



---

**Universidad de Valladolid**

# **Análisis de la validez del cuestionario "Repercusión de la corrección refractiva en la calidad de vida" (QIRC) en lengua española.**

---

**Realizado por:** Elena Sanz Sanz  
**Tutor:** Dr. Miguel José Maldonado López

**Máster en Investigación en Ciencias de la Visión**

**Universidad de Valladolid**

**Curso 2015/2016**



# 1.Índice

1.ÍNDICE .....	3
2.CURRÍCULUM VITAE .....	4
3. RESUMEN.....	5
4. INTRODUCCIÓN .....	6
4.1 CALIDAD DE VIDA .....	6
4.2 CUESTIONARIOS DE CALIDAD DE VIDA .....	7
4.3 QIRC .....	8
4.4 ANÁLISIS RASCH.....	9
4.5 VALIDACIÓN DE UN CUESTIONARIO.....	10
5. HIPÓTESIS .....	13
6. OBJETIVOS.....	13
7. MATERIALES Y MÉTODOS.....	14
8. RESULTADOS .....	18
8.1 DATOS DE LA MUESTRA .....	18
8.2 RESULTADOS DEL CUESTIONARIO .....	19
8.3 ANÁLISIS RASCH.....	22
8.4 RETEST.....	31
9. DISCUSIÓN.....	34
10. CONCLUSIONES .....	38
11. AGRADECIMIENTOS .....	38
12. BIBLIOGRAFÍA .....	39
13. ANEXOS.....	41
13.1 AUTORIZACIÓN DEL TUTOR .....	41
13.2 ACEPTACIÓN DEL ESTUDIO POR PARTE DE LA COMISIÓN DE INVESTIGACIÓN .....	42
13.3 CONSENTIMIENTO INFORMADO .....	43
13.4 CUESTIONARIO QIRC EN ESPAÑOL .....	46
13.5 TABLA DE RESULTADOS.....	49

## 2.Currículum Vitae

### Información personal:

Nombre: Elena Sanz Sanz

Fecha de nacimiento: 19/11/1991

### Estudios académicos:

- Estudios de Educación Infantil, Primaria, Secundaria y Bachillerato en el colegio Nuestra Señora de Lourdes (Valladolid). 1993-2009.
- Diplomatura en Óptica y Optometría en la facultad de Ciencias de la Universidad de Valladolid (UVa). 2009-2012. Nota media: 7,447
- Máster en Optometría Avanzada y Ciencias de la Visión en la facultad de Ciencias de la Universidad de Valencia (UV). Curso 2012/13. Nota media: 8,22
- Máster en Investigación en Ciencias de la Visión, en la Universidad de Valladolid (UVa). Curso 2015/2016.

### Experiencia laboral:

- Prácticas durante un mes (febrero de 2013) en la clínica optométrica Contacvisión, en Valencia.
- Óptico-optometrista en óptica Vision&Co, en León, desde enero de 2014 hasta junio de 2014.
- Óptico-optometrista en óptica Puppiloptica, en Valladolid, desde julio de 2014 hasta octubre de 2015.

### Situación actual y méritos durante el curso académico:

- Máster en Investigación en Ciencias de la Visión, en la Universidad de Valladolid (UVa). Nota media (a falta del TFM): 9,57.
- Beca del CIBER-BBN (mayo 2016/mayo 2017) en el proyecto EYEPOC 2.
- Obtención del nivel B2 (First) en inglés en los exámenes de Cambridge, con nivel C1 en Reading y Use of English.

### 3. Resumen

**Introducción:** Desde un punto de vista clínico y de investigación, es de gran interés conocer cómo afecta la corrección refractiva de una persona no presbita, ya sea mediante el uso de gafas, lentes de contacto o de cirugía refractiva, a su calidad de vida. Para ello, una forma de averiguarlo es mediante un cuestionario en el que se analicen diferentes aspectos que afectan a la vida cotidiana. Por este motivo, se creó el test QIRC (Quality of Life Impact of Refractive Correction), un test de 20 preguntas realizado en lengua inglesa.

**Objetivos:** El objetivo de este estudio consiste en realizar las pruebas y el análisis necesario para validar el cuestionario QIRC sobre la repercusión que tiene la corrección refractiva de una persona no presbita en su calidad de vida en lengua española.

**Material y métodos:** Se realiza el cuestionario a un total de 91 pacientes, divididos entre los tres grupos. Se repite el cuestionario para valorar la fiabilidad a 31 pacientes. Después se hizo un análisis estadístico incluyendo un análisis Rasch para evaluar los resultados.

**Resultados:** La media total del cuestionario fue de  $51,88 \pm 7,05$ , con resultados significativamente mejores en el grupo de cirugía refractiva. Los ajustes al modelo con el análisis Rasch fueron inicialmente peores a lo esperado, con un Cronbach's  $\alpha$  de 0,643, un test de Andersen significativo, mapa ítem-persona desajustado, umbrales desordenados, y valores de infit y outfit fuera de lo esperado. Estos resultados mejoraron cuando se redujo a un total de diez preguntas con menos categorías de respuesta. Finalmente se consigue Cronbach's  $\alpha$  de 0,68, test de Andersen no significativo, mapa ítem-persona mejorado, umbrales ordenados, infit de  $0,93 \pm 0,05$ , y el outfit de  $0,92 \pm 0,07$ , y PCA (análisis de componentes principales) por debajo de 2. El re-test nos indica resultados significativamente peores en el segundo pase, con distribución normal pero en el análisis de Bland-Altman con límites de acuerdo (LoA) superior de 10,65, e inferior de -6,05, aunque el ICC es aceptable, de 0,762.

**Conclusiones:** Aunque la muestra es limitada y los datos no totalmente concluyentes, se puede considerar la validez del cuestionario QIRC en español. Además, se ha creado un nuevo cuestionario reducido eliminando preguntas cuyos resultados no se adaptaban al modelo esperado, pasando de las 20 preguntas originales a un total de 10, y disminuyendo también las categorías de respuesta a las imprescindibles.

## 4. Introducción

Las ametropías en edad previa a la presbicia, es decir, miopía, hipermetropía y astigmatismo, afectan a una parte importante de la población. Sólo en España, según el Libro Blanco de la Visión, en 2013 el 53,4% necesitaba algún tipo de corrección, y sin tener en cuenta las personas mayores de 40 años, la cifra se queda en un 42% (1).

Estos defectos visuales pueden ser compensados mediante el uso de gafas, de lentes de contacto (o ambas), o realizando un procedimiento de cirugía refractiva que lo corrija de forma permanente. Aunque lo más extendido sea el uso de gafas, el 7,4% de la población española utiliza lentes de contacto, es decir, unos 2,5 millones de personas. Este porcentaje aumenta entre la población de 12-24 años, en cuyo caso el 14,11% de la gente en esta franja de edad utiliza lentes de contacto (1). El tratamiento mediante cirugía refractiva va mejorando con los años, con cirugías cada vez más seguras y exactas. En 2012, hubo 125000 intervenciones de cirugía refractiva, y anteriormente, en 2009, se llegaron a 280000 operaciones (1).

Por lo tanto, viendo los datos anteriores, se puede comprobar el gran número de personas que ven o han visto su vida afectada, en mayor o menor medida, su calidad de vida debido a sus defectos visuales.

### 4.1 Calidad de vida

No hay una definición clara y exacta, y han sido múltiples autores los que han buscado la manera de entender este concepto. La Organización Mundial de la Salud definió la calidad de vida como la percepción de cada individuo de su situación de vida en el contexto de la cultura en la que vive y su relación con sus objetivos, expectativas, criterios y preocupaciones. Es un concepto multidimensional que se ve afectado por la salud física, estado psicológico, nivel de independencia, relaciones sociales, creencias personales y su relación con el entorno (2).

Se puede categorizar en cinco dimensiones: bienestar físico, material, social y emocional, y desarrollo y actividad. Esta división no es empírica, si no una forma de clasificar el contenido, derivada del estudio de la literatura previa (3). En el caso de calidad de vida relacionada con la salud, también se tienen en cuenta dimensiones como los síntomas, la percepción de la salud, y la satisfacción (4).

La calidad de vida de un individuo o subgrupo se puede localizar comparando su posición respecto a la distribución de la población total (3).

La visión está relacionada con la calidad de vida, ya que dependemos de ella para entender el entorno y recibir información del exterior, y si se ve afectada influye en el individuo en múltiples factores, tanto físicos (poder interactuar con el entorno), como psicológicos (autoconfianza, dependencia o la imagen de uno mismo) o incluso económicos (gasto en ayudas ópticas). En este estudio se analiza la calidad de vida relacionada con la corrección refractiva (5).

## 4.2 Cuestionarios de calidad de vida

La satisfacción del paciente es un parámetro complejo y multidimensional. Está influenciado por la calidad de visión subjetiva, expectativas personales o tipo de personalidad, entre otros. La evaluación de la satisfacción del paciente y los resultados de autopercepción requieren un instrumento psicométrico sensible que tenga en cuenta estos parámetros (6).

La clasificación de los pacientes según su calidad de vida es muy útil para investigación y práctica clínica. Para poder medir y cuantificar la calidad de vida se utilizan cuestionarios. Esta es una herramienta eficiente para reunir gran cantidad de datos rápidamente (5,7). En el caso de la investigación oftalmológica, nos ayuda a complementar y mejorar la comprensión del estado visual del paciente, siendo una prueba suplementaria a otros test de valoración clínica (7). Su utilización se empezó a desarrollar y extenderse desde 1980 (8).

Un cuestionario es un conjunto de preguntas sobre ciertas actividades específicas, que se deben responder mediante una clasificación que representa el nivel de dificultad que experimenta la persona en la actividad descrita. Sirven como valoración funcional de la calidad de vida (8). La habilidad es una característica de una persona. Diferentes personas tienen diferentes niveles de habilidad, y eso es exactamente lo que se quiere medir. La habilidad visual es una componente de la habilidad que puede ser modulada por cambios en la función visual (9). Los resultados de un cuestionario proporcionan una representación numérica del resultado de un paciente (10).

Un cuestionario o instrumento, nombre utilizado en la bibliografía, debe ser desarrollado en una población comparable a la población objetivo. Debe incluir preguntas relevantes para todos los tipos de pacientes. En nuestro caso, para todos los tipos de corrección (11).

Las preguntas se pueden agrupar en dominios o subescalas. Preguntas, también llamadas ítems, del mismo dominio valoran la misma variable (por ejemplo, salud general); aunque algunos ítems pueden valorar más de una variable (8).

Las respuestas pueden ser dicotómicas (sí/no), o múltiples (en una escala) (8). Las respuestas abiertas reducen la sensibilidad y crean problemas de validez (3). Las preguntas deben ser simples y cortas, para asegurarse una asimilación rápida y fácil. No hay que usar dobles negativos, ni términos técnicos o médicos. Las preguntas deben agruparse en un orden lógico. Completar el cuestionario no debe suponer un esfuerzo emocional para el paciente (6).

Para que un instrumento de calidad de vida sea clínicamente útil tiene que reunir un conjunto de características como: validez, fiabilidad, sensibilidad (detectar cambios en el tiempo y entre grupos), aceptabilidad e interpretabilidad (6,12).

Anterior al test QIRC (Quality of Life Impact of Refractive Correction), aparecieron otros dos test de calidad de vida en relación con la refracción. Estos son el RSVP (Refractive Status and Vision Profile) y NEI-RQL (National Eye Institute-Refractive Error Correction Quality of Life). Son test tradicionales, desarrollados con el sistema de análisis Likert.

El test RSVP se creó en 1999. Se compone de 42 preguntas. Fue diseñado para medir funcionamiento, síntomas, percepción de salud y expectativas de individuos con error refractivo; aunque fue desarrollado mayoritariamente con pacientes que se querían someter a cirugía refractiva o ya operados (92,4%), por lo que el cuestionario puede hallarse sesgado hacia

este grupo de pacientes (4). Este test está más correlacionado con la satisfacción postoperatoria que con el error refractivo o la AV. Además, es posible que el resultado final demuestre mejora tras la operación incluso cuando hay una disminución significativa en alguna de las subescalas. Hay diferencias significativas entre usuarios de gafas, lentes de contacto y emétopes, y es sensible a los operados mediante cirugía refractiva, sin embargo no distingue el tipo (LASIK, PRK...). Entre sus problemas, se encuentra que es insensible a diferentes tipos de lentes de contacto y a diferentes grados de ametropía, y que no es capaz de diferenciar entre un grupo de lentes de contacto y gafas de otro que se quiere someter a cirugía refractiva (4,5,8,13,14,15).

El test NEI-RQL fue desarrollado en 2001. Consta de 42 preguntas. Su propósito era medir la calidad de vida de personas con error refractivo, cuya visión corregida sea de 20/30 o mejor. Sus resultados indicaron que dichos pacientes experimentan un rango de problemas relacionados con la visión y el uso de la corrección (12). Este test discrimina los diferentes tipos de corrección, el grado de ametropía, y es capaz de discriminar entre ambos grupos y sujetos emétopes, cuya calidad de vida es mejor. Su fiabilidad es mayor que la del test RSVP. Uno de los problemas de este test es que sobre enfatiza en la pérdida de función visual y síntomas, porque es lo más frecuente cuando un paciente es examinado en clínica (5,7).

### **4.3 QIRC**

En 2004, Pesudovs et al desarrolló otro cuestionario para calidad de vida dependiendo del error refractivo, intentando resolver los problemas de los anteriores. Este test (Quality of life Impact of Refractive Correction) está adecuado para los tres tipos de corrección (gafas, lentes de contacto y cirugía refractiva), con preguntas adaptadas para ello. Pero el gran avance de este test respecto a los otros es la utilización de análisis Rasch. Fue desarrollado mediante procedimientos habituales, y después mejorado mediante reducción de preguntas y aplicación de una calificación procesada por este programa, el cual proporciona diferentes pesos a las respuestas y permite una escala más exacta (5).

La obtención de preguntas se realizó a partir de seis fuentes: literatura de calidad de vida en general, literatura de calidad de vida relacionada con la visión, literatura de calidad de vida en relación con la cirugía cosmética, análisis representativo de casos de la University of Bradford Eye Clinic, 63 expertos de diferentes campos y grupos de atención.

De todas las preguntas obtenidas se resumieron en un cuestionario piloto de 90 preguntas, que fue reducido hasta obtener el cuestionario final de 20 preguntas. Está dividido en dos grupos: valoración de la severidad (preguntas 1-13), y valoración de la incidencia (preguntas 14-20). A cada una de las preguntas se puede responder con cinco categorías de respuesta (5).

En la realización del cuestionario, tanto el piloto como el final, fue analizado con el mismo número de personas por cada uno de los tres grupos, para así evitar que alguno de ellos tuviera mayor influencia que los otros (5).

Un cuestionario con un resultado mayor indicará una mayor calidad de vida (13).

Por ejemplo, este cuestionario ha demostrado que pacientes que se quieren someter a cirugía refractiva obtienen peor resultado de calidad de vida que aquellos que prefieren continuar con gafas o lentes de contacto (13). En otro estudio se comprobó que la calidad de vida de operados por cirugía refractiva es mejor que la de usuarios de lentes de contacto, y la de estos mejor que la de usuarios de gafas. Además, hay diferencias significativas en el grupo de usuarios de gafas entre pacientes de alta y baja graduación, siendo los primeros los que tienen peores resultados (14).

#### **4.4 Análisis Rasch**

Para analizar los resultados tradicionalmente se ha usado la calificación Likert. Es el caso de los cuestionarios RSVP y NEI-RQL. La escala Likert asume que igual distancia entre las opciones de respuesta representa igual distancia en la dimensión medida. Además, asume que todas las preguntas representan la misma dificultad. Lo que realiza es una sumación de respuestas, por lo que si las preguntas no tienen la misma importancia, el resultado final no será preciso (5,11).

Rasch es un modelo de análisis, nombrado por el matemático danés Georg Rasch, creado a finales de 1950 (9,15).

El modelo Rasch asume que todas las categorías pertenecen a una misma escala. Transforma resultados ordinales a una escala lineal de intervalos que permite el uso de estadísticas paramétricas, y facilita la interpretación. Después realiza ponderaciones, calibrando de forma diferente a las categorías de respuesta y a las diferentes preguntas. Esto es debido a que las habilidades visuales varían en dificultad y tienen que ser calificadas acorde a ello. Además, este sistema detecta malas respuestas que puedan influir en el resto de datos (5,11,15,16).

Asume que la probabilidad de un encuestado respondiendo a un ítem es una función logística de la distancia relativa entre la localización del ítem (dificultad) y la localización del encuestado (capacidad) en una escala lineal. Se espera que la probabilidad de usar una categoría en particular aumentará proporcionalmente con la diferencia entre el nivel de la persona al realizar una actividad y el nivel de dificultad requerido para una tarea en particular. De esta forma, cuando la habilidad de una persona y la dificultad de un ítem están localizados, la respuesta se podrá estimar a partir de los datos disponibles (15,17,18).

La nueva escala se mide en logits, que son intervalos que representan igual cantidad de construcción medido a lo largo del rango completo. Como puede ser una unidad negativa, se suele escalar de 0 a 100, para facilitar su lectura (logits escalados). Esto permite que cambios en la medida de la función visual representen la misma cantidad de cambio en la construcción subyacente a lo largo del rango, lo cual no sucedía con la escala Likert (19).

Requiere que todos los ítems reflejen cantidades variables de la característica medida, y que exista una única característica subyacente o unidimensionalidad (17).

El resultado es un método con mayor precisión que la calificación Likert, que evita mejor el ruido, mejora la sensibilidad al cambio y las correlaciones con otras variables (11). Además,

mejora la consistencia interna del cuestionario y ajusta la dificultad de cada pregunta a la habilidad de la población (5).

Los cuestionarios que han utilizado el modelo antiguo, con la escala Likert, pueden ser mejorados realizando un análisis Rasch posterior, que mediante un proceso iterativo, elimina preguntas innecesarias y determina el número óptimo de categorías de respuesta en cada pregunta (5,11).

## 4.5 Validación de un cuestionario

Para que un cuestionario pueda ser utilizado debe cumplir una serie de criterios, que ayuden a conocer la exactitud de la medida (16). A esto se le llama validación de un test, y requiere de un análisis estadístico a partir de una muestra recogida para ello. Con esos datos se podrá comprobar que se adapta a los requisitos requeridos. Esta validación está dividida en diferentes tipos, según el criterio que se mida en cada caso. Aunque no siempre se calculan todos ellos, y no todos son cumplidos de la misma forma, cuanto mayor sea el número de ellos y mejores sean los resultados, más válido será el cuestionario creado. Además, aparte de válido, tiene que ser fiable y sensible, tanto a cambios en el tiempo como a diferencias entre grupos (7).

- Validación aparente: Un cuestionario debería medir lo que tiene intención de valorar. Se debe considerar si el concepto está bien medido, si la escala de medida parece razonable y si el nivel de comprensión es el adecuado. Esta validación no analiza los ítems incluidos ni si falta alguno, por lo que se puede considerar la forma más débil de validación (11).
- Unidimensionalidad: Es la demostración de que todos los ítems incluidos en el instrumento se ajustan a una única construcción subyacente, es decir, que mide la característica que nos interesa. No se debe confundir con los diferentes subconjuntos de un mismo dominio. Es un requisito para permitir la apropiada suma de un conjunto de ítems, y proporciona un importante valor de si se ha obtenido una medida significativa. Para obtenerla se utilizan métodos estadísticos como la estadística descriptiva convencional, Cronbach's  $\alpha$ , análisis de factores, análisis Rasch o PCA (análisis de componentes principales) (7,19,23,24).
- Validación de contenido: Es el grado en el que los ítems reflejan por completo el concepto que se mide. Es complicado probar que las preguntas elegidas sean las más representativas del total, aunque métodos como la identificación y reducción de ítems ayudan a este tipo de validación. Un factor crítico es la amplitud de contenido. Hay que llamar al instrumento conforme al concepto que se pretende medir. Por ejemplo, en nuestro caso, para medir la calidad de vida respecto a la visión, no sólo hay que tener en cuenta la limitación en ciertas actividades, sino otras dimensiones como la emocional o la económica. Todo esto se utiliza para justificar la elección de preguntas y las escalas de

respuesta. Para todo ello, se utilizan test de consistencia interna, como la correlación entre ítems, el coeficiente Cronbach's  $\alpha$  o el coeficiente Guttman de reproductibilidad (11,8).

- Validación de construcción: Se refiere a si un instrumento mide la construcción no observable, en este caso, la calidad de vida, que se pretende medir. No está muy clara la distinción entre ésta y la de contenido. Es usado para apoyar el argumento de que el resultado del instrumento realmente representa una cantidad de la variable de interés, es decir, si las expectativas teóricas del cuestionario se corresponden con los resultados empíricos. No es posible demostrarlo por un solo test, y hay varios tipos específicos de validaciones que contribuyen a ello: concurrente, convergente, predictiva y discriminante. Por ejemplo, en este caso el instrumento debería diferenciar entre tipos de error refractivo o entre tipos de corrección. Para medirlo se usan el análisis de componentes principales (validación factorial), e intercorrelaciones (validación convergente y divergente), comparando la distribución de los resultados. También es evidenciado cuando hay ausencia de DIF (Differential Item Functioning), usando el Kruskal-Wallis test o el coeficiente de correlación de Pearson (4,7,8,9,19,24,25).
- Validación de criterio: El instrumento se correlaciona con un estándar existente que mida lo mismo, aceptado y externo, comparando su habilidad para discriminar o predecir. Se usa para conocer si es lo suficientemente sensible y preciso como para ser una medida útil. El problema en este campo es que no existe un gold-standard para medir la calidad de vida ni la calidad de vida relacionada con la visión. Se realiza con análisis criterio-dependientes tradicionales de sensibilidad y especificidad, y con correlaciones con variables gold-standard (4,7,9,25).
- Validación convergente: Para correlacionar con algo que mide una estructura relacionada. Por ejemplo, en instrumentos de limitación de la actividad visual se suele correlacionar con la AV. También se demuestra con fuertes correlaciones de cada ítem con un dominio perteneciente a la medida. Los análisis estadísticos utilizados son el coeficiente de correlación de Pearson para variables continuas, y el análisis Chi cuadrado con coeficiente Phi para datos dicotómicos, o se puede valorar con análisis multimuestra. (4,7).
- Validación discriminante: Es el grado con el que un instrumento diverge de otros que deberían ser diferentes. Además, si un cuestionario tiene múltiples dominios, existe validez discriminante si cada ítem está más correlacionado con la medida en un dominio en particular que con otros dominios. Es poco utilizado, y se mide con el coeficiente de correlación de Pearson, o con análisis multimuestra (4,7).
- Validación predictiva: Determina si un instrumento puede hacer predicciones precisas con futuros resultados. Se mide mejor con un estudio prospectivo. Un test apropiado es el coeficiente de correlación de Pearson, y para respuestas dicotónicas el odd ratio o  $\chi^2$  (11).

- Validación concurrente: Explica la habilidad del instrumento para diferenciar grupos que debería ser capaz de distinguir, cuando ambos son medidos a la vez. Los resultados serán mejores cuando se discriminen grupos parecidos. Es la forma más sencilla para contribuir a la validación de construcción (11).
- Fiabilidad: Es la consistencia que tiene un instrumento al medir la misma construcción en diferentes momentos. Examina la proporción de varianza total que es atribuible a diferentes sujetos, incluyendo las verdaderas diferencias y el error de medida. Valora y minimiza estos errores de medida, que pueden ser al azar o sistemáticos. Es una medida importante, pero que no mide validez, ya que no asume que la construcción medida fuera la correcta. Se puede dividir en dos categorías:
  - Métodos de administración única: Incluye test de partición a la mitad y de consistencia interna, aunque a veces esto indica unidimensionalidad más que fiabilidad. Es alto cuando los ítems que constituyen la escala están relacionados entre ellos. Por ejemplo se usa el Cronbach's  $\alpha$ .
  - Métodos de administración múltiple o fiabilidad externa: Incluye el re-test, que detecta errores de medida al azar y parcialidad, formas alternativas (intermodo) y fiabilidad interobservador. Se puede calcular con el coeficiente de correlación producto-momento Pearson ( $r$ ), el coeficiente de correlación intraclass (ICC), los límites de acuerdo Bland-Altman (LoA), las estadísticas Kappa o el índice de concordancia Kendall's.

El análisis Rasch da límites de fiabilidad de separación de ítems y sujetos, que indica el rendimiento total del instrumento, gracias al índice de separación de personas (PSI) (7,8,9).

- Sensibilidad: Para saber si el instrumento puede detectar cambios clínicamente importantes a lo largo del tiempo, algo importante por ejemplo si se quiere valorar una intervención quirúrgica. Se puede medir lo que constituye un cambio importante gracias al MID (Minimum clinically Important Difference), o con análisis pareados (11,4).
- Interpretación: Indica si el grado en el que los resultados de una medida pueden considerarse significativos, asignando un significado cualitativo a los resultados de un instrumento cuantitativo. Para ello debe ser probado en una población representativa. Se valora mediante correlaciones entre la escala de resultados y condiciones clínicas reconocidas (6,7).

## 5. Hipótesis

Cuando se desarrolla un cuestionario en un idioma concreto, es necesario validarlo y valorarlo para conocer sus características y hasta qué punto cumple las expectativas que se quieren conseguir. Una vez cumplido este trámite, es posible utilizar dicho cuestionario para la práctica clínica e investigadora, pero sólo en dicho idioma, ya que es donde se ha comprobado su utilidad. Este es el caso del cuestionario QIRC, que fue creado y validado en Reino Unido en inglés. Sin embargo, la utilización en lengua española no será posible hasta que se conozca si cumple los requisitos necesarios.

Por lo tanto, la hipótesis de este estudio se basa en que, teniendo en cuenta la existencia del cuestionario validado en otro idioma, se puede comprobar que su uso también es válido en lengua española, extendiendo la posibilidad de su utilización en un gran número de pacientes.

## 6. Objetivos

El objetivo general de este estudio consiste en realizar las pruebas y el análisis necesario para validar el test QIRC sobre la repercusión que tiene la corrección refractiva de una persona no presbita en su calidad de vida, y pueda ser usado en español.

Para conseguirlo, es necesario cumplir una serie de objetivos específicos que nos permitan alcanzar dicho objetivo:

- Analizar los datos de una muestra recogida de pacientes distribuidos en tres grupos diferentes: usuarios de gafas, usuarios de lentes de contacto, y operados mediante cirugía refractiva.
- Realizar un análisis Rasch de los resultados del cuestionario, comprobando los diferentes tipos de validación existentes para conseguir que, en conjunto, el test sea considerado válido para su utilización en lengua española.
- Comprobar la fiabilidad del cuestionario, mediante la repetición del test en una parte de la muestra después de cierto número de semanas y el análisis de estos resultados, lo que nos proporcionará más información sobre sus características.

## 7. Materiales y métodos

Antes de realizar las pruebas, este trabajo fue revisado y admitido por la Comisión de Investigación del IOBA y por el Comité Ético de la Universidad de Valladolid, asegurando que se cumplen los requisitos requeridos (anexo 1).

Para este estudio prospectivo, se necesitó una muestra de pacientes, distribuida de forma parecida en tres grupos: usuarios de gafas, usuarios de lentes de contacto, y operados mediante cirugía refractiva.

Según los criterios de inclusión, los pacientes debían:

- Ser mayores de 18 años sin síntomas de presbicia.
- Con ametropía corregida con gafas o lentes de contacto, o sometidos a cirugía refractiva tras un tiempo de estabilización mínimo de 6 meses.
- Entender y hablar la lengua española.
- Aceptar la participación en el estudio mediante la firma del consentimiento informado (anexo 2).

Como criterios de exclusión, no podían participar en el estudio, aparte de los pacientes que no cumplan lo citado anteriormente, aquellos que:

- Padeciesen alguna patología ocular.
- Hubiesen sido sometidos a alguna otra cirugía ocular, que no fuera cirugía refractiva.

Para la participación en el estudio, se reclutaron pacientes voluntarios de la clínica del IOBA y otros voluntarios externos que cumplieren los criterios dados.

El material utilizado, al no tener ninguna prueba clínica, consistió únicamente en el cuestionario, que se puede encontrar en los anexos finales (anexo 3). La traducción del mismo fue realizada por un grupo de la Universidad Nacional Mayor de San Marcos, Lima (Perú), y mejorada en el IOBA para su próxima utilización en estudios del centro.

Se recogió además una serie de datos complementarios, que incluían:

- Edad.
- Sexo.
- Refracción actual de cada ojo.
- Tipo de corrección utilizado: gafas, lentes de contacto o cirugía refractiva.
- Tipo de cirugía realizada o tipo de lentes de contacto
- Tiempo desde la operación.
- Tiempo de la gafa actual y de las lentes de contacto actuales, refiriéndose al último cambio de graduación o tipo de lentes, no de recambio.
- Horas de uso de gafas y lentes de contacto, en el caso de uso de ambas o de uso parcial.

En la visita, una vez explicado el estudio y firmado el consentimiento, se procedió a contestar el cuestionario de 20 preguntas en español, con cinco categorías de respuesta y un apartado extra para los casos no aplicables. Para ello, se le fueron leyendo y explicando al paciente cada una de las preguntas y las respuestas. Se realizó de esta forma con motivo de asegurarse que todos los pacientes entendían de forma correcta lo que se les preguntaba, evitando así falsos resultados.

En el caso de uso tanto de gafas y lentes de contacto por parte de una misma persona, se le pidió al paciente que contestara el cuestionario respondiendo cada pregunta dos veces, una por cada tipo de corrección. En el momento de analizar, sólo se tuvo en cuenta una de las dos respuestas, por lo que se eligió el método que tuviera mayor número de horas de uso. Debido a que la mayor parte de los pacientes usaban a lo largo del día durante más horas las gafas, para poder igualar grupos se optó como usuarios de lentes de contacto los que hubieran referido mayor uso de las mismas hasta igualar ambos grupos en número.

En un subgrupo de los pacientes se repitió la realización del cuestionario para valorar la fiabilidad. En esos casos, se volvió a contactar con los sujetos y se les pasó de nuevo el cuestionario. Esta vez, la mayoría fue mediante vía telefónica, ya que su presencia ya no era necesaria y facilitaba el contacto, además de que todos ellos estaban familiarizados de la visita anterior con el cuestionario.

Para el análisis estadístico, se utilizó el programa R versión 3.3.1. (21/06/2016) [R Core Team (2016)].

En los datos de la muestra, se midió la media y su desviación estándar para variables cuantitativas, y porcentajes para las cualitativas.

A la hora de comparar grupos, en las variables cuantitativas se aplicó ANOVA. Se calculó la normalidad con el test de Shapiro-Wilk y la homogeneidad gracias al contraste de Levene. Si se asume normalidad, se utilizó el ANOVA de Welch para contrastar la hipótesis, y si el resultado era significativo, se realizó un análisis *post hoc* comparando los grupos dos a dos con el contraste t-Student. Si no se asume normalidad, se utilizó la alternativa no paramétrica, el ANOVA de Kruskal-Wallis, y para las comparaciones dos a dos en los casos que sean necesarios se aplicó el contraste U de Mann-Whitney. Para las variables no cuantitativas se aplicó el test chi-cuadrado o el test exacto de Fisher.

El análisis Rasch se enmarca dentro de una familia de métodos conocida como Teoría de Respuesta al Item (Item Response Theory, IRT). Para aplicarlo se utilizó el *package eRM* [Mair & Hatzinger, 2007] [Mair et al, 2015] y el *package TAM* [Kiefer et al, 2016] de R.

Para conocer el resultado del total de los pacientes y de cada uno de los subgrupos, se aplicó, según el análisis Rasch, una ponderación para cada una de las respuestas y cada una de las preguntas. La tabla de ponderaciones está incluida en el anexo 4. De esta forma, se consigue que cada pregunta se valore según su dificultad, y permite agrupar respuestas con un mismo resultado cuando se ha comprobado que reduciendo el número de categorías mejora el cuestionario. Para obtener el resultado de un paciente se suma lo obtenido en cada una de las preguntas según la tabla, y se divide entre el número de preguntas respondidas (se eliminan las de categoría 0: No sabe/ No es aplicable). La transformación y obtención de resultados se puede realizar gracias a una tabla de Excel dada por el creador del cuestionario, Pesudovs, disponible en su página web: <http://www.pesudovs.com/konrad/questionnaire.html>.

Para saber si la muestra se ajustaba a un modelo, se realizó una prueba de bondad de ajuste, en este caso el test de Andersen LR. El resultado ideal de los residuos es de  $0 \pm 1$ . Los residuos o residual es la diferencia entre la respuesta esperada por el modelo y la respuesta obtenida (16).

Utilizando el modelo de crédito parcial (PCM), se calcularon los umbrales de cada ítem, es decir, la localización del paso de una categoría a otra, donde un individuo tiene la misma probabilidad de contestar en una categoría o la siguiente. Se supone que un individuo con mayor habilidad caerá en el umbral 2 superando el umbral 1.

Para recalculer las categorías se observó la curva característica de cada ítem (ICC). Esta curva representa la relación entre la probabilidad de acertar un ítem y los valores de la variable latente. Cada curva corresponde con una categoría. Interesa la relevancia de cada curva (el pico de la curva no sea tapado por otro) y su orden dentro del intervalo -1 y 1, y estas características nos indicarán como combinar las categorías.

Para eliminar ítems redundantes se utilizaron dos estadísticos que comparan los datos con los valores esperados del modelo, ayudando a la identificación de los ítems que contribuyen más a la medida de la característica latente.

Se trata del infit (ajuste de la información ponderada) mean square, que nos proporciona información del patrón de respuestas de individuos cuya habilidad está cerca de la dificultad del ítem. El valor esperado es de 1, y se acepta un rango entre 0,80 y 1,20. Valores menores representan preguntas que se sobreajustan al modelo y son muy predecibles y redundantes; y valores mayores representan desajustes, es decir, que el ítem mide algo diferente a la escala total (5,21,23).

Por otro lado encuentra el outfit (ajuste sensible de valores atípicos) mean square, que es sensible a respuestas diferentes a lo esperado y informa de las diferencias entre los valores observados y los esperados para ítems que están lejos de la habilidad de la persona. El valor esperado es 1, y se acepta un rango entre 0,70 y 1,30. Valores menores representan preguntas que se sobreajustan al modelo y son muy predecibles y redundantes; y valores mayores representan desajustes, es decir, que el ítem mide algo diferente a la escala total (5,21).

Una vez ajustado, se debía valorar si la unidimensionalidad del cuestionario, es decir, si mide una única característica subyacente. Para ello se aplicó el análisis de componentes principales (PCA). Se miden los residuos que quedan cuando el factor Rasch o la característica subyacente ha sido suprimida, intentando que no haya ninguna estructura en ellos. Existe unidimensionalidad si los valores son menores de 2.

Otro método de evaluar la unidimensionalidad es a través del coeficiente Cronbach's  $\alpha$ . Sus valores óptimos estarían entre 0,7 y 0,9 (15).

Para representar los resultados, se utilizó el mapa persona-ítem. Este mapa muestra los niveles de habilidad de cada persona, en la parte superior, de izquierda a derecha de menor a mayor habilidad, según el resultado total en el cuestionario. En la parte de abajo la dificultad de los ítems, a la izquierda los más sencillos y a la derecha los más complicados, mostrando además los umbrales de cada pregunta. De esta forma, se puede ver si hay muchos ítems que representen la misma dificultad, o si cubren todos los niveles de habilidad de los individuos. El mapa puede estar medido en logits, con valores positivos y negativos, o logits escalados de 0 a 100 para

facilitar su lectura. Si el cuestionario es bueno y los ítems están bien orientados a los sujetos, la media de los ítems y de las personas está cerca (5,20).

Para medir la fiabilidad del cuestionario a través del re-test se comparó el resultado de los dos pases en cada ítem, usando el contraste de Wilcoxon, que contrasta la hipótesis de igualdad de medianas para dos muestras pareadas.

Se midió la diferencia entre los dos pases, y, para conocer si los datos seguían una distribución normal, se midió la curtosis y la asimetría u oblicuidad. Ambos valores debían encontrarse entre -2,00 y +2,00, con mejor resultado cuanto más se acercaran dichos valores a 0. La normalidad se midió con el test Shapiro-Wilk, y gracias a un histograma y un QQ-plot se pudo comprobar estos resultados de forma gráfica.

Se calculó la desviación estándar de la media de cada individuo y el coeficiente de variación intra-sujeto, para conocer el porcentaje de variación respecto a la media. Cuanto más pequeños sean estos valores, mejor serán los resultados.

Para comparar las diferencias entre las dos mediciones respecto a su media se utilizó el análisis de Bland-Altman. Se utilizan los límites de acuerdo (LoA), que es el rango de valores donde el 95% de las diferencias entre las dos medidas pueden caer.

Finalmente se midió el coeficiente de correlación intraclass (ICC). Este coeficiente nos indica la proporción de la variabilidad total debida a la variabilidad de los sujetos. Su rango es entre 0 y 1. Un valor alto indica que la variabilidad se debe a las diferencias entre sujetos, y un valor bajo indicaría respuestas al azar (11).

## 8. Resultados

### 8.1 Datos de la muestra

El tamaño final de la muestra fue de 91 pacientes, divididos en: 33 usuarios de gafas, 32 de lentes de contacto y 26 de cirugía refractiva. En la tabla 1 se puede observar algunos de los datos sobre la muestra. La media de edad total fue de  $28,91 \pm 7,75$  años, con un rango desde los 19,05 a los 50,27 años. Se observó que el grupo de cirugía refractiva tiene una media de edad mayor al de los otros dos grupos, con una diferencia estadísticamente significativa respecto a los otros dos grupos ( $p=0,0025$  con el grupo de usuarios de gafas, y  $p<0,0001$  respecto al de lentes de contacto). No hubo diferencias significativas entre el grupo de gafas y el de lentes de contacto ( $p=0,172$ ).

En cuanto al género, el 51,1% de la muestra total fueron mujeres. Aunque el grupo de lentes de contacto fue el que contiene mayores diferencias (61,3% de mujeres), el análisis chi cuadrado indicó que los tres grupos eran parecidos ( $p=0,3746$ ).

El error refractivo de los operados con cirugía refractiva era nulo. En el resto de pacientes, la refracción de la esfera abarcaba desde  $-13,25D$  hasta  $+5,75D$  con una media de  $-2,59 \pm 2,63D$  en el ojo derecho, y desde  $-14,00D$  hasta  $+2,25D$  en el ojo izquierdo, con una media de  $-2,75 \pm 2,57D$ . No hubo diferencias significativas entre el grupo de gafa y el de lentes de contacto en ninguno de los ojos ( $p=0,159$  y  $p=0,2772$ , respectivamente).

**Tabla 1.** Datos de la muestra.

Muestra		Gafas	LC	CR	Total
	N (%)	33 (36,26%)	32 (36,16%)	26 (28,57%)	91 (100%)
Edad:	Media±DE	27,7±6,65	25,07±5,29	34,84±8,12	28,91±7,75
	Rango	19,05-45,83	19,36-43,03	20,64-50,27	19,05-50,27
Género:	Mujer n (%)	15 (45,5%)	19 (61,3%)	12 (46,2%)	46 (51,1%)
	Varón (%)	18 (54,5%)	12 (38,7%)	14 (53,8%)	44 (48,9%)
Esfera OD:	Media±DE	-2,38D±3,11	-2,9D±2,03	-	-2,59D±2,63
	Rango	-13,25/+5,75	-7,00/+3,75	-	-13,25/+5,75
Esfera OI:	Media±DE	-2,65D±2,99	-2,96D±2,09	-	-2,75D±2,57
	Rango	-14,00/+2,25	-8,75/+1,00	-	-14,00/+2,25

Sólo 30 pacientes tuvieron componente cilíndrica en el error refractivo. El rango fue de -0,25D a -2,25D en el ojo derecho, con una media de  $-1,07 \pm 0,72$ D, y de -0,75D a -3,75D en el ojo izquierdo, con una media de  $-1,16 \pm 0,89$ D. En este caso tampoco hubo diferencias significativas entre el grupo de gafas y el de lentes de contacto ( $p=0,2185$  en ojo derecho y  $p=0,0787$  en el ojo izquierdo).

Dentro del grupo de cirugía, la mitad de ellos (13 pacientes) fue operado mediante la técnica LASEK, 8 mediante LASIK y 5 de ellos mediante el implante de ICL (lente fáquica). El tiempo pasado desde la operación transcurre desde los 6 meses (mínimo indicado en los criterios de inclusión, para asegurar que el paciente ha alcanzado la estabilidad en cuanto a su visión) hasta los 18 años.

Entre los usuarios de lentes de contacto (31 pacientes), la opción elegida más frecuentemente fueron las lentes blandas hidrofílicas mensuales (54,8%), seguido de las blandas de recambio diario (22,6%), lentes mensuales de hidrogel de silicona (9,7%), y finalmente un usuario de hidrogel de silicona diarias, uno de rígidas de ortoqueratología, uno de RPG y otro de RPG bianual (3,2% cada uno). Su uso aproximado en número de horas a lo largo del día iba desde las 4 horas hasta las 13, con una media de  $7,71 \pm 2,37$  horas.

## **8.2 Resultados del cuestionario**

En la tabla 2 se recogen los resultados para cada uno de los ítems y del global en los 91 pacientes y también en cada uno de los subgrupos. El resultado está escalado de 0 a 100, y cuanto mayor sea, mejor será la calidad de vida.

La media total del cuestionario fue de  $51,88 \pm 7,05$ . El grupo con mejor resultado fue el de cirugía refractiva, con  $56,95 \pm 5,9$ . Los grupos de gafas y lentes de contacto fueron muy parecidos, con  $49,72 \pm 7,05$  y  $49,99 \pm 5,88$ , respectivamente.

**Tabla 2.** Resultado del cuestionario QIRC en la muestra analizada, media±desviación estándar.

Ítem	Gafas	LC	CR	Total
1	45,63±13,19	44,32±13,36	46,35±14,35	45,49±13,47
2	50,6±12,18	49,18±9,99	52,63±10,72	50,68±11
3	33,87±13,87	55,68±3,92	56,71	49,67±12,81
4	49,2±13,8	46,4±14,41	59,59±5,03	51,18±13,27
5	50,42±12,23	45,93±13,9	58,13±4,2	51,16±12,12
6	43,8±12,5	39,01±11,05	62,68±6,18	47,43±14,32
7	35,39±13,02	52,51±8,33	-	44,58±13,7
8	52,06±13,27	46,75±13,08	55,7±11,7	51,22±13,14
9	46,11±13,34	46,62±13,79	47,55±12,88	46,7±13,24
10	42,52±11,67	45,03±13,46	58,38±12,87	47,73±14,16
11	44,07±12,7	42,93±12,37	49,69±14,79	45,23±13,35
12	49,19±13,75	42,11±12,25	42,85±13,77	44,89±13,5
13	41,81±10,17	39,58±7,85	37,57±6,79	39,84±8,6
14	57,54±20,92	62,44±17,47	69,88±13,81	62,91±18,32
15	62,95±19,37	57,51±18,11	70,08±15,2	63±18,28
16	69,42±17,06	63,27±16,05	71,72±17,63	67,85±17,05
17	57,84±18,84	52,14±15,2	58,36±15,45	55,92±16,66
18	54,33±16,8	55,32±17,46	57,38±18,52	55,57±17,39
19	43,2±20,22	52,82±15,13	60,73±11,29	51,68±17,56
20	62,45±17,41	60,89±15,45	68,63±11,97	63,66±15,5
Total	49,72±7,05	49,99±5,88	56,95±5,9	51,88±7,05

Una vez obtenidos los resultados, hay que compararlos para conocer si existen diferencias entre los grupos en cada ítem o no. En la tabla 3 hay un resumen de las comparaciones. Nueve de los ítems mostraron diferencias significativas ( $p < 0,05$ ) entre alguno de los grupos, y dos de ellos próximos a la significación ( $p < 0,1$ ;  $> 0,05$ ).

El uso de gafas sin graduación (pregunta 3) tuvo un peor resultado estadísticamente significativo en el grupo de usuarios de gafas respecto a los otros dos. En las preguntas 4 (preocupación en la preparación antes de alguna actividad), 5 (dificultad para ver por la noche), 6 (dificultad para ver nadando o en la playa) y 10 (preocupación por la dependencia de algún tipo de corrección), el grupo de cirugía refractiva obtuvo resultados significativamente mejores que los otros dos. La pregunta 7 (problema que generan gafas o lentes de contacto al realizar ejercicio) fue asumida como no aplicable para el grupo de cirugía, y mostró diferencias significativas de las gafas respecto a las lentes de contacto, con mejor resultado en el segundo método de compensación.

**Tabla 3.** Análisis estadístico de cada uno de los ítems.

	p-valor K-W <sup>1</sup>	CR-Gafas <sup>2</sup>	CR-LC <sup>2</sup>	Gafas-LC <sup>2</sup>
1	0,8751	1	1	1
2	0,4824	1	0,6127	1
3	<b>&lt;0,0001</b>	<b>&lt;0,0001</b>	0,4824	<b>&lt;0,0001</b>
4	<b>0,0006</b>	<b>0,0041</b>	<b>0,0005</b>	1
5	<b>0,0008</b>	<b>0,0138</b>	<b>0,0006</b>	0,5797
6	<b>&lt;0,0001</b>	<b>&lt;0,0001</b>	<b>&lt;0,0001</b>	0,2465
7	-	-	-	<b>&lt;0,0001</b>
8	<b>0,0333</b>	0,9123	<b>0,0314</b>	0,3377
9	0,7452	1	1	1
10	<b>0,0001</b>	<b>0,0002</b>	<b>0,0024</b>	1
11	0,2329	0,4481	0,2438	1
12	<i>0,0638</i>	0,2311	1	<i>0,0806</i>
13	0,129	0,1395	0,5474	1
14	<i>0,0574</i>	<i>0,0621</i>	0,2877	1
15	<b>0,0352</b>	0,5721	<b>0,0262</b>	0,7386
16	0,7797	1	0,2124	0,5234
17	0,2169	1	0,4175	0,4051
18	0,7152	1	1	1
19	<b>0,0007</b>	<b>0,0006</b>	<i>0,0599</i>	0,17
20	0,1276	0,4973	0,1248	1
Total	<b>&lt;0,0001<sup>3</sup></b>	<b>0,0001<sup>4</sup></b>	<b>0,0002<sup>4</sup></b>	1 <sup>4</sup>

En negrita si es estadísticamente significativo ( $p < 0,05$ ), y en cursiva si está próximo a la significación ( $p < 0,1$ ;  $> 0,05$ ). <sup>1</sup> Prueba de Kruskal-Wallis para variables no paramétricas. <sup>2</sup> Comparaciones entre grupos dos a dos con el contraste U de Mann-Whitney. <sup>3</sup> ANOVA de Welch para variables paramétricas. <sup>4</sup> Comparaciones entre grupos dos a dos con el contraste t-Student.

Para la pregunta 8 (preocupación por el gasto en financiación) y la 15 (frecuencia de ser visto bien por otra persona) el grupo de cirugía refractiva tuvo mejores resultados significativos que el de lentes de contacto. En la pregunta 19 (frecuencia de poder hacer lo que se quiere hacer), fue significativa la diferencia entre el grupo de cirugía, con mejor resultado, que el de gafas, y próximo a la significación entre el grupo de cirugía y lentes de contacto ( $p = 0,0599$ ).

Cercano a ser estadísticamente significativo ( $p < 0,1$ ;  $> 0,05$ ) se encontró el ítem 12 (preocupación por las complicaciones médicas) entre el grupo de gafas y lentes de contacto ( $p = 0,0806$ ), y el ítem 14 (frecuencia de sentir buena apariencia) entre el grupo de cirugía y los usuarios de gafas ( $p = 0,0621$ ).

Finalmente, en el resultado global de los 20 ítems, se mostraron diferencias estadísticamente significativas entre el grupo de cirugía refractiva respecto al de gafas ( $p = 0,0001$ ) y de lentes de contacto ( $p = 0,0002$ ). Entre estos dos últimos no se observaron diferencias significativas.

### 8.3 Análisis Rasch

Para realizar el análisis Rasch, primero hubo que transformar las categorías de respuesta para igualarlas, ya que en los ítems 1-13, la categoría 1 indicaba la mejor calidad de vida y la 5 la peor, y en los ítems 14-20 era al revés. Por ello, se realizó una inversión en los primeros ítems de forma que mayor categoría indicara mejor calidad de vida (1=5, 2=3, 3=3, 4=2, 5=1).

Como las categorías 1 y 2 fueron elegidas por pocos pacientes, se decidió unir las en una sola. Además, al analizar los resultados, por motivos del programa, se necesitaba empezar por la categoría 0, con lo cual las categorías finales se distribuyeron de la siguiente manera: 0=1/2, 1=3, 2=4 y 3=5, considerando la conversión anterior.

El resultado inicial del test Andersen LR no fue bueno, con un resultado significativo ( $p < 0,0001$ ). La media  $\pm$  desviación estándar de los residuos fue de  $-0,86 \pm 1,43$  para los ítems y  $-0,37 \pm 1,15$  para los individuos, lejos del resultado ideal.

Se utilizó el modelo de crédito parcial (PCM) para calcular los umbrales. En este caso, al existir 4 categorías, se obtuvieron 3 umbrales:

**Tabla 4.** Localización y umbrales de los ítems.

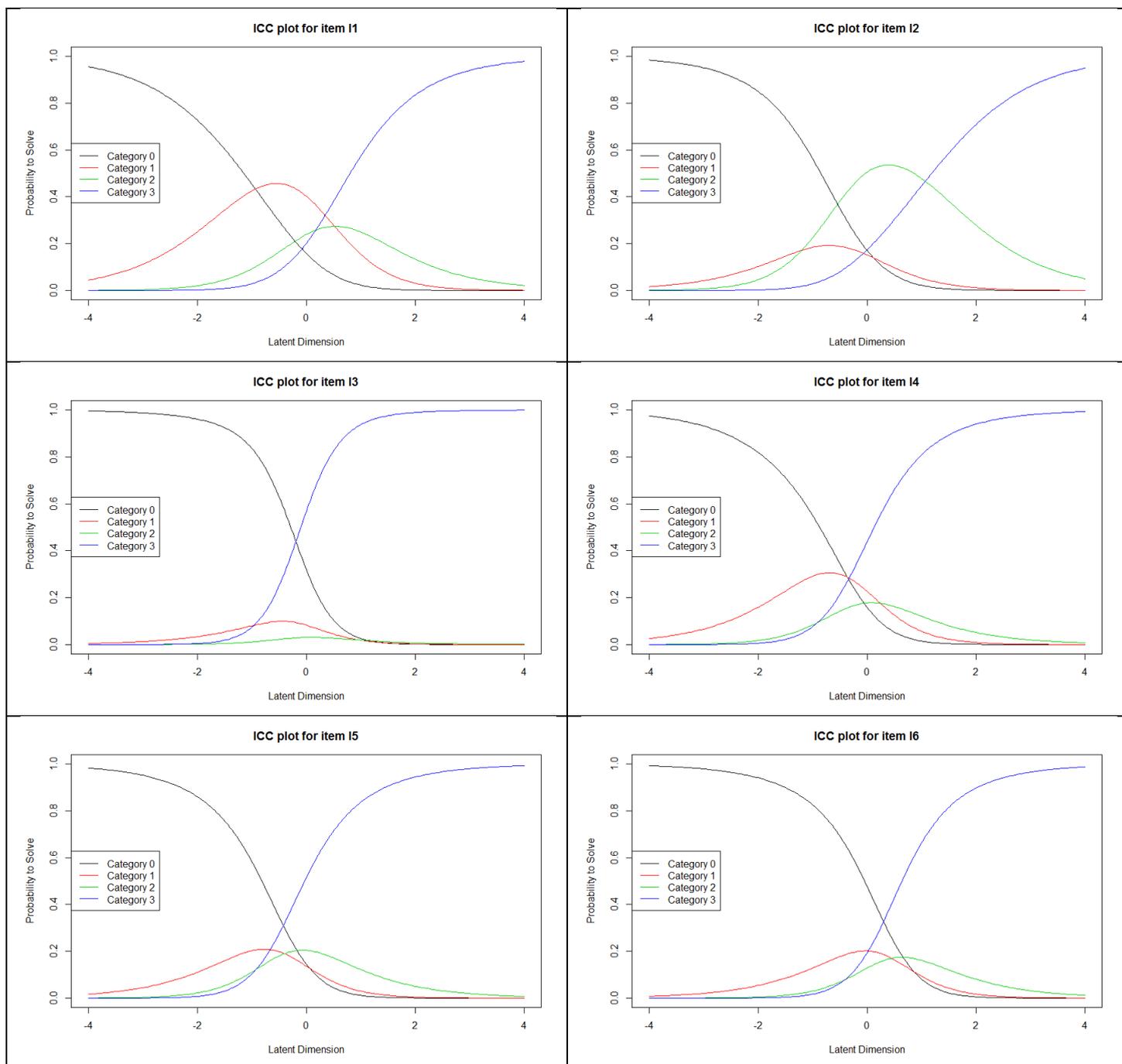
	Localización	Umbral 1	Umbral 2	Umbral 3
1*	-0,07974	-0,93273	0,52353	0,16998
2*	-0,00385	0,11995	-1,20103	1,06954
3	-0,19045	1,35287	0,99898	-2,92320
4*	-0,33804	-0,36016	0,24823	-0,90219
5	-0,42132	0,05900	-0,39980	-0,92315
6	0,30259	0,86361	0,45397	-0,40981
7	-0,29261	0,40627	-0,02747	-1,25663
8*	-0,06684	0,04569	-0,25061	0,00440
9*	0,00751	0,20229	-0,23306	0,05330
10*	0,33305	0,56806	0,70607	-0,27497
11*	0,27881	-0,46754	1,10581	0,19816
12*	0,14937	0,80095	-0,36727	0,01143
13	1,21172	0,34887	1,46547	1,82084
14*	-0,15608	-0,26310	0,11577	-0,32091
15*	0,00950	-0,36297	0,27580	0,11565
16*	-0,07379	-1,13432	0,71750	0,19546
17	-0,02407	-0,87978	0,25366	0,55389
18	-0,11229	-0,46298	-0,06354	0,18966
19*	-0,30192	-0,28007	-0,06594	-0,55976
20*	-0,61463	-1,19802	0,04431	-0,69018

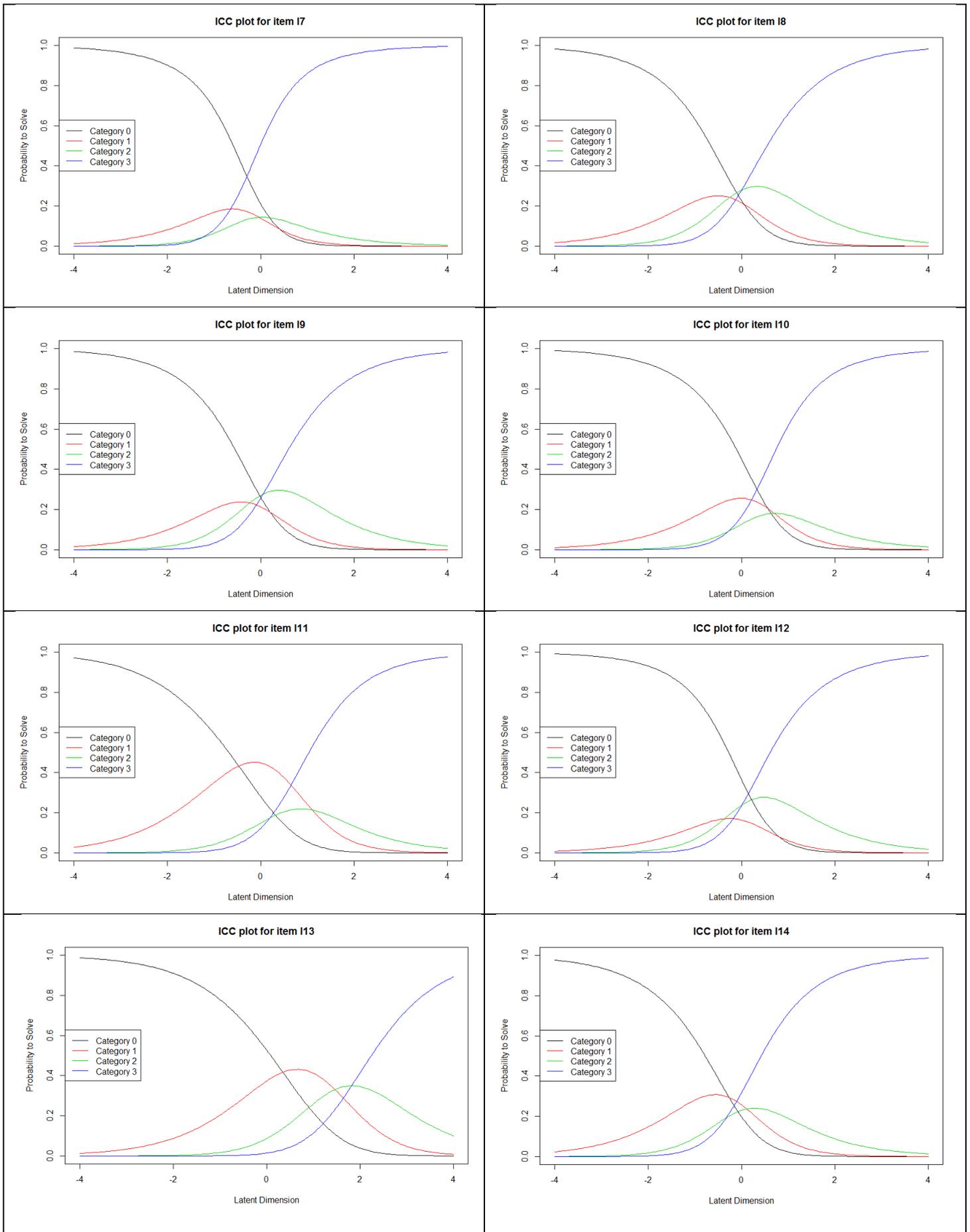
\*Ítems con elementos desordenados.

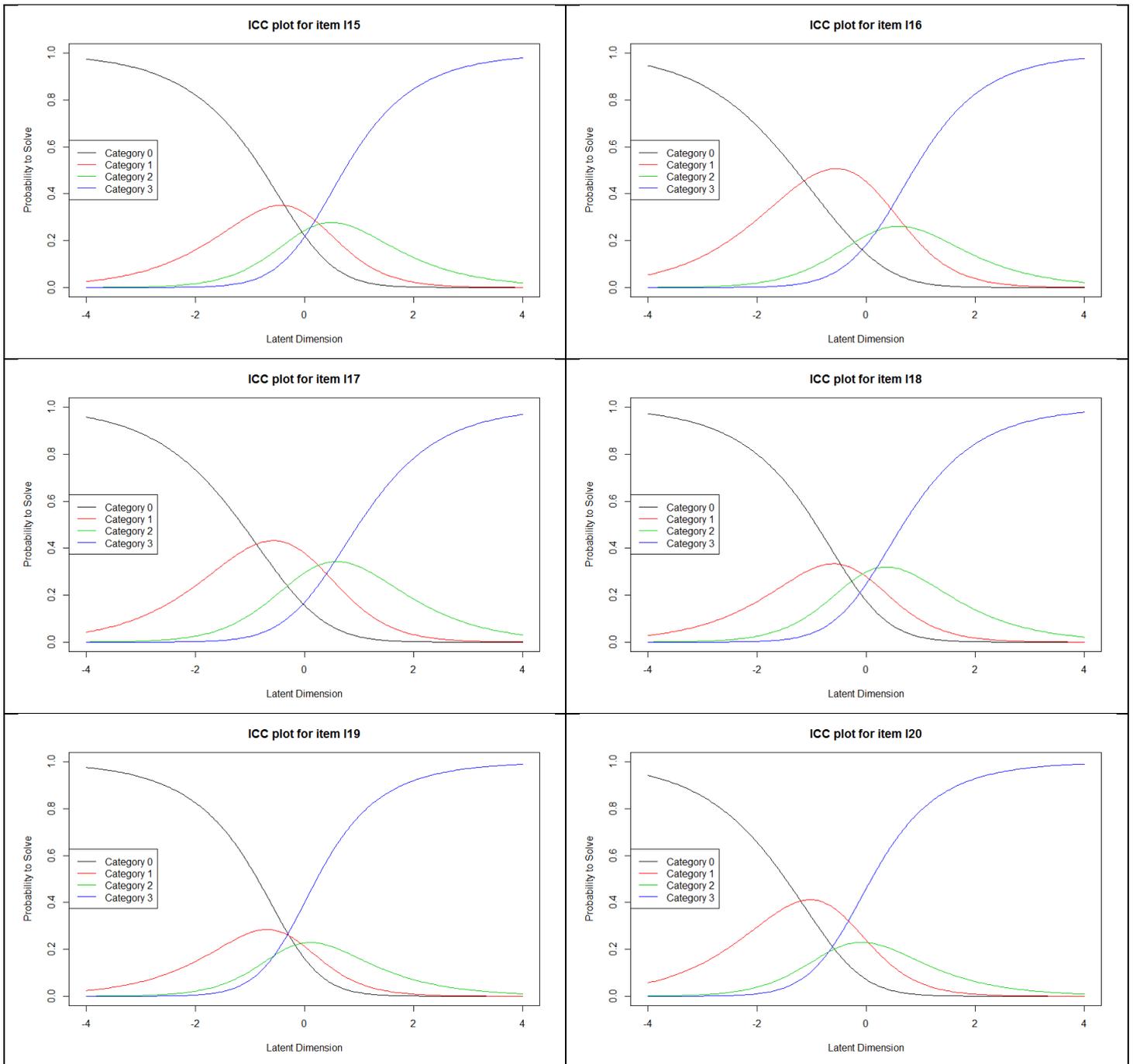
Al comprobar los umbrales de los ítems, se pudo observar que la mayoría estaban desordenados, y los únicos que siguieron el orden normal fueron el 3, 5, 6, 7, 13, 17 y 18. Cuando existe este desorden indica que categorías supuestamente más fáciles se relacionan con un nivel de habilidad mayor, o al revés. Hay varios motivos por lo que puede ocurrir esto, como que haya

demasiadas categorías de respuesta, que una categoría tenga un gran rango de habilidad, o que las opciones sean similares y haya problemas al contestarlas.

El primer cambio realizable para mejorar los umbrales y el ajuste fue agrupar categorías, tal como se hizo anteriormente con la 1 y 2. Para ello se debía observar la curva característica de cada ítem (ICC).







**Gráfico 1.** Curvas características de los ítems 1-20.

Se mantuvieron iguales los ítems 13, 17 y 18, cuyas curvas eran correctas. Se combinaron las categorías 0 y 1 en los ítems 12 (categoría 0 desordenada), 2, 8 y 9 (categoría 1 desordenada), 14 (categorías 0 y 1 superpuestas). Combinación de las categorías 0, 1 y 2 en los ítems 3, 5, 6 y 7 (las dos categorías centrales, 1 y 2, no aportan al modelo). Se unieron las categorías 2 y 3 en los ítems 11 (categoría 2 desordenada), 1, 10, 14, 15, 16, 19 y 20 (categoría 3 desordenada).

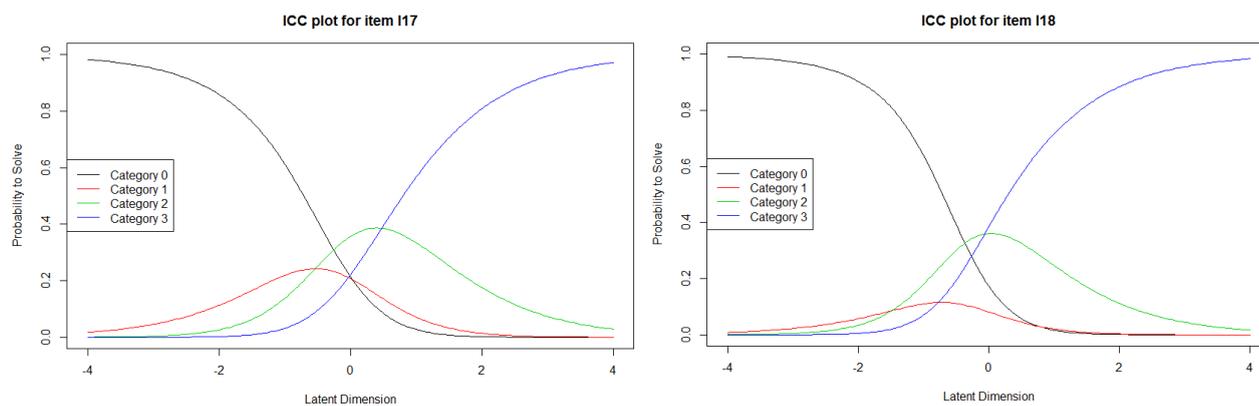
Se repitieron los pasos anteriores para comprobar cómo quedaban los umbrales después de las combinaciones realizadas:

**Tabla 5.** Localización y umbrales de los ítems tras ser reducidas las categorías.

	Localización	Umbral 1	Umbral 2	Umbral 3
1	-0,65530	-1,10105	-0,20956	-
2	0,31023	-0,62826	1,24872	-
3	-0,83808	-0,83808	-	-
4	-0,10029	0,72293	-0,92350	-
5	-0,34477	-0,34477	-	-
6	0,87136	0,87136	-	-
7	-0,36620	-0,36620	-	-
8	0,06110	1,31616	-1,19396	-
9	0,06311	1,21126	-1,08504	-
10	-0,29641	1,43179	-2,02462	-
11	-0,13475	0,16773	-0,43723	-
12	-0,03726	1,07359	-1,14810	-
13	1,10205	0,58948	0,88051	1,83616
14	-0,90558	2,62460	-4,43577	-
15	-0,95016	-0,21986	-1,68047	-
16	-0,70811	-0,67383	-0,74239	-
17*	-0,01115	0,02959	-0,53839	0,47536
18*	-0,26372	0,76001	-1,48833	-0,06283
19	-1,07214	0,63875	-2,78302	-
20	-0,70496	0,28143	-1,69135	-

\*Ítems con elementos desordenados.

En los ítems 17 y 18 los umbrales continuaron desordenados, y con el ICC se comprobó que la categoría 1 estaba desordenada, por lo que se unieron las categorías 0 y 1.



**Gráfico 2.** Curvas características de los ítems 17 y 18.

**Tabla 6.** Localización y umbrales de los ítems 17 y 18 tras ser reducidos.

	Localización	Umbral1	Umbral 2	Umbral 3
17	0,12758	0,17370	0,08147	-
18	0,12270	1,02828	-0,78287	-

Se volvió a realizar el test de Andersen LR, pero el resultado siguió sin ajustarse al modelo ( $p < 0,0001$ ), aunque la media de los residuos (desviación estándar) se acercaba más al ideal, con  $-0,25 \pm 1,31$  para los ítems, y  $-0,11 \pm 1,11$  para los individuos.

Como aún no se ajustaba a lo esperado, se procedió a eliminar ítems, observando el infit y outfit de cada uno de ellos.

El resultado total del infit fue de  $0,90 \pm 0,18$ , y el del outfit de  $0,88 \pm 0,27$ . En la tabla 7 se puede observar el resultado en cada uno de los ítems. Con los resultados obtenidos se comprueba que no son los esperados, tanto por encima como por debajo del margen otorgado. Se realizó un proceso iterativo eliminando primero el ítem más desajustado, volviendo a calcular los datos y eliminando el siguiente peor, y así sucesivamente hasta que los ítems restantes tuvieran unos valores de infit y outfit que se ajustasen al modelo.

**Tabla 7.** Infit y outfit de cada uno de los ítems.

	Infit MSQ	Outfit MSQ
1	0,998	0,917
2	1,101	1,097
3	0,953	0,915
4	0,949	0,869
5	0,898	0,844
6	0,928	0,904
7	1,043	1,043
8	<b>0,731</b>	<b>0,690</b>
9	0,819	0,814
10	<b>0,780</b>	1,001
11	<b>0,798</b>	0,774
12	1,003	1,067
13	<b>1,357</b>	<b>1,667</b>
14	<b>0,776</b>	<b>0,523</b>
15	0,959	0,752
16	<b>0,736</b>	<b>0,673</b>
17	0,999	1,105
18	0,987	1,015
19	<b>0,684</b>	<b>0,551</b>
20	<b>0,513</b>	<b>0,386</b>

En negrita se destacan los resultados que se salen del ajuste esperado.

De esta forma, se eliminaron uno a uno los ítems número 13, 20, 19, 14, 15, 8, 16, 11 y 10, en este orden. Una vez acabado este proceso, todos los ítems cumplieron los criterios, como se puede ver en la tabla 8. El infit total fue de  $0,94 \pm 0,06$ , y el outfit de  $0,93 \pm 0,08$ .

**Tabla 8.** Infit y outfit de cada uno de los ítems restantes.

	Infit MSQ	Outfit MSQ
1	0,961	0,853
2	0,984	0,983
3	0,958	0,879
4	0,899	0,818
5	0,945	0,870
6	0,866	0,848
7	1,042	1,024
8	-	-
9	0,849	0,914
10	-	-
11	-	-
12	0,935	0,993
13	-	-
14	-	-
15	-	-
16	-	-
17	0,897	0,969
18	1,001	1,039
19	-	-
20	-	-

Si se volvía a realizar el test de Andersen LR, el resultado siguió siendo insuficiente ( $p=0,012$ ). Para mejorarlo, se eliminó a uno de los individuos, del grupo de lentes de contacto, ya que su patrón de respuestas era raro, con residuo  $>2,5$  (desajuste con el modelo). El p valor mejoró hasta  $p=0,021$ . Como siguió siendo un mal resultado, se decidió eliminar el ítem número 7, ya que era el que más respuestas tenía como no aplicable (categoría 0 en el test inicial). Ahora el resultado del test era  $p=0,226$ , y al ser no significativo indicó que ahora estaba bien ajustado al modelo. El infit final fue de  $0,93\pm 0,05$ , y el outfit de  $0,92\pm 0,07$ .

Al estar ya ajustado, se midió la unidimensionalidad, con el análisis de componentes principales (PCA). Como se puede ver en la tabla 9, se cumplió la unidimensionalidad en todos los valores.

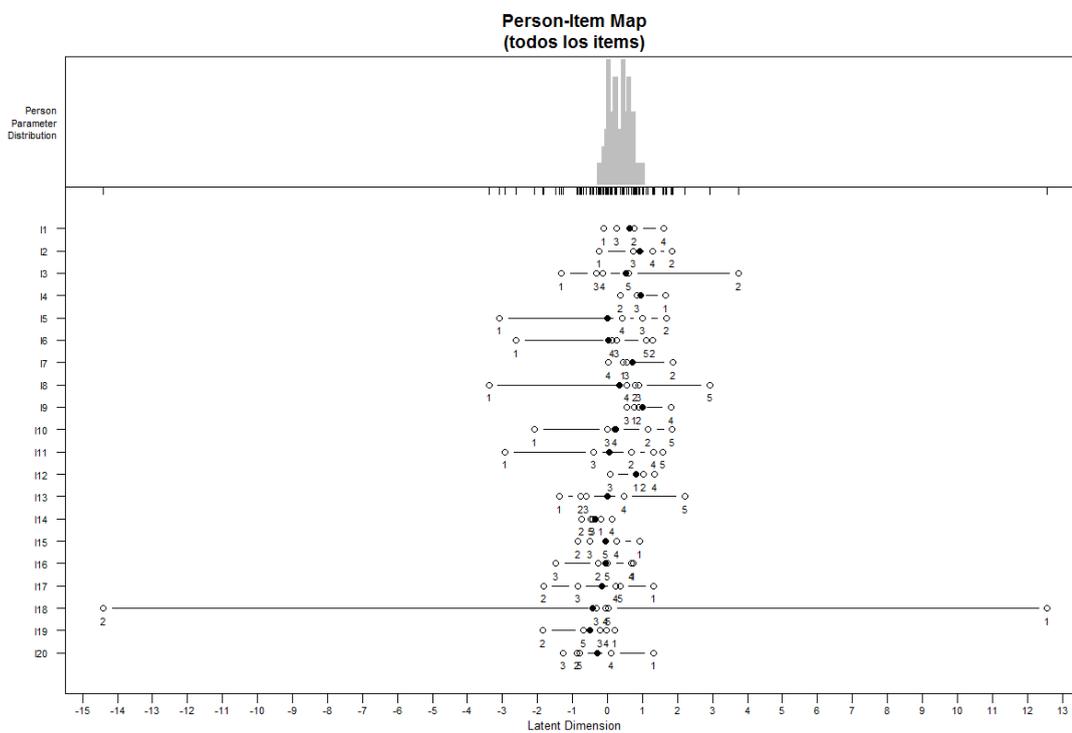
**Tabla 9.** PCA de los ítems finales.

Factores	1	2	3	4	5	6	7	8	9	10
Residuos	1,47	1,26	1,15	1,12	1,03	0,97	0,85	0,78	0,57	0,13

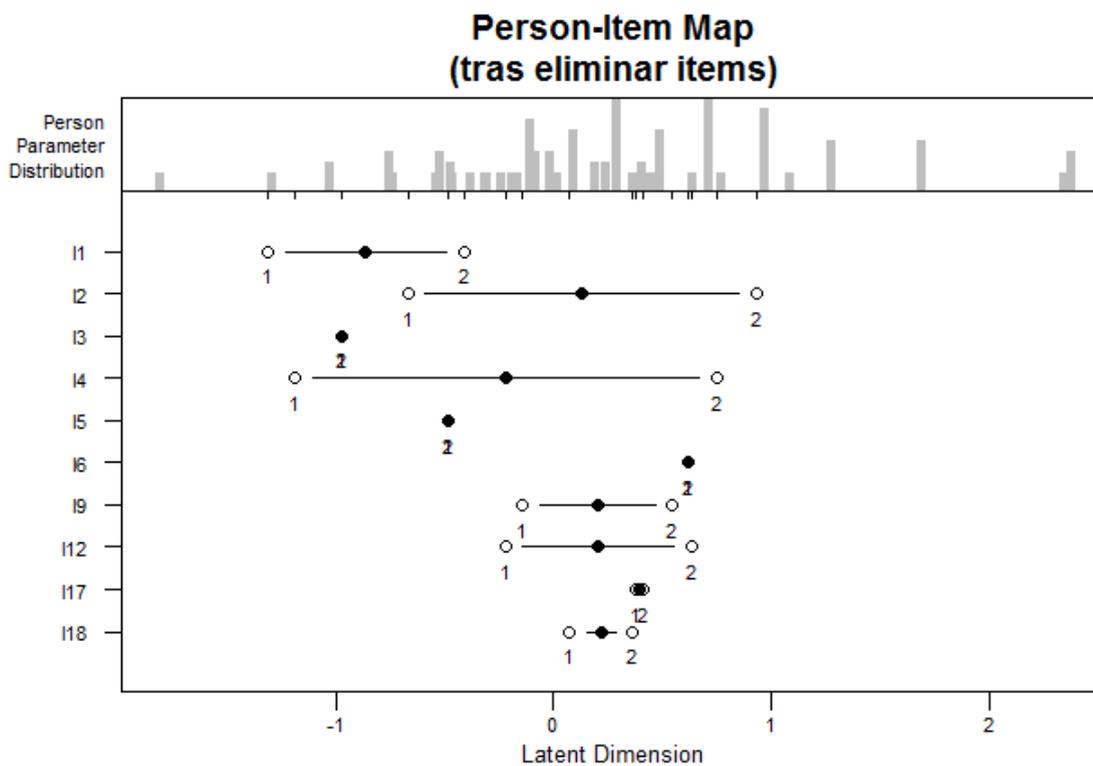
El valor inicial del coeficiente Cronbach's  $\alpha$  fue de 0,643, y el del cuestionario reducido de 0,68. Aunque mejoró el resultado, no llegó a ser del todo bueno, entre 0,7 y 0,9.

Se observó gráficamente la puntuación de los ítems y los pacientes con el mapa persona-ítem. En el gráfico 3 se puede observar el mapa inicial, y en el 4 el mapa final tras la reducción de ítems. Comparando los dos mapas, se observó como en el segundo mejoraba la distribución. En el segundo mapa se pudo comprobar de forma gráfica como el cuestionario en general fue sencillo para la habilidad de la muestra. Los ítems 1 y 3 fueron los más fáciles, y el 6 el más complicado. La media de los ítems se encontraba en 0 (50 si la escala fuera de 0 a 100). Para los encuestados, se obtuvo una media de  $0,33\pm 0,8$ , con resultados entre -1,81 y 2,37 (había cuatro

personas por encima de 2, un valor de calidad de vida muy alto). El test de Shapiro-Wilk dio  $p=0,6621$ , lo que indica que la distribución de la muestra era normal.



**Gráfico 3.** Mapa persona-ítem antes de la reducción de ítems.



**Gráfico 4.** Mapa persona-ítem tras la reducción de ítems.

Finalmente, el nuevo cuestionario tendría un total de 10 preguntas, con dos o tres categorías de respuesta.

**Tabla 10.** Cuestionario final. Entre paréntesis el número de pregunta original.

	Categorías	Puntos
<b>I1.</b> (1) ¿Qué dificultad tiene al conducir en condiciones de deslumbramiento?	Ninguna	2
	Un poco /Moderada	1
	Mucho / No puedo	0
<b>I2.</b> (2) Durante el mes pasado, con qué frecuencia Ud. ha sentido sus ojos cansados o tensos?	Nunca	2
	Ocasionalmente	1
	Seguido / Muy seguido / Siempre	0
<b>I3.</b> (3) ¿Qué problema le origina a Ud. el no poder usar gafas de sol sin prescripción?	Ninguno	1
	Un poco / Moderadamente / Mucho / Extremadamente	0
<b>I4.</b> (4) ¿Cuánta molestia le genera pensar en el estado de sus gafas / lentes de contacto / ojos después de la intervención antes de realizar una actividad como viajar, hacer deporte, nadar?	Ninguno	2
	Un poco	1
	Moderadamente / Mucho / Extremadamente	0
<b>I5.</b> (5) ¿Cuánta dificultad tiene usted para ver cuando se despierta por la noche? Ejemplos: para ir al baño, atender a un bebe, ver la hora	Ninguno	1
	Un poco / Moderadamente / Mucho / Extremadamente	0
<b>I6.</b> (6) ¿Cuánta dificultad tiene para ver cuando está en la playa o nadando en una piscina, si en ese momento no tiene sus lentes o lentes de contacto?	Ninguno	1
	Un poco / Moderadamente / Mucho / Extremadamente	0
<b>I7.</b> (9) ¿Cuánta preocupación le genera el gasto en mantenimiento por daños inesperados de su cirugía refractiva / gafas / lentes de contacto; por ejemplo: ruptura, pérdida, un nuevo problema de visión?	No mucho	2
	Un poco	1
	Moderadamente / Mucho / Extremadamente	0
<b>I8.</b> (12) ¿Cuánto le preocupan posibles complicaciones médicas referentes a su elección de corrección óptica? Ejemplos: Cirugía refractiva, gafas o lentes de contacto.	No mucho	2
	Un poco	1
	Moderadamente / Mucho / Extremadamente	0
<b>I9.</b> (17) Durante el último mes, ¿Cuántas veces ha sentido mayor confianza en usted mismo?	Nunca / Ocasionalmente / Seguido	2
	Muy seguido	1
	Siempre	0
<b>I10.</b> (18)Durante el último mes, ¿Cuántas veces Ud. se ha sentido feliz?	Nunca / Ocasionalmente / Seguido	0
	Muy seguido	1
	Siempre	2

Para obtener los resultados, se sumarían los puntos totales y se calcula según la tabla 11:

**Tabla 11.** Puntuación final según los resultados del cuestionario.

Suma	0	1	2	3	4	5	6	7	8
Resultado	0	4,56	11,31	18,31	24,14	31,08	37,31	43,18	55,74
Suma	9	10	11	12	13	14	15	16	
Resultado	55,74	62,01	67,85	74	80,88	86,57	93,72	100	

Se analizaron los resultados de los pacientes con el nuevo cuestionario. El resultado global fue de  $61,33 \pm 19,7$ . Por grupos, el grupo de gafas obtuvo  $58,92 \pm 19,05$ , el de lentes de contacto  $53,97 \pm 18,48$ , y por último, el de cirugía refractiva  $73,17 \pm 18,01$ .

Comparando grupos dos a dos, aplicando ANOVA, fue estadísticamente significativo la diferencia entre el grupo de cirugía y los otros dos,  $p=0,013$  respecto al grupo de gafas, y  $p=0,0006$  respecto a los usuarios de lentes de contacto. No hubo diferencias significativas entre gafas y lentes de contacto ( $p=0,8697$ ).

## 8.4 Retest

Para estudiar la fiabilidad del cuestionario, se volvió a realizar el test a 31 de los 91 pacientes. El tiempo transcurrido entre ambos pases comprendió desde una semana y media hasta 10 semanas y media, con una media de  $6,13 \pm 2,88$  semanas. Esta forma de fiabilidad es por métodos de administración múltiple, es decir, lo que se mide es la fiabilidad externa.

Primero se comparó el resultado de los dos pases en cada ítem, usando el contraste de Wilcoxon.

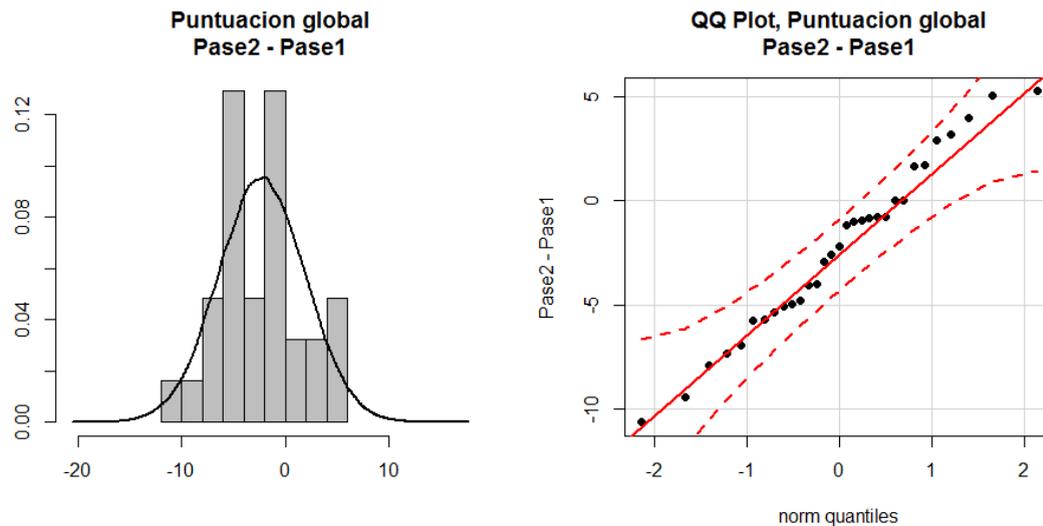
**Tabla 12.** Contraste de Wilcoxon, mostrando el p-valor de cada ítem.

Ítem	1	2	3	4	5	6	7	8	9	10
p-valor Wilcoxon	0,2234	0,4154	1	0,3088	0,8244	1	0,3405	<b>0,0087</b>	0,1319	0,7464
Ítem	11	12	13	14	15	16	17	18	19	20
p-valor Wilcoxon	<i>0,0594</i>	<i>0,0575</i>	1	<b>0,0335</b>	0,212	0,2093	0,1057	1	0,6754	0,2152

En negrita los valores significativos ( $p < 0,05$ ), y en cursiva los valores próximos a la significación ( $p < 0,1$ ;  $> 0,05$ ).

Se observó que la comparación de los dos pases para la mayoría de las preguntas no tuvo diferencias significativas. En las preguntas 8 (preocupación por el gasto económico de financiación) y 14 (sentir que se tiene buena apariencia) sí que existieron diferencias significativas, con peores resultados para ambas en el segundo pase. Las preguntas 11 (preocupación porque la visión no sea tan buena como debería) y 12 (preocupación por las complicaciones médicas) se quedaron próximas a la significación, también ambas con resultados peores en la segunda visita.

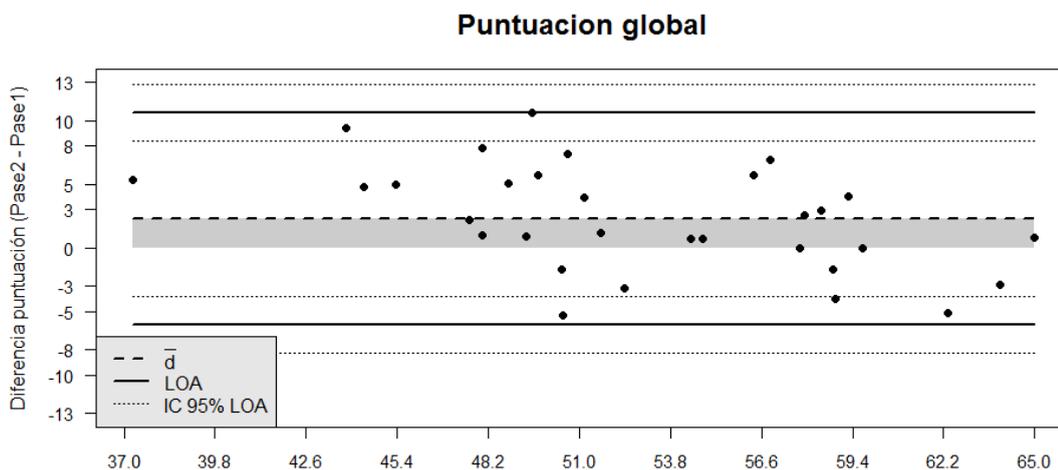
La media y desviación estándar en cada uno de los pases (en el primero incluyendo sólo a los 31 pacientes, no al total) fue de  $54,16 \pm 5,68$  para la primera visita, y  $51,86 \pm 7,65$  para la segunda, siendo éste un resultado peor en cuanto a calidad de vida. La diferencia entre ambos (pase2-pase1) es de -2,3. Se obtuvo un valor de curtosis de -0,86 ( $p=0,9748$ ) y de asimetría de 0,01 ( $p=0,3299$ ). En el test de Shapiro-Wilk el p-valor fue de 0,7935. Todo ello indicó que la distribución era normal. En el gráfico 5 se pueden observar estos resultados.



**Gráfico 5.** Histograma (izquierda) y QQ-plot (derecha) de los datos de normalidad.

Además, se obtuvo una desviación estándar de la media de cada individuo de 2,72 y un coeficiente de variación intra-sujeto del 5,4%.

La diferencia de medias en el análisis de Bland-Altman ( $\bar{d}$ ) fue de 2,3. El límite de acuerdo (LoA) superior era de 10,65, con un intervalo de confianza del 95% inferior de 8,449 y superior de 12,858. El límite de acuerdo inferior era de -6,05, con un intervalo de confianza del 95% inferior de -8,253 y superior de -3,844.



**Gráfico 6.** Análisis de Bland-Altman, incluyendo los límites de acuerdo, los límites de confianza y la diferencia de medias, cuyo sesgo se muestra con la zona sombreada.

Se evaluó el sesgo de mediciones con la T-student para muestras pareadas, comparando las medias. El p-valor fue de 0,0045, por lo que la diferencia entre las dos muestras resultó estadísticamente significativa.

Por último, el valor del coeficiente de correlación intraclase (ICC) obtenido fue de 0,762, lo que indica un resultado bueno de correlación.

## 9. Discusión

En este trabajo cuyo propósito es validar el cuestionario QIRC en lengua española, se pueden valorar principalmente dos cuestiones: analizar los resultados obtenidos para determinar cómo la muestra ha respondido al test, y comprobarlo y mejorarlo a través del análisis Rasch, para conocer si se ajusta al modelo esperado. Para ello se analizan todos los resultados obtenidos, y se compara con los datos de la bibliografía y del cuestionario QIRC original.

Este proceso, sobre todo la segunda parte, tiene sus limitaciones. Primero por el tamaño muestral, que en este caso es pequeño, debido a la limitación en el tiempo. Además, no existen unos criterios universales a la hora de la validación. Existen una gran multitud de pruebas que se pueden realizar, así como múltiples tipos de validación (como se describió en la introducción), y ningún cuestionario responde a todas ellas o elige los mismos análisis estadísticos, aunque sí que hay algunos más utilizados por la mayoría. Con lo cual, cada trabajo busca ser lo más completo posible, pero es difícil comparar con exactitud los resultados.

La muestra de 91 pacientes obtuvo tres grupos muy parecidos en número, ligeramente menor en el grupo de cirugía, y en edad, en la que en ese mismo grupo había diferencias significativas con los otros dos. El número total de hombres y mujeres también fue parecido, por lo que la muestra se puede considerar representativa.

La media de las respuestas fue de  $51,88 \pm 7,05$ , siendo significativa la diferencia entre el grupo de cirugía refractiva, cuyos resultados son mejores que los otros dos grupos, cuyas diferencias eran mínimas. Estos resultados confirman el resultado de otros trabajos (14,20,21), y demuestran cómo la calidad de vida de los sujetos se ve afectada por su error refractivo, y cómo, al operarse y dejar de depender de ello, mejora. Sí que se ha demostrado en otro estudio (14) que existen diferencias significativas también entre el grupo de gafas y de lentes de contacto, no encontrado aquí, que puede ser debido al tamaño muestral o al grado de error refractivo de los sujetos.

Respecto al análisis Rasch, al realizarlo se comprobó que los resultados no se ajustaron al modelo esperado. A diferencia del artículo original (5), los datos son peores y hay que realizar modificaciones. Cronbach's  $\alpha$  consiguió un valor de 0,643, lejos del 0,78 original. Este factor nos indica unidimensionalidad, y aunque un valor muy alto, por encima de 0,9, puede implicar redundancia, se suele aceptar como bueno un valor a partir de 0,7. Esto nos indicó que faltaba unidimensionalidad y que había que considerar reducir el número de ítems (5,11). El mapa persona-ítem tampoco estaba bien ajustado y se observó como los ítems y la muestra de sujetos no coincidían en habilidad.

Por todo ello, se realizó una reducción iterativa de categorías y de ítems. Para la reducción de ítems hay que tener en cuenta que aunque en el original se muestran cinco categorías por pregunta, si se observa la tabla de análisis de resultados, muchos de ellos representan la misma puntuación. En el QIRC original (5), en las preguntas 1-13, las categorías 3, 4 y 5 (0, 1 y 2 para el análisis Rasch) tienen la misma puntuación, por lo que en realidad se está valorando el cuestionario en tres categorías. En las preguntas 14-20, se produce una reducción a cuatro categorías, uniendo 1 y 2 (o y 1 en análisis Rasch). Por lo tanto, la unión de categorías realizada en este trabajo no fue tan diferente a la original y las diferencias en número de categorías observada no era tan drástica, si no que en el original mantuvo un número constante de

categorías para facilitar la respuesta del paciente, aunque a la hora de valorarlo fuera diferente. Aquí, como se puede ver tras la reducción de ítems (tabla 10), las categorías de respuesta en la mayoría de preguntas se disminuyó a tres, excepto en una de ellas que sólo se mantuvieron dos opciones.

Gracias a las estadísticas de ajuste al modelo (infit y outfit), se consiguió eliminar los ítems que peor se ajustaron a lo esperado, obteniendo datos muy positivos, infit de  $0,93 \pm 0,05$  (rango entre 0,849 y 1,042), y outfit de  $0,92 \pm 0,07$  (rango entre 0,818 y 1,039), muy cercanos todos ellos a 1, que representaría el resultado ideal.

Los resultados finales obtenidos fueron mejorados, y aunque el nuevo valor de Cronbach's  $\alpha$  no terminó de ser un resultado positivo (0,68), otro método como el análisis de componentes principales sí que indicó la existencia de unidimensionalidad en el cuestionario. También se pudo observar que el nuevo mapa persona-ítem mejoró y se ajustaba mejor.

El nuevo resultado del cuestionario a nivel global fue de  $61,33 \pm 19,7$ . Las diferencias entre grupos se mantuvieron, con resultados estadísticamente significativos entre el grupo de cirugía refractiva respecto a los otros dos, sin diferencias significativas entre ellos. Sí que se produjo un cambio debido a que el grupo de gafas obtuvo mayor puntuación que el de lentes de contacto. Era un resultado lógico teniendo en cuenta las diferencias mínimas en el original y que al reducir ítems muchos de los que permanecieron tenían mayor puntuación para el grupo de gafas (7 de 10).

De todas formas, la forma de puntuación en este nuevo cuestionario no se puede valorar de la misma forma que el anterior. Si se observa la tabla de puntuación original (anexo 3), se puede comprobar que cada ítem tenía como resultado valores diferentes, debido a las diferencias de dificultad entre ellos, lo que suponía una mejora en la puntuación respecto a otros cuestionarios anteriores. En este cuestionario reducido se volvió a perder dicha mejora, valorando cada pregunta entre 0 y 2 y sumando los resultados para obtener una puntuación, al igual que se hacía en la escala Likert anterior al análisis Rasch. Haría falta un análisis más profundo y detallado para ponderar las preguntas y tener resultados más exactos.

Para comprobar la fiabilidad se realizó un re-test. La muestra también fue muy pequeña, 31 pacientes, por lo que los resultados podrían ser mejorables con una muestra mayor. Al realizarlo se obtuvieron unos resultados significativamente peores en el segundo pase. Aunque todos los sujetos ya conocían el cuestionario, es posible que el realizarlo esta segunda vez de forma mayormente telefónica en vez de presencial, en el que fue explicado más ampliamente, haya podido influenciar en esta diferencia. La distribución obtenida en esta muestra siguió la normalidad. El test de Bland-Altman mostró la amplitud en la diferencia de respuestas, con una media de 2,3, pero con diferencias de hasta un máximo de 10,65 puntos sobre 100. Es un rango muy amplio, por lo que sería preferible uno más reducido. El coeficiente de correlación intraclass (ICC) fue de 0,762, un resultado aceptablemente bueno aunque menor que el del QIRC original (5), que es de 0,88. Teniendo en cuenta todos los datos, se puede pensar que el test es fiable, pero que los resultados se ven afectados por la muestra y especialmente por ciertos sujetos cuyas medidas empeoran los resultados totales, ya que la media en general era buena.

Como se explicó anteriormente, existen muchos tipos de validación, explicados en la introducción, pero normalmente no se responde a todos ellos. A continuación se expone un resumen de ellos indicando como se ha procedido en este trabajo:

- Validación aparente: Es una valoración más bien subjetiva. Se puede considerar que si que se consigue medir lo que se pretende, que la escala es buena, pero que la comprensión de las preguntas a veces es más complicada de lo que se esperaría.
- Unidimensionalidad: Se mide de dos maneras, y si que se considera que exista una única dimensión subyacente, ya que aunque el índice Cronbach's  $\alpha$  no llega a ser del todo bueno, el PCA final sí que tiene todos sus valores dentro de los límites.
- Validación de contenido: Sí que se valoran diferentes aspectos de la calidad de vida, pero no se ha realizado ningún test específico de consistencia interna, aparte de Cronbach's  $\alpha$ , el cual no sería el más indicado.
- Validación de construcción: Aunque hay muchas formas de valorarla, si que se ha realizado correlaciones para comparar los diferentes tipos de corrección (gafas, lentes de contacto y cirugía), que era uno de los objetivos buscados.
- Validación de criterio: No se ha podido valorar, debido a la falta de un gold-standard para cuestionarios de calidad de vida según el error refractivo.
- Validación convergente: No se ha valorado, aunque en un futuro se podría hacer un estudio correlacionando el cuestionario con el nivel refractivo, la AV o algún otro valor relacionado.
- Validación discriminante: No se ha comparado el cuestionario con ningún otro existente.
- Validación predictiva: Aunque se necesitaría un análisis más profundo de los datos, se puede considerar que el cuestionario es capaz de diferenciar sujetos sometidos a cirugía refractiva respecto de los demás. Y en otro estudio (13), se indica que los pacientes que se quieren operar obtienen peores resultados.
- Validación concurrente: Al igual que la anterior, se considera que al medir a la vez a varios sujetos, se podrá diferenciar si ha sido sometido a cirugía refractiva o no. Serían necesarios más estudios para conocer que más grupos puede diferenciar.
- Fiabilidad: Por el método de administración única, el índice de Cronbach's  $\alpha$  obtiene un resultado aceptable pero no bueno; y mediante métodos de administración múltiple, el ICC es bueno aunque el análisis de Bland-Altman muestre un rango demasiado amplio.
- Sensibilidad: No se ha realizado, pero sí que hay algún estudio (21) que muestra que existen diferencias entre pre y postoperados.
- Interpretación: Sí que se ha valorado el cuestionario en una población representativa, aunque sería necesario tomar más medidas y realizar más correlaciones para comprobar los resultados obtenidos.

En resumen, aunque los resultados no son del todo concluyentes, se ha conseguido adaptar el cuestionario QIRC a la lengua española. Los resultados son peores a los esperados, pero podrían mejorarse con un mayor tamaño muestral. Sin embargo, si se reduce el cuestionario a un total de 10 ítems, se demuestra una mejora en el resultado final, aunque con detalles que se podrían mejorar aún. Para un trabajo más completo se hubiera podido añadir alguna prueba más, como por ejemplo el análisis factorial o intercorrelaciones entre otros datos.

También hay que tener en cuenta las dificultades y problemas que han surgido a lo largo del estudio y que limitan sus conclusiones:

- Tamaño muestral. Como se ha comentado a lo largo de este trabajo, el número de sujetos ha sido bajo comparado con estudios del mismo tipo, por lo que nos ha impedido tener resultados más concluyentes. Esto es debido primero a la limitación de tiempo para recoger las muestras y a la obtención de pacientes de un único centro (especialmente importante para el grupo de pacientes operados). Por ejemplo, en el estudio QIRC original (5) se pasó el cuestionario en un total de 18 centros, lo que finalmente significó 306 y 312 pacientes en las dos pruebas realizadas.
- Al obtener peores resultados de los esperados, se tuvo que realizar una reducción del cuestionario a la mitad de las preguntas. Aunque se suponía que no existían preguntas redundantes, los datos recogidos mostraban lo contrario.
- En la pregunta 7, hubo un gran número de no contestaciones, ya que al preguntar por el problema que generan las gafas y lentes de contacto al hacer ejercicio, se consideró que el grupo de cirugía era no aplicable. Este problema influyó en los resultados e hizo que se eliminara dicho ítem. Este estudio no es el único con dificultades al contestar dicha pregunta (21).
- Las preguntas sobre bienestar (14-20) también tuvieron problemas a la hora de ser respondidas. En este y otros estudios (14,20,21) ha habido quejas y dificultades al contestarlas, entre otros motivos porque las consideraban irrelevantes o influidas por otros factores y circunstancias que no tienen relación con la visión. Finalmente después de la reducción sólo quedaron dos preguntas de las ocho iniciales.
- Se podrían haber realizado más pruebas, pero los resultados a los que se ha llegado y a la gran cantidad de posibilidades disponibles han impedido su realización, dejando sólo las necesarias.

## 10. Conclusiones

El estudio realizado demuestra la validez limitada del cuestionario QIRC en lengua española, pero acepta una posible mejora del mismo disminuyendo el número de preguntas a un total de 10, y reduciendo el número de categorías. Este nuevo cuestionario permitiría optimizar los resultados y aumentar su validez en este idioma.

Existen tres partes diferenciadas en este trabajo, que se pueden desgranar en las siguientes conclusiones específicas:

- En una muestra representativa de nuestro entorno social, la calidad de vida de los pacientes operados de cirugía refractiva es claramente superior a la expresada por los sujetos que compensan su ametropía mediante el uso de gafas o lentes de contacto, lo que reafirma el carácter funcional de este tipo de cirugía.
- El análisis del cuestionario confirma su unidimensionalidad, pero con resultados menos concluyentes para el resto de validaciones. Un nuevo cuestionario reducido gracias al análisis Rasch incrementa notablemente su validez. La realización de estudios secundarios permitirían analizar otros tipos de validaciones no estudiadas en este trabajo.
- La fiabilidad obtenida es aceptable, considerando las limitaciones mencionadas, aunque los resultados en el segundo análisis sugieren que existen otros aspectos que deberían mejorar.

## 11. Agradecimientos

Este trabajo ha requerido la ayuda y colaboración de muchas personas. Primero agradecer a mi tutor, el Dr. Miguel Maldonado, por la orientación y las facilidades dadas, sin las que este trabajo no sería posible. En segundo lugar, es necesario nombrar a la estadista del IOBA, Itziar Fernández, cuyo análisis estadístico es una parte indispensable de este estudio. También nombrar al resto de profesionales de la consulta del IOBA, gracias a los cuales el reclutamiento de pacientes ha sido más sencillo. Y, por último, agradecer el apoyo a familiares y amigos, y a todos los participantes del estudio por su colaboración.

## 12. Bibliografía

1. FEDAO, Asociación Visión y Vida, CGCOO. Libro Blanco de la Visión 2013; 2013.
2. World Health Organization. Measuring quality of life: The World Health Organization Quality of Life Instruments (the WHOQOL-100 and the WHOQOL-BREF). World Health Organization, Division of mental health and prevention of substance abuse; 1997.
3. Felce D, Perry J. Quality of life: its definition and measurement. *Research in Developmental Disabilities*. 1995; 16(1): p. 51-74.
4. Schein OD. The measurement of patient-reported outcomes of refractive surgery: The Refractive Status and Vision Profile. *Transactions of the American Ophthalmological Society*. 2000; 98: p. 439-469.
5. Pesudovs K, Garamendi E, Elliot DB. The Quality of Live Impact of Refractive Correction (QIRC) Questionnaire: Development and validation. *Optometry and Vision Science*. 2004 October; 81(10): p. E769.
6. Brunette I, Gresset J, Boivin JF, Boisjoly H, Makni H. Functional outcomes and satisfaction after PRK. *American Academy of Ophthalmology*. 2000 September; 107(9): p. 1783-1789.
7. Nichols JJ, Mitchell L, Saracino M, Zadnik K. Reliability and validity of refractive error-specific quality-of-life instruments. *Archives of Ophthalmology*. 2003; 121: p. 1289-1296.
8. Massof RW, Rubin GS. Visual function assessment questionnaires. *Survey of Ophthalmology*. 2001 May-June; 45(6): p. 531-548.
9. Massof RW. The measurement of vision disability. *Optometry and Vision Science*. 2002 August; 79(8): p. 516-552.
10. Pesudovs K, Garamendi E, Keeves JP, Elliott DB. The Activities of Daily Vision Scale for cataract surgery outcomes: Re-evaluating validity with Rasch analysis. *Investigative Ophthalmology & Visual Science*. 2003 July; 44(7): p. 2892-2899.
11. Pesudovs K, Burr JM, Harley C, Elliot DB. The development, assesment, and selection of questionnaires. *Optometry and Vision Science*. 2007 August; 84(8): p. 663-674.
12. Berry S, Mangione CM, Lindblad AS, McDonnell PJ. Development of the National Eye Institute Refractive error correction Quality of Life questionnaire. *American Academy of Ophthalmology*. 2003 December; 110(12): p. 2285-2291.
13. Garamendi E, Pesudovs K, Elliott DB. Changes in quality of life after LASIK for myopia. *J Cataract Refractive Surgery*. 2005 August; 31: p. 1537-1543.
14. Pesudovs K, Garamendi E, Elliott DB. A quality of life comparison of people wearing spectacles or contact lenses or having undergone refractive surgery. *Journal of Refractive Surgery*. 2006 January/February; 22: p. 19-27.

15. Lamoureux EL, Pallant JF, Pesudovs K, Hassell JB, Keeffe JE. The Impact of Vision Impairment questionnaire: An evaluation of its measurement properties using Rasch analysis. *Investigative Ophthalmology & Vision Science*. 2006 November; 47(11): p. 4732-4741.
16. Stelmack J, Szlyk JP, Stelmack T, Babcock-Parziale J, Demers-Turco P, Williams T, et al. Use of Rasch person-item map in exploratory data analysis: A clinical perspective. *Journal of Rehabilitation, Research & Development*. 2004 March-April; 41(2): p. 233-242.
17. Mallinson T, Stelmack J, Velozo C. A comparison of the Separation Ratio and the Coefficient  $\alpha$  in the creation of minimum item sets. *Medical Care*. 2004 January; 42(1): p. 17-24.
18. Tennant A, McKenna SP, Hagell P. Application of Rasch analysis in the development and application of quality of life instruments. *Value in Health*. 2004; 7(1): p. 22-26.
19. Mallinson T. Why measurement matters for measuring patient vision outcomes. *Optometry and Vision Science*. 2007 August; 84(8): p. 675-682.
20. Jeong A, Rubin GS, Allan BDS. Quality of life in high myopia: Implantable Collamer Lens implantation versus contact lens wear. *American Academy of Ophthalmology*. 2009 February; 116(2): p. 275-280.
21. Jeong A, Hau SCH, Rubin GS, Allan BDS. Quality of life in high myopia before and after Implantable Collamer Lens implantation. *American Academy of Ophthalmology*. 2010 December; 117(12): p. 2295-2300.
22. Nichols JJ, Twa MD, Mitchell GL. Sensitivity of the National Eye Institute Refractive error Quality of Life instrument to refractive surgery outcomes. *J Cataract Refractive Surgery*. 2005 December; 31: p. 2313-2318.
23. Garamendi E, Pesudovs K, Stevens MJ, Elliott DB. The RSVP: Evaluation of psychometric properties and comparison of Rasch and summated Likert-scaling. *Vision Research*. 2006; 46: p. 1375-1383.
24. Nichols JJ, Mitchell GL, Zadnik K. The performance of the Refractive Status and Vision Profile survey in a contact lens clinical trial. *American Academy of Ophthalmology*. 2001 June; 108: p. 1160-1166.
25. Vitale S, Schein OD, Meinert CL, Steinberg EP. The Refractive Status and Vision Profile: A questionnaire to measure vision-related quality of life in persons with refractive error. *American academy of Ophthalmology*. 2000 August; 107(8): p. 1529-1539.
26. McAlinden C, Pesudovs K, Moore JE. The development of an instrument to measure quality of vision: The Quality of Vision (QoV) questionnaire. *Investigative Ophthalmology & Visual Science*. 2010 November; 51(11): p. 5537-5545.
27. de Boer MR, Moll AC, de Vet HCW, Terwee CB, Völker-Dieben HJM, van Rens GHMB. Psychometric properties of vision-related quality of life questionnaires: A systematic review. *Ophthalmic and Physiological Optics*. 2004; 24(4): p. 257-273.

# 13. Anexos

## 13.1 Autorización del tutor

### AUTORIZACIÓN DEL TUTOR PARA LA EXPOSICIÓN PÚBLICA DEL TRABAJO DE FIN DE MÁSTER

*(Art. 6.2 del Reglamento de la UVA sobre la Elaboración y Evaluación del Trabajo Fin de Máster)*

D./Dña. Miguel José Maldonado López

en calidad de Tutor/a del alumno/a

D. /Dña. Elena Sanz Sanz

del Máster en: Investigación en Ciencias de la Visión

Curso académico: 2015/2016

CERTIFICA haber leído la memoria del Trabajo de Fin de Máster titulado

“Análisis de la validez del cuestionario "Repercusión de la corrección refractiva en la calidad de vida" (QIRC) en lengua española “.

y estar de acuerdo con su exposición pública en la convocatoria de septiembre

En Valladolid a 2 de septiembre de 2016

Vº Bº

Fdo.: Miguel José Maldonado López

El/La Tutor/a



## 13.2 Aceptación del estudio por parte de la Comisión de Investigación



Universidad de Valladolid



COMISION DE INVESTIGACION

Dña. M<sup>a</sup> Paz García García como Secretaria de la Comisión de Investigación del Instituto Universitario de Oftalmobiología Aplicada (IOBA) de la Universidad de Valladolid,

CERTIFICA

Que el TFM titulado "Análisis de la validez del cuestionario "Repercusión de la corrección refractiva en la calidad de vida" (QIRC) en lengua española." con número de registro 01/2016 de Dña. Elena Sanz Sanz, se encuentra en el momento de la última reunión de la Comisión de Investigación de 4 de marzo de 2016

- Aprobado  
 Pendiente de

Y para que así conste expido el presente certificado.

En Valladolid, a 15 de marzo de 2016

Fdo.: M<sup>a</sup> Paz García García  
Secretaria de la Comisión de Investigación

### 13.3 Consentimiento informado



Universidad de Valladolid

## CONSENTIMIENTO INFORMADO

**Análisis de la validez del cuestionario "Repercusión de la corrección refractiva en la calidad de vida" (QIRC) en lengua española.**

### HOJA DE INFORMACIÓN AL PACIENTE

Le ofrecemos participar en el estudio titulado: *Análisis de la validez del cuestionario "Repercusión de la corrección refractiva en la calidad de vida" (QIRC) en lengua española.*

**Promotor del Estudio:** IOBA

**Duración del Estudio:** Febrero-Julio 2016

**Responsables del Estudio:** Miguel Maldonado (IP), Elena Sanz

**Centro:** Instituto Universitario de Oftalmobiología Aplicada (IOBA), Pº de Belén 17; 47011 Valladolid

Tf. 983 184734

#### Propósito del estudio

Este estudio tiene como objetivo la validación en español de un test de 20 preguntas que analiza la calidad de vida de una persona dependiendo de cómo se ve afectada por su corrección refractiva, ya sea mediante el uso de gafas, lentes de contacto o que se haya sometido a refractiva.

#### Participación voluntaria

Debe saber que su participación en este estudio es voluntaria y que puede decidir no participar o cambiar su decisión y retirar el consentimiento en cualquier momento, sin que por ello se altere la relación con el personal del centro.

## **Condiciones del estudio**

Si se decide participar, usted accede a la realización de un cuestionario tipo test, cuyas respuestas serán analizadas para obtener los objetivos del estudio.

## **Pruebas que se realizarán durante el estudio**

La prueba consistirá en un test llamado "*Repercusión de la corrección refractiva en la calidad de vida*", cuyo contenido es un cuestionario 20 preguntas tipo test. No supone ningún procedimiento invasivo.

## **Riesgos que entraña el presente estudio**

Usted será tratado siempre según los postulados para la investigación clínica en seres humanos recogidos en la Declaración de Helsinki. No se trata de un ensayo clínico, sino de un estudio de recogida de datos de manera sistematizada, que permitirá validar un test sobre la calidad de vida según la corrección refractiva.

## **Confidencialidad**

Sus datos formarán parte de la base de datos del IOBA, cuyo responsable es el Gerente de la Fundación General de la Universidad de Valladolid. En todo momento se seguirán las normativas establecidas en la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal (BOE núm. 298, de 14-12-1999, pp. 43088-43099), lo cual garantiza el que su identificación será siempre confidencial fuera del equipo oftalmológico que cuidará su salud. Sólo se emplearán los datos de su historia clínica para correlacionarlos con los obtenidos en los análisis de los datos, estando en todo momento de este estudio desvinculado su nombre de los mismos. Los resultados de este estudio podrán ser divulgados en revistas científicas, congresos y otro tipo de reuniones.

## **Otra información relevante**

Si usted decide retirar el consentimiento para participar en este estudio, ningún dato nuevo será añadido a la base de datos. También debe saber que puede ser excluido del programa si los responsables del estudio lo consideran oportuno.

## CONSENTIMIENTO INFORMADO

**Análisis de la validez del cuestionario "Repercusión de la corrección refractiva en la calidad de vida" (QIRC) en lengua española.**

**Promotor del Estudio:** IOBA

**Responsables del Estudio:** Dr. Miguel Maldonado, (IP) ; Elena Sanz (Óptico-optometrista).

**Centro:** Instituto Universitario de Oftalmobiología Aplicada (IOBA), Pº de Belén 17; 47011 Valladolid  
Tf. 983 184734

**Yo, (nombre y apellidos):**

He leído la Hoja de Información que se me ha entregado, he podido hacer preguntas sobre el estudio y he recibido suficiente información sobre el mismo.

**He hablado con, (nombre y apellidos del Facultativo):**

Comprendo que mi participación es voluntaria y que puedo retirarme del estudio cuando quiera , sin tener que dar explicaciones, sin que esto repercuta en mis cuidados médicos.

Presto libremente mi conformidad para participar en el estudio y doy mi consentimiento para el acceso y utilización de mis datos en las condiciones detalladas en la hoja de información.

<b>Firma del paciente</b>	<b>Firma del facultativo</b>
<b>Nombre:</b>	<b>Nombre:</b>
<b>Fecha:</b>	<b>Fecha:</b>

### 13.4 Cuestionario QIRC en español

Nº	PREGUNTA	Respuesta elegida
1	¿Qué dificultad tiene al conducir en condiciones de deslumbramiento?	No conduzco por razones ajenas a mi visión Ninguna Un poco Moderada Mucho Tanto que no puedo hacer esta actividad
2	Durante el mes pasado, con qué frecuencia Ud. ha sentido sus ojos cansados o tensos?	No sabe / No es aplicable Nunca Ocasionalmente Seguido Muy seguido Siempre
3	¿Qué problema le origina a Ud. el no poder usar gafas de sol sin prescripción?	No sabe / No es aplicable Ninguno Un poco Moderadamente Mucho Extremadamente
4	¿Cuánta molestia le genera pensar en el estado de sus gafas / lentes de contacto / ojos después de la intervención antes de realizar una actividad como viajar, hacer deporte, nadar?	No sabe / No es aplicable Ninguno Un poco Moderadamente Mucho Extremadamente
5	¿Cuánta dificultad tiene usted para ver cuando se despierta por la noche? Ejemplos: para ir al baño, atender a un bebe, ver la hora.	No sabe / No es aplicable Ninguno Un poco Moderadamente Mucho Extremadamente
6	¿Cuánta dificultad tiene para ver cuando está en la playa o nadando en una piscina, si en ese momento no tiene sus lentes o lentes de contacto?	No sabe / No es aplicable Ninguno Un poco Moderadamente Mucho Extremadamente
7	¿Qué problema le originan a Ud. sus gafas o lentes de contacto cuando los usa mientras esta en el gimnasio, en un circuito de entrenamiento, en deportes aeróbicos, etc.?	No sabe / No es aplicable Ninguno Un poco Moderadamente Mucho Extremadamente

8	¿Cuánto le preocupa el gasto inicial y futuros gastos para financiar su cirugía refractiva/sus actuales gafas y/o lentes de contacto?	No sabe / No es aplicable
		Ninguno
		Un poco
		Moderadamente
		Mucho
9	¿Cuánta preocupación le genera el gasto en mantenimiento por daños inesperados de su cirugía refractiva / gafas / lentes de contacto; por ejemplo: ruptura, pérdida, un nuevo problema de visión?	Extremadamente
		Mucho
		Moderadamente
		Un poco
		No mucho
10	¿Cuánta preocupación le genera tener que depender cada vez más de sus gafas o lentes de contacto desde que empezó a utilizarlas?	No sabe / No es aplicable
		No mucho
		Un poco
		Moderadamente
		Mucho
11	¿Cuánto le preocupa que su visión no sea tan buena como debería ser?	Extremadamente
		Mucho
		Moderadamente
		Un poco
		No mucho
12	¿Cuánto le preocupan posibles complicaciones médicas referentes a su elección de corrección óptica? Ejemplos: Cirugía refractiva, gafas o lentes de contacto.	No sabe / No es aplicable
		No mucho
		Un poco
		Moderadamente
		Mucho
13	¿Cuánto le preocupa proteger sus ojos de la radiación ultravioleta (UV)?	Extremadamente
		Mucho
		Moderadamente
		Un poco
		No mucho

Ahora, nosotros estamos interesados en el efecto que su corrección óptica tiene en la forma en la que Ud. se siente consigo mismo. El efecto en su estado sería directo (e.j.Ud. siente que tiene mejor aspecto con gafas nuevas) o sería indirecto (e.j., Ud. siente más confianza desde que usa LC o realizó cirugía por que tiene mejor apariencia).

Nº	PREGUNTA	Respuesta elegida
14	Durante el último mes , ¿Cuántas veces ha sentido que usted tiene muy buena apariencia?	No sabe / No es aplicable
		Nunca
		Ocasionalmente
		Seguido
		Muy seguido
		Siempre
15	Durante el último mes, ¿Cuántas veces ha sentido que las personas le ven como a usted le agrada que lo vean (por ejemplo: inteligente, sofisticado, exitoso, etc.)?	No sabe / No es aplicable
		Nunca
		Ocasionalmente
		Seguido
		Muy seguido
		Siempre
16	Durante el último mes, ¿Cuántas veces se ha sentido aprobado por los demás, satisfecho / halagado?	No sabe / No es aplicable
		Nunca
		Ocasionalmente
		Seguido
		Muy seguido
		Siempre
17	Durante el último mes, ¿Cuántas veces ha sentido mayor confianza en usted mismo?	No sabe / No es aplicable
		Nunca
		Ocasionalmente
		Seguido
		Muy seguido
		Siempre
18	Durante el último mes, ¿Cuántas veces Ud. se ha sentido feliz?	No sabe / No es aplicable
		Nunca
		Ocasionalmente
		Seguido
		Muy seguido
		Siempre
19	Durante el último mes, ¿Cuántas veces ha sentido que puede hacer las cosas que usted quiere hacer?	No sabe / No es aplicable
		Nunca
		Ocasionalmente
		Seguido
		Muy seguido
		Siempre
20	Durante el último mes, ¿Cuántas veces se ha sentido con ganas de tratar de hacer cosas nuevas?	No sabe / No es aplicable
		Nunca
		Ocasionalmente
		Seguido
		Muy seguido
		Siempre

### 13.5 Tabla de resultados

Item number	Response category				
	1	2	3	4	5
1	60.51	45.06	29.61	29.61	29.61
2	65.11	49.66	34.21	34.21	34.21
3	56.71	41.26	25.81	25.81	25.81
4	61.37	45.92	30.47	30.47	30.47
5	59.32	43.87	28.42	28.42	28.42
6	63.92	48.48	33.03	33.03	33.03
7	55.17	39.72	24.27	24.27	24.27
8	64.61	49.16	33.71	33.71	33.71
9	60.62	45.18	29.73	29.73	29.73
10	65.46	50.01	34.56	34.56	34.56
11	65.14	49.69	34.24	34.24	34.24
12	59.49	44.04	28.59	28.59	28.59
13	66.62	51.17	35.72	35.72	35.72
14	28.25	28.25	45.52	60.79	79.18
15	31.72	31.72	48.99	64.26	82.65
16	37.28	37.28	54.55	69.82	88.21
17	25.4	25.4	42.67	57.94	76.34
18	22.34	22.34	39.61	54.88	73.27
19	14.39	14.39	31.66	46.92	65.32
20	23.95	23.95	41.22	56.48	74.88

