesos de respuesta.

1. Respecto a la validez de contenido, para la elaboración de la Bayley-III se ha contado con un grupo de 71 expertos. En el apartado referido a la validez, se remite a los capítulos anteriores, donde se detalla el contenido y la estructura del test. No se describe en detalle el proceso sistematizado utilizado en las consultas con dichos expertos.

2. Evidencia basada en los procesos de respuesta. Durante el desarrollo y revisión de los ítems, el equipo de investigación y los expertos evaluaron que las tareas se centran en las habilidades y los procesos que se pretenden evaluar y no requieren habilidades aún no adquiridas.

3. Evidencias de relación entre las puntuaciones del test y otras variables.

Los datos relativos a este grupo de evidencias se dividen en dos apartados. En el primero, se presentan los estudios llevados a cabo para analizar las relaciones entre las puntuaciones de la Bayley-III y otras medidas de rendimiento intelectual y de aptitud cognitiva, de lenguaje y de motricidad (6 estudios). En el segundo apartado, se muestran las evidencias para grupos especiales (10 estudios).

Los test marcadores utilizados en el primer grupo de estudios (Escala McCarthy de aptitudes y psicomotricidad (MSCA) y Batería de evaluación de Kaufman para niños (K-ABC) en muestras de población española; y Wechsler Preschool and Primary Scale of Intelligence-Third Edition (WPPSI-III); Preschool Language Scale-Fourth Edition (PLS-4) y Peabody Developmental Motor Scales, Second Edition (PDMS-2) en muestras de población estadounidense, tienen propiedades psicométricas satisfactorias pero no se presentan tales propiedades ni se justifica su elección.

Los criterios utilizados en el segundo grupo de estudios son discapacidad intelectual, trastorno del espectro autista, parálisis cerebral, trastorno del lenguaje, retraso en el desarrollo, déficit de atención con hiperactividad, riesgo de asfixia al nacer, exposición prenatal al alcohol, tamaño pequeño para su edad gestacional, y prematuridad o bajo peso al nacer. Se han realizado comparaciones de medias entre las muestras de grupos especiales y las muestras de control. El tamaño del efecto promedio ha sido grande.

4. Evidencia basada en la estructura interna. En primer lugar, se han realizado estudios de intercorrelación entre las pruebas del test. En segundo lugar, se han utilizado modelos de análisis factorial confirmatorio. En ambos casos se aporta suficiente información.

No se aporta apenas información sobre el funcionamiento diferencial de los ítems, aunque en el manual se habla de que se han utilizado metodologías contemporáneas para examinar el potencial sesgo de los ítems.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Consistencia interna, error típico de medida y coeficientes de estabilidad	

2.12.2. Equivalencia formas paralelas:

Contenido	Valoración	Puntuación
2.12.2.1. Tamaño de las muestras	No se aporta	--
2.12.2.2. Puesta a prueba de los supuestos de paralelismo	No se aporta	--
2.12.2.3. Promedio de coeficientes de equivalencia	No se aporta	--

2.12.3. Consistencia interna:

Contenido	Valoración	Puntuación
2.12.3.1. Tamaño de las muestras	Buena	4
2.12.3.2. Coeficientes de consistencia interna presentados	Dos mitades, corregidos mediante Spearman-Brown	
2.12.3.3. Promedio de coeficientes de consistencia	Excelente	5

2.12.4. Estabilidad (test-retest):

Contenido	Valoración	Puntuación
2.12.4.1. Tamaño de las muestras	Adecuada	2
2.12.4.2. Coeficientes de estabilidad	Buena	4

2.12.5. Cuantificación de la precisión mediante TRI:

Contenido	Valoración	Puntuación
2.12.5.1. Tamaño de las muestras	No se aporta	--
2.12.5.2. Coeficientes proporcionados	No se aporta	
2.12.5.3. Tamaño de los coeficientes	No se aporta	--

2.12.6. Fiabilidad inter-jueces:

Contenido	Valoración	Puntuación
2.12.6.1. Tipos de coeficientes presentados	No se aporta	
2.12.6.2. Promedio de los coeficientes	No se aporta	--

2.12.7. Comentarios generales sobre evidencias de fiabilidad:

Se han realizado análisis de consistencia interna, calculado los errores típicos de medida e intervalos de confianza, y analizado la estabilidad temporal para cada una de las escalas.

Con respecto a la consistencia interna, se han utilizado dos muestras grandes: la muestra de tipificación estadounidense de N=1700 y una muestra de poblaciones especiales (síndrome de Down, riesgo de retraso en el desarrollo, trastorno específico del lenguaje, parálisis cerebral, trastorno generalizado en el desarrollo, exposición prenatal al alcohol, prematuros o de bajo peso al nacer, pequeños para su edad gestacional y asfixia al nacer) de N=668. Se ha utilizado el método de las dos mitades: se ha calculado el coeficiente de correlación entre las puntuaciones de las dos mitades corregida mediante la fórmula de Spearman-Brown para cada prueba. El promedio de los índices calculado para todas las pruebas y las distintas edades en la muestra de tipificación es de .90 y en la muestra de poblaciones especiales de .96, que pueden considerarse Excelentes.

El análisis de la estabilidad test-retest se ha llevado a cabo con una muestra de 197 niños estadounidenses. Se calculó la correlación de Pearson entre las puntuaciones de las dos aplicaciones llevadas a cabo con un intervalo que osciló entre los 2 y los 15 días con un intervalo medio de 6 días. También se calcularon los coeficientes de correlación corregidos de acuerdo con la variabilidad de la muestra de tipificación. Promediando para las distintas pruebas y edades, el coeficiente de correlación fue de .77 (que indica buena estabilidad) y el coeficiente de correlación corregido de .82 (que cabe considerarse excelente).

Como punto de mejora cabe destacar que se deberían realizar los análisis de la fiabilidad con muestras españolas.

2.13. Baremación e interpretación de las puntuaciones:

2.13.1. Interpretación normativa de las puntuaciones:

Contenido	Valoración	Puntuación
2.13.1.1. Calidad de las normas	Adecuada con algunas carencias	2
2.13.1.2. Tamaño de las muestras	Excelente	5
2.13.1.3. Aplicación de estrategia de tipificación continua	Sí	
2.12.1.4. Procedimiento de selección de las muestras	Incidental	
2.12.1.5. Actualización de baremos	Excelente	5

2.13.2. Interpretación referida a criterio:

Contenido	Valoración	Puntuación
2.13.2.1. Adecuación del establecimiento de los puntos de corte	No se aporta	--
2.13.2.2. Procedimiento empleado para fijar los puntos de corte	No se aporta	
2.13.2.3. Procedimiento de obtención del acuerdo inter-jueces	No se aporta	
2.13.2.4. Valor del coeficiente de acuerdo inter-jueces	No se aporta	--

2.13.3. Comentarios generales sobre baremación e interpretación de las puntuaciones:

El manual ofrece una amplia variedad de información sobre la interpretación de las puntuaciones obtenidas a partir de un baremo estadounidense que es el que se aplica igualmente en España. Las puntuaciones obtenidas comprenden desde la propia puntuación directa, hasta las puntuaciones de desarrollo, lo que resulta de enorme interés dado el tipo de constructo y población medida por la escala. Además, ofrece perfiles, lo que permite comprobar la evolución intra e interindividual de los sujetos.

Sin embargo, la baremación asume la de Estados Unidos. A este respecto, la única información observada en el manual dice textualmente "...teniendo en cuenta la solidez del test, el tamaño de tipificación estadounidense (N=1700), la naturaleza de las pruebas y el uso de los baremos estadounidense en las versiones previas, se valoró la pertinencia de la elaboración de baremos españoles". Más adelante también se indica que "Durante la traducción y adaptación de los ítems se tuvo en cuenta que las modificaciones realizadas no alterarían el nivel de dificultad de los ítems originales, ni el constructo que se pretendía medir, de forma que dichas modificaciones no interfiriesen en la utilización de los baremos españoles". Aunque la rigurosidad en el proceso de traducción y adaptación de los ítems es un aspecto muy positivo, esto no garantiza la adecuación de los baremos para muestras españolas ni su generalización.

Seguindo las recomendaciones de la *International Test Commission* para la traducción y adaptación de pruebas (directriz C3), es necesario proporcionar evidencias que apoyen el uso de los baremos en la población objetivo. Esto solo se garantiza cuando las muestras empleadas para la baremación forman parte de la población objetivo. Solo en casos muy excepcionales, para constructos y pruebas que no tengan carga cultural ni existan diferencias lingüísticas podría justificarse el empleo de los baremos originales. Incluso en este caso, antes de usar los baremos originales, estudios rigurosos deberían verificar la invarianza de las medidas y la igualdad de las distribuciones de las distintas poblaciones en el constructo evaluado (siendo en cualquier caso deseable contar con muestras locales)