



**Universidad Autónoma de Baja California**

Instituto de Investigación y Desarrollo Educativo

**Desarrollo e implementación de una metodología para la adaptación de instrumentos de medición en contextos educativos**

TESIS

Que para obtener el grado de

***MAESTRO EN CIENCIAS EDUCATIVAS***

Presenta:

***Carlos David Díaz López***

Director de tesis

***Joaquín Caso Niebla***



INSTITUTO DE  
INVESTIGACIÓN  
Y DESARROLLO  
EDUCATIVO

Ensenada, Baja California, febrero de 2015



**Universidad Autónoma de Baja California**  
Instituto de Investigación y Desarrollo Educativo  
Maestría en Ciencias Educativas



**“Adaptación de un instrumento para la medición de la  
convivencia escolar”**

**TESIS**

Que para obtener el grado de

**MAESTRO EN CIENCIAS EDUCATIVAS**

Presenta

***Carlos David Díaz López***

APROBADO POR:

**Dr. Joaquín Caso Niebla**  
Director(a) de tesis

**Dr. Luis Ángel Contreras Niño**  
Sinodal

**Dr. Luis Lizasoain Hernández**  
Sinodal





Ensenada, B.C. a 27 de enero de 2015

**ASUNTO:** Voto aprobatorio sobre trabajo de tesis de grado de Maestría.

**Alicia Aleli Chaparro Caso López**  
**Coordinadora de la Maestría en Ciencias Educativas**  
**Presente.**

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por la C. **CARLOS DAVID DÍAZ LÓPEZ** para poder presentar la defensa de su examen y obtener el grado de Maestría en Ciencias Educativas, me permito comunicarle que he dado mi VOTO APROBATORIO, sobre su trabajo intitulado:

***“DESARROLLO E IMPLEMENTACIÓN DE UNA METODOLOGÍA PARA LA ADAPTACIÓN DE INSTRUMENTOS DE MEDICIÓN EN CONTEXTOS EDUCATIVOS”.***

Esperando reciba el presente de conformidad, quedo de Usted.

Atentamente

Dr. Joaquín Caso Niebla



Ensenada, B.C. a 27 de enero de 2015

**ASUNTO:** Voto aprobatorio sobre trabajo de tesis de grado de Maestría.

**Alicia Alelí Chaparro Caso López**  
**Coordinadora de la Maestría en Ciencias Educativas**  
**Presente.**

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por el **C. CARLOS DAVID DÍAZ LÓPEZ** para poder presentar la defensa de su examen y obtener el grado de Maestría en Ciencias Educativas, me permito comunicarle que he dado mi VOTO APROBATORIO, sobre su trabajo intitulado:

***“DESARROLLO E IMPLEMENTACIÓN DE UNA METODOLOGÍA PARA LA ADAPTACIÓN DE INSTRUMENTOS DE MEDICIÓN EN CONTEXTOS EDUCATIVOS”.***

Esperando reciba el presente de conformidad, quedo de Usted.

Atentamente

Una firma manuscrita en tinta que parece ser "L. Contreras Niño".

---

Dr. Luis Ángel Contreras Niño



Ensenada, B.C. a 27 de enero de 2015

**ASUNTO:** Voto aprobatorio sobre trabajo de tesis de grado de Maestría.

**Alicia Alelí Chaparro Caso López**  
**Coordinadora de la Maestría en Ciencias Educativas**  
**Presente.**

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por la **C. CARLOS DAVID DÍAZ LÓPEZ** para poder presentar la defensa de su examen y obtener el grado de Maestría en Ciencias Educativas, me permito comunicarle que he dado mi VOTO APROBATORIO, sobre su trabajo intitulado:

***“DESARROLLO E IMPLEMENTACIÓN DE UNA METODOLOGÍA PARA LA ADAPTACIÓN DE INSTRUMENTOS DE MEDICIÓN EN CONTEXTOS EDUCATIVOS”.***

Esperando reciba el presente de conformidad, quedo de Usted.

Atentamente

---

Dr. Luis Lizasoain Hernández

## Índice

Resumen	1
Introducción	2
Antecedentes	4
Objetivos	7
Objetivo general	7
Objetivos específicos	7
Justificación	8
Capítulo 1. La adaptación de instrumentos de medición psicológica y educativa	10
Validez y confiabilidad	13
Equivalencia y Sesgo	19
Capítulo 2. Estándares y directrices que orientan el desarrollo y adaptación de instrumentos de medición	28
Standards for Educational and Psychological Testing	29
Standards for Educational and Psychological Testing de 1999	32
Standards for Educational and Psychological Testing de 2014	34
Guidelines for Adapting Educational and Psychological Tests	38
Directrices para la traducción y adaptación de PISA 2012	43
Capítulo 3. Aproximaciones en la adaptación/traducción de test	47
Procedimientos de validación en la adaptación de un test	53
Validación del constructo en la población objetivo	54
Validación del constructo entre la población original y la población objetivo	55
Capítulo 4. Método	63
Etapa 1. Traducción	64
Etapa 2. Validación lingüística	65
Etapa 3. Validación de contenido	67
Etapa 4. Prueba piloto	70
Etapa 5. Aplicación a gran escala	71
Etapa 6. Análisis de equivalencia y sesgo	74
Análisis iniciales	78
Análisis de equivalencia de constructo	78
Análisis de dimensionalidad	79

Análisis del funcionamiento diferencial de los ítems	80
Capítulo 5. Resultados	81
Etapa 1: Traducción	81
Etapa 2: Validación lingüística	82
Etapa 3: Validación de contenido	86
Etapa 4: Prueba piloto	92
Etapa 5: Aplicación a gran escala	94
Etapa 6. Análisis de equivalencia y sesgo	95
Análisis iniciales	95
Análisis 1: Equivalencia de constructo	100
Análisis 2: Análisis de dimensionalidad	105
Análisis 3: Análisis del funcionamiento diferencial de los ítems	110
Capítulo 6. Discusión	117
Referencias	124
Anexos	135
Anexo A. Procedimiento general para de la adaptación de la Escala de Convivencia Escolar	136
Anexo B. Traducción	137
Anexo C. Validación lingüística	138
Anexo D. Validación de contenido	143
Anexo E. Prueba piloto	158
Anexo F. Aplicación	159
Anexo G. Frecuencias y porcentajes de las respuestas a los ítems	169
Anexo H. Sintaxis de los análisis de equivalencia y sesgo	182
Anexo I. Análisis de normalidad y matriz de correlaciones policóricas	188

## Índice de tablas

Tabla 1. Tipos y fuentes de sesgo .....	22
Tabla 2. Estructura de los estándares para las pruebas psicológicas y educativas (1985) .....	30
Tabla 3. Estándares aplicables en los procesos de adaptación de test (1985).....	31
Tabla 4. Estructura de los estándares para las pruebas psicológicas y educativas (1999) .....	33
Tabla 5. Estándares de la edición de 1999 aplicables en los procesos de adaptación de test. ....	34
Tabla 6. Estructura de los estándares para pruebas psicológicas y educativas (2014).....	36
Tabla 7. Clasificación de los índices de ajuste en el análisis factorial confirmatorio.....	56
Tabla 8. Modelos anidados en análisis factorial confirmatorio multigrupo.....	57
Tabla 9. Perfil de los jueces del proceso de validación de contenido .....	68
Tabla 10. Criterios de validación del contenido de la adaptación de la ECE .....	69
Tabla 11. Características de la población participante en la prueba piloto de la adaptación de la ECE.....	70
Tabla 12. Estimación muestral para la aplicación de la ECE-A con respecto a la muestra original.....	72
Tabla 13. Características de los estudiantes que conformaron la muestra .....	75
Tabla 14. Dimensiones y número de ítems que conforman la ECE-A .....	77
Tabla 15. Modificaciones realizadas a los ítems de la versión original de los ítems de la ECE tras su traducción .....	82
Tabla 16. Índice de discrepancia intra-jueces de la etapa de validación lingüística .....	83
Tabla 17. Puntuaciones totales e índice de equivalencia de la etapa de validación lingüística de la traducción de la ECE.....	83
Tabla 18. Relación de ítems modificados en la etapa de validación lingüística .....	85
Tabla 19. Discrepancia intra-jueces de la etapa de validación de contenido .....	87
Tabla 20. Promedio de las puntuaciones obtenidas por cada ítem en los rasgos considerados en la etapa de validación de contenido.....	88
Tabla 21. Promedio de las puntuaciones obtenidas en el rasgo de suficiencia por cada dimensión en la etapa de validación de contenido .....	89
Tabla 22. Relación de ítems modificados en la etapa de validación de contenido de la ECE .....	91
Tabla 23. Tiempos de respuesta de la ECE.....	92
Tabla 24. Estadísticos descriptivos de la muestra utilizada en la prueba piloto .....	93
Tabla 25. Tiempos de respuesta en minutos de la aplicación ECE-A .....	94
Tabla 26. Incidencias registradas en la aplicación de la ECE-A, por pregunta y por grupo.....	95
Tabla 27. Estadísticos descriptivos de la ECE-A.....	96

Tabla 28. Coeficientes de confiabilidad alfa ordinal y correlación punto biserial de los ítems de la ECE-A .....	98
Tabla 29. Covarianzas entre los factores de la ECE-A .....	101
Tabla 30. Parámetros estimados para cada ítem .....	102
Tabla 31. Estadísticos e índices de bondad de ajuste del modelo de rasgos correlacionados .....	103
Tabla 32. Estadísticos de ajuste e índices de proximidad asociados con los modelos propuestos .....	107
Tabla 33. Cargas factoriales estimadas asociadas con los modelos propuestos.....	109
Tabla 34. Estadísticos descriptivos de los ítems de la ECE-A en función del sexo de los estudiantes ....	111
Tabla 35. Estadísticos descriptivos de los ítems de la EECE-A en función del nivel educativo de los estudiantes .....	113
Tabla 36. Modelos de regresión logística utilizados para detectar DIF .....	114
Tabla 37. Análisis del funcionamiento diferencial de los ítems en función del sexo y nivel educativo de los estudiantes .....	116

## Índice de figuras

Figura 1. Proceso de Traducción.....	50
Figura 2. Ejemplo de DIF uniforme y DIF no uniforme.....	60
Figura 3. Fases y etapas de la metodología para la adaptación de la ECE.....	64
Figura 4. Esquema de los modelos de medida empleados para estudiar la dimensionalidad de la ECE ....	80
Figura 5. Modelo factorial de rasgos correlacionados .....	100
Figura 6. Parámetros estimados en el modelo.....	104
Figura 7. Modelo unidimensional, Modelo factorial de segundo orden y Modelo bifactorial del análisis de dimensionalidad de la ECE-A .....	106

## Resumen

El propósito del presente estudio fue desarrollar e implementar una metodología para la adaptación de instrumentos en contextos educativos, lo que permitió la adaptación de la Escala de Convivencia Escolar (ECE) al contexto de Baja California. La revisión de los fundamentos, estándares y directrices que orientan los procesos de traducción y adaptación de pruebas psicológicas y educativas (APA, AERA, NCME, 2014; ITC, 2012) permitió proponer una metodología compuesta por seis etapas: traducción, validación lingüística, validación de contenido, prueba piloto, aplicación y análisis de equivalencia y sesgo. A lo largo de este trabajo se documentan los objetivos, procedimientos y productos registrados en cada una de estas etapas, en las que se contó con la participación de un total de 10 jueces y 927 estudiantes de primaria y secundaria. Tanto los análisis factorial confirmatorio y de dimensionalidad, como el funcionamiento diferencial de los ítems, confirman que la versión adaptada de la ECE mide las mismas dimensiones que la versión original y que sus ítems se encuentran libres de sesgo. La evidencia registrada en cada una de las etapas de esta metodología respalda el uso de la ECE en escuelas de Baja California, con lo que se sientan así las bases para la adaptación instrumentos en otros contextos y en otras latitudes.

*Palabras clave:* Adaptación de test, Equivalencia, Sesgo, Validez, Convivencia Escolar.

## **Introducción**

La convivencia escolar es un fenómeno educativo que ha ido adquiriendo relevancia en las agendas de los sistemas educativos, convirtiéndose en un tema que habitualmente forma parte de sus esquemas y estrategias evaluativas. Al igual que sucede con otros constructos, el hecho de que existan instrumentos válidos y confiables en otros países que abordan su estudio con apego a marcos de referencia sólidos, no compromete el que estos instrumentos midan necesariamente lo mismo en otras poblaciones. Con base en ello, en el presente trabajo se propuso desarrollar e implementar una metodología para la adaptación de instrumentos de medición en contextos educativos en apego a lineamientos, procedimientos y estándares exigidos por organismos internacionales, lo que derivó en la adaptación de la Escala de Convivencia Escolar (ECE) al contexto de Baja California.

El contenido de este documento se organizó de la siguiente manera:

En el Capítulo 1 se describen los principales conceptos empleados por la literatura especializada asociados con la adaptación de pruebas psicológicas y educativas (validez, confiabilidad, equivalencia y sesgo).

En el Capítulo 2 se presenta una revisión de los principales documentos que orientan las prácticas de desarrollo y adaptación y traducción de instrumentos de medición. En lo particular, se presentan los Standards for Educational and Psychological Testing, de la American Educational Research Association (AERA), la American Psychological Association (APA) y el National Council on Measurement in Education (NCME) en sus ediciones de 1985, 1999 y 2014, mismos que permiten identificar aquellos estándares susceptibles de emplearse en la traducción y

adaptación de instrumentos. Posteriormente se presentan las Guidelines for Adapting Educational and Psychological Tests, lineamientos propuestos por la International Test Commission (2012) que en la actualidad representan al principal referente en los procesos de adaptación y traducción de instrumentos. Concluye el capítulo con la presentación de las Directrices para la traducción y adaptación de PISA 2012, documento que representa un claro ejemplo en la definición y operación de procedimientos rigurosos y estandarizados en los procesos de adaptación de test.

En el Capítulo 3 se caracterizan los principales enfoques reconocidos en el proceso de adaptación, específicamente los de traducción, adaptación y ensamblaje. Aquí se presentan los procedimientos y análisis psicométricos recomendados para cada una de los componentes que los conforman, haciendo énfasis en aquéllos orientados a la validación del constructo (análisis factorial confirmatorio y funcionamiento diferencial de los ítems).

En el capítulo 4 se presenta el método, donde se describe a los participantes, el instrumento que se utilizó, el procedimiento y las particularidades de cada una de las etapas de la metodología propuesta. Por su parte, en los capítulos 5 y 6 se presentan y discuten los resultados de cada una de las etapas que conformaron la metodología y se discuten sus implicaciones.

Finalmente, se incluye un apartado de anexos donde se presentan los materiales utilizados en cada una de las etapas (guías de instrucción, hojas de respuesta, presentaciones, etc.), así como el detalle de los resultados del conjunto de análisis estadísticos realizados.

### Antecedentes

La convivencia escolar ha ido adquiriendo relevancia tanto en el campo de investigación científica como en las agendas de los sistemas educativos y los medios de comunicación, convirtiéndose en preocupación de docentes, directivos, padres de familia y otros sectores de la población (Fierro y Caso, 2013; Ortega-Ruiz, Del Rey y Casas, 2013). De acuerdo con Fierro y Caso (2013), el uso del término *convivencia escolar* se generalizó en la literatura y en los programas de política pública a partir del *Informe Delors*, en el cual se recupera el sentido profundo del concepto al asumirlo como uno de los cuatro pilares de la educación (Delors, 1996).

Las distintas e incluso contradictorias aproximaciones y perspectivas empleadas en torno al estudio de la convivencia escolar han contribuido a diversificar el significado del término, lo cual deriva en dificultades para su delimitación conceptual y obstaculiza las bases para el diseño y desarrollo de instrumentos de medición que permitan evaluar este constructo en contextos educativos (Caso, Díaz y Chaparro, 2012; Fierro y Caso, 2013; Ortega-Ruiz et al., 2013).

A dicha problemática se agrega la falsa creencia de que un instrumento que es útil en un contexto o cultura dada, puede resultar útil en otro contexto o cultura. Esta práctica, una de las más graves e ineficaces en el campo de las pruebas psicológicas y educativas, consiste en “pedir prestado” un test de una cultura y adoptarlo en otra (Mereda, 2005).

A este respecto, los enfoques más recientes asociados con la adaptación de un test buscan contribuir a la equidad de las comparaciones entre individuos y grupos de contextos culturales o lingüísticos distintos, facilitando el desarrollo de estudios comparativos en diversos grupos culturales, y permitiendo la comparación de los estudiantes en diferentes países (Cook et al., 2005).

La adaptación de un test incluye desde la decisión de si éste puede o no medir el mismo constructo en un idioma o cultura diferente, hasta la selección de traductores, la preparación del instrumento para su uso en un segundo idioma, y la revisión y aseguramiento de la equivalencia entre la forma adaptada y la versión original del test (Hambleton, 2005).

Así, en el presente trabajo surge una doble oportunidad. Por un lado, siendo la convivencia escolar un campo del conocimiento en construcción (Fierro, 2013), y teniendo en cuenta el papel central que ha adquirido dentro de los conceptos de prácticas y cultura educativa (Ortega-Ruiz, Del Rey y Sánchez, 2012), resulta indispensable contar con instrumentos válidos que permitan medirla, y así contribuir a establecer las bases para su caracterización y para el diseño e implementación de intervenciones específicas en los centros escolares (Caso et al., 2013).

Por otro lado, si bien en Baja California existe registro de estudios orientados a la medición de la convivencia escolar (Caso et al., 2012), los instrumentos utilizados en estos estudios no se han acompañado de los procedimientos y cuidados específicos que aseguren la equivalencia cultural, métrica, lingüística y de constructo, que exige este tipo de evaluaciones, particularmente cuando se utilizan instrumentos validados en una población distinta a la población original.

Por consiguiente y teniendo como marco un *Acuerdo bilateral de colaboración entre el Laboratorio de Estudios sobre Convivencia y Prevención de la Violencia* de la Universidad de Córdoba (España) y el Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California, se decidió utilizar un instrumento desarrollado por Rosario Del Rey, Rosario Ortega Ruiz y José Antonio Casas, académicos adscritos a dicha institución española. Si bien es cierto que el instrumento en cuestión representa un insumo importante en la presente investigación, el foco principal de este proyecto se encuentra en desarrollar, implementar y evaluar

una metodología para la adaptación de instrumentos, que en este caso particular permitirán medir de manera válida y confiable la convivencia escolar en las escuelas de Baja California.

El instrumento comprometido en dicho acuerdo se denomina Escala de Convivencia Escolar (Ortega-Ruiz et al., 2013), el cual se conforma por 50 ítems agrupados en 8 factores que explican distintas dimensiones de la convivencia escolar desde la percepción de los estudiantes, el cual evalúa las siguientes dimensiones: a) Gestión interpersonal positiva; b) Victimización; c) Disruptividad; d) Red social de iguales; e) Agresión; f) Indisciplina; y g) Desidia docente. La escala ha sido validada mediante rigurosos procesos estadísticos que dan cuenta de su idoneidad y consistencia interna. (Ortega-Ruiz et al., 2013).

Con base en lo anterior, el presente estudio comprometió los siguientes objetivos:

## **Objetivos**

### **Objetivo general**

Desarrollar una metodología para la adaptación de instrumentos de medición utilizados en contextos educativos.

### **Objetivos específicos**

- Revisar los fundamentos, estándares y directrices que orientan los procesos de traducción y adaptación de pruebas psicológicas y educativas.
- Proponer e implementar una metodología para la adaptación de instrumentos de medición en contextos educativos.
- Documentar la evidencia registrada al aplicar cada una de las etapas que conforman la metodología propuesta.

### **Justificación**

La necesidad de contar con instrumentos de medición válidos que permitan obtener información sobre procesos, prácticas y resultados educativos generalmente trae consigo la disyuntiva entre diseñar y desarrollar un nuevo instrumento, o bien adaptar alguno que haya sido utilizado en otro contexto o cultura. Ante las ventajas que representa la adaptación de instrumentos (bajos costos, la equivalencia en las puntuaciones, la coyuntura para realizar estudios comparativos, etc.) y la oportunidad de contribuir al desarrollo de esta línea de investigación en nuestro país, se propone el presente estudio.

En lo particular se propone la adaptación de la Escala de Convivencia Escolar desarrollada por un grupo de investigadores españoles, la cual es aplicada a población infantil y adolescente en dicho país. Lo anterior requerirá desarrollar una metodología que incorpore directrices, lineamientos y estándares propuestos por especialistas en la materia, y por organismos internacionales que han emitido recomendaciones específicas a este respecto, entre los que destacan la International Test Commission (ITC), la American Psychological Association (APA), el National Council on Measurement in Education (NCME) y la American Education Research Association (AERA). Por lo tanto, la operacionalización de estas directrices y lineamientos en pautas de acción específicas, que apoyen y orienten a los usuarios y desarrolladores de test, representará una aportación importante a esta línea de investigación aún incipiente en nuestro país.

Finalmente, la adaptación de la ECE permitirá en el mediano plazo caracterizar la convivencia escolar en Baja California, respetando el marco conceptual en que se fundamenta, y posibilitando la realización de estudios comparativos entre México y España.



## **Capítulo 1. La adaptación de instrumentos de medición psicológica y educativa**

Cada vez es más común la aplicación e intercambio de instrumentos de evaluación psicológica o educativa de un país a otro y la utilización de exámenes traducidos a diversos idiomas, lo que ha exigido adaptar pruebas de rendimiento y aptitudes para su uso en múltiples culturas e idiomas (Hambleton, 1996; Klerk, 2008). La adaptación de instrumentos para uso en contextos lingüísticos y culturales diferentes a aquellos en donde fueron construidos, es una práctica que se remonta a los inicios del siglo XX (Muñiz, Elosua y Hambleton, 2013). Actualmente, la utilización de test adaptados se ha incrementado debido a la creciente globalización y a la facilidad y rapidez de uso de los medios de comunicación, pues aumenta la posibilidad de que los instrumentos generados en un país puedan utilizarse con prontitud en uno distinto rápidamente (Muñiz y Hambleton, 1996).

Lo anterior se ve reflejado en el creciente interés por los estudios comparativos internacionales y en investigaciones interculturales, pues en la mayoría de estos se utilizan instrumentos y otros métodos de recolección de información que requieren adaptarse para su uso en otras culturas (Byrne, Oakland, Leong, van de Vijver, Hambleton, Cheung y Bartram, 2009; Hambleton, Mereda y Spielberger, 2005). Tal es el caso del Programa Internacional de Evaluación de los Alumnos (PISA, por sus siglas en inglés), en el cual participan un gran número de países con idiomas, culturas y sistemas educativos diferentes. PISA evalúa las competencias en lectura, matemáticas, ciencia y resolución de problemas, en estudiantes de 15 años de edad, a través de pruebas, encuestas, cuestionarios y manuales traducidos y adaptados en cada uno de los idiomas de instrucción de los países participantes (Organization for Economic Co-operation and Development, 2010). PISA, por la importancia y relevancia de sus resultados en el contexto

internacional, probablemente represente el examen al que más recursos técnicos y presupuestales se le canalizan, lo que en cierta medida garantiza la calidad técnica de sus resultados.

Así, la comunidad científica dedicada a la medición es cada vez más consciente de la necesidad de perfeccionar los métodos para traducir y adaptar estas pruebas. En lo particular, Solano-Flores, Backhoff y Contreras-Niño (2009), han señalado la importancia de comprender cómo estos instrumentos han sido desarrollados en un idioma, país y cultura, a fin de que puedan utilizarse de manera apropiada en otro idioma, país y cultura.

No obstante, es común que los investigadores que utilizan test de otros países no se encuentren familiarizados con los métodos psicométricos necesarios para llevar a cabo un adecuado proceso de adaptación (Byrne et al., 2009).

La decisión de adaptar un test de una cultura a otra, supone que (a) el constructo medido por el test guiará hacia las mismas, o similares, interpretaciones y predicciones en la nueva cultura; y (b) que la población donde se adaptará el test difiere de la población original en términos de cultura o idioma (Geisinger, 1994; Tanzer y Sim, 1999). De este modo, el proceso de adaptación de pruebas permite la obtención de mediciones válidas en cada contexto o cultura y facilita la realización de estudios comparativos entre los grupos involucrados, además, representa una reducción considerable del costo, pues resulta más oneroso crear un nuevo test que adaptar uno existente (Klerk, 2008). Para realizar una adaptación válida se deben considerar aspectos psicológicos, psicométricos, culturales, interculturales y lingüísticos, así como el conocimiento de todas las idiosincrasias específicas de un idioma o cultura que pueda causar sesgo en la evaluación (Tanzer y Sim, 1999).

De acuerdo con el Committee on International Relations in Psychology (APA, 2013), el objetivo de una adaptación es transformar un instrumento existente (test fuente), diseñado para su uso en una población determinada (población fuente), en otro instrumento (test meta) que puede utilizarse de manera equivalente en una población distinta (población meta). Se considera que un test adaptado es sólido psicométricamente si cuenta con pruebas exhaustivas que demuestren su validez, equivalencia lingüística y de constructo con referencia al instrumento y la población originales, y que ha sido normalizado para su uso en la población meta. Una adaptación deficiente puede cambiar significativamente la validez de las puntuaciones o puede hacer que el test resulte más fácil o más difícil en otro idioma o cultura (Hambleton, 1996). Por ejemplo, no se debe suponer que un test desarrollado para su uso en un idioma y país determinado, medirá automáticamente lo mismo en otro idioma o país. Es decir, no se puede asumir que las evidencias de validez y propiedades psicométricas del test fuente prevalecerán en otra cultura o idioma (APA, 2013; Byrne, Stewart y Lee, 2004; Van de Vijver y Hambleton, 1996).

Los test deben adaptarse en función de las diferencias interculturales entre los idiomas y las culturas en cuestión (Muñiz y Hambleton, 1996). De acuerdo a Klerk (2008), se requiere explorar la influencia de la cultura en la medición del constructo, para poder ajustar las mediciones y darles significado en la cultura y población meta, y obtener mediciones equivalentes y comparables. De esta manera, el término adaptación de test se utiliza para describir una aproximación más avanzada en el desarrollo y uso de instrumentos traducidos (Hambleton et al., 2005). La adaptación de test involucra una serie integral y rigurosa de procedimientos que examinan la validez de las puntuaciones obtenidas y su equivalencia estructural y de medida, dentro del nuevo contexto cultural del test objetivo (Byrne et al., 2009).

Lo anterior deja ver la necesidad de contar con procedimientos que permitan asegurar que se atienden todas estas recomendaciones y particularidades asociadas con los procesos de adaptación. Para ello, existen procedimientos y técnicas psicométricas que permiten analizar estos procesos e intentar determinar si los test están midiendo el mismo constructo, si existe equivalencia entre las puntuaciones, si está presente el sesgo a nivel de ítem o del test, y si las puntuaciones son interpretadas de la misma manera en ambas poblaciones (Muñiz y Fernández-Hermida, 2010).

### **Validez y confiabilidad**

Un test con propiedades psicométricas estables es aquel que ha sido examinado rigurosamente para asegurar la medición precisa y verdadera del constructo subyacente para cuya medición fue diseñado, y el cual produce resultados consistentes en cada aplicación a distintas muestras de la misma población (APA, 2013). La Comisión Internacional de los Test (International Test Commission, 2010), sostuvo que los test deben respaldarse por evidencias de confiabilidad y validez, las cuales permitan dar fundamento a las inferencias que se puedan realizar a partir de las puntuaciones obtenidas. Asimismo, sí el instrumento es aplicado en una nueva población, se requiere demostrar que este continua evaluando las mismas cualidades con el mismo grado de precisión y exactitud (Geisinger, 1994).

La validez se considera la propiedad más importante y fundamental de un test, y hace referencia tanto al proceso de validación de las puntuaciones como al desarrollo y usos del mismo. A través del análisis de las variables definidas en el test y su relación con otras variables, los estudios de validez aportan significado a las puntuaciones obtenidas y nos permiten conocer si es adecuado el uso que se pretender hacer de ellas (Abad, Olea, Ponsoda y García, 2011; Elosua, 2003). Actualmente, la validez es entendida como el grado en que la teoría y los datos apoyan la

interpretación de las puntuaciones de un test para un uso concreto (Abad et al., 2011; American Education Research Association et al., 1999), por lo que se habla de la validez de las puntuaciones de un test y no de la validez de un test, ya que las propiedades de éste dependen del contexto de evaluación y de la muestra.

El proceso de validación debe iniciar con (a) el desarrollo de un argumento interpretativo, en donde se propongan con detalle las interpretaciones y usos de las puntuaciones y se especifiquen los objetivos del test, los componentes del constructo, las diferencias con otros constructos y su relación con otras variables, (b) posteriormente se debe desarrollar el argumento de validez, en donde se evalúan las interpretaciones propuestas a través de estudios empíricos y la obtención de diferentes tipos de evidencias de validez (Abad et al., 2011).

A este respecto, los *Standards for Educational and Psychological Testing* (American Education Research Association et al., 1999), mismos que han normado el diseño y desarrollo de instrumentos de medición educativa y psicológica durante los últimos 15 años, plantearon el estudio de evidencias relacionadas con el contenido, la estructura interna, la relación con otras variables, el proceso de respuesta y las consecuencias de aplicación del test, mismas que se comentan a continuación:

- *Evidencias basadas en el contenido del test.* Se busca garantizar que el conjunto de ítems que conforman el test sean relevantes y representativos del constructo que se pretende medir. Los estudios de validación del contenido están conformados por la definición operacional del constructo, realizada generalmente a través de una tabla de especificaciones de las áreas que se pretenden evaluar. Por su parte, el estudio de la

representación del constructo evalúa la adecuación general del test y el grado en que cada ítem se adecua con el contenido que representa el test. Estos estudios requieren del trabajo conjunto de jueces o expertos los cuales emiten juicios sobre el grado de adecuación observado entre los ítems y el contenido definido.

- *Evidencias basadas en la estructura interna del test.* Evalúan el grado de relación que tienen los ítems para medir el constructo subyacente al test. A través de estudios de dimensionalidad y funcionamiento diferencial de los ítems, se busca comparar la estructura observada del test con la estructura teórica previamente definida, e identificar aquellos ítems donde los sujetos que pertenecen a grupos distintos y tienen el mismo nivel de rasgo medido por el test, presentan distinta probabilidad de acertar o estar de acuerdo con éstos.
- *Evidencias basadas en la relación con otras variables.* También es imprescindible analizar si las relaciones entre las puntuaciones obtenidas por el test y otras variables externas son consistentes con la interpretación propuesta. De acuerdo a Abad et al. (2011), las variables externas pueden ser otras puntuaciones de test que miden el mismo constructo (evidencia convergente), puntuaciones en constructos diferentes pero que se encuentran dentro del modelo teórico del test (evidencia discriminante), o bien puntuaciones en algún tipo de variable que se pretenda predecir a partir de las puntuaciones obtenidas (evidencia referida a un criterio).
- *Evidencias basadas en los procesos de respuesta a los ítems.* Se analizan los procesos involucrados en la resolución de los ítems que conforman un test, lo que permite obtener información sobre las inferencias que se pueden realizar a partir de las puntuaciones obtenidas. En específico, el estudio de los procesos cognitivos debe orientarse por un

modelo explicativo de tales procesos que debe guiar el desarrollo del test y ser útil para predecir el rendimiento diferencia de los ítems.

- *Evidencias basadas en las consecuencias de la aplicación del test.* Existe una controversia entre autores a favor (Cronbach, 1988; Messick, 1995) y en contra (Borsboom, Mellenbergh y van Heerden, 2004), en torno a la consideración de este tipo de información como evidencias de validez. Los *Standards for Educational and Psychological Testing* de 1999 plantearon analizar las consecuencias del uso de los test, diferenciando entre aquellas que tienen que ver con su validez y las que no. Si la evidencia empírica permite mantener las interpretaciones, la decisión sobre el uso del test puede considerar aspectos sociales o políticos que no forman parte del proceso de validación (Abad et al., 2011).

En lo particular, la validez de un test adaptado debe estudiarse después de haber concluido su adaptación en un nuevo grupo lingüístico o cultural, para determinar si mide el mismo constructo (Geinsinger, 1994). Tanzer y Sim (1999) utilizaron el término validez ecológica para hacer referencia a la adecuación del test en distintas poblaciones, destacando la importancia de la familiaridad de los sujetos con el idioma y con el diseño y contenido del test y los ítems, así como la correspondencia entre el nivel de dificultad del test y el nivel de habilidad de los sujetos.

La confiabilidad es otra propiedad indispensable de un test. Se refiere al grado en que las puntuaciones obtenidas son consistentes para los mismos sujetos en diferentes condiciones, momentos y formas del test (Abad et al., 2011; Hambleton et al., 2005). El término confiabilidad ha sido utilizado en dos sentidos por la literatura en medición. Por un lado se ha utilizado para referirse a los coeficientes de confiabilidad de la Teoría Clásica de los Test (TCT) entendidos como (a) la correlación entre formas paralelas del test si se replican las mismas puntuaciones al aplicar

el test con ítems distintos; (b) la estabilidad temporal del mismo, es decir, si se obtienen las mismas puntuaciones cuando se miden en distintos momentos; o (c) la consistencia interna del conjunto de ítems, si se analiza el grado en que los distintos ítems miden el constructo en cuestión, estudiada principalmente a través del coeficiente alfa de Cronbach (American Education Research Association et al., 2014). Por otra parte, el término también ha sido empleado de una manera más general para referirse a la consistencia de las puntuaciones a través de la réplica de los procedimientos de aplicación, sin importar cómo es estimada la consistencia (por ejemplo, en términos de errores estándar, coeficientes de generalizabilidad, y funciones de información de la Teoría de la Respuesta al Ítem) (American Education Research Association et al., 2014). De acuerdo a la más reciente edición de los *Standards* (American Education Research Association et al., 2014), para mantener un vínculo entre las diferentes nociones de confiabilidad y evitar ambigüedad en el uso del término, se recomendó utilizar el término *Confiabilidad/Precisión* para indicar una noción más general de la consistencia de las puntuaciones a través de diversas aplicaciones y procedimientos de prueba, y el término *Coefficiente de confiabilidad* para referirse a los coeficientes de confiabilidad de la Teoría Clásica de los Test.

En este sentido, existen métodos alternativos para estimar el coeficiente de confiabilidad (Abad et al., 2011), los cuales superan las limitaciones que tienen los métodos clásicos (equivalencia entre formas, dos mitades y consistencia interna). Específicamente, existe debate en relación a la utilidad del coeficiente alfa para informar acerca de la consistencia interna de escalas compuestas por variables ordinales (Gadermann, Guhn y Zumbo, 2012), y de escalas con una estructura multidimensional (Reise, Moore y Haviland, 2010). En consecuencia, se propone estimar el coeficiente de consistencia interna, ya sea a través del coeficiente alfa ordinal (para

variables ordinales) o del análisis factorial (coeficiente omega) para escalas con estructura multidimensional.

El concepto de confiabilidad entendido desde la Teoría Clásica de los Test (TCT), se concibe bajo los supuestos de que las mediciones contienen error y que el error es aleatorio (Hambleton et al., 2005), en donde este último hace referencia a los cambios en la puntuación de una persona de una medición a otra (Abad et al., 2011). De esta manera, los errores pueden causarse por (a) factores transitorios consecuencia de los cambios entre una aplicación y otra; (b) debido a la especificidad del contenido de los ítems; o (c) por errores causados por factores aleatorios.

Otro indicador de confiabilidad o precisión es el error típico de medida (ETM), el cual indica el grado en que un test proporciona a cada persona una puntuación cercana a su nivel de rasgo (Abad et al., 2011; Hambleton, 2005). Con base en el ETM se pueden generar intervalos de confianza sobre las puntuaciones obtenidas, y cuando se tiene el objetivo de interpretar puntuaciones, resulta más informativo que los coeficientes de confiabilidad o generalizabilidad pues permite conocer el grado de error que contienen las puntuaciones obtenidas (American Education Research Association et al., 2014; Elosua, 2012). Por último, resulta importante estimar el efecto del ETM de modo diferenciado, sobre todo si el test aporta criterios de interpretación distintos en función de alguna variable relevante como el sexo o la edad (Elosua, 2012).

Al igual que en los estudios de evidencias de validez, la confiabilidad/precisión, los índices de confiabilidad y el ETM deben calcularse y estudiarse cuando un test es traducido o adaptado para su uso en un nuevo grupo lingüístico o cultural (American Education Research Association et al., 2014; Geisinger, 1994).

### **Equivalencia y Sesgo**

El sesgo y la equivalencia son conceptos centrales en el proceso de adaptación de un test (Van de Vijver y Tanzer, 2004). De acuerdo a Klerk (2008), el test y las adaptaciones utilizadas en diferentes grupos culturales deben encontrarse tan libres de sesgo como sea posible para asegurar su equivalencia y, en consecuencia, maximizar la comparabilidad de las puntuaciones.

Aunque existen disparidades en la conceptualización de lo que se entiende por equivalencia, usualmente es entendida desde un enfoque de medición con un fuerte énfasis en análisis estadísticos (Byrne et al., 2009; Drasgow, 1984). Las directrices para la traducción y adaptación de los tests propuestas por Muñiz et al., (2013), exigen documentar evidencias sobre la equivalencia de constructo, de método y de los ítems (segunda directriz de confirmación, p. 154). En este mismo sentido, He y van de Vijver (2012) presentaron una taxonomía de la equivalencia, que aborda cuestiones de sesgo y comparabilidad de constructos y de puntuaciones, destacando tres tipos de equivalencia: de constructo, de medición y escalar.

La equivalencia de constructo, también entendida como equivalencia estructural, equivalencia funcional (Byrne et al., 2009) o equivalencia conceptual (Van de Vijver y Tanzer, 2004), se refiere al grado en que el constructo es medido a través de todos los grupos culturales, sin importar si la medición está basada en instrumentos idénticos (Byrne et al., 2009; Tanzer y Sim, 1999; Van de Vijver y Tanzer, 2004). Lo anterior implica que tanto la definición conceptual como la estructura del constructo se pueden generalizar entre las poblaciones de interés (Tanzer y Sim, 1999). Por lo tanto, si no se asegura la equivalencia de constructo, no existen bases para ninguna comparación intercultural (He y van de Vijver, 2012). Los estudios de equivalencia de constructo se centran en las variables no observadas o variables latentes, y se interesan en las

relaciones entre factores, las covarianzas de los factores y los errores de covarianzas residuales (Byrne, 2008).

Por su parte, la equivalencia de medición se interesa por el grado en el cual el contenido del ítem y las propiedades psicométricas del test son similares entre los grupos (Byrne et al., 2009), es decir, el grado en que el contenido de cada ítem está siendo percibido e interpretado de la misma manera entre los grupos considerados (Byrne y Watkins, 2003). Las pruebas de equivalencia en la medición se enfocan en las variables observadas, por lo que se puede examinar tanto el funcionamiento invariante de los ítems entre poblaciones, como la forma en la cual el contenido de cada ítem es interpretado entre las muestras.

Así, la equivalencia de medición considera tanto la equivalencia métrica como la equivalencia de unidades de medición. De acuerdo con Drasgow (1984), se habla de equivalencia métrica cuando las relaciones entre las puntuaciones observadas y la variable latente (medida por el test) son idénticas entre las poblaciones consideradas. Cabe mencionar que la equivalencia métrica no puede reducirse a la comparación de coeficiente de confiabilidad, de validez o de estructuras factoriales (Elosua y López, 1999). Por otra parte, cuando dos mediciones tienen las mismas unidades de medida pero diferente origen se debe obtener la equivalencia de las unidades de medición (He y van de Vijver, 2012; Van de Vijver y Tanzer, 2004). Al asegurar este tipo de equivalencia, resulta válido realizar comparaciones directas de las puntuaciones entre el test original y el test adaptado, al interior de cada grupo cultural (He y van de Vijver, 2012).

Por último, la equivalencia escalar, es considerada el máximo nivel de equivalencia y puede obtenerse cuando dos escalas tienen las mismas unidades de medida y están libres de sesgo. Por lo tanto las puntuaciones del instrumento original y el adaptado pueden compararse directamente (He

y van de Vijver, 2012; Tanzer y Sim, 1999; Van de Vijver y Tanzer, 2004). Para la obtención de la equivalencia escalar es necesario (a) que la equivalencia de constructo y de medición sean establecidas, (b) que la relación entre el rasgo latente y las puntuaciones del test sean idénticas en ambas poblaciones y (c) eliminar o reducir (a una cantidad razonable) todos los factores que perturben la medición del constructo.

Por otro lado, diversos autores identifican a la equivalencia intercultural como indispensable dentro de los estudios interculturales y en el contexto de adaptación de test (Byrne, 2008; Byrne et al., 2009; Muñiz et al., 2013; Tanzer y Sim, 1999; Tanzer, 1995). Este nivel de equivalencia se refiere a la adecuación de todos los componentes de un test para todas las poblaciones de interés, el cual resalta la importancia de la familiaridad de los sujetos con el idioma y lenguaje utilizado en el test, así como el grado en que el contenido del ítem y la estructura del constructo son percibidos de manera similar entre grupos.

A este respecto, de acuerdo a Byrne (2008), las diferencias en las normas y valores sociales y las prácticas de socialización pueden crear diferencias culturales en el significado o estructura de un constructo y en la percepción del contenido asociado a cada ítem. Cuando un constructo psicológico es medido en diferentes grupos culturales, no se puede asumir que el significado de las puntuaciones será idéntico y que estas pueden compararse entre sí. De esta manera, los fenómenos sociales pueden contribuir a la presencia (o no) de equivalencia de medición o equivalencia estructural (Byrne et al., 2009).

La equivalencia puede estudiarse a través del empleo de varias fuentes, como el análisis del significado de los comportamientos medidos, las propiedades diferenciadas de la escala o el test y los diferentes tipos de sesgo. Dichos procedimientos, tal como lo señaló Elosua (2012),

pueden involucrar tanto una aproximación cualitativa (procedimientos de juicio, análisis de contenido, entrevistas), como una aproximación cuantitativa (índice de congruencia de Tucker, modelos de invarianza derivados de TRI o de modelos de ecuaciones estructurales).

El proceso de adaptación de instrumentos de medida de una cultura a otra, es un proceso complejo que se encuentra expuesto a una gran cantidad de fuentes de error (Hambleton, 1996; Muñiz y Hambleton, 1996; Sireci, Patsula y Hambleton, 2005; Van de Vijver y Hambleton, 1996; Van de Vijver y Poortinga, 2002).

Por su parte, el sesgo es un término que hace referencia a los factores que ponen en riesgo la validez del instrumento, aplicado en diferente culturas, y que comprometen las comparaciones de las puntuaciones obtenidas (APA, 2013; He y van de Vijver, 2012; Van de Vijver y Tanzer, 2004). De acuerdo a Van de Vijver y Leung (1997) el sesgo ocurre si las diferencias de las puntuaciones en los indicadores de un constructo no corresponden con las diferencias en la habilidad o rasgo subyacente (como se citó en Klerk, 2008).

El sesgo es un error sistemático de medición que produce una distorsión en el significado de las puntuaciones y altera las interpretaciones propuestas (Elosua, 2003; He y van de Vijver, 2012). Los usos del término sesgo hacen referencia a la utilización de un instrumento de medición dentro del marco de una aplicación particular de comparaciones interculturales. Así, los problemas de sesgo pueden ser de tres tipos (a) sesgo de constructo, (b) sesgo de método y (c) sesgo de ítem (Byrne y Watkins, 2003; He y van de Vijver, 2012); y pueden ser determinados o producidos por diversas fuentes (Tabla 1).

Tabla 1  
*Tipos y fuentes de sesgo*

Tipo de sesgo	Descripción	Fuente de sesgo
Sesgo de constructo	Superposición incompleta del constructo en los grupos culturales	-Diferencias en las definiciones del constructo entre culturas. -Apropiación diferencial de las conductas asociadas con el constructo.
Sesgo de método	Término general para todos los factores de ruido que surgen de aspectos del método	
Sesgo de muestra	Falta de comparabilidad de las muestras	-Muestras no comparables. -Familiaridad diferencial con los estímulos/materiales.
Sesgo de instrumento	Características del instrumento, no relacionadas con el constructo, que provocan diferencias en las puntuaciones.	-Familiaridad diferencial con los procedimientos de respuesta. -Estilos de respuestas diferenciales. -Diferencias, físicas o sociales, en las condiciones del entorno de aplicación.
Sesgo de aplicación	Errores de comunicación entre los aplicadores del test y los examinados.	
Sesgo de ítem	Anomalías del ítem	-Pobre traducción o ítems ambiguos. -Factores de ruido asociados con el ítem.

Nota: Elaboración propia basada en “Conceptual and Methodological Issues in Adapting Tests” por F. van de Vijver y Y. Pootinga, en “Adapting Educational and Psychological Tests for Cross-Cultural Assessment” por R. Hambleton, P. Mereda y C. Spielberger, 2005, Lawrence Erlbaum Associates, New Jersey, EUA.

El sesgo de constructo hace referencia a que el constructo medido por un instrumento presenta diferencias importantes entre los grupos culturales considerados (APA, 2013; He y van de Vijver, 2012; Van de Vijver y Hambleton, 1996; Van de Vijver y Tanzer, 2004). Este tipo de sesgo puede surgir como consecuencia de (a) las diferencias de los comportamientos considerados como indicador del constructo entre los grupos culturales, debidas a que las conductas no se encuentran representadas apropiadamente en cada cultura (APA, 2013; Byrne y Watkins, 2003; He y van de Vijver, 2012; Van de Vijver y Hambleton, 1996); y (b) el grado en el cual las dimensiones del constructo, del que se desprende la formulación del contenido de los ítems, varían entre los grupos culturales (APA, 2013; Byrne y Watkins, 2003; Van de Vijver y Hambleton, 1996).

Por su parte, el sesgo de método es un término general que refiere al sesgo de muestra, de instrumento y de aplicación, relacionados con la estrategia metodológica empleada (APA, 2013; Van de Vijver y Hambleton, 1996). El sesgo de muestra se asocia a las limitaciones que se presentan en hacer comparables las muestras, como las variaciones interculturales en las características de la muestra (p. e. variaciones en la residencia urbana o rural, afiliación a un grupo religioso, etc.) que inciden en la medición del constructo. En tanto, el sesgo de instrumento se refiere a los problemas asociados a las características de los instrumentos de evaluación. Algunas fuentes comunes de sesgo de instrumento son (a) cuando un grupo de comparación tiene mayor familiaridad con la presentación de los estímulos considerados en el test, (b) cuando los sujetos emplean patrones de respuesta para contestar el test, y (c) cuando los sujetos seleccionan puntos de la escala que llevan a dar una impresión favorable de sí mismos. En el mismo sentido, el sesgo de administración deriva de las diferencias en la aplicación de un instrumento entre los grupos considerados. Este tipo de sesgo se puede presentar si existen diferencias en el formato de presentación de los ítems, en el formato de administración del test, o en el formato de las instrucciones para resolver el instrumento (APA, 2013; He y van de Vijver, 2012; Van de Vijver y Tanzer, 2004).

Finalmente, el sesgo de ítem se refiere a las distorsiones a nivel de ítem (APA, 2013; Byrne y Watkins, 2003; Van de Vijver y Hambleton, 1996; Van de Vijver y Tanzer, 2004). Un ítem se encuentra sesgado si las personas de diferentes grupos pero con el mismo nivel de rasgo no tienen la misma probabilidad de emitir la misma respuesta al ítem (APA, 2013; Van de Vijver y Hambleton, 1996). El sesgo de ítem se debe principalmente a que se le da un significado diferencial al contenido del ítem (Byrne y Watkins, 2003; He y van de Vijver, 2012); o bien, puede ser resultado de un pobre fraseo del ítem, de una traducción inadecuada (Van de Vijver y Hambleton,

1996), de ambigüedades en el ítem original, de idiosincrasias lingüísticas, o de la familiaridad o proximidad entre los sujetos y el contenido del ítem (He y van de Vijver, 2012; Van de Vijver y Tanzer, 2004).

Ignorar el impacto del sesgo pone en riesgo la validez de las inferencias (Van de Vijver y Tanzer, 2004). Por lo tanto, el sesgo y la validez se convierten en aspectos afines, ya que el sesgo supone falta de validez y la falta de validez puede ser el origen del sesgo (Elosua, 2003). Es decir, debe demostrarse la ausencia de sesgo para poder realizar inferencias válidas.

Con el fin de maximizar la validez del instrumento de medición, se han empleado distintas estrategias que permiten minimizar o inclusive remover el sesgo, mismas que se encuentran apoyadas en (a) procedimientos derivados de la Teoría Clásica de los Test, que permiten explicar las respuestas dentro de una cultura; (b) procedimientos de la Teoría de la Respuesta al Ítem, que proveen información acerca de la calidad de la traducción, el significado de los ítems en relación al constructo que se desea medir, la equivalencia del test y de las muestras entre diferentes grupos culturales, y la identificación de ítems que necesitan algún tipo de revisión; y (c) aproximaciones enfocadas en el desarrollo e implementación de los test (Hulin y Mayer, 1986; Van de Vijver y Tanzer, 2004).

Van de vijver y Tanzer (2004) identificaron las técnicas y procedimientos más sobresalientes para tratar con cada uno de los tres tipos de sesgo. Así, para tratar con el sesgo de constructo, se destacan el “descentramiento” cultural y el enfoque de “convergencia”. En el primero, las palabras y conceptos que son específicos de una cultura o idioma particular son eliminados, y posteriormente se desarrolla de manera simultánea el mismo instrumento en cada cultura. Por su parte, el enfoque de convergencia consiste en el desarrollo independiente de

instrumentos dentro de la cultura y en la posterior aplicación intercultural de todos los instrumentos (He y van de Vijver, 2012; Van de Vijver y Tanzer, 2004). De acuerdo con Van de vijver y Hambleton (1996), para evitar de manera eficiente las tendencias etnocéntricas de los instrumentos, se puede emplear un equipo multicultural y multilingüístico con experiencia en el constructo bajo estudio.

En este sentido, Van de Vijver y Tanzer (2004) identificaron los siguientes procedimientos para tratar simultáneamente con el sesgo de constructo y el sesgo de método: (a) uso de informantes con un conocimiento profundo de la cultura y el idioma local; (b) aplicación del instrumento a grupos de sujetos bilingües; y (c) uso de encuestas locales para conocer el grado en el cual los estímulos considerados se asocian con el constructo que se desea estudiar (si el instrumento va a aplicarse por primera vez en un grupo cultural). También, la aplicación no estandarizada del instrumento puede brindar información acerca de cómo interpretan los sujetos las instrucciones y los ítems, en particular las alternativas de respuesta, y las motivaciones para responder (Van de Vijver y Hambleton, 1996). Por último, se recomienda comparar el marco nomológico desde diferentes métodos que permitan examinar la consistencia de los hallazgos entre grupos culturales y la validez convergente y discriminante del constructo (Van de Vijver y Tanzer, 2004).

Por su parte, los procedimientos o estrategias que tratan con el sesgo de método son: (a) capacitación o entrenamiento de las personas encargadas de aplicar los instrumentos para asegurar que la administración se lleve a cabo de la misma manera; (b) el desarrollo de un manual o protocolo que ayude a resolver problemas comunes de aplicación; (c) la elaboración de instrucciones detalladas que permitan una comunicación precisa entre todos los involucrados; (d) la consideración de variables contextuales y del sujeto que pueden influir en la variación del

constructo bajo estudio; (e) el uso de información adicional como el tiempo de administración o el estudio de los factores de motivación y disposición de los sujetos para realizar el test; y (f) la evaluación de los estilos de respuesta y el empleo de estudios test-retest, de entrenamiento o de intervención, que permitan identificar cambios en los patrones de respuesta (Van de Vijver y Tanzer, 2004).

Por último, los procedimientos que permiten evaluar el sesgo de ítem son: (a) el empleo de métodos de jueceo lingüístico o psicológico (para identificar ítems potencialmente sesgados); (b) el uso de métodos psicométricos (para identificar ítems sesgados); (c) el análisis de los errores o distractores (que permite dar una idea de los procesos cognitivos involucrados en los procesos de solución); (d) y la utilización de un banco de ítems (para sustituir los ítems que presentan sesgo).

Los procedimientos mencionados proveen información acerca de la viabilidad, comparabilidad y posibles defectos del diseño metodológico empleado. De acuerdo a Hambleton (2001) conocer más sobre la naturaleza de los problemas, cuando se adapta un test de un idioma o cultura a otro, puede ayudar a minimizar los problemas que ocurren durante el proceso de adaptación.

## **Capítulo 2. Estándares y directrices que orientan el desarrollo y adaptación de instrumentos de medición**

La internacionalización del uso de los test ha puesto de manifiesto la necesidad de contar con un marco general de evaluación que recoja las buenas prácticas de evaluación (Hambleton, 2004). De manera particular, los desarrollos técnicos, metodológicos y analíticos en torno al desarrollo y adaptación de instrumentos de medición psicológica y educativa son un referente imprescindible en las prácticas de evaluación. Sin embargo, tanto la literatura técnica como los métodos de medición no son bien conocidos por los investigadores que realizan desarrollos o adaptaciones de instrumentos (Hambleton, 1996). En consecuencia, organismos como la International Test Commission (ITC), la American Educational Research Association (AERA), la American Psychological Association (APA) y el National Council on Measurement in Education (NCME), han impulsado y diseminado el empleo de estándares y directrices que orientan las prácticas de evaluación en general, y en particular las de desarrollo, adaptación, validación, administración y uso de instrumentos de medición psicológica y educativa.

Así, los Standards for Educational and Psychological Testing (American Education Research Association, American Psychological Association y National Council on Measurement in Education, 1985, 1999, 2014) y las Guidelines for Adapting Educational and Psychological Tests (International Test Commission, 2010), son los principales documentos que prescriben los procesos de adaptación y desarrollo de instrumentos. A su vez, estos documentos han motivado el desarrollo de directrices específicas para la traducción y adaptación de pruebas utilizadas en evaluaciones educativas internacionales como PISA (Organization for Economic Co-operation and Development, 2010).

No obstante, dichos documentos son y han sido susceptibles de modificación a la luz de los avances metodológicos y sociales (Elosua, 2012; Muñiz, Fernández-Hermida, Fonseca-Pedrero, Campillo-Álvarez y Peña-Suárez, 2011), pues han presentado limitaciones para guiar los distintos tipos de razonamiento utilizados por los traductores o revisores de traducción de pruebas (Instituto Nacional para la Evaluación de la Educación, 2011; Solano-Flores, Backhoff y Contreras-Niño, 2009), y para orientar los métodos y técnicas necesarios para el establecimiento de la equivalencia intercultural (Muñiz, Elosua y Hambleton, 2013).

A continuación, se analizarán los documentos antes mencionados resaltando las aportaciones de éstos a los procesos de adaptación de instrumentos de medición psicológica y educativa.

### **Standards for Educational and Psychological Testing**

En respuesta a los problemas relacionados con los instrumentos de medición y con base en los avances técnicos en materia de medición, en los nuevos y emergentes usos de los test, y en la creciente preocupación social sobre el papel que juegan los test en el cumplimiento de políticas públicas, en 1985 la American Educational Research Association (AERA), la American Psychological Association (APA) y el National Council on Measurement in Education (NCME), coordinaron el trabajo de más de 125 especialistas para actualizar y sustituir los estándares, directrices y principios que orientan el diseño, desarrollo, adaptación y validación de instrumentos. Como consecuencia de este trabajo se obtuvieron 180 estándares divididos en 16 capítulos o secciones, que tratan la mayoría de los usos, aspectos técnicos y un amplio espectro de cuestiones legales y sociales asociadas a los test (Tabla 2). El objetivo de la publicación de los estándares fue proporcionar los criterios mínimos para la evaluación de los test, orientar las prácticas de

evaluación y regular el efecto del uso de los test. Sin embargo, no prescriben el uso de un método estadístico específico (American Education Research Association et al., 1985).

Tabla 2

*Estructura de los estándares para las pruebas psicológicas y educativas (1985)*

Sección	Capítulo	Número de estándares
I. Estándares técnicos sobre la construcción y evaluación de los test	1. Validez	25
	2. Confiabilidad y errores de medida	12
	3. Desarrollo y revisión de un test	25
	4. Escalamiento, normalización, comparabilidad de puntuaciones y equiparación	9
	5. Publicación de test: Manuales técnicos y Guías de usuario	11
II. Estándares profesionales para el uso de los test	6. Principios generales del uso de los test	13
	7. Pruebas clínicas	6
	8. Pruebas psicológicas y educativas	12
	9. Uso de los test en psicoterapia	9
	10. Pruebas utilizadas en contextos laborales	9
	11. Acreditación y certificación profesional y ocupacional	5
III. Estándares para aplicaciones particulares	12. Evaluación de programas	8
	13. Pruebas para minorías lingüísticas	7
IV. Estándares para procedimientos administrativos	14. Pruebas para personas que tienen una condición de discapacidad	8
	15. Aplicación, calificación y reporte de un test	11
	16. Protección de los derechos de los examinados	10
Total		180

Nota: Elaboración propia, basada en Standards for Educational and Psychological Testing (AERA, APA y NCME, 1985)

Estos estándares sostienen que no se puede asumir que al atender procedimientos de traducción se genera una versión del test equivalente en contenido, nivel de dificultad, confiabilidad y validez. De igual modo, no se debe suponer que las propiedades psicométricas de dos versiones de un instrumento sean comparables entre idiomas o dialectos. En este sentido, la edición comprometió 9 estándares distribuidos en 4 capítulos (Tabla 3), que deben considerarse en los procesos de adaptación de test y evaluaciones interculturales (Geisinger, 1994).

Tabla 3

*Estándares aplicables en los procesos de adaptación de test (1985)*

Sección	Capítulo	No. de estándares
I. Estándares técnicos sobre la construcción y evaluación de los test	2. Confiabilidad y errores de medida	1
	3. Desarrollo y revisión de un test	2
II. Estándares profesionales para el uso de los test	6. Principios generales del uso de los test	2
III. Estándares para aplicaciones particulares	13. Pruebas para minorías lingüísticas	4
Total		9

Nota: Elaboración propia, basada en Standards for Educational and Psychological Testing (AERA, APA y NCME, 1985)

En cuanto a la confiabilidad y el error típico de medida, se propuso que cuando existan razones teóricas o empíricas para esperar que estos indicadores difieran sustancialmente en diferentes poblaciones, se deben presentar las estimaciones para cada población para las cuales el test es recomendado [Estándar 2.9 p. 22]. Asimismo, para el desarrollo y revisión de los test se estableció que el tipo y contenido de los ítems debe realizarse con reconocimiento a los antecedentes culturales y a las experiencias previas de las poblaciones involucradas (grupos étnicos, edad, género, etc.) [Estándar 3.5, p. 26]. De igual modo se debe prever la realización de aquellos estudios que sean necesarios cuando los resultados de las investigaciones señalen la necesidad de estudiar las diferencias en el funcionamiento del ítem, o del test, para miembros de distintos grupos (edad, étnicos, género, etc.). Tales estudios deben diseñarse para detectar y eliminar aspectos del diseño, contenido o formato del test que puedan sesgar las puntuaciones de un grupo en particular [Estándar 3.10, p.27].

Por otro lado, se establecieron principios generales para el uso de los test, mismos que recomiendan que cuando se realizan cambios sustanciales en el formato del test, modo de aplicación, instrucciones, idioma o contenido, se debe validar el uso del test para las condiciones

modificadas, o bien tener un sustento racional de que una validación adicional no es necesaria [Estándar 6.2, p.41]. Además, cuando un test se va utilizar para un propósito para el que no ha sido previamente validado, o para el cual no existe evidencia de su validez, los responsables del test deben proporcionarla [Estándar 6.3, p. 42].

Finalmente, el capítulo que establece el uso de test en sujetos que forman parte de una minoría lingüística, exige que cuando un test sea traducido de un idioma o dialecto a otro, se debe establecer la validez y confiabilidad del test para los usos esperados en el grupo lingüístico por examinar [Estándar 13.4, p.75]. Mientras que cuando se busca que las versiones del test en dos idiomas del test sean comparables, se deben reportar evidencias de la comparabilidad [Estándar 13.6, p75]. A este respecto, tanto los responsables del test como los editores deben proporcionar la información necesaria para realizar un uso e interpretación apropiada del test [Estándar 13.3, p.75], y describir en el manual del test todas las modificaciones lingüísticas recomendadas por los editores [Estándar 13.2, p.75].

### **Standards for Educational and Psychological Testing de 1999**

Al igual que la edición de 1985, los estándares de la edición de 1999 contaban con un carácter prescriptivo y por sí solos no incluían mecanismos de aplicación. Fueron formulados con la pretensión de ser consistentes con otros estándares, directrices y códigos de conductas publicados por las organizaciones patrocinadoras; para promover el uso racional y ético de los test, y para proporcionar las bases para la evaluación de la calidad de las prácticas de evaluación. En relación con la edición anterior, la estructura y capítulos de los *Standards* 1999 se modificaron y se incrementaron notablemente (Tabla 4), debido principalmente a lo siguiente: (a) los nuevos tipos

de test y usos para los test existentes; (b) la inclusión de aspectos que procuran la evitación de los conflictos de interés y (c) el tratamiento equitativo de todos los examinados.

Tabla 4

*Estructura de los estándares para las pruebas psicológicas y educativas (1999)*

Sección	Capítulos	No. de estándares
I. Construcción, evaluación y documentación del test.	1. Validez	24
	2. Confiabilidad y errores de medida	20
	3. Construcción y revisión de test	27
	4. Escalas, normas y comparabilidad de puntuaciones	21
	5. Aplicación, calificación y reporte de resultados del test	16
	6. Documentación de apoyo de los test	15
II. Equidad de los test	7. Equidad en las pruebas y en el uso de las mismas	12
	8. Derechos y responsabilidades de los examinados	13
	9. Pruebas en sujetos de diversos contextos lingüísticos	11
	10. Pruebas en sujetos con discapacidad	12
III. Aplicaciones de los test	11. Responsabilidades de los responsables del test	24
	12. Evaluación y pruebas psicológicas	20
	13. Evaluación y pruebas educativas	19
	14. Pruebas de empleo y acreditación	17
	15. Pruebas en programas de evaluación y de política pública	13
Total		264

Nota: Elaboración propia, basada en Standards for Educational and Psychological Testing (AERA, APA y NCME, 1999)

En la edición de 1999 se propusieron nueve estándares que versan sobre los procesos de adaptación de test y de las evaluaciones interculturales, distribuidos en 4 capítulos (Tabla 5). En esta edición se propusieron nuevos planteamientos entre los que destacan: (a) la realización del cálculo de la confiabilidad y de los errores de medida cuando se enfatice la diferencia entre dos puntuaciones observadas de sujetos o de grupos [Estándar 2.3, p. 32]; (b) la inclusión de la función de información del test (FIT), la apropiación del contenido, la estructura interna de las respuestas, la relación de las puntuaciones del test con otras variables y los procesos de respuesta utilizados por los examinados como criterio de variación y comparación de los subgrupos que conforman la muestra [Estándar 2.11, p. 34 y Estándar 7.1, p.80]; (c) la participación de los responsables de la

construcción del test en la identificación y eliminación del vocabulario, símbolos, palabras, frases y contenido que se consideren ofensivos por parte de miembros de un grupo racial, étnico, de género u otros grupos, excepto cuando se juzgue necesario para una adecuada representación del dominio [Estándar 7.4, p. 82]; y (d) la descripción y presentación de evidencia empírica y lógica del método utilizado para establecer la adecuación de la traducción [Estándar 9.7, p.99].

Tabla 5

*Estándares de la edición de 1999 aplicables en los procesos de adaptación de test.*

Sección	Capítulos	No. de estándares
I. Construcción, evaluación y documentación de test.	1. Validez.	1
	2. Confiabilidad y errores de medida.	2
	7. Equidad en las pruebas y en el uso.	3
II. Equidad de los test	9. Pruebas en sujetos de diversos contextos lingüísticos.	3
	<b>Total</b>	<b>9</b>

Nota: Elaboración propia, basada en los Standards for Educational and Psychological Testing (AERA, APA y NCME, 1999)

Es importante destacar que ésta versión retoma el concepto de sesgo expuesto con anterioridad e introduce los términos de equivalencia de constructo, funcionamiento diferencial del ítem y equivalencia de traducción; conceptos centrales en el proceso de adaptación de test y de las evaluaciones interculturales.

### **Standards for Educational and Psychological Testing de 2014**

Desde la edición de 1999, diversas organizaciones, empresas y especialistas en el campo de la medición psicológica y educativa han propuesto nuevos enfoques y desarrollos relacionados con el diseño de pruebas de medición, y así como con las técnicas del análisis, normalización, interpretación y uso de las puntuaciones que de ellas se desprenden. Como respuesta, en 2005 un comité conformado por miembros de la AERA, APA y NCME determinó que la edición de 1999 demandaba una revisión y actualización. Posteriormente, en 2008 un segundo comité conformado

por 15 miembros de diferentes organizaciones, desarrolló la nueva versión de los Estándares y en 2011 fueron revisados tanto por las organizaciones patrocinadoras (AERA, APA y NCME) como por 35 asociaciones profesionales, 10 empresas orientadas a la medición, 4 centros de investigación, 3 organismos acreditadores y 6 instituciones independientes. La revisión se orientó hacia cuatro áreas principales: avances tecnológicos en medición, incremento en el uso de pruebas para la rendición de cuentas y política pública, inclusión, temas vinculados con el uso de pruebas en el contexto laboral.

Al igual que las ediciones anteriores, estos nuevos Estándares son prescriptivos y no contienen reglas específicas para su aplicación. Estos fueron formulados con la intención de ser consistentes con otros estándares, directrices y códigos de conducta publicados por estas organizaciones. Los *Standards* 2014, buscan promover las buenas prácticas de medición y proporcionar las bases para evaluar la calidad de las mismas. A diferencia de las ediciones anteriores, esta propuesta organiza los estándares de una manera diferente (Tabla 6). Así, la sección de Fundamentos concentra los capítulos de validez, confiabilidad/precisión, errores de medida y equidad de las pruebas. La sección de Operación concentra los estándares asociados al diseño, desarrollo, aplicación, calificación y normalización de los test, así como los derechos y responsabilidades de los examinados y usuarios de los test. Por último, la sección de Aplicaciones de pruebas contiene los estándares asociados con los distintos contextos de aplicación y con el uso de las pruebas (evaluación psicológica, ámbito laboral y acreditación, evaluación educativa y evaluación de programas, política pública y rendición de cuentas).

Tabla 6

*Estructura de los estándares para pruebas psicológicas y educativas (2014)*

Sección	Capítulo	Número de estándares
Fundamentos	1. Validez	25
	2. Confiabilidad/Precisión y errores de medida	20
	3. Equidad en las pruebas	20
Operación	4. Diseño y desarrollo de test	25
	5. Puntuaciones escalas, normas, vinculación de puntuaciones y puntos de corte	23
	6. Aplicación de test, calificación, reporte e interpretación	16
	7. Documentación de apoyo para los test	14
	8. Derechos y responsabilidades de los examinados	12
	9. Derechos y responsabilidades de los usuarios del test	23
Aplicaciones de pruebas	10. Pruebas y evaluación psicológica	18
	11. Pruebas en el ámbito laboral y de acreditación	16
	12. Pruebas y evaluación educativa	19
	13. Usos de los test en evaluación de programas, política pública y rendición de cuentas.	9
Total		240

Nota: Elaboración propia, basada en Standards for Educational and Psychological Testing (AERA, APA y NCME, 2014)

En lo particular, los principales estándares que ayudan a orientar las prácticas en traducción y adaptación de pruebas se encuentran en los capítulos 3 y 7 (Equidad en las pruebas y Documentación de apoyo para los test, respectivamente). En comparación con las anteriores, esta edición sitúa a la equidad como una cuestión inherente y fundamental para la validez de un test, misma que requiere atenderse en todas las etapas del diseño, desarrollo y uso del instrumento.

La equidad es entendida como la sensibilidad hacia las características individuales y contextuales de las pruebas y los sujetos, de manera que las puntuaciones que se desprenden de ellas permitan interpretaciones válidas para los usos esperados. En síntesis, las características de todos los individuos de la población bajo estudio (raza, etnicidad, género, edad, nivel

socioeconómico y contexto lingüístico o cultural), deben considerarse durante todas las etapas de desarrollo, aplicación, calificación, interpretación y uso de las pruebas.

De hecho, los nuevos estándares reconocen la existencia de situaciones donde el test no es apropiado para todos los sujetos bajo estudio, aun cuando éste haya sido desarrollado bajo los principios y prácticas de equidad. En este sentido, se establece que algunas adaptaciones pueden ser necesarias para aquellos individuos cuyas características impedirían su participación en el proceso de evaluación. Las adaptaciones. Dependiendo de las circunstancias de aplicación, los objetivos del test y las características de los sujetos, redundan en cambios en el contenido o presentación de los ítems, en las condiciones de aplicación o en los procesos de respuesta, a fin de incrementar el acceso al test a todos los sujetos.

No obstante, es importante diferenciar entre cambios que resultan en puntuaciones comparables (acomodaciones) y cambios que pueden no producir puntuaciones que sean comparables con las obtenidas por el test original (modificaciones). La presente edición de los Estándares propone entender las adaptaciones de pruebas como un continuo donde en un extremo se encuentran las acomodaciones (cambios relativamente menores a la presentación o formato de los test, aplicación del test, o de los procesos de respuesta que mantienen el constructo original y resulta en puntuaciones comparables con el test original), y en el extremo opuesto, las modificaciones (cambios que transforman el constructo que está siendo medido, incluyendo el contenido del test o las condiciones de evaluación, para obtener una medida diferente pero que mide el constructo de manera apropiada en los sujetos esperados).

En cuanto a la eficacia de las adaptaciones, ello depende del grado en que incrementan el acceso de los sujetos al constructo medido por el test. Dicha evidencia puede recolectarse a través

de estudios cuantitativos y cualitativos, donde el juicio de expertos tiene un rol determinante en las decisiones acerca de los cambios a realizar en el test o en las condiciones de aplicación. Aquellos que informan, interpretan o utilizan las puntuaciones de un test adaptado, requieren desarrollar un completo entendimiento de la utilidad y limitaciones de los procedimientos de diseño de los test, así como reconocer cuáles adaptaciones proporcionan puntuaciones comparables con las puntuaciones del test original y cuáles adaptaciones no las proporcionan.

Concretamente, esta nueva edición de los estándares sugiere que: (a) los diseñadores de pruebas son responsables del desarrollo de test y de minimizar la posibilidad de que este se vea afectado por características lingüísticas, comunicativas, cognitivas, culturales o físicas irrelevantes al constructo [estándar 3.2, p. 64]; (b) cuando un test es traducido y adaptado de un idioma a otro, los diseñadores y los usuarios del test son responsables de describir los métodos utilizados para establecer la idoneidad de la adaptación y documentar la evidencia empírica, o lógica para determinar la validez de las interpretaciones y el uso de las puntuaciones [estándar 3.12, p. 68]; y (c) cuando un test está disponible en más de un idioma, la documentación debe brindar información de los procedimientos empleados para traducir y adaptar el test y, cuando sea factible, la información también debe contener evidencias de confiabilidad/precisión y validez para la forma adaptada [estándar 7.6, p. 127].

### **Guidelines for Adapting Educational and Psychological Tests**

Uno de los principales propósitos de la Comisión Internacional de los Test (ITC, por sus siglas en inglés), ha sido preparar y diseminar un conjunto de directrices para la adaptación de test e instrumentos de medición psicológica y educativa, y para el establecimiento de la equivalencia de puntuaciones entre los grupos lingüísticos y culturales considerados (Hambleton, 1996). Se

entiende por directriz para la adaptación de instrumentos, a aquellas prácticas que se consideran importantes a realizar y evaluar dentro del proceso de adaptación o de construcción paralela de instrumentos de medición psicológica o educativa, para uso en poblaciones que difieren en aspectos culturales o lingüísticos (Hambleton, 1996).

El empleo de las directrices de la ITC permite detectar problemas potenciales en el proceso de adaptación y problemas que necesitan tratarse antes de que la versión adaptada se pueda considerar aceptable en el nuevo contexto (Hambleton, Yu y Slater, 1999).

Así, la ITC puso en marcha en 1992 un proyecto para la elaboración de las directrices, el cual se conformó por un comité de doce representantes de distintas organizaciones (European Association of Psychological Assessment, European Test Publishers Group, International Association for Cross-Cultural Psychology, International Association of Applied Psychology, International Association for the Evaluation of Educational Achievement, International Language Testing Association and International Union of Psychological Science). Dicho comité trabajó durante varios años definiendo 22 directrices que posteriormente fueron sometidas a prueba en trabajos de campo (Hambleton, Mereda y Spielberger, 2005; Hambleton et al., 1999; Hambleton, 2001; Tanzer y Sim, 1999).

En una primera edición, las directrices para la traducción y adaptación de test fueron integradas y ordenadas en cuatro áreas generales que orientan el proceso de adaptación de test: contexto, construcción y adaptación, aplicación e interpretación (Hambleton, 1996). Estos lineamientos se describen a continuación.

- Las directrices de contexto tratan con la equivalencia de los constructos medidos en las poblaciones de interés. Buscan minimizar los efectos de las diferencias culturales y evaluar el solapamiento de los constructos en las poblaciones, lo cual permite decidir si es adecuado adaptar el test a la nueva población.
- El área de construcción y adaptación de test busca asegurar una correcta adaptación, cuidando que los contenidos, formatos, estímulos y que todos los demás componentes del test original tengan la misma familiaridad para los sujetos de las dos culturas implicadas. Esta área enfatiza la necesidad de utilizar juicios analítico-rationales y técnicas estadísticas que permitan garantizar que el test es válido para las dos poblaciones.
- Las directrices del área de aplicación incluyen todos aquellos aspectos necesarios para realizar una aplicación correcta de los test.
- Las directrices propuestas para el área de interpretación de puntuaciones se centran en detectar la importancia de realizar las interpretaciones con base en las evidencias contenidas en toda la documentación generada en los procesos de las áreas anteriores.

En general, las veintidós directrices buscan prevenir las diferentes fuentes de error que se presentan en el proceso de adaptación de test y, al mismo tiempo, ofrecen acciones para controlarlas (Muñiz et al., 2013).

Se pueden distinguir dos contextos de aplicación de dichas directrices. Uno de ellos se refiera a la adaptación de test existentes y el otro al desarrollo de nuevos test para comparaciones a nivel internacional. La evidencia recopilada durante una década sobre la aplicabilidad de las directrices confirmó su utilidad en una amplia variedad de contextos (Hambleton, 2001; International Test Commission, 2010). No obstante, el análisis de contenido de las directrices

llevado a cabo por Tanzer y Sim (1999), identificó que las directrices trataban principalmente con principios fundamentales en el proceso de adaptación de test y que había algunas inconsistencias y ambigüedades en las mismas, de tal manera que algunas necesitaban ampliarse, clarificarse, combinarse, o requerían una mayor orientación sobre su aplicación; y (c) resultaba necesario indagar sobre más tipos de evidencias para establecer la equivalencia de los test, y sobre ideas, diseños y técnicas estadísticas que permitieran investigar el posible sesgo a nivel de constructo, de método o de ítem de los test traducidos y adaptados.

Es importante observar que dichas directrices eran normativas y no prescriptivas (Solano-Flores et al., 2009; Tanzer y Sim, 1999). Es decir, no proveían un método específico para obtener test multiculturales o multi-lingüísticos válidos. Por otro lado, para que los principios normativos provistos por las directrices pudieran transformarse en aplicaciones prácticas, era necesario desarrollar un compendio de las dificultades y soluciones en el proceso de adaptación de test, y también se requiere la institucionalización de programas de formación para los constructores y usuarios de los mismos (Tanzer, 2005).

Tanto los análisis de contenido de las directrices, como los avances metodológicos, psicométricos y sustantivos en el campo de la adaptación de test, hicieron necesaria una revisión de las directrices originales (Muñiz et al., 2013). De esta manera, la ITC coordinó un nuevo grupo de trabajo interdisciplinar compuesto por representantes de asociaciones de psicólogos de 5 países (Reino Unido, Turquía, Bélgica, España y Holanda) para realizar la revisión y modificación de las directrices originales, a la luz de los nuevos desarrollos.

La segunda edición de las directrices (Muñiz et al., 2013), propuso un marco integral que busca que la adaptación consiga, con respecto al test original, el máximo nivel de equivalencia

lingüística, cultural, conceptual y métrica posible. Éstas son entendidas como un esquema que orienta a los investigadores y profesionales en el proceso de adaptación. A través de veinte directrices, agrupadas en seis categorías, se propone un proceso ordenado y global que considera todas las fases y cuestiones relevantes en la adaptación de test. Así, las nuevas directrices (Muñiz et al., 2013), se agrupan de la siguiente manera:

- Directrices previas. Resaltan el respeto a los derechos de propiedad intelectual y a los acuerdos sobre el uso de los test vigentes en los países implicados. Además, proponen el estudio de la influencia de las diferencias culturales o lingüísticas en la medición, y de la relevancia del constructo en la población de interés.
- Directrices sobre el desarrollo del test. Sugieren que el proceso de adaptación sea el más pertinente para las poblaciones de interés y que considere la utilización de juicios de expertos para valorar si el proceso de adaptación considera las diferencias lingüísticas, psicológicas y culturales suficientes. Asimismo, se interesan por que las instrucciones, contenido, formato, escalas, formas de aplicación y demás aspectos relacionados con el test y los ítems, tengan el mismo significado y sean pertinentes en las poblaciones implicadas.
- Directrices de confirmación. Fijan su atención en ofrecer información empírica acerca de la equivalencia de constructo, equivalencia de método y equivalencia entre los ítems en todas las poblaciones implicadas. De la misma manera, señalan la necesidad de obtener evidencias sobre la confiabilidad y validez del test adaptado y de establecer el nivel de comparabilidad entre las puntuaciones obtenidas por el test en sus dos versiones.
- Directrices sobre la aplicación. Se enfocan en la disminución de las diferencias culturales y lingüísticas derivadas de los procedimientos de aplicación o de los formatos de respuesta;

y en la especificación de las condiciones de aplicación de manera que no se vea amenazada la validez de las inferencias derivadas de las puntuaciones del test.

- Directrices sobre puntuación e interpretación. Comunican la necesidad de considerar información demográfica pertinente al momento de realizar las interpretaciones de las diferencias de las puntuaciones entre los grupos bajo estudio. También, resaltan la importancia de utilizar el nivel de invarianza establecido para la escala de puntuación sobre la que se hacen las comparaciones.
- Directrices sobre documentación. Señalan la importancia de proporcionar toda la documentación técnica acerca de las modificaciones del test adaptado, donde además se incluyan las evidencias de la equivalencia con la versión original e información a los usuarios sobre el uso correcto del test en la población a la que va dirigido.

### **Directrices para la traducción y adaptación de PISA 2012**

Las directrices para la traducción y adaptación de PISA son un claro ejemplo de la importancia de contar con procesos rigurosos y procesos específicos en traducción y adaptación de pruebas de medición psicológica o educativa. PISA es un estudio realizado en un gran número de países con diferentes idiomas, culturas y sistemas escolares, orientados a la evaluación de las competencias de los estudiantes en lectura, matemáticas, ciencia y resolución de problemas. Estas evaluaciones incluyen cuestionarios de contexto que ayudan a explicar las diferencias observadas en los resultados obtenidos entre y a través de los países, así como manuales de operaciones y de procedimientos, que buscan asegurar la estandarización de la aplicación de los instrumentos.

Dado que uno de los objetivos de esta evaluación es realizar comparaciones internacionales, la equivalencia de los instrumentos y materiales utilizados es un requisito

imprescindible. Es por ello que el consorcio responsable de la implementación de PISA, elaboró un conjunto de recomendaciones y directrices para asegurar la equivalencia de todas las versiones de cada país, a través de un proceso de traducción que no introduzca sesgo o comprometa las comparaciones internacionales entre las pruebas (Organization for Economic Co-operation and Development, 2010).

Las directrices para la traducción y adaptación de PISA están basadas en la revisión de la literatura internacional de test, en directrices similares desarrolladas en estudios previos y en la experiencia adquirida de la verificación de los materiales de PISA. Dichas normas se organizan en seis secciones:

- Directrices generales de traducción/adaptación. Establecen la necesidad de construir los test, cuestionarios de contexto y otros materiales de apoyo a través del trabajo de dos traductores independientes y de la posterior revisión de una tercera persona quien concilia una sola versión. Asimismo, se debe documentar cualquier adaptación, cultural o de uso local, implementada a los materiales de evaluación. Este conjunto de directrices también considera el análisis de los ítems para determinar las cualidades psicométricas de los mismos.
- Instrucciones generales. Incluyen la descripción de los materiales utilizados en la evaluación, así como de las obligaciones de las personas responsables del proyecto en cada país, los requerimientos de seguridad, y los perfiles y las obligaciones de los traductores y del conciliador. En el mismo sentido, se resalta la importancia de hacer un seguimiento meticuloso de cualquier error identificado y de las modificaciones a cualquier material de evaluación.

- Traducción de instrumentos, incluye recomendaciones que tratan con los obstáculos comunes en el proceso de traducción de los materiales, tales como la variación de una versión a otra, de la presentación de las ilustraciones y los elementos gráficos, la inclusión de expresiones propias del idioma, técnicas, afectivas, metafóricas, etcétera.
- Adaptación de material cognitivo; se incluye un listado de cuidados que se consideran deseables, aceptables u obligatorios cuando se traducen los test cognitivos de PISA. Se trata de evitar que los estudiantes se enfrenten a expresiones o conceptos que no les son familiares, pero sin alterar el significado sustantivo del texto o del ítem. De manera general, no se deben adaptar o alterar los formatos de los ítems, o incluir notas explicativas o instrucciones adicionales; y se debe valorar la pertinencia de adaptar nombres de personas, lugares, símbolos, abreviaciones y fórmulas matemáticas.
- Directrices específicas sobre la traducción de cuestionarios de contexto. Para traducir adecuadamente los cuestionarios de contexto, la cuestión central es mantener el mismo significado, en todas las versiones de los países participantes, de las preguntas sobre el contexto educativo y las instrucciones de ejecución del cuestionario. Para ello, los responsables de la evaluación deben encontrar las palabras o expresiones de mayor equivalencia con la versión original.
- Directrices específicas sobre la traducción de los manuales de operación. Consideran procedimientos que pueden, o no, modificarse en los manuales de aplicación. Además, se hace hincapié en la elaboración de un registro detallado y completo de todos los cambios realizados a cada estímulo, ítem o procedimiento, pues esta información es necesaria para la interpretación de los resultados y en la revisión de la calidad psicométrica de los ítems.

Como parte de los procedimientos de control de calidad implementados por PISA, se incluyó un listado de verificación que es utilizado por los revisores internacionales para identificar posibles problemas de traducción, y sugerir modificaciones y correcciones.

El objetivo principal de estas directrices es asegurar que el proceso de traducción no introduzca sesgos que puedan distorsionar las comparaciones internacionales, como pueden ser:

- (a) la alteración de la dificultad de textos, gráficos o tablas utilizadas como estímulos en los test;
- (b) la alteración de la dificultad de un ítem de manera que se modifique el tipo de estrategia mental requerido para responder;
- (c) la introducción de ambigüedades que perjudiquen la medición de las variables de los cuestionarios de contexto; o
- (d) la inadecuada adaptación manuales al contexto nacional de forma que se modifiquen los procesos de muestreo, aplicación o calificación.

### Capítulo 3. Aproximaciones en la adaptación/traducción de test

El principal objetivo de la adaptación de un test es asegurar la validez de constructo, de contenido y ecológica, lo que proporciona sustento a las inferencias que se realizan en poblaciones específicas, independientemente de su contexto lingüístico y cultural. Cuando se utiliza un test en una cultura distinta a la cultura donde fue desarrollado y estandarizado, se asume que las características psicológicas, o al menos las raíces de estas características, son universales para todos los grupos involucrados. De acuerdo a Poortinga y Van der Flier (1988) (como se citó en Klerk, 2008), para utilizar un test en poblaciones con diferente cultura se tiene que asumir que (a) el dominio conductual -habilidad o rasgo- medido por los ítems tiene, aproximadamente, el mismo significado en cada cultura; (b) la habilidad o el rasgo medido juega el mismo rol en la organización de la conducta en cada cultura, (c) la puntuación tiene el mismo significado para quienes responden el test independiente de su contexto cultural; y (d) para referir la comparabilidad entre las puntuaciones del test de diferentes grupos culturales, los términos sesgo y equivalencia necesitan emplearse.

Para atender los supuestos anteriores es necesario realizar modificaciones y adecuar el instrumento con el fin a establecer equivalencias y realizar comparaciones entre las poblaciones implicadas (Tanzer y Sim, 1999). Actualmente, se reconocen tres principales enfoques en el proceso de adaptación: (a) adopción o traducción; (b) adaptación; y (c) ensamblaje (He y van de Vijver, 2012; Van de Vijver y Hambleton, 1996).

El primero de esos enfoques, llamado adopción o traducción, equivale a la traducción de un instrumento al idioma objetivo. Supone que el constructo subyacente que mide el instrumento

es adecuado para cada grupo cultural y que la traducción directa es suficiente para obtener un instrumento que mide el mismo constructo en el grupo objetivo. Este enfoque es el más utilizado en la investigación empírica pues resulta simple de implementar, es económico y brinda la oportunidad de comparar las puntuaciones entre las poblaciones. Sin embargo, solo se puede utilizar cuando los ítems, en ambos idiomas o grupos culturales, tienen una cobertura adecuada del constructo medido en ambas poblaciones. El enfoque de adopción se debe considerar sí, y solo sí, las características del instrumento y del constructo son pertinentes en dichos grupos culturales (He y van de Vijver, 2012; Tanzer y Sim, 1999; Van de Vijver y Hambleton, 1996; Van de Vijver y Tanzer, 2004). A este respecto, la APA (2013) ha cuestionado la contundencia y suficiencia del enfoque de traducción. Para dicho organismo la adaptación representa el enfoque ideal para la transformación de escalas de medición.

El enfoque de adaptación tiene como base la noción de que una adopción producirá instrumentos sesgados ya sea por razones lingüísticas, culturales o psicométricas. El proceso de adaptación requiere tanto de habilidades lingüísticas como del conocimiento de las culturas e idiomas involucrados para poder evaluar la relevancia y pertinencia del instrumento en otro contexto. Esta alternativa busca maximizar la equivalencia y posibilita la realización de comparaciones entre los grupos considerados (He y van de Vijver, 2012; Tanzer y Sim, 1999; Van de Vijver y Tanzer, 2004).

El tercer enfoque, denominado ensamblaje, equivale a la compilación de un nuevo instrumento. Este enfoque maximiza la pertinencia cultural del instrumento, pues integra aspectos del constructo que son importantes para la cultura objetivo que no están siendo considerados por el instrumento original. Sin embargo, no permite la comparación de puntuaciones entre las culturas

pues no garantiza la equivalencia de constructo ni de medición. Es una opción que emerge cuando la traducción o la adaptación de un instrumento no producen un test con la precisión lingüística, cultural y psicométrica que se requiere (He y van de Vijver, 2012; Tanzer y Sim, 1999; Van de Vijver y Hambleton, 1996; Van de Vijver y Tanzer, 2004).

El término adaptación de test reemplazó al término traducción de test para hacer referencia al proceso de preparación de un test para su uso en otro idioma o cultura (Geisinger, 1994; Hambleton, 1996). La adaptación de test es un proceso más amplio y exhaustivo que refleja la manera más precisa de transformar y producir un test que sea igualmente válido en dos o más idiomas o contextos culturales (APA, 2013; Hambleton, 1996). A diferencia de la traducción de test, la adaptación estudia si la nueva versión del instrumento mide o no el mismo constructo, considerando aspectos del proceso de evaluación tales como la equivalencia de los procedimientos de aplicación y el formato utilizado en los ítems (Klerk, 2008). De acuerdo con Muñiz, Elosua y Hambleton (2013), la adaptación de un test no se trata meramente de una cuestión lingüística, (...) exige la conjunción de aspectos culturales, conceptuales y métricos que han de acometerse desde perspectivas de análisis analítico-rationales y empíricas. El proceso de adaptación proporciona el medio más sistemático y exhaustivo para transformar con éxito una escala de medida. De acuerdo con la APA (2013), este proceso se apoya en tres fases enfocadas en asegurar la validez e interpretación de las puntuaciones en la población objetivo: traducción, validación y establecimiento de normas.

La traducción de un test, validado previamente en una cultura e idioma específico, ha sido el procedimiento más comúnmente utilizado en la adaptación de un test (Tanzer y Sim, 1999). Así, la tarea del traductor es decodificar un mensaje presentado en un código (código A), y codificar el

mensaje en un segundo código (código B), de tal forma que los dos mensajes sean aproximadamente equivalentes (Figura 1). La equivalencia perfecta, en el sentido de que los mensajes evoquen respuestas idénticas en los locutores de dos idiomas, es probablemente imposible de alcanzar (Casagrande, 1954).

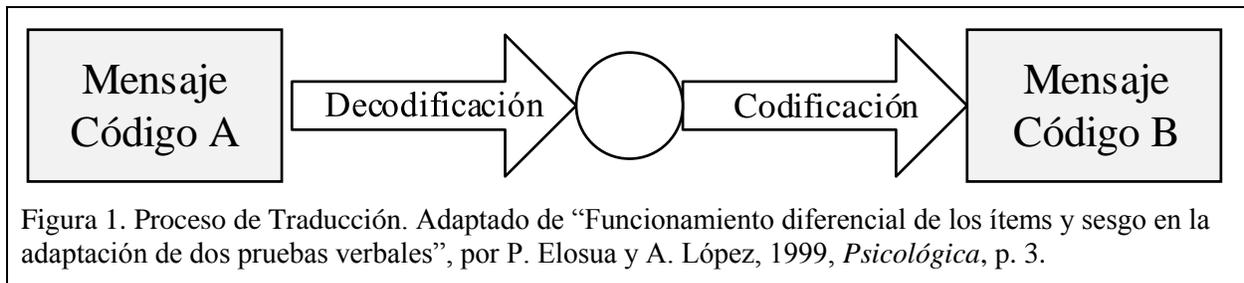


Figura 1. Proceso de Traducción. Adaptado de “Funcionamiento diferencial de los ítems y sesgo en la adaptación de dos pruebas verbales”, por P. Elosua y A. López, 1999, *Psicológica*, p. 3.

Existen dos aproximaciones principales para la traducción de test: la traducción directa y la traducción inversa. En la traducción directa, un solo traductor o un grupo de traductores decodifican el idioma del instrumento original y lo codifican en el idioma de uso dentro de la nueva cultura. Otro grupo de traductores evalúa la equivalencia de estas dos versiones del instrumento y, si existen discrepancias, se realizan los cambios en la versión objetivo (APA, 2013; Byrne et al., 2009).

Por su parte, el enfoque de traducción inversa inicia con la traducción directa de la versión original a otro idioma y un traductor diferente traduce de nuevo la versión objetivo en el idioma original. Posteriormente, las versiones originales y traducidas se comparan por un grupo independiente de traductores, a fin de evaluar su equivalencia e identificar cualquier discrepancia entre las versiones (APA, 2013; Byrne et al., 2009; Hambleton, 2005; Hulin y Mayer, 1986; Instituto Nacional para la Evaluación de la Educación, 2011).

Como ya se estableció, la traducción de instrumentos de medición para su uso en otra cultura o grupo lingüístico es más compleja que la simple traducción del texto en otro idioma debido, principalmente, a las fuentes de sesgo que pueden amenazar su precisión y grado de adecuación. En la traducción de un test, se deben considerar los significados dependientes de la cultura o las connotaciones de las palabras en cada contexto (Klerk, 2008; Van de Vijver y Hambleton, 1996). Las diferencias entre los idiomas juegan un papel importante en la conformación del significado del entorno en que vivimos. Así, la presencia de factores lingüísticos específicos del contexto y la falta de correspondencia perfecta entre los idiomas, hace necesario evaluar la ejecución de los test con el fin de asegurar la correcta interpretación de las puntuaciones obtenidas (Elosua y López, 1999; Hulin y Mayer, 1986; Instituto Nacional para la Evaluación de la Educación, 2011). Dado que una traducción lingüística precisa de los ítems es condición necesaria pero no suficiente para generar test equivalentes, se requiere de un proceso balanceado de consideraciones psicológicas, lingüísticas y culturales, que posteriormente permita realizar comparaciones interculturales válidas (Hambleton, 1996; Van de Vijver y Hambleton, 1996; Van de Vijver y Tanzer, 2004).

En la actualidad se reconoce que en general los procedimientos que se utilizan en los enfoques basados en la traducción no son suficientemente rigurosos para asegurar la validez de un instrumento (Byrne et al., 2009). En lo particular, tanto el enfoque de traducción directa como el de traducción inversa, presentan las siguientes limitaciones: (a) no consideran las pruebas de campo, lo cual reduce la posibilidad de identificar ítems defectuosos que puedan reducir la confiabilidad y validez del test (Hambleton, 2005); (b) no tratan con la comparabilidad del constructo subyacente que está siendo medido, por lo que solo pueden asumir que el constructo en ambas poblaciones es equivalente en cuanto al número de dimensiones consideradas o la cantidad

y tipo de conductas consideradas en la medición del mismo (Van de Vijver y Hambleton, 1996); y (c) no requieren del análisis de la confiabilidad del instrumento traducido y la validez de las puntuaciones en la población objetivo (Mereda, 2005).

Recientemente se han desarrollado enfoques complementarios que dan respuesta a las debilidades de la traducción directa y la traducción inversa. A este respecto, Van de Vijver y Tanzer (2004) recomendaron el empleo de un enfoque de comité, donde un grupo de personas con diferentes tipos de especialización, evalúan la calidad de la traducción. A este comité se suma un segundo grupo de expertos quienes determinan si la traducción ha alterado el constructo que se desea medir con el test (Instituto Nacional para la Evaluación de la Educación, 2011).

Lo anterior sugiere que los responsables de conducir procesos de adaptación deben establecer y gestionar con éxito equipos con áreas de especialización en el constructo a evaluar, con especialistas en técnicas de construcción de test y psicometría, y con sensibilidad a las especificidades culturales, lingüísticas y contextuales de su propia cultura e idioma (Tanzer y Sim, 1999). Además, de acuerdo con Byrne et al. (2009), estos comités también deben incorporar, en medida de lo posible, a los autores originales del test.

En resumen, la traducción lingüística no garantiza la equivalencia métrica. Resulta indispensable realizar análisis del sesgo a través de procedimientos empíricos y analítico-rationales, que permitan confirmar la igualdad de significado de las puntuaciones obtenidas a través de la versión original y de la versión adaptada (Elosua y López, 1999; Muñoz et al., 2013).

### **Procedimientos de validación en la adaptación de un test**

En el proceso de validación de un test que pretende adaptarse se identifican tres procedimientos empíricos independientes y lógicamente ordenados: (a) el estudio piloto y la prueba de campo, (b) la validación del constructo dentro del contexto objetivo, y (c) la validación del constructo entre el contexto original y objetivo (APA, 2013).

En el caso del estudio piloto y de la prueba de campo, estos se realizan una vez obtenida la versión traducida del test, con el propósito de recolectar información que permita conocer el funcionamiento del test y de sus ítems, de identificar los ítems que no están midiendo el constructo de la manera esperada (Geisinger, 1994), y de conocer el efecto de factores extraños en las puntuaciones de los ítems (APA, 2013). De esta manera, el estudio piloto considera la aplicación del test a una muestra pequeña de sujetos a manera de revisión inicial del funcionamiento del instrumento.

Dentro del estudio piloto se puede utilizar la técnica de entrevista cognitiva, para asegurar que el contenido de cada ítem del test tenga el mismo sentido para todos los sujetos, sea aplicable al contexto de vida de los sujetos, y su significado sea interpretado y comprendido de la manera esperada (Miller, Mont, Maitland, Altman y Madans, 2010). En este sentido, la técnica *pensar en voz alta*, permite conocer los procesos cognitivos involucrados en la respuesta al ítem (Van de vijver y Tanzer, 2004). Una vez identificados los ítems defectuosos, así como los factores que amenazan la validez de las puntuaciones del instrumento, se deben modificar y corregir para posteriormente realizar la prueba de campo, donde el test se aplica a una muestra más amplia y semejante a aquella con la que fue validado originalmente (APA, 2013).

**Validación del constructo en la población objetivo.** El test adaptado requiere de un proceso de validación de constructo para asegurar que el mismo constructo está siendo medido de la misma manera que en el caso del test original. La validación de constructo analiza la estructura factorial del instrumento adaptado para evaluar la estructura hipotética con respecto al número de factores subyacentes, las cargas factoriales de los ítems asignados a cada dimensión y a la relación entre los factores. De esta manera, en la medida en que el número de factores subyacentes y las cargas factoriales sean iguales entre la muestra original y la muestra objetivo, ello constituye una evidencia de validez de constructo (APA, 2013).

El análisis factorial confirmatorio (AFC), es el método más apropiado y estadísticamente más riguroso para obtener evidencias de validez de constructo de un test. Considera la estructura factorial postulada por el instrumento original como la estructura hipotética y la contrasta con la estructura observada en la muestra objetivo. Los resultados son descritos en términos de bondad de ajuste entre la estructura hipotética del instrumento y el conjunto de datos obtenidos en la muestra (puntuaciones de las respuestas a los ítems), donde altos índices de bondad de ajuste sugieren una fuente confiable de evidencia de validez de constructo del instrumento adaptado (APA, 2013).

En contraparte, cuando se obtienen niveles bajos de bondad de ajuste puede deberse a: (a) que los ítems de un test pertenezcan a más de un factor subyacente, (b) que el contenido de un ítem se sobrepone con el de otro ítem, o (c) que las correlaciones entre los factores subyacentes son inconsistentes con la estructura hipotética (APA, 2013).

**Validación del constructo entre la población original y la población objetivo.** Una vez que se hayan obtenido evidencias suficientes que den cuenta de la validez del constructo del instrumento adaptado en la población objetivo, se debe determinar el grado en el cual la estructura factorial y las puntuaciones del instrumento adaptado son equivalentes con las obtenidas por la versión original. Tanto el AFC como el análisis del funcionamiento diferencial de ítems (DIF, por su acrónimo en inglés) son enfoques empleados para detectar evidencias de sesgo en un ítem y para estudiar la equivalencia del instrumento adaptado con relación al instrumento original (APA, 2013; Geisinger, 1994; He y van de Vijver, 2012). Ambos procedimientos se comentan a continuación.

*Análisis factorial confirmatorio.* El empleo del AFC para el estudio de la equivalencia de instrumentos de medición tiene como objetivo determinar el grado en el cual los ítems están operando de manera similar en las poblaciones consideradas, y estudiar la estructura subyacente del instrumento (APA, 2013). De manera específica, se busca conocer si: (a) el contenido del ítem es interpretado y percibido exactamente de la misma manera; (b) los ítems miden el constructo subyacente en el mismo grado; (c) el formato de los ítems opera de la misma manera; (d) el constructo medido por cada dimensión (del test) tiene el mismo significado; (e) el constructo tiene la misma estructura dimensional; y si (f) las relaciones entre las dimensiones son las mismas.

Si el modelo del AFC muestra un ajuste aceptable (Tabla 7), significa que la estructura factorial asumida no puede rechazarse y que pueden establecerse diferentes niveles de equivalencia (He y Van de vijver, 2012).

Tabla 7

*Clasificación de los índices de ajuste en el análisis factorial confirmatorio*

Inferencial/ Descriptivo	Absoluto/ comparativo	Índice de ajuste	Criterios de interpretación
Inferencial	Ajuste absoluto	Estadístico $X^2$	Se mantiene el modelo si $p \geq 0.05$ ; no se recomienda
	Ajuste comparativo	Comparación de modelos anidados ( $X^2$ )	Los modelos no difieren si $p \geq 0.05$
Descriptivo	Ajuste absoluto	$X^2/gf$	Ratios menores que 2 ó 3 suelen considerarse indicadores de ajuste aceptable; no se recomienda.
		RMR	Cuanto más pequeño, mejor ajuste; 0 indica ajuste perfecto; no se recomienda porque es difícil de interpretar.
		SRMR	$\leq 0.08$ , el modelo se ajusta.
		RMSEA	$\leq 0.06$ , el modelo se ajusta
		GFI/AGFI	$\geq 0.95$ , el modelo se ajusta; no se recomienda.
	Ajuste comparativo	NFI	$\geq 0.95$ : el modelo se ajusta. Tiende a sobrestimarse en muestras pequeñas
		NNFI o TLI	$\geq 0.95$ (para variables continuas o $\geq 0.996$ (para variables categóricas): el modelo se ajusta.
RNI CFI		$\geq 0.95$ (para variables continuas o variables categóricas): el modelo se ajusta.	

Nota: Adaptado de "Medición de ciencias sociales y de la salud" por F. Abad, J. Olea, V. Ponsoda y C. García, 2011, *Síntesis*, Madrid, España

Para presentar evidencia de los diferentes niveles de equivalencia (equivalencia de constructo, equivalencia de medición y equivalencia de puntuaciones), se puede implementar un modelo de AFC multigrupo para contrastar las hipótesis respectivas. He y van de Vijver (2012) propusieron cinco modelos que brindan indicadores importantes sobre la equivalencia, su operacionalización e interpretación (Tabla 8).

Tabla 8

*Modelos anidados en Análisis Factorial confirmatorio multigrupo*

Modelos	Operacionalización	Interpretación del nivel de equivalencia
Invarianza de la configuración	Mismo patrón de constructos observables y latentes	El mismo constructo latente es medido utilizando los mismos indicadores (no equivalencia métrica)
Pesos de medición	Las cargas factoriales en las partes de medición de cada grupo cultural son idénticas	Los mismos factores latentes son medidos entre los grupos, indicando equivalencia de constructo y equivalencia métrica
Invarianza de intercepto	Ítems tienen el mismo intercepto (media latente) entre las culturas	Todos los grupos representan la misma diferencia entre grupos, indicando estar libre de sesgo de ítem y una equivalencia completa en la puntuación
Residuales estructurales	La varianza de error del factor latente es idéntica	El rango de puntuaciones en los factores latentes no varía entre las culturas, lo cual indica equivalencia completa de puntuación
Residuales de medición	La varianza de error de los ítems observados es idéntica	Los grupos utilizan el mismo rango del continuo del constructo, lo cual indica equivalencia completa de puntuación

Nota: Adaptado de "Bias and Equivalence in Cross-Cultural Research" por J. He y F. van de Vijver, 2012, *Online Readings in Psychology and Culture*, 2,2.

Dentro de esta técnica, los estudios de validez analizan la invarianza de los parámetros calculados a través de los grupos. Es decir, se busca comprobar que las puntuaciones, en los grupos considerados, tienen el mismo significado o están libres de sesgo (Abad et al., 2011). De acuerdo con Geisinger (1994), cuando se identifica que los instrumentos adaptados miden las mismas dimensiones entre los grupos culturales o lingüísticos, pero las cargas de los ítems difieren, es un indicador de que efectivamente dichas dimensiones están presentes en ambas poblaciones, pero que las dimensiones difieren interculturalmente de alguna manera.

Además de la utilización del AFC, distintos autores recomiendan la implementación de otras técnicas que ayudan en el estudio de validez y de la equivalencia entre los instrumentos

considerados (Byrne et al., 2009; Geisinger, 1994; Van de Vijver y Watkins, 2006). Así, el modelamiento de ecuaciones estructurales (SEM, por su acrónimo en inglés) y las técnicas factoriales analíticas de máxima verosimilitud permiten desarrollar pruebas de hipótesis sobre si el instrumento de medición mide las mismas cualidades, en las mismas proporciones y con la misma estructura factorial que el instrumento original. Como parte de otros diseños de validación de constructo, resulta conveniente utilizar la matriz multi-rasgo multi-métodos, el estudio del cambio ante manipulaciones experimentales, la implementación del análisis de conglomerados y el escalamiento multidimensional.

*Análisis del funcionamiento diferencial de ítems.* De acuerdo a Elosua, López y Torres (1999), dentro de cualquier proceso de adaptación es imprescindible realizar un análisis detallado de las relaciones entre cada uno de los ítems y el rasgo o habilidad que pretenden medir. Un test bien construido [o adaptado] no debería penalizar a cierto grupo de personas por razones ajenas al atributo que se mide. Por lo tanto, dentro del proceso de validación de cualquier instrumento de medición, debe comprobarse que las puntuaciones que reflejan distintos niveles del rasgo que se pretenden medir, no están determinadas por variables ajenas al constructo bajo estudio (Abad et al., 2011).

Para el análisis de sesgo, resulta relevante comprender la diferencia entre los términos: *impacto del ítem, funcionamiento diferencial del ítem y sesgo de ítem*. En este sentido, el impacto del ítem es evidente cuando los examinados de grupos diferentes presentan diferentes probabilidades de adscribirse (o responder correctamente) a un ítem, debido a que existen diferencias verdaderas entre los grupos en el rasgo subyacente medido por el ítem. Por su parte, el DIF tiene lugar cuando los examinados de distintos grupos presentan diferentes probabilidades de

adscribirse (o responder correctamente) a un ítem, después de equipararse en función del nivel de rasgo que se pretende medir. En tanto, el sesgo de ítem ocurre cuando los examinados de un grupo tienen menos probabilidades de adscribirse (o responder correctamente) a un ítem a causa de algunas características del, o de las condiciones de la aplicación, que no sea relevante para los propósitos del test (Zumbo, 1999).

El DIF es una condición necesaria pero no suficiente para identificar sesgo en los ítems, por lo que la presencia de DIF no es suficiente para señalar sesgo en un ítem y se deben implementar estudios de sesgo posteriores (análisis de contenido, entrevistas cognitivas, etc.) para poder determinar la presencia de sesgo (Zumbo, 1999).

Existen distintos procedimientos y análisis estadísticos para evaluar el DIF (Delta, chi-cuadrada, Mantel-Haenszel, comparación de parámetros de las curvas característica de los ítems, modelos log-lineales y regresión logística) (Elosua y López, 1999; Elosua, 2003; He y van de Vijver, 2012). En lo general, estos ayudan a identificar aquellos ítems que presentan DIF, mientras que otros permiten conocer si el funcionamiento diferencial es de tipo *uniforme* (cuando el efecto es constante) o *no uniforme* (si el efecto varía en función del nivel de rasgo) (Figura 2). Algunos de estos procedimientos emplean el análisis de regresión logística (Zumbo, 1999), permitiendo conocer la magnitud del funcionamiento diferencial y valorar si ésta es insignificante, moderada o severa.

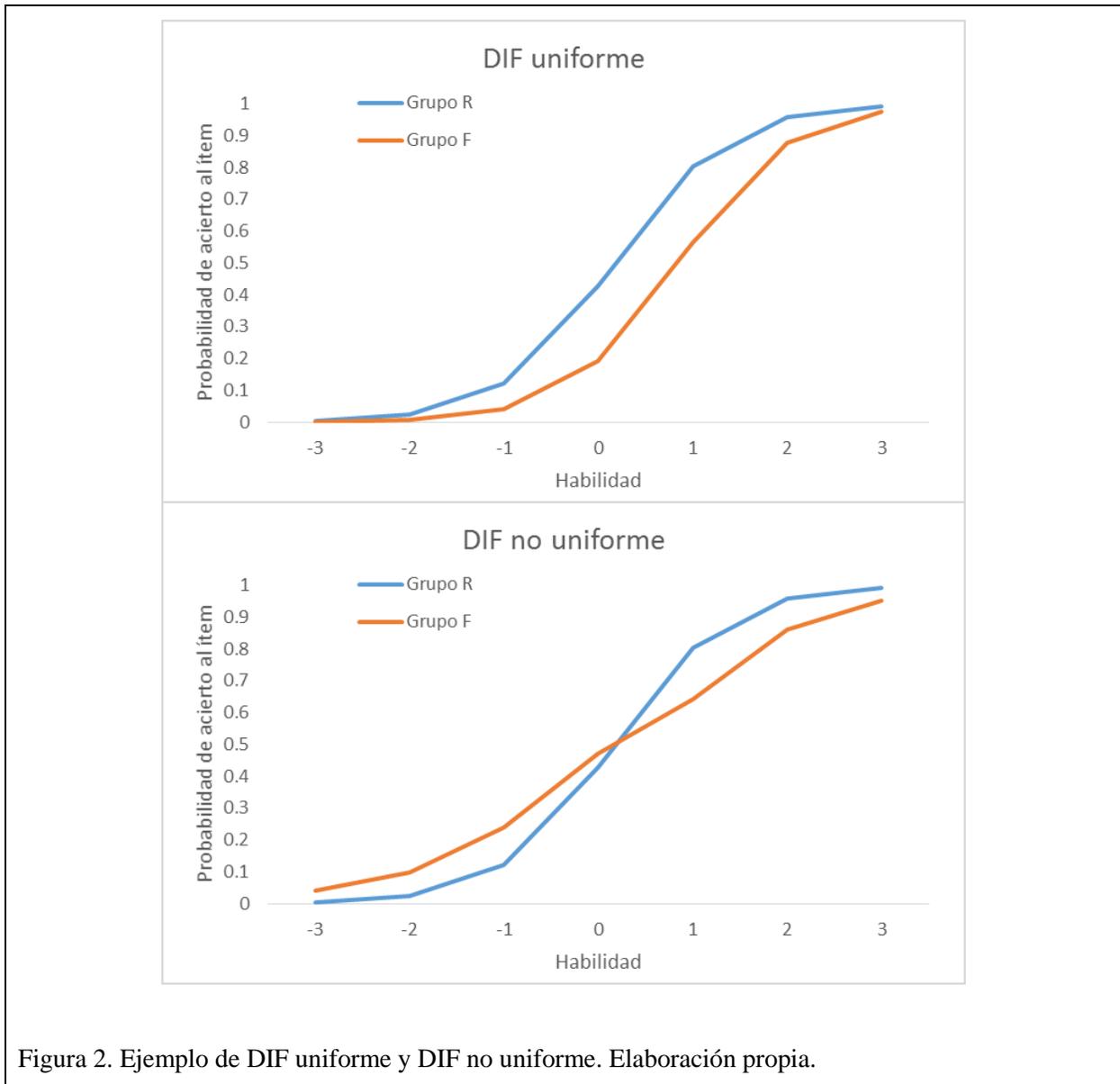


Figura 2. Ejemplo de DIF uniforme y DIF no uniforme. Elaboración propia.

La utilidad de estos procedimientos ha quedado demostrada al permitir determinar la medida en que los ítems son equivalentes entre los grupos de sujetos bajo estudio (APA, 2013; Klerk, 2008).

La fase final del proceso de adaptación se refiere al establecimiento de normas para el instrumento adaptado. De acuerdo con Geinsinger (1994), las normas proporcionan información

acerca del lugar que ocupa la puntuación de un individuo dentro de la población de sujetos que respondieron al mismo instrumento; al mismo tiempo que facilita la interpretación de las puntuaciones y ubicar, clasificar o diagnosticar a los sujetos bajo estudio.

La calidad métrica de los instrumentos es una condición necesaria pero no suficiente para alcanzar resultados rigurosos, ya que el uso inapropiado de un instrumento puede comprometer la validez y confiabilidad del mismo (Evers et al., 2013). En este sentido, los instrumentos de medición deben contar con manuales de uso e interpretación que informen sobre el conjunto de evidencias que garantizan que se puedan realizar inferencias de manera válida y confiable (Muñiz y Fernández-Hermida, 2010).

Así, podría afirmarse que un test vale tanto como su manual (Muñiz, Fernandes-Harmida, 2010), pues en estos debe documentar el procedimiento llevado a cabo durante el desarrollo del test, las evidencias de validez y confiabilidad del instrumento, el procedimiento de aplicación del test, de interpretación (normativa o criterial) de las puntuaciones, así como cualquier tarea asociada con la transformación de puntuaciones y el establecimiento de criterios de selección (Byrne et al., 2009; Elosua, 2012; International Test Commission, 2010).

Por lo tanto, cuando se adapta un instrumento en una nueva población resulta inapropiado interpretar las puntuaciones a la luz de la información normativa asociada al instrumento original. Así, la adaptación de instrumentos de medición debe considerar la recolección de información necesaria que permita establecer las normas para valorar el desempeño de la población objetivo en el nuevo instrumento (APA, 2013; Geinsinger, 1994). Del mismo modo, el manual del test adaptado requiere contener más documentación que el original (Tanzer y Sim, 1999), pues deberá incluir (a) toda la información relativa al proceso de traducción y adaptación que se llevó a cabo

(Muñiz y Fernandez-Harmida, 2010); (b) información detallada de todos los cambios realizados acompañados de una justificación de tipo lingüística o psicométrica; (c) evidencias de equivalencia de constructo y de medición dentro de cada población; y (d) evidencias de equivalencia de puntuaciones entre las poblaciones (Van de Vijver y Tanzer, 2004).

#### Capítulo 4. Método

Con base en la revisión de los fundamentos, estándares y directrices que orientan los procesos de traducción y adaptación de instrumentos de medición psicológica y educativa, se propuso una metodología que permitió adaptar la Escala de Convivencia Escolar (ECE) (Ortega-Ruiz et al., 2013) al contexto de Baja California. Esta metodología incluye aspectos conceptuales, técnicos, instrumentales y operativos que comúnmente suelen considerarse en el proceso de adaptación de instrumentos.

Así, la presente metodología se conforma por dos fases: (a) la *fase empírico-racional* que comprende las etapas de traducción, validación lingüística, validación de contenido y prueba piloto; y (b) la *fase psicométrica*, misma que considera las etapas de aplicación y análisis de equivalencia y sesgo. Cada una de las seis etapas compromete objetivos, procedimientos y productos específicos (Anexo A), y cada uno de estos productos son indispensables en el desarrollo de las etapas subsecuentes (Figura 3).

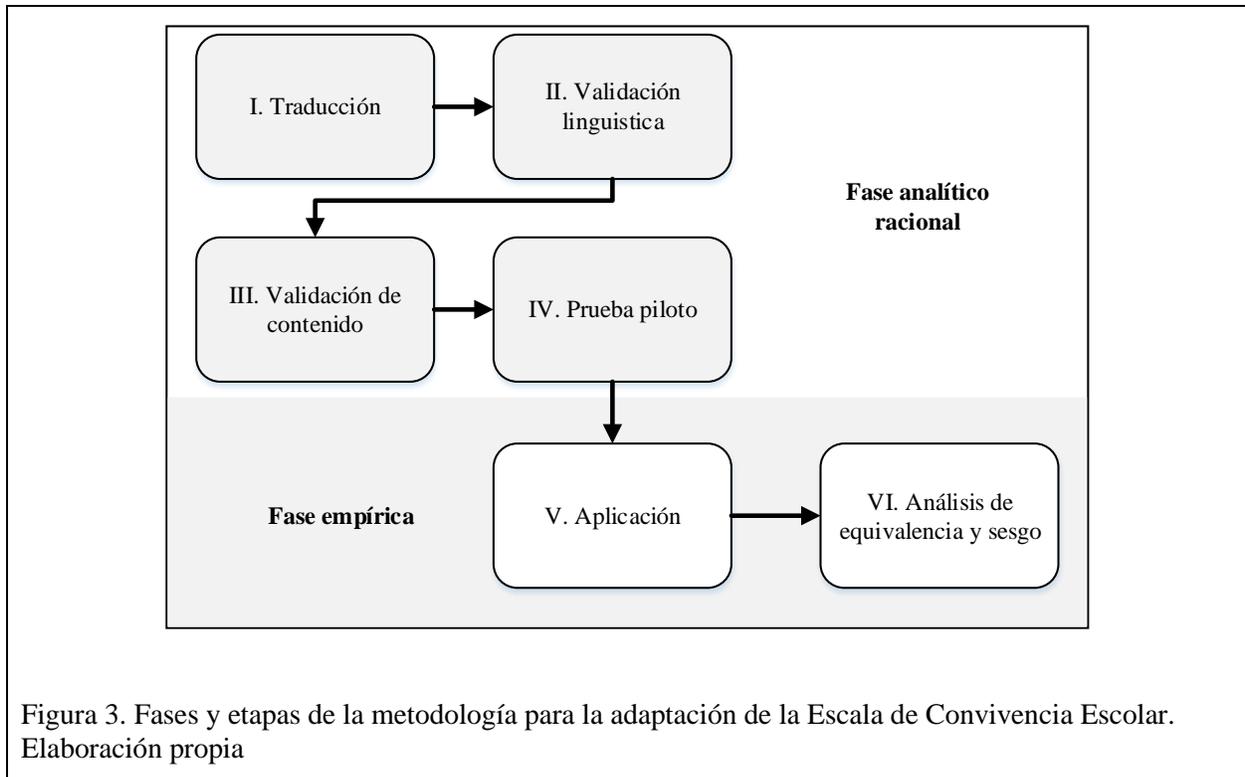


Figura 3. Fases y etapas de la metodología para la adaptación de la Escala de Convivencia Escolar. Elaboración propia

A continuación se describen las etapas que se siguieron en el presente estudio.

### **Etapas 1. Traducción**

**Objetivo:** Transformar las instrucciones, términos y opciones de respuesta que conforman los ítems, al contexto de Baja California.

**Participantes.** La transformación de las instrucciones, términos y opciones de respuesta que conforman los ítems fue elaborada por un especialista, de nacionalidad española, con experiencia en el diseño, desarrollo y validación de escalas de medición, tanto en su país (España) como en México.

**Materiales.** El principal insumo de trabajo de este especialista fue un listado de los 50 ítems que conforman la versión española de la ECE, en su formato original. También se le

proporcionaron las instrucciones utilizadas en dicho instrumento, y se le presentaron los objetivos y materiales a emplear en esa etapa. Esta información estaba contenida en una hoja electrónica donde se le presentaba en una columna el ítem en su formato original, una segunda columna donde se le solicitaba incorporar la traducción correspondiente, y una tercera columna para describir el tipo de ajuste o modificación realizada (ver Anexo B)

**Procedimiento.** El procedimiento empleado fue el siguiente: (1) se entregó al traductor los materiales correspondientes; (2) se solicitó al traductor que, con base en su experiencia, propusiera las modificaciones que considerara pertinentes a los ítems, instrucciones y opciones de respuesta, tomando en cuenta las características de la población objeto (estudiantes de primaria y secundaria de escuelas de Baja California); (3) se recopilaron los materiales con los ítems traducidos y el listado de observaciones y recomendaciones realizados por el traductor; (4) se atendieron cada una de las modificaciones propuestas y se solicitó el visto bueno final del traductor; (5) finalmente, se elaboró el informe correspondiente y se obtuvo la versión 1 de la adaptación de la Escala de Convivencia Escolar.

## **Etapa 2. Validación lingüística**

**Objetivo:** Asegurar la equivalencia lingüística de la Escala de Convivencia Escolar a través de las valoraciones de un grupo de expertos.

**Participantes.** Se contó con la participación de los siguientes especialistas: Un licenciado en traducción de nacionalidad española, que a su vez cuenta con 10 años viviendo en México, quien actualmente se desempeña como docente en la Facultad de Idiomas de la Universidad Autónoma de Baja California; un doctor en sociolingüismo, de nacionalidad mexicana, con residencia de 10 años en España; y un especialista en convivencia escolar.

Materiales. Para realizar el proceso de validación lingüística, se utilizaron los materiales siguientes: (a) Archivo electrónico de evaluación, con el formato de evaluación electrónico que fue desarrollado dentro del programa Excel®, el cual permitió capturar las valoraciones de los especialistas. En las primeras columnas se presentaron los ítems en su forma original y en su forma traducida. Posteriormente se dispuso de cuatro columnas para la captura de las valoraciones. Por último, se presentó una columna que permitió dar seguimiento al proceso de valoración; (b) Manual de elaboración de la tarea. Documento en donde se describen los antecedentes, objetivos y metodología del proyecto, así como una descripción de las tareas a realizar como parte del proceso de validación lingüística; (c) Correo electrónico. Insumo empleado para enviar y recibir los materiales que comprometió el proceso de validación lingüística (Ver Anexo C).

Procedimiento. Para valorar la calidad de la traducción y la equivalencia lingüística de cada ítem, se realizaron las siguientes acciones: Se entregó a cada uno de los jueces un manual para apoyar la elaboración de la actividad solicitada, así como un archivo electrónico de evaluación. Cada uno de los jueces evaluó de manera independiente la equivalencia lingüística de la traducción de los ítems considerando los siguientes rasgos: (a) Gramática. La traducción del ítem se encuentra libre de errores gramaticales y la sintaxis es simple y común para la población destinataria; (b) Semántica. Las ideas y el significado transferidos al ítem traducido son iguales a los del ítem en el idioma fuente; y (c) Equivalencia. El ítem traducido y el ítem en el idioma original son equivalentes en términos gramaticales y semánticos.

Para evaluar cada uno de los ítems, cada juez asignó “1” si consideraba que el ítem cumplía de manera satisfactoria con el rasgo a evaluar (Gramática, Semántica o Equivalencia), “-1” si no cumplía de manera satisfactoria con el rasgo a evaluar, y “0” si el juez tenía dudas con respecto al

grado de cumplimiento del rasgo a evaluar. Se solicitó a los jueces que en caso de asignar “-1” o “0” emitieran alguna recomendación o propuesta de modificación a la traducción del ítem, misma que habría de incluirse en la sección correspondiente del formato de evaluación. Una vez concluida dicha tarea, cada uno de los jueces envió el formato de evaluación a través del correo electrónico.

Se integró la información que proporcionó cada uno de los jueces en un solo archivo y se analizaron las valoraciones que realizaron cada uno ellos de los criterios evaluados. Para ello, se calculó un índice de discrepancia para cada juez con el fin de estudiar el nivel de congruencia de sus valoraciones. Asimismo, se calculó la valoración total para cada uno de los rasgos evaluados (gramática, semántica y equivalencia). Posteriormente, se calculó el índice de equivalencia general (IEG) que se refiere al promedio de las valoraciones en los tres rasgos evaluados.

Así, para considerar un ítem equivalente, el IEG debía presentar un valor igual o mayor a 2, al igual que cada uno de los rasgos que lo componen. Finalmente, se analizaron y atendieron las modificaciones y observaciones propuestas por los jueces.

### **Etapas 3. Validación de contenido**

Objetivo: Evaluar la congruencia, relevancia, claridad y suficiencia de los ítems que componen la Escala de Convivencia Escolar, a fin de asegurar que el contenido sustantivo de los ítems de la escala no se vio comprometido por el proceso de traducción.

Participantes. Participaron un director y un docente de educación primaria, un director y un docente de educación secundaria, un especialista en convivencia escolar y un especialista en psicometría (Tabla 9).

Tabla 9

*Perfil de los jueces del proceso de validación de contenido*

Perfil general
Docente de educación primaria
Docente de educación secundaria
Director de escuela primaria
Director de escuela secundaria
Especialista en convivencia escolar
Especialista en psicometría

Fuente: Elaboración propia

**Materiales.** Los materiales empleados en el proceso de validación de contenido fueron los siguientes: (a) Guía y formato individual de evaluación. Documentos impresos que describen los objetivos e instrucciones de las tareas de evaluación a realizar, y donde se presentan y describen los ítems de cada una de las ocho dimensiones que conforman la escala. (b) Presentación electrónica en PowerPoint®, presentación de apoyo para orientar el proceso de validación de contenido, en la que se exponen los objetivos del proyecto, la descripción de las actividades a realizar y las dimensiones e ítems de la ECE. (c) Hoja electrónica de evaluación. Formato electrónico dentro del programa Excel ® que permitió la captura y sistematización de los juicios emitidos por cada uno de los jueces respecto a las dimensiones e ítems que conforman la ECE.

**Procedimiento.** El proceso se dividió en dos sesiones: la primera compuesta por un conjunto de tareas asociadas a la valoración del contenido de cada ítem, y la segunda orientada a la discusión y el establecimiento de acuerdos acerca de las modificaciones y observaciones que emergieron de la sesión anterior. En lo particular, en la primera sesión se realizaron las siguientes actividades: (a) se convocó a una sesión de trabajo a todos los jueces; (b) se realizó una presentación del proyecto de investigación, que consideraba la descripción general de la Escala de Convivencia Escolar, así como de las dimensiones que la conforman y del modelo teórico que la sustenta; (c) se entregó a cada uno de los jueces el material con el que trabajarían (Anexo D); (d)

se les solicitó otorgaran el número que mejor reflejara su opinión con relación a los criterios de congruencia, relevancia y claridad de cada ítem, así como el grado de suficiencia de los ítems considerados para evaluar cada dimensión (Tabla 10); y e) se les solicitó que emitieran las sugerencias u observaciones que consideraran pertinentes.

Tabla 10

*Criterios de validación del contenido de la adaptación de la Escala de Convivencia Escolar*

Criterio	Descripción	Escala
Congruencia	Grado en que el contenido del ítem es consecuente con la dimensión correspondiente	(1) No congruente (2) Poco congruente (3) Congruente (4) Muy congruente
Relevancia	Grado en que la información contenida en el ítem es importante para evaluar la dimensión correspondiente	(1) No relevante (2) Poco relevante (3) Relevante (4) Muy relevante
Claridad	Grado en que el ítem comunica de manera clara y directa la información; es decir que no admite más de una interpretación posible	(1) No claro (2) Poco claro (3) Claro (4) Muy claro
Suficiencia	Grado en que se incluyen los ítems necesarios para evaluar cada dimensión	(1) No suficiente (2) Poco suficiente (3) Suficiente (4) Muy suficiente

Fuente: Elaboración propia.

Posteriormente, se realizaron las siguientes actividades: (a) se integró el material y se capturó la información en un archivo electrónico; (b) se regresó el material correspondiente a cada uno de los jueces; (c) con base en las valoraciones obtenidas se calcularon las medias y medianas de los ítems y dimensiones (d) se identificaron y modificaron aquellos ítems que no cumplieron de manera satisfactoria con al menos uno de los criterios de evaluación (media o mediana inferior a 3); (e) se valoró el grado en que cada una de las dimensiones se encontraba suficientemente representada por los ítems que las conforman (media o mediana inferior a 3); (f) se calculó el

índice de discrepancia intra-jueces, y (g) se atendieron las observaciones y recomendaciones de cada juez.

#### **Etapa 4. Prueba piloto**

Objetivo. Valorar el funcionamiento inicial de la escala, ítems e instrucciones, a fin de realizar en su caso las modificaciones correspondientes.

Participantes. En la aplicación piloto participó, como aplicador de la ECE, un licenciado en psicología con experiencia en la coordinación y aplicación de evaluaciones a gran escala en contextos escolares de Baja California. Asimismo, participaron 21 estudiantes (12 mujeres y 9 hombres) que cursaban el quinto grado de primaria en el turno matutino, con un rango de edad de 10 a 11 años y una media de 10.3 (Tabla 11)

Tabla 11

*Características de la población participante en la prueba piloto de la adaptación de la Escala de Convivencia Escolar*

		Frecuencia	Porcentaje
Edad	10	14	66.7
	11	7	33.3
Sexo	Femenino	12	57.1
	Masculino	9	42.9
Total		21	100

Materiales. Los materiales utilizados en la prueba piloto fueron los siguientes: (a) Escala de Convivencia Escolar adaptada, que contiene las instrucciones de respuesta y los 50 ítems que conforman la ECE-A. (b) Hoja de respuesta, compuesta por 4 secciones: sección de instrucciones la cual orienta el llenado de la misma; sección de Datos generales que permite registrar la edad y sexo de los estudiantes; sección de Datos de la escuela donde el estudiante registra el nombre del centro escolar de pertenencia, el grado, el grupo y el turno; y la sección de la Escala de Convivencia

Escolar en donde los estudiantes cuentan con el espacio para emitir sus respuestas a cada uno de los 50 ítems que la conforman (ver anexo E).

Procedimiento. La aplicación piloto de la escala se realizó de la siguiente manera: (1) Se estableció contacto con la directora del centro escolar para solicitar el acceso al grupo y posteriormente se acudió al centro escolar, en la fecha y hora establecida por la directora. (2) Las autoridades de la escuela acompañaron al aplicador al aula y lo presentaban ante el grupo a evaluar. (3) El aplicador proporcionó a cada uno de los estudiantes el material necesario (cuadernillo de la Escala de Convivencia Escolar, hoja de respuesta y lápiz). (4) Se registró la hora en que todos los estudiantes comenzaron a responder en la hoja de respuesta, así como el momento en que finalizaron el primer estudiante y el último. (5) Finalmente, una vez que todos los estudiantes terminaron de responder en la hoja de respuesta, se les cuestionó acerca de la claridad de las instrucciones de la escala, del formato de respuesta y del contenido de los ítems.

### **Etapa 5. Aplicación a gran escala**

Objetivo: Aplicar la Escala de Convivencia Escolar a una muestra de estudiantes de quinto y sexto grado de primaria, y del primero, segundo y tercer grado de secundaria.

Participantes. Participaron 12 aplicadores capacitados que se encontraban estudiando los últimos semestres de la licenciatura en psicología o la licenciatura en ciencias de la educación. Por su parte, la selección de la muestra de estudiantes se estimó teniendo como referencia las características de la muestra de la población origen (Tabla 12).

Tabla 12

*Estimación muestral para la aplicación de la ECE-A con respecto a la muestra original*

		Población origen		Muestra esperada	
		N	Porcentaje	N	Porcentaje
Primaria	5°	631	8.97	250	22.36
	6°	647	9.19	215	20.13
Educación secundaria obligatoria	1°	1191	16.92	200	19.17
	2°	1079	15.33	200	19.17
	3°	1022	14.52	200	19.17
	4°	853	12.12	0	0.00
Bachillerato	1°	896	12.73	0	0.00
	2°	718	10.20	0	0.00
Total		7037	100.0	1065	100.00

Como puede observarse, en la selección de la muestra no se consideraron estudiantes de bachillerato debido a que la estrategia evaluativa de la unidad de evaluación educativa (UEE) de la que formó parte esta aplicación, no incluía a dicho nivel educativo. Es importante destacar que en el sistema educativo español las edades de los estudiantes del 4° grado de la educación secundaria obligatoria corresponden con las edades de los estudiantes de los primeros dos semestres de bachillerato en el sistema educativo mexicano.

Materiales. A continuación se describen los materiales utilizados dentro de cada proceso:

(a) Guía de aplicación en la cual se presentan los objetivos de la aplicación y las tareas a realizar antes, durante y después de aplicar el instrumento. Asimismo, se utilizó una presentación electrónica que permitió estructurar y guiar la sesión de capacitación. Tanto la guía como la presentación fueron diseñadas por el grupo de investigadores adscritos a la Unidad de Evaluación Educativa. (b) Cuadernillo de aplicación, mismo que concentraba una presentación de la ECE, y las instrucciones y sus ítems; (c) Hojas de respuesta, mismas que permitieron capturar las valoraciones que los estudiantes hicieron en cada uno de los ítems, así como sus datos generales. (d) Reporte de incidencias, documento impreso en donde cada aplicador responsable capturó la

información relacionada con el grupo de estudiantes, y cualquier eventualidad que pudo presentarse y comprometer la aplicación en un estudiante o en todos. (f) Finalmente, para la captura y elaboración de base de datos se diseñaron los formatos necesarios en el programa Excel®, que permitieron registrar, por un lado, cada una de las hojas de respuestas y, por otro, los reportes de incidencias de cada aplicación.

Procedimiento. La aplicación estuvo conformada por cuatro procesos: (1) definición del diseño muestral y selección de la muestra; (2) capacitación al grupo de aplicadores; (3) aplicación a gran escala; y (4) captura y elaboración de la base de datos. Es necesario indicar que los procesos de Capacitación y Aplicación se realizaron dentro de una estrategia evaluativa más amplia administrada por la Unidad de Evaluación Educativa (UEE) del Instituto de Investigación y Desarrollo Educativo, por encargo del Sistema Educativo Estatal de Baja California.

- Definición del diseño muestral y selección de la muestra. La muestra fue determinada con base en las características de la muestra de estudiantes de la Comunidad Autónoma de Andalucía, España (Ortega-Ruiz et al., 2013), misma que buscó respetar las proporciones de estudiantes en cuanto a su nivel de estudios.
- Capacitación de aplicadores. La capacitación tuvo por objetivo preparar a los colaboradores para aplicar la ECE. Para ello se organizó una reunión de capacitación con una duración aproximada de 2 horas, en donde se expusieron los objetivos generales de la estrategia evaluativa, y los objetivos específicos de cada uno de los instrumentos que la conformaban. Una vez concluida la sesión de capacitación se realizó la asignación de cada uno de los aplicadores a las escuelas y grupos.

- Aplicación a gran escala. Las tareas asociadas con la aplicación se organizaron de la siguiente manera: (a) el coordinador y los aplicadores asistieron al centro escolar correspondiente con media hora de anticipación de la hora pactada; (b) el coordinador de la aplicación presentó a los aplicadores y entabló acuerdo con las autoridades de cada centro escolar para realizar dicha aplicación; (c) las autoridades escolares acompañaban a cada aplicador al grupo y los presentaban ante los estudiantes; (d) el aplicador realizó la presentación de la aplicación y repartió el material a cada uno de los estudiantes; (e) los estudiantes respondieron de manera simultánea la sección de datos generales guiados por el aplicador; (f) el aplicador leía las instrucciones del cuadernillo e indicaba el momento en que todos los estudiantes podían comenzar a responder la escala, al mismo tiempo que registraba la información requerida por el reporte de incidencias; (g) el aplicador recogió los materiales a medida que los estudiantes terminaban y agradeció su participación; (h) cuando todos los aplicadores concluyeron su aplicación, se reunieron con el coordinador y éste informó de ello a las autoridades del centro escolar, para posteriormente retirarse del mismo.
- Captura y elaboración de base de datos. Una vez realizada la aplicación, se realizó la captura de las hojas de respuestas y de los reportes de incidencia. La integración y recodificación de la base de datos se realizó en el programa estadístico SPSS versión 21.

### **Etapa 6. Análisis de equivalencia y sesgo**

Objetivo: Obtener evidencias de confiabilidad, validez, equivalencia y ausencia de sesgo de la Escala de Convivencia Escolar adaptada, mediante el empleo de un conjunto de técnicas estadísticas.

Participantes. Participaron 906 estudiantes de educación básica (320 de primaria y 586 de secundaria), de los cuales el 49.5 % eran mujeres y el resto hombres. Las edades de los estudiantes fluctuaron entre los diez y diecisiete años, con una media de 12.4 años (Tabla 13).

Tabla 13

*Características de los estudiantes que conformaron la muestra*

		Frecuencia	Porcentaje	
Edad	10	90	10.01	
	11	158	17.58	
	12	204	22.69	
	13	216	24.03	
	14	161	17.91	
	15	65	7.23	
	16	3	.33	
	17	2	.22	
Sexo	Femenino	447	49.50	
	Masculino	456	50.50	
Nivel y grado escolar	Primaria	Quinto grado	153	47.81
		Sexto grado	167	52.19
	Secundaria	Primer grado	296	50.51
		Segundo grado	208	35.49
		Tercer grado	82	13.99
Turno escolar	Matutino	517	57.06	
	Vespertino	389	42.94	
Modalidad escolar*	General	500	85.32	
	Técnica	86	14.68	

Nota: \* Solo en el nivel secundaria

Materiales. Se utilizaron los resultados de la aplicación de la Escala de Convivencia Escolar Adaptada (ECE-A) resultado del proceso descrito con antelación, compuesta por 50 ítems con 5 opciones de respuesta (0=Nunca, 1=Casi nunca, 2=Algunas veces, 3=Casi siempre y 4=Siempre). El entramado conceptual en que se fundamenta propone la agrupación de estos ítems en 8 dimensiones (Ortega-Ruiz et al., 2013) que explican diferentes aspectos de la convivencia escolar desde la percepción de los estudiantes (Tabla 14):

- Gestión interpersonal positiva. Manera en que los profesores gestionan las relaciones interpersonales en el centro educativo, haciendo referencia a la conducta social ejemplar,

la buena comunicación con el alumnado y sus familias y a la atención de las dificultades que éstos pueden encontrar.

- Victimización. Mide la percepción de ser maltratado, atacado, golpeado, etc.
- Disruptividad. Valora aspectos relativos al vandalismo, los destrozos e impedir el normal desarrollo de las clases.
- Red social de iguales. Considera elementos de las buenas relaciones entre los iguales.
- Agresión. Registra información sobre la acción de agredir físicamente, ser insultado y robado, etc.
- Indisciplina. Incluye aspectos relativos al incumplimiento de las normas, castigos, interrupciones de clase o aburrimiento.
- Desidia docente. Incluye cuestiones asociadas al interés y apoyo que el docente tiene hacía con los estudiantes.

La escala ha sido validada mediante rigurosos procesos estadísticos que dan cuenta de su idoneidad y consistencia interna. Así, se ha obtenido índices de confiabilidad con rangos de .78 al .90 para las 8 dimensiones, la explicación de un 63 % de la varianza e índices de bondad de ajuste óptimos (Ortega-Ruiz et al., 2013).

Tabla 14

*Dimensiones y número de ítems que conforman la Escala de Convivencia Escolar Adaptada*

Dimensiones	Reactivos	k
1. Gestión interpersonal positiva	1,4,7,10,14,18,21,25,27,36,41	11
2. Victimización	2,6,11,15,19,22	6
3. Disruptividad	3,8,13,16,20,24	6
4. Red social de iguales	5,12,17,26,28,33,38,42,46	9
5. Agresión	30,35,44,49	4
6. Ajuste normativo	9,29,34,39,45	5
7. Indisciplina	31,37,43,47	4
8. Desidia docente	23,32,40,48,50	5
Escala de Convivencia Escolar Adaptada	1 al 50	50

Nota. Elaboración propia basada en “La Convivencia Escolar: clave en la predicción del Bullying” por R. Ortega-Ruiz, R. Del Rey y J. Casas, 2013, *Revista Iberoamericana de Evaluación Educativa*, 6,2

Procedimiento: La presente etapa consideró una serie de análisis cuyos resultados orientaron la toma de decisiones en los análisis posteriores, entre los que destacan: análisis iniciales, análisis de equivalencia de constructo, análisis de dimensionalidad y análisis del funcionamiento diferencial de los ítems.

**Análisis iniciales.** Se obtuvieron las frecuencias y porcentajes de respuesta de cada uno de los ítems que conforman la ECE-A y se calcularon las medidas de tendencia central y dispersión (mediana, media y desviación estándar). Asimismo, se contrastó la hipótesis de normalidad de cada variable a través de la prueba Shapiro-Wilk y de los estadísticos de simetría y curtosis de cada ítem. Por otra parte, se estudió el comportamiento de los ítems en función de su discriminación y consistencia interna a través del cálculo del coeficiente de correlación punto biserial y del alfa ordinal, respectivamente.

**Análisis de equivalencia de constructo.** Con la finalidad de determinar si los ítems que componen la escala identifican las dimensiones que conforman el constructo convivencia escolar, se estudió el ajuste de las respuestas de los ítems al modelo de 8 dimensiones que subyace a la escala a través de un análisis factorial confirmatorio (AFC). Posteriormente se calcularon los índices de consistencia interna de las dimensiones identificadas.

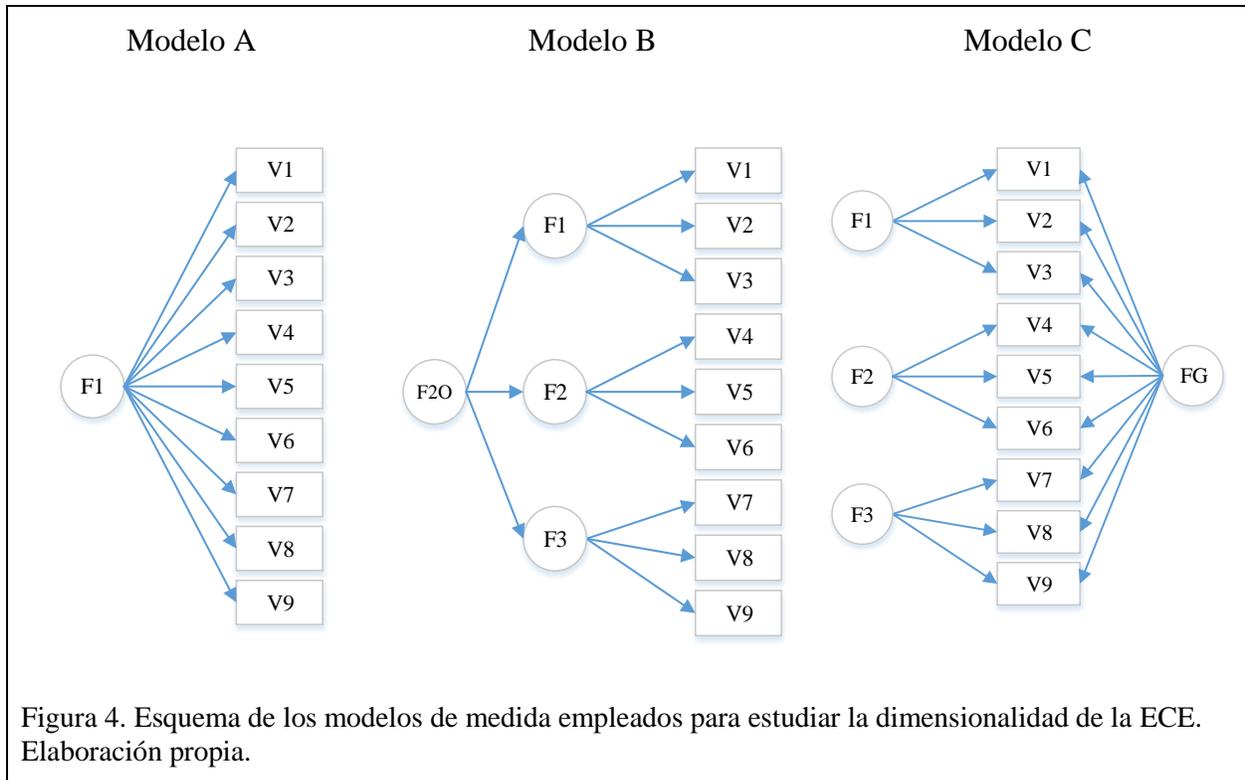
Dada la naturaleza ordinal de los ítems el AFC se calculó con base en la matriz de correlación polícórica y se utilizó el método de estimación de mínimos cuadrados no ponderados (ULS) (Luo, 2011). La estimación del modelo se elaboró dentro del paquete *lavaan* del programa estadístico R (Anexo H). Para poder establecer el grado en el cual el modelo propuesto reproduce los datos bajo estudio, se consideraron los índices de bondad de ajuste SRMR ( $\leq 0.08$ ), RMSEA ( $\leq 0.06$ ), TLI ( $\geq 0.95$ ) y CFI ( $\geq 0.95$ ) (Abad et al., 2011). En relación al estudio de la consistencia interna, se calculó el coeficiente Omega (Hancock y Mueller, 2001) para cada una de los 50 ítems, tanto en función de la escala y en su conjunto como de la dimensión de pertenencia. En comparación con el alfa de Cronbach, las estimaciones de la consistencia interna derivadas de un enfoque de variable latente [enfoque del coeficiente Omega] son más precisas y versátiles pues

representan de mejor manera los modelos multifactoriales (Gignac, Palmer y Stough, 2007; Reise et al., 2010).

**Análisis de dimensionalidad.** Con el objetivo de estudiar la estructura dimensional que subyace a la escala, se procedió a confirmar la hipótesis de unidimensionalidad, o en su caso, acumular evidencia de su estructura multidimensional.

Para ello, se estudió el ajuste de tres modelos de medida: modelo unidimensional, modelo factorial de segundo orden y modelo bifactor (Figura 4). Lo anterior permitió identificar cuál de ellos explica mejor la matriz de correlaciones de los ítems (Gignac et al., 2007).

De acuerdo con Reise et al. (2010), la estructura dimensional de una escala puede estudiarse a través de tres modelos de medida. En primera instancia el modelo unidimensional (Modelo 1) asume que cada ítem es influenciado [explicado] por un solo factor común. De acuerdo a los autores, se cree que este modelo es el más razonable, pues se supone que la suma de puntuaciones en cada ítem de cada sujeto proporciona una interpretación directa de la posición del sujeto en el constructo bajo estudio. Por su parte, el modelo factorial de segundo orden (Modelo 2) supone que el constructo medido es una dimensión de orden superior que explica la relación de las dimensiones de primer orden. Sin embargo, en este modelo no se hace explícita la relación que guarda cada ítem con el constructo que mide y ésta solo puede obtenerse a través de transformaciones posteriores al análisis. Por último se encuentra el modelo bifactor (Modelo 3), que permite conocer la relación de cada ítem con un factor general y con la dimensión de la que también forma parte. Es decir, es posible estudiar si se está midiendo solo un constructo una vez que se controla el ruido producido por la agrupación de los ítems en dimensiones.



**Análisis del funcionamiento diferencial de los ítems.** A fin de comprobar si las diferencias observadas en los niveles de rasgo en cada escala y dimensión no se encuentran determinadas por variables ajenas al constructo medido, tales como el sexo y el nivel escolar, se realizó un análisis del Funcionamiento Diferencial de los Ítems (DIF, por su acrónimo en inglés).

Para estudiar el DIF se empleó la técnica de regresión logística (RL) para ítems politómicos, misma que permite detectarlo en los ítems con base en la puntuación de cada sujeto estimada desde un enfoque de la Teoría de la Respuesta al Ítem, y con base en los resultados obtenidos en el análisis de dimensionalidad. El uso de RL permite detectar aquellos ítems que presentan un comportamiento diferencial (uniforme o no uniforme) y cuantificar su magnitud (Choi, Gibbons y Crane, 2011; Zumbo, 1999). Zumbo (1999) sugirió utilizar las diferencias en el coeficiente de determinación  $R^2$  de cada modelo como medida y lineamiento para clasificar la magnitud del DIF (Insignificante  $<0.13$ ; Moderado entre 0.13 y 0.26; y Severo  $>0.26$ ).

## Capítulo 5. Resultados

En ese apartado se presentan los resultados obtenidos en cada una de las fases y etapas correspondientes, del presente estudio.

### **Etapa 1: Traducción**

De los 50 ítems que conforman la Escala de Convivencia Escolar, 14 presentaron al menos un tipo de modificación (Tabla 15). Así, estos ítems fueron modificados y se logró obtener la versión traducida de la escala. Las opciones de respuesta (Nunca, Casi nunca, Algunas veces, Casi siempre y Siempre) no registraron ningún tipo de modificación.

Dada la naturaleza de las modificaciones sugeridas, éstas se pudieron organizar en tres categorías: convenciones, precisiones y omisiones. Así, en cuanto a la categoría *convenciones*, los ítems 8, 13, 17, 20, 33, 38, 44 y 50 sufrieron modificaciones asociadas con los usos de los términos en ambas culturas (p. e., “pega” por “golpea”, “manía” por “tirria”, “del alumnado” por “de los alumnos”, “metido miedo” por “intimidar”). Por su parte, las modificaciones dentro de la categoría *precisiones* consideraron la incorporación de algunas palabras, específicamente en los ítems 5, 10, 34 y 40 (p. e., “entre nosotros”, “actividades” por “actividades escolares”). De manera opuesta, se realizaron modificaciones asociadas a omitir algún término redundante de los ítems 20, 44 y 48 (p. e., “metidos”, “son”). Finalmente, se realizó una modificación sustancial en la versión original del ítem 37, pues el contenido del ítem estaba expresado en forma de pregunta y no de afirmación como en el resto de los ítems de la escala.

Tabla 15

*Modificaciones realizadas a los ítems de la versión original de los ítems de la Escala de Convivencia Escolar tras su traducción*

Ítem Original	Modificación	Ítem Traducido
5. Los alumnos/as nos llevamos bien	Se agrega al final: <i>entre nosotros</i>	5. Los alumnos(as) nos llevamos bien entre nosotros
8. Hay peleas en las que la gente se pega	"pega" por "golpea"	8. Hay peleas en las que la gente se golpea
10. Las familias del alumnado se implican en las actividades	"del alumnado" por "de los alumnos(as)"; "actividades" por "actividades escolares"; "implica" por "participan"	10. Las familias de los alumnos(as) participan en las actividades escolares
13. Hay niños/as que no dejan dar clase	"niños" por "alumnos (as)"; "dejan" por "permiten"	13. Hay alumnos(as) que no permiten dar clase
17. Me uno a las actividades que realizan los demás	"Me uno a" por "Me integro en"	17. Me integro en las actividades que realizan los demás
20. Hay niños/as que siempre están metidos en peleas	"niños/as" por "alumnos(as)"; se omite "metidos"	20. Hay alumnos(as) que siempre están en peleas
33. Caigo bien entre mis compañeros/as	"entre" por "a"	33. Caigo bien a mis compañeros(as)
34. Pido la palabra y espero turno para hablar	Se agrega "mi"	34. Pido la palabra y espero mi turno para hablar
37. Cuántas veces te han castigado	Se modifica el sentido de la expresión, de pregunta a afirmación	37. Me han castigado
38. Me gusta trabajar en grupo	"grupo" por "equipo"	38. Me gusta trabajar en equipo
40. Hay profesores que castigan siempre a los mismos	Se agrega "alumnos(as)"	40. Hay profesores(as) que castigan siempre a los mismos alumnos(as)
44. He amenazado a otra persona o metido miedo a otra persona	"metido miedo" por "intimidar"; se elimina la redundancia: "a otra persona"	44. He amenazado o intimidado a otra persona
48. Los profesores hacen actividades son aburridas	se quita "son"	48. Los profesores(as) hacen actividades aburridas
50. Hay niños/as a los que los profesores/as les tienen manía	"manía" por "tirria"	50. Hay alumnos(as) a los que los profesores(as) les tienen tirria

## **Etapa 2: Validación lingüística**

Se calculó el índice de discrepancia intra-jueces que equivale a la sumatoria de los valores absolutos de las diferencias de las respuestas (valoraciones) de cada juez, con respecto a la media de cada ítem (Tabla 16).

Tabla 16

*Índice de discrepancia intra-jueces de la etapa de validación lingüística*

Juez	Índice de discrepancia
MH	24.3
VC	18.7
CF	32.3

A este respecto, si se tiene en cuenta que el valor teórico de la discrepancia es igual a 149 (el número de valoraciones realizadas por el juez multiplicado por número de valores de la escala, menos 1), se pueden considerar válidas las valoraciones de los tres jueces.

Posteriormente se obtuvieron las puntuaciones totales de los ítems en cada uno de los rasgos evaluados y el índice de equivalencia general de cada ítem (IEG). Así, se puede observar que la mayoría de los ítems se pueden considerar equivalentes en términos lingüísticos, a excepción de los ítems 9, 20, 23, 28, 33, 37, 42, 44, 47 y 48, pues éstos registraron un IEG inferior a 2 (Tabla 17).

Tabla 17

*Puntuaciones totales e índice de equivalencia de la etapa de validación lingüística de la traducción de la Escala de Convivencia Escolar*

Ítems traducidos de la Escala de Convivencia Escolar	Puntuación total			IEG
	Gramática	Semántica	Equivalencia	
1. Los profesores(as) se llevan bien entre ellos	3	3	3	3.0
2. He tenido miedo de venir a la escuela	3	3	3	3.0
3. Dentro de la escuela hay problemas de vandalismo	3	3	3	3.0
4. Hay buenas relaciones entre los profesores(as) y los alumnos(as)	3	3	3	3.0
5. Los alumnos(as) nos llevamos bien entre nosotros	3	3	2	2.7
6. Algún compañero(a) me ha golpeado	3	3	3	3.0
7. Los padres de familia se llevan bien con los profesores(as)	3	2	3	2.7
8. Hay peleas en las que la gente se golpea	3	3	3	3.0
9. Aprendo	<b>1</b>	<b>1</b>	<b>1</b>	1.0
10. Las familias de los alumnos(as) participan en las actividades escolares	3	2	2	2.3
11. Algún compañero(a) me ha insultado	3	3	3	3.0
12. Mis compañeros(as) me ayudan cuando lo necesito	3	3	3	3.0
13. Hay alumnos(as) que no permiten dar clase	3	3	<b>1</b>	2.3
14. Los profesores(as) son respetados	3	2	3	2.7
15. Me he sentido amenazado(a)	3	3	3	3.0
16. Hay alumnos(as) que no respetan las normas	3	3	3	3.0

Ítems traducidos de la Escala de Convivencia Escolar	Puntuación total			IEG
	Gramática	Semántica	Equivalencia	
17. Me integro en las actividades que realizan los demás	3	2	2	2.3
18. Los profesores(as) son ejemplo de buenas relaciones	3	1	2	2.0
19. Me han robado	3	3	2	2.7
20. Hay alumnos(as) que siempre están en peleas	2	1	1	1.3
21. Los profesores(as) evitan que nos burlemos unos de otros	3	3	3	3.0
22. Me he sentido excluido, aislado o rechazado por compañeros(as)	3	3	3	3.0
23. Los profesores(as) sólo explican para los listos(as) de la clase	2	1	2	1.7
24. Algunos alumnos(as) destruyen el material y las instalaciones	3	3	3	3.0
25. Los profesores(as) nos ayudan a resolver nuestros problemas	3	3	3	3.0
26. Mis compañeros(as) se interesan por mi	3	3	3	3.0
27. Mis profesores(as) me ayudan cuando lo necesito	3	3	1	2.3
28. Expreso y defiendo mis opiniones sin dañar a los demás	3	1	0	1.3
29. Dejo trabajar a los demás sin molestarlos	3	3	3	3.0
30. He golpeado a algún compañero(a)	3	3	3	3.0
31. Sólo cumplo las normas que me convienen	3	3	2	2.7
32. Las normas de los profesores(as) son injustas	3	3	2	2.7
33. Caigo bien a mis compañeros(as)	1	3	1	1.7
34. Pido la palabra y espero mi turno para hablar	3	3	3	3.0
35. He excluido o rechazado a algún compañero	3	1	3	2.3
36. Mis padres se llevan bien con mis profesores(as)	3	3	3	3.0
37. Me han castigado	3	0	0	1.0
38. Me gusta trabajar en equipo	3	3	3	3.0
39. Cumplo las normas	3	3	3	3.0
40. Hay profesores(as) que castigan siempre a los mismos alumnos(as)	3	3	3	3.0
41. Los profesores(as) valoran mi trabajo y me animan a mejorar	1	3	3	2.3
42. Ayudo a mis compañeros(as) en lo que necesitan	3	1	1	1.7
43. Interrumpo la clase porque me aburro	3	3	3	3.0
44. He amenazado o intimidado a otra persona	3	1	1	1.7
45. Respeto la opinión de los demás aunque no la comparta	3	3	3	3.0
46. Siento que tengo amigos(as)	3	3	3	3.0
47. Me aburro	3	0	0	1.0
48. Los profesores(as) hacen actividades aburridas	1	2	1	1.3
49. He insultado a algún compañero(a)	3	3	3	3.0
50. Hay alumnos(as) a los que los profesores(as) les tienen tirria	3	3	3	3.0

Nota: IEG = Índice de equivalencia general

Posteriormente, se atendieron todas las observaciones o modificaciones propuestas (Tabla 18) para los ítems no equivalentes con excepción del ítem 37 (*Me han castigado*), debido a que éste fue formulado como pregunta en la versión original de la escala. De igual manera, se atendieron las observaciones de los ítems equivalentes que indicaban cambios en ciertas expresiones de los ítems orientados a mejorar la comprensión de su significado en el nuevo

contexto de uso (p. e., en el ítem 18 se cambió el término “*son ejemplo*” por “*dan ejemplo*”), o que indicaran la necesidad de hacer uniforme el contenido y expresiones de los ítems (p. e., en el ítem 35 “*He excluido o rechazado a algún compañero*”, se agregó la alternativa en femenino “*compañero(a)*”).

Tabla 18

*Relación de ítems modificados en la etapa de validación lingüística*

#	Ítem original	Ítem traducido	Observaciones	Ítems modificados
9	Aprendo	Aprendo	-Me parece demasiado general como ítem, comparado con los otros. -Es vaga la idea falta completar la idea	Aprendo en mi escuela
10	Las familias del alumnado se implican en las actividades	Las familias de los alumnos(as) participan en las actividades escolares	-Implicarse no significa lo mismo que participar. Implicar lleva implícito un compromiso	Las familias de los alumnos(as) se comprometen con las actividades escolares
13	Hay niños/as que no dejan dar clase	Hay alumnos(as) que no permiten dar clase	-No dejan	Hay alumnos(as) que no dejan dar clase
17	Me uno a las actividades que realizan los demás	Me integro en las actividades que realizan los demás	-"Me uno" o "me integro" no me suenan bien. ¿Participo?	Participo en las actividades que realizan los demás
18	Los profesores/as son ejemplo de buenas relaciones	Los profesores(as) son ejemplo de buenas relaciones	-Dan ejemplo.	Los profesores(as) dan ejemplo de buenas relaciones
20	Hay niños/as que siempre están metidos en peleas	Hay alumnos(as) que siempre están en peleas	-Se meten.	Hay alumnos(as) que siempre se meten en peleas
23	Los profesores/as sólo explican para los listos/as de la clase	Los profesores(as) sólo explican para los listos(as) de la clase	-No estoy seguro que en México la palabra "listo" signifique lo mismo que en México. -A los más listos.	Los profesores(as) sólo explican para los más listos(as) de la clase
27	Mis profesores/as me ayudan cuando lo necesito	Mis profesores(as) me ayudan cuando lo necesito	-Se los pido	Mis profesores(as) me ayudan cuando se lo pido
28	Expreso y defiendo mis opiniones sin dañar a los demás	Expreso y defiendo mis opiniones sin dañar a los demás	-¿Dañar o molestar? -Digo y defiendo	Digo y defiendo mis opiniones sin dañar o molestar a los demás
33	Caigo bien entre mis compañeros/as	Caigo bien a mis compañeros(as)	-Le caigo	Le caigo bien a mis compañeros(as)

#	Ítem original	Ítem traducido	Observaciones	Ítems modificados
35	He excluido o rechazado a algún compañero/a	He excluido o rechazado a algún compañero	-Falta la alternativa en femenino: "Compañera"	He excluido o rechazado a algún compañero(a)
42	Ayudo a mis compañeros/as en lo que necesitan	Ayudo a mis compañeros(as) en lo que necesitan	-Cuando me lo piden o cuando me necesitan	Ayudo a mis compañeros(as) cuando me lo piden
44	He amenazado a otra persona o metido miedo a otra persona	He amenazado o intimidado a otra persona	-Intimidad puede ser una palabra fuera del vocabulario	He amenazado a otra persona
47	Me aburro	Me aburro	-¿En qué situaciones? -Situación espacio aula o escuela	Me aburro en la escuela
48	Los profesores hacen actividades son aburridas	Los profesores(as) hacen actividades aburridas	-¿Hacen o proponen? -Ponen actividades	Los profesores(as) ponen actividades aburridas

Finalmente, se obtuvieron observaciones en ítems valorados como equivalentes que expresaban algún tipo de duda en cuanto al grado de comprensión de los términos en la población objetivo. Por ejemplo, se indicó la posibilidad de que la palabra *vandalismo* [del ítem 3] no fuera entendida por los estudiantes de primaria. También un juez expresó dudas con respecto al uso de la palabra *normas* en vez del uso de la palabra *reglas* en los ítems 31, 32 y 29. Por lo tanto, se decidió llevar este tipo de observaciones a la etapa de validación de contenido.

### **Etapas 3: Validación de contenido**

De manera inicial, cabe destacar que el índice de discrepancia intra-jueces confirmó la inclusión de todas las valoraciones de los expertos participantes en el análisis de la información, puesto que los valores absolutos de los índices no difieren de manera importante entre sí, y no se observan valores demasiado elevados (Tabla 19).

Tabla 19

*Discrepancia intra-jueces de la etapa de validación de contenido*

Juez	Índice de discrepancia
Juez 1	148.2
Juez 2	129.9
Juez 3	118.4
Juez 4	127.4
Juez 5	129.9
Juez 6	135.6

Si bien el juez 1 (J1) es el que presenta un valor de discrepancia más elevado (148.2) en comparación con el resto, el valor teórico de la discrepancia sería de 599 (el número de valoraciones realizadas por dicho juez multiplicado por el número de valores de la escala, menos 1), por lo que la totalidad de las respuestas de los jueces se consideran homogéneas.

Posteriormente, se calcularon las medias de las puntuaciones para los rasgos de congruencia, relevancia y claridad. Así, se obtuvo que la totalidad de los ítems que componen la escala cumplieron de manera satisfactoria en los tres rasgos evaluados, pues todas las medias obtenidas fueron superiores a 3. De manera adicional, se calculó el índice general de contenido (IGC) para cada uno de los rasgos evaluados (congruencia, relevancia y claridad), observándose que ningún ítem registró valores inferiores a 3.5 (Tabla 20).

Tabla 20

*Promedio de las puntuaciones obtenidas por cada ítem en los rasgos considerados en la etapa de validación de contenido*

Ítems	Congruencia	Relevancia	Claridad	IGC
<b>1. Gestión interpersonal positiva</b>				
1. Los profesores(as) se llevan bien entre ellos	3.8	4.0	4.0	3.9
4. Hay buenas relaciones entre los profesores(as) y los alumnos(as)	3.8	3.8	3.5	3.7
7. Los padres de familia se llevan bien con los profesores(as)	3.5	4.0	3.3	3.6
10. Las familias de los alumnos(as) se comprometen con las actividades escolares	3.3	4.0	3.5	3.6
14. Los profesores(as) son respetados	3.8	4.0	3.3	3.7
18. Los profesores(as) dan ejemplo de buenas relaciones	4.0	4.0	3.7	3.9
21. Los profesores(as) evitan que nos burlemos unos de otros	4.0	4.0	4.0	4.0
25. Los profesores(as) nos ayudan a resolver nuestros problemas	3.8	4.0	3.5	3.8
27. Mis profesores(as) me ayudan cuando se lo pido	4.0	4.0	3.5	3.8
36. Mis padres se llevan bien con mis profesores(as)	3.8	3.6	3.6	3.7
41. Los profesores(as) valoran mi trabajo y me animan a mejorar	4.0	4.0	4.0	4.0
<b>2. Victimización</b>				
2. He tenido miedo de venir a la escuela	4.0	4.0	3.5	3.8
6. Algún compañero(a) me ha golpeado	4.0	4.0	3.7	3.9
11. Algún compañero(a) me ha insultado	4.0	4.0	3.7	3.9
15. Me he sentido amenazado(a)	4.0	4.0	3.7	3.9
19. Me han robado	3.8	3.7	3.3	3.6
22. Me he sentido excluido, aislado o rechazado por compañeros(as)	4.0	4.0	3.7	3.9
<b>3. Disruptividad</b>				
3. Dentro de la escuela hay problemas de vandalismo	4.0	4.0	4.0	4.0
8. Hay peleas en las que la gente se golpea	4.0	4.0	3.2	3.7
13. Hay alumnos(as) que no dejan dar clase	4.0	4.0	3.3	3.8
20. Algunos alumnos(as) destruyen el material y las instalaciones	4.0	4.0	4.0	4.0
24. Hay alumnos(as) que siempre se meten en peleas	4.0	4.0	4.0	4.0
29. Dejo trabajar a los demás sin molestarlos	4.0	4.0	4.0	4.0
<b>4. Red social de iguales</b>				
5. Los alumnos(as) nos llevamos bien entre nosotros	3.8	3.8	3.5	3.7
12. Mis compañeros(as) me ayudan cuando lo necesito	4.0	4.0	4.0	4.0
17. Participo en las actividades que realizan los demás	4.0	4.0	3.5	3.8
26. Mis compañeros(as) se interesan por mí	4.0	4.0	4.0	4.0
28. Le caigo bien a mis compañeros(as)	3.8	3.8	3.7	3.8
33. Me gusta trabajar en equipo	3.8	4.0	4.0	3.9
38. Ayudo a mis compañeros(as) cuando me lo piden	4.0	4.0	4.0	4.0
42. Siento que tengo amigos(as)	3.8	4.0	3.5	3.8
46. Digo y defiendo mis opiniones sin dañar o molestar a los demás	3.8	4.0	3.7	3.8
<b>5. Agresión</b>				
30. He insultado a algún compañero(a)	4.0	4.0	3.7	3.9
35. He golpeado a algún compañero(a)	4.0	4.0	3.7	3.9
44. He amenazado a otra persona	4.0	4.0	3.5	3.8
49. He excluido o rechazado a algún compañero(a)	4.0	4.0	3.6	3.9
<b>6. Ajuste normativo</b>				
31. Cumpló las normas	4.0	4.0	3.0	3.7
32. Sólo cumpló las normas que me convienen	4.0	4.0	3.7	3.9
34. Pido la palabra y espero mi turno para hablar	4.0	4.0	4.0	4.0
39. Respeto la opinión de los demás aunque no la comparto	4.0	4.0	3.7	3.9
45. Las normas de los profesores(as) son injustas	3.8	4.0	3.5	3.8
<b>7. Indisciplina</b>				
16. Me han castigado en la escuela	3.8	4.0	3.3	3.7
37. Me aburro en la escuela	3.5	3.5	3.8	3.6
43. Interrumpo la clase porque me aburro	3.8	3.8	3.8	3.8
47. Hay alumnos(as) que no respetan las normas	4.0	4.0	3.2	3.7
<b>8. Desidia docente</b>				
9. Hay alumnos(as) a los que los profesores(as) les tienen tirria	3.8	4.0	3.2	3.7
23. Los profesores(as) sólo explican para los más listos(as) de la clase	3.8	4.0	3.7	3.8
40. Hay profesores(as) que castigan siempre a los mismos alumnos(as)	3.8	4.0	4.0	3.9
48. Aprendo en mi escuela	3.5	3.5	3.7	3.6
50. Los profesores(as) ponen actividades aburridas	3.8	3.8	3.5	3.7

Nota: IGC=Índice General de Contenido

Por su parte, a fin de analizar el grado en que los ítems son necesarios para evaluar cada uno de las dimensiones que componen la ECE-A, se calcularon la mediana y la media de las valoraciones registrados por los jueces con respecto al rasgo de suficiencia (Tabla 21).

Tabla 21

*Promedio de las puntuaciones obtenidas en el rasgo de suficiencia por cada dimensión en la etapa de validación de contenido*

Dimensiones	k	Suficiencia	
		Mediana	Media
1. Gestión interpersonal positiva	11	3	3.40
2. Victimización	6	4	3.75
3. Disruptividad	6	3	3.33
4. Red social de iguales	9	4	3.60
5. Agresión	4	3	3.20
6. Ajuste normativo	5	3.5	3.50
7. Indisciplina	4	3	3.20
8. Desidia docente	5	3	2.60

Nota: k= Número de ítems

De acuerdo con los criterios establecidos, solo la dimensión *Desidia docente* no cumple satisfactoriamente con el rasgo de suficiencia (media inferior a 3), ya que los jueces consideraron que los ítems que la conforman, son poco suficientes o insuficientes para evaluar esta dimensión. A este respecto, los jueces recomendaron agregar más ítems que permitan registrar un mayor número de comportamientos que reflejen el interés y apoyo de los docentes hacia los estudiantes (p. e., si los profesores se ausentan de clases, si los profesores llegan tarde, o si los profesores faltan a clases). Si bien el valor en el rasgo suficiencia es inferior a 3, se decidió no agregar más ítems a la escala puesto que podría comprometer la comparación con la escala original.

Por otra parte, los jueces identificaron que los ítems 7. *Los padres de familia se llevan bien con los profesores(as)* y 36. *Mis padres se llevan bien con mis profesores(as)*, de la dimensión

Gestión interpersonal positiva, aparentemente evalúan conductas similares y en consecuencia propusieron eliminar alguno de ellos. Sin embargo, se decidió dejar ambos ítems pues la dimensión busca medir la forma en que los profesores gestionan las relaciones interpersonales en el centro educativo, a través de una conducta social ejemplar, buena comunicación con los estudiantes y sus familias, y de la atención de las dificultades que éstos puedan tener. Además, esta dimensión obtuvo una calificación aceptable en el rasgo de suficiencia (media= 3.40).

En este mismo sentido, respecto a la dimensión Agresión, los jueces propusieron agregar más ítems que permitieran medir más comportamientos de agresión dentro del centro escolar (p. e., si el estudiante se ha burlado de algún compañero o compañera, si el estudiante ha comenzado una pelea, etcétera.). Sin embargo, en esta dimensión no se agregaron más ítems por las razones antes expuestas (Suficiencia con media=3.20).

También se realizaron modificaciones atendiendo las observaciones de los jueces (Tabla 22). El cambio más sustancial se hizo en el ítem 10. *Las familias de los alumnos(as) se comprometen con las actividades escolares*), puesto que de acuerdo con los jueces, este ítem alude a comportamientos poco observados en el contexto de Baja California. En consecuencia, se modificó el contenido del ítem para que hiciera alusión a comportamientos observados en los padres que demuestran algún nivel de compromiso con la educación de sus hijos (*Mis padres acuden a la escuela cuando son llamados por mis profesores*). Otro cambio importante se realizó en el ítem 17. *Participo en las actividades que realizan los demás*, pues de acuerdo con los jueces éste puede comprenderse de mejor manera en su forma original (17. *Me uno a las actividades que realizan los demás*). Asimismo, se cambió la palabra *normas* por la palabra *reglas* de los ítems 16,

31, 32 y 33 ya que este término resulta más familiar para los estudiantes de educación primaria y secundaria.

Tabla 22

*Relación de ítems modificados en la etapa de validación de contenido de la Escala de Convivencia Escolar*

#	Ítems equivalentes	Observaciones/Modificaciones	ítem definitivos
10	Las familias de los alumnos(as) se comprometen con las actividades escolares	El compromiso de las familias con las actividades escolares no es muy frecuente en el nuevo contexto.	Mis padres acuden a la escuela cuando son llamados por mis profesores
27	Mis profesores(as) me ayudan cuando se lo pido	"cuando se lo pido" por "cuando lo necesito"	Mis profesores(as) me ayudan cuando lo necesito
17	Participo en las actividades que realizan los demás	"Me uno"... como el original	Me uno a las actividades que realizan los demás
15	Me he sentido amenazado(a)	"Me he sentido" por "Me siento", se agrega "en la escuela"	Me siento amenazado(a) en la escuela
19	Me han robado	Se agrega "en la escuela"	Me han robado en la escuela
44	He amenazado a otra persona	"otra persona" por "algún compañero"	He amenazado a algún compañero
31	Sólo cumplo las normas que me convienen	"normas" por "reglas"	Sólo cumplo las reglas que me convienen
32	Las normas de los profesores(as) son injustas	"normas" por "reglas"	Las reglas de los profesores(as) son injustas
39	Cumplo las normas	"normas" por "reglas", se agrega "de mi escuela y de mi grupo"	Cumplo las reglas de mi escuela y de mi grupo
45	Respeto la opinión de los demás aunque no la comparto	"aunque no la comparto" por "aunque no esté de acuerdo"	Respeto la opinión de los demás aunque no esté de acuerdo
16	Hay alumnos(as) que no respetan las normas	"normas" por "reglas"	Hay alumnos(as) que no respetan las reglas
37	Me han castigado en la escuela.	se agregar "en la escuela"	Me han castigado en la escuela

Por último, se hicieron precisiones a algunos ítems para facilitar su comprensión. Así, en los ítems 19 y 37 se agregó al final del ítem la expresión "en la escuela", y en el ítem 39 se agregó la expresión "de mi escuela y de mi grupo", para hacer el contenido de los ítems más preciso y homogéneo al resto.

#### **Etapa 4: Prueba piloto**

Como parte de la prueba piloto se registraron los tiempos en los que se respondía la escala. En la tabla 23 se observa que la escala puede responderse entre 12 y 27 minutos. Estos registros constituyeron un insumo importante en la planeación de la aplicación.

Tabla 23

*Tiempos de respuesta de la Escala de Convivencia Escolar*

Etapa	Hora
Inicio de aplicación	10:40 am
Terminó el primer examinado	10:52 am
Terminó el último examinado	11:07 am

Una vez concluida la aplicación se preguntó a los estudiantes si habían encontrado alguna dificultad para comprender las instrucciones, opciones de respuesta o contenido de los ítems. Éstos respondieron que no comprendían el significado de la palabra *tirria* en el ítem 50. *Hay alumnos(as) a los que los profesores(as) les tienen tirria*. En consecuencia el ítem fue modificado de la siguiente manera: *Hay alumnos(as) que les caen mal a los profesores(as)*.

Se obtuvieron los estadísticos descriptivos asociados al conjunto de ítems que componen la ECE-A (Tabla 24). En general, se observan ítems con valores medios con un rango entre 1.19 y 4.90, siendo el ítem 30. *He golpeado a algún compañero(a)*, el que registró la media más baja y el ítem 9. *Aprendo en mi escuela*, el que registra la media más alta.

Tabla 24

*Estadísticos descriptivos de la muestra utilizada en la prueba piloto*

Ítem	N	Media	DE	Mediana
1. Mis profesores(as) se llevan bien entre ellos.	21	4.52	.81	5
2. He tenido miedo de venir a la escuela.	21	1.48	.81	1
3. Dentro de la escuela hay problemas de vandalismo.	21	1.86	1.01	2
4. Hay buenas relaciones entre los profesores(as) y los alumnos(as).	21	4.29	1.10	5
5. Los alumnos(as) nos llevamos bien entre nosotros.	21	3.33	.91	3
6. Algún compañero(a) me ha golpeado.	21	1.71	.96	1
7. Los padres de familia se llevan bien con los profesores(as).	21	4.14	1.06	4
8. Hay peleas en las que la gente se golpea.	21	2.14	1.28	2
9. Aprendo en mi escuela.	21	4.90	.30	5
10. Mis padres acuden a la escuela cuando son llamados por mis profesores.	21	4.14	1.11	5
11. Algún compañero(a) me ha insultado.	21	2.76	1.30	3
12. Mis compañeros(as) me ayudan cuando lo necesito.	21	3.86	1.01	4
13. Hay alumnos(as) que no dejan dar clase.	21	3.29	1.35	3
14. Los profesores(as) son respetados.	21	4.19	1.25	5
15. Me siento amenazado(a) en la escuela.	21	1.81	1.21	1
16. Hay alumnos(as) que no respetan las reglas.	21	3.24	1.14	3
17. Me uno a las actividades que realizan los demás.	21	4.38	.74	5
18. Los profesores(as) dan ejemplo de buenas relaciones.	21	4.43	1.08	5
19. Me han robado en la escuela.	21	1.67	1.02	1
20. Hay alumnos(as) que siempre se meten en peleas.	21	2.71	1.15	3
21. Los profesores(as) evitan que nos burlemos unos de otros.	21	4.76	.89	5
22. Me he sentido excluido, aislado o rechazado por compañeros(as).	21	1.76	1.04	1
23. Los profesores(as) sólo explican para los más listos(as) de la clase.	21	1.67	1.15	1
24. Algunos alumnos(as) destruyen el material y las instalaciones.	21	1.76	1.09	1
25. Los profesores(as) nos ayudan a resolver nuestros problemas.	21	3.86	1.31	4
26. Mis compañeros(as) se interesan por mí.	21	3.33	1.32	3
27. Mis profesores(as) me ayudan cuando lo necesito.	21	4.38	1.02	5
28. Digo y defiendo mis opiniones sin dañar o molestar a los demás.	21	3.48	1.33	4
29. Dejo trabajar a los demás sin molestarlos.	21	4.33	1.15	5
30. He golpeado a algún compañero(a).	21	1.19	.51	1
31. Sólo cumplo las reglas que me convienen.	21	2.29	1.65	1
32. Las reglas de los profesores(as) son injustas.	21	1.67	1.32	1
33. Le caigo bien a mis compañeros(as).	21	3.86	1.11	4
34. Pido la palabra y espero mi turno para hablar.	21	4.67	.58	5
35. He excluido o rechazado a algún compañero(a).	21	1.38	.86	1
36. Mis padres se llevan bien con mis profesores(as).	21	4.43	.87	5
37. Me han castigado en la escuela.	21	1.67	1.11	1
38. Me gusta trabajar en equipo.	21	4.29	1.10	5
39. Cumplo las reglas de mi escuela y de mi grupo.	21	4.48	.75	5
40. Hay profesores(as) que castigan siempre a los mismos alumnos(as).	21	2.10	1.37	1
41. Mis profesores(as) valoran mi trabajo y me animan a mejorar.	21	4.29	1.27	5
42. Ayudo a mis compañeros(as) cuando me lo piden.	21	4.19	1.08	5
43. Interrumpo la clase porque me aburro.	21	1.43	1.03	1
44. He amenazado a algún compañero.	21	1.38	1.07	1
45. Respeto la opinión de los demás aunque no esté de acuerdo.	21	3.62	1.32	4
46. Siento que tengo amigos(as).	21	4.24	1.22	5
47. Me aburro en la escuela.	21	1.71	1.06	1
48. Los profesores(as) ponen actividades aburridas.	21	1.71	1.10	1
49. He insultado a algún compañero(a).	21	2.00	1.45	1
50. Hay alumnos(as) a los que los profesores(as) les tienen tirria.	21	1.38	1.07	1

Nota: N= Número de sujetos; DE=Desviación Estándar

### **Etapa 5: Aplicación a gran escala**

Una de las actividades consideradas por parte de los aplicadores fue registrar el tiempo en que terminaron de responder el primer y último examinados, lo que permitió calcular el tiempo total aproximado en que la escala se respondió (Tabla 25). En promedio, los 50 ítems que conforman la escala se contestaron en 12.3 minutos, con un rango entre los 10.5 y los 22.8 minutos.

Tabla 25

*Tiempos de respuesta en minutos de la aplicación Escala de Convivencia Escolar Adaptada*

	Media
Tiempo en que terminó el primer examinado	10.5
Tiempo en que terminó el último examinado	22.8
Tiempo total	12.3

Nota: Las cantidades expresan minutos

Los aplicadores también registraron todas aquellas incidencias que pudieran comprometer la aplicación. Como se puede observar en la tabla 26, no se presentó ningún accidente que pudiera poner en riesgo la aplicación. De manera general, las incidencias registradas se asocian con la comprensión de algún término (palabra *vandalismo*, del ítem 3), la confusión entre dos ítems (ítem 7 y 36 con contenidos similares) y la no ocurrencia del comportamiento explorado por el ítem (ítem 10 no aplica si los profesores no llaman a los padres).

Posteriormente, se realizó la captura de las hojas de respuesta y de los reportes de incidencia en sus formatos correspondientes. Para la captura de las hojas de respuesta se consideraron las incidencias registradas, lo que ayudó a prescindir de la información de tres estudiantes pues uno de ellos respondía de forma desordenada y sin leer los ítems, y los otros dos presentaban problemas severos de atención.

Tabla 26

*Incidencias registradas en la aplicación de la Escala de Convivencia Escolar Adaptada, por pregunta y por grupo*

Ítems	Nivel	Grupos	Incidencia	Número de estudiantes
Ítem 10	Secundaria	1D	“Si nunca han llamado a mis papás, ¿qué contesto?”	1
		2D	“Siempre, ¿aunque no los llamen?”	1
	Primaria	6C	Solicitó explicación un alumno	1
Ítem 3	Primaria	5B	Sobre el significado de la palabra <i>vandalismo</i>	1
		5A	No comprendían el significado de <i>vandalismo</i>	*
		6B	No entendía que es vandalismo.	1
		5C	No saben qué es vandalismo	4
		5B	“¿Qué es vandalismo?”	1
		6C	“¿Qué es vandalismo?”	1
	Secundaria	5A	“¿Qué significa vandalismo?”	1
		1D	“¿Qué es vandalismo?”	1
Ítem 7 y 36	Primaria	5A	Confundían las preguntas, creían que eran las mismas.	*
		5C	Se confunden piensan que es la misma.	*

Nota: \*= El aplicador no registró el número de estudiantes

### **Etapas 6. Análisis de equivalencia y sesgo**

A continuación se describen resultados del conjunto de análisis que conforman esta etapa.

**Análisis iniciales.** Se obtuvieron los estadísticos descriptivos de los 50 ítems que conforman la ECE (Tabla 27). Se observa que el ítem 15 obtiene la puntuación media más baja (.31). Es decir, *Nunca* o *Casi nunca*, los alumnos se sienten amenazados en la escuela. Por su parte, el ítem 9 (*Aprendo en mi escuela*) registra la media más alta (3.48), lo cual indica que el grupo de sujetos bajo estudio se encuentra *Casi siempre* o *Siempre* en esta situación. Adicionalmente, se calcularon las distribuciones de frecuencias de las opciones de respuesta de cada ítem (Anexo G).

Tabla 27

*Estadísticos descriptivos de la Escala de Convivencia Escolar Adaptada*

Ítems	N	Media	Mediana	DE
1. Mis profesores(as) se llevan bien entre ellos.	906	3.29	4	.94
2. He tenido miedo de venir a la escuela.	906	.43	0	.88
3. Dentro de la escuela hay problemas de vandalismo.	906	1.56	2	1.25
4. Hay buenas relaciones entre los profesores(as) y los alumnos(as).	905	2.81	3	1.06
5. Los alumnos(as) nos llevamos bien entre nosotros.	906	2.79	3	1.00
6. Algún compañero(a) me ha golpeado.	906	.62	0	1.04
7. Los padres de familia se llevan bien con los profesores(as).	906	3.06	3	1.06
8. Hay peleas en las que la gente se golpea.	903	1.61	2	1.30
9. Aprendo en mi escuela.	904	3.48	4	.80
10. Mis padres acuden a la escuela cuando son llamados por mis profesores.	905	3.42	4	1.10
11. Algún compañero(a) me ha insultado.	907	1.39	1	1.22
12. Mis compañeros(as) me ayudan cuando lo necesito.	907	2.80	3	1.13
13. Hay alumnos(as) que no dejan dar clase.	900	2.49	3	1.29
14. Los profesores(as) son respetados.	907	3.04	3	1.01
15. Me siento amenazado(a) en la escuela.	906	.31	0	.83
16. Hay alumnos(as) que no respetan las reglas.	906	2.33	2	1.11
17. Me uno a las actividades que realizan los demás.	903	2.87	3	1.04
18. Los profesores(as) dan ejemplo de buenas relaciones.	906	3.01	3	1.14
19. Me han robado en la escuela.	902	.99	1	1.20
20. Hay alumnos(as) que siempre se meten en peleas.	906	1.87	2	1.27
21. Los profesores(as) evitan que nos burlemos unos de otros.	905	2.98	4	1.34
22. Me he sentido excluido, aislado o rechazado por compañeros(as).	906	.91	0	1.18
23. Los profesores(as) sólo explican para los más listos(as) de la clase.	905	.82	0	1.24
24. Algunos alumnos(as) destruyen el material y las instalaciones.	905	1.38	1	1.23
25. Los profesores(as) nos ayudan a resolver nuestros problemas.	903	2.70	3	1.25
26. Mis compañeros(as) se interesan por mí.	904	2.38	2	1.20
27. Mis profesores(as) me ayudan cuando lo necesito.	906	2.98	3	1.16
28. Digo y defiendo mis opiniones sin dañar o molestar a los demás.	906	2.78	3	1.20
29. Dejo trabajar a los demás sin molestarlos.	903	3.24	4	1.04
30. He golpeado a algún compañero(a).	904	.70	0	.97
31. Sólo cumplo las reglas que me convienen.	900	1.10	1	1.35
32. Las reglas de los profesores(as) son injustas.	901	1.15	1	1.25
33. Le caigo bien a mis compañeros(as).	903	3.02	3	1.01
34. Pido la palabra y espero mi turno para hablar.	904	3.04	3	1.13
35. He excluido o rechazado a algún compañero(a).	903	.91	1	1.11
36. Mis padres se llevan bien con mis profesores(as).	904	3.20	4	1.14
37. Me han castigado en la escuela.	901	.97	1	1.12
38. Me gusta trabajar en equipo.	902	3.13	4	1.15
39. Cumplo las reglas de mi escuela y de mi grupo.	902	3.06	3	1.05
40. Hay profesores(as) que castigan siempre a los mismos alumnos(as).	905	2.28	2	1.35
41. Mis profesores(as) valoran mi trabajo y me animan a mejorar.	905	2.95	3	1.20
42. Ayudo a mis compañeros(as) cuando me lo piden.	905	2.99	3	1.05
43. Interrumpo la clase porque me aburro.	905	.54	0	.95
44. He amenazado a algún compañero.	905	.45	0	.97
45. Respeto la opinión de los demás aunque no esté de acuerdo.	901	2.87	3	1.22
46. Siento que tengo amigos(as).	901	3.34	4	1.11
47. Me aburro en la escuela.	900	1.45	1	1.22
48. Los profesores(as) ponen actividades aburridas.	903	1.45	1	1.19
49. He insultado a algún compañero(a).	902	.93	1	1.10
50. Hay alumnos(as) que les caen mal a los profesores(as).	902	1.78	2	1.43

Nota: N=Número de sujetos; DE=Desviación Estándar

Asimismo, con el fin de valorar el funcionamiento inicial de los ítems en términos de su consistencia interna y niveles de discriminación, se calcularon los coeficientes de confiabilidad alfa ordinal y la correlación punto biserial corregida, tanto para la escala como para las dimensiones que la conforman (Tabla 28). El cálculo de los estadísticos se realizó en el paquete *psych* (Revelle, 2014) dentro del programa R versión 3.1.1. La sintaxis utilizada de éste y de los siguientes análisis puede consultarse en el anexo H.

Tabla 28

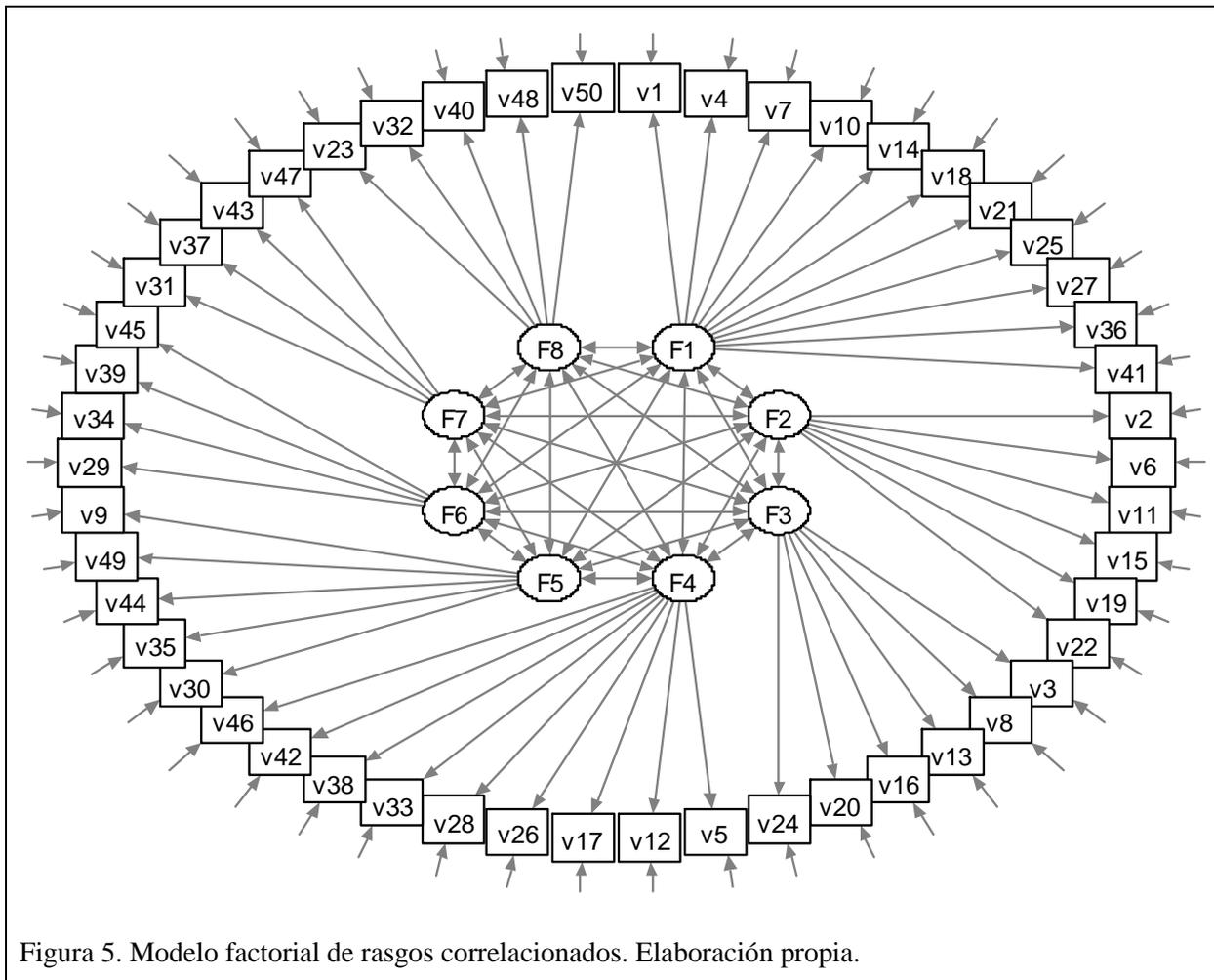
*Coefficientes de confiabilidad alfa ordinal y correlación punto biserial de los ítems de la Escala de Convivencia Escolar Adaptada*

Ítem	ECE-A	F1	F2	F3	F4	F5	F6	F7	F8
v1	0.45	0.50							
v4	0.50	0.55							
v7	0.50	0.61							
v10	0.30	0.31							
v14	0.56	0.57							
v18	0.55	0.64							
v21	0.43	0.46							
v25	0.52	0.63							
v27	0.65	0.76							
v36	0.60	0.70							
v41	0.60	0.65							
v2	0.28		0.37						
v6	0.37		0.60						
v11	0.47		0.68						
v15	0.46		0.62						
v19	0.40		0.42						
v22	0.36		0.57						
v3	0.39			0.58					
v8	0.46			0.59					
v13	0.27			0.55					
v16	0.33			0.60					
v20	0.39			0.61					
v24	0.49			0.63					
v5	0.36				0.39				
v12	0.39				0.61				
v17	0.43				0.50				
v26	0.31				0.58				
v28	0.46				0.37				
v33	0.47				0.69				
v38	0.42				0.51				
v42	0.50				0.50				
v46	0.39				0.58				
v30	0.58					0.66			
v35	0.45					0.57			
v44	0.57					0.71			
v49	0.59					0.76			
v9	0.60						0.55		
v29	0.48						0.61		
v34	0.52						0.49		
v39	0.54						0.63		
v45	0.41						0.51		
v31	0.38							0.39	
v37	0.33							0.44	
v43	0.56							0.75	
v47	0.53							0.59	
v23	0.54								0.56
v32	0.49								0.61
v40	0.27								0.44
v48	0.54								0.61
v50	0.56								0.70
Alfa ordinal	0.93	0.85	0.73	0.77	0.77	0.78	0.71	0.65	0.74

Con relación a los coeficientes de correlación punto biserial calculados, se observa que todos los ítems obtuvieron valores superiores a .20, ya sea cuando se utiliza la puntuación total de la escala o cuando se utiliza la puntuación del conjunto de ítems del que forma parte el ítem. Asimismo, los coeficientes alfa ordinal obtenidos señalan altos índices de consistencia interna de los 50 ítems que conforman la ECE-A, así como de cada una de sus dimensiones. Si bien la dimensión *Indisciplina* presenta el índice más bajo ( $\alpha = .65$ ), éste puede considerarse como aceptable dado que los índices de correlación punto biserial de los ítems que la conforman presentan valores entre 0.39 y 0.75.

La anterior información obtenida forma parte de las evidencias iniciales que señalan que los cincuenta ítems, se encuentran midiendo la convivencia escolar o alguna de las ocho dimensiones que conforman este constructo.

**Análisis 1: Equivalencia de constructo.** Para estudiar el ajuste de las respuestas de los sujetos al modelo teórico/hipotético en el que se sustenta la ECE-A, se especificó el siguiente modelo (Figura 5), en el cual el constructo convivencia escolar se encuentra dividido en 8 dimensiones distintas (variables latentes) relacionadas entre sí, y éstas a su vez explican el comportamiento de un conjunto específico de ítems (variables observadas).



Antes de realizar el análisis propuesto, se examinó la distribución de los ítems para revisar su normalidad. Observándose que 47 de los 50 ítems presentaron niveles de simetría y curtosis por debajo de 2.0 y 4.0 (en valores absolutos), respectivamente. Los ítems 2, 15 y 44 presentaron

niveles altos de simetría y curtosis (2.2 y 4.7, 3.0 y 8.9, y 2.4 y 4.8, respectivamente). Sin embargo, la media de los valores absolutos de éstos indicadores (simetría y curtosis) fue de 1.0 y 1.1 respectivamente (Anexo I).

Al realizar el análisis factorial confirmatorio (AFC) utilizando el método de estimación de mínimos cuadrados no-ponderados (ULS) sobre la matriz de correlaciones polícóricas, dada la naturaleza ordinal de las variables observadas (Luo, 2011), se observó que las cargas factoriales, los coeficientes de consistencia interna Omega ( $\Omega$ ) y las covarianzas obtenidas son apropiadas y estadísticamente significativas (Tabla 29 y Tabla 30). Así, se observan cargas factoriales superiores a .300 para todos los ítems, coeficientes de consistencia interna con valores de .662 a .856, y covarianzas del orden de .295 (entre las dimensiones *Ajuste normativo* y *Disruptividad*) a .935 (entre las dimensiones *Indisciplina* y *Agresión*).

Tabla 29.  
Covarianzas entre los factores de la Escala de Convivencia Escolar Adaptada

	F1	F2	F3	F4	F5	F6	F7	F8
F1. Gestión interpersonal positiva	1							
F2. Victimización	.440	1						
F3. Disruptividad	.511	.479	1					
F4. Red social de iguales	.726	.573	.268	1				
F5. Agresión	.586	.647	.483	.620	1			
F6. Ajuste normativo	.623	.365	.295	.719	.803	1		
F7. Indisciplina	.569	.428	.419	.449	.935	.794	1	
F8. Desidia docente	.793	.408	.711	.359	.680	.446	.742	1

Tabla 30

*Parámetros estimados para cada ítem*

Ítem	Dimensiones								Error estándar	Valor Z	P(> z )
	F1	F2	F3	F4	F5	F6	F7	F8			
v1	.509								0.011	44.419	0.00
v4	.558								0.012	48.269	0.00
v7	.569								0.012	49.207	0.00
v10	.324								0.011	28.858	0.00
v14	.617								0.012	52.868	0.00
v18	.627								0.012	53.626	0.00
v21	.473								0.011	41.469	0.00
v25	.599								0.012	51.481	0.00
v27	.751								0.012	62.628	0.00
v36	.680								0.012	57.568	0.00
v41	.687								0.012	58.059	0.00
v2		.379							0.015	25.782	0.00
v6		.549							0.015	35.439	0.00
v11		.683							0.016	41.541	0.00
v15		.655							0.016	40.402	0.00
v19		.560							0.016	36.016	0.00
v22		.520							0.015	33.937	0.00
v3			.596						0.015	38.585	0.00
v8			.695						0.016	43.322	0.00
v13			.432						0.015	29.298	0.00
v16			.511						0.015	33.974	0.00
v20			.598						0.015	38.698	0.00
v24			.732						0.016	44.876	0.00
v5				.430					0.013	33.666	0.00
v12				.486					0.013	37.557	0.00
v17				.533					0.013	40.689	0.00
v26				.410					0.013	32.258	0.00
v28				.575					0.013	43.291	0.00
v33				.596					0.013	44.577	0.00
v38				.525					0.013	40.114	0.00
v42				.612					0.013	45.593	0.00
v46				.503					0.013	38.702	0.00
v30					.663				0.014	48.873	0.00
v35					.522				0.013	41.302	0.00
v44					.652				0.013	48.359	0.00
v49					.687				0.014	49.977	0.00
v9					.665				0.014	48.975	0.00
v29						.619			0.015	40.461	0.00
v34						.674			0.016	42.323	0.00
v39						.710			0.016	43.317	0.00
v45						.530			0.014	36.726	0.00
v31							.472		0.015	32.339	0.00
v37							.429		0.014	30.357	0.00
v43							.728		0.018	39.624	0.00
v47							.662		0.017	38.489	0.00
v23								.649	0.014	45.017	0.00
v32								.616	0.014	43.598	0.00
v40								.344	0.012	27.652	0.00
v48								.662	0.015	45.537	0.00
v50								.693	0.015	46.727	0.00
$\Omega$	0.850	0.733	0.765	0.767	0.775	0.720	0.662	0.730			

Nota: v1=ítem 1, v2=ítem2, ..., v50=ítem50; ECE-A= Escala de Convivencia Escolar Adaptada; F1= Gestión interpersonal positiva; F2= Victimización; F3= Disruptividad; F4= Red social de iguales; F5= Agresión; F6= Ajuste normativo; F7= Indisciplina; F8= Desidia docente.  $\Omega$ = Coeficiente Omega

El ajuste del modelo fue determinado a través de los índices de bondad de ajuste absoluto RMSEA ( $\leq 0.06$ ) y SRMSR ( $\leq 0.08$ ) y los índices de bondad de ajuste comparativo CFI ( $\geq 0.96$ ) y TLI ( $\geq 0.95$ ) (Hu y Bentler, 1999; Abada et al., 2011). En la tabla 31, se despliegan los valores de los índices de bondad de ajuste, mismos que indican un ajuste óptimo (RMSEA=0.054 y SRMR=0.060) aunque en algunos casos con valores cercanos a los valores esperados (CFI=0.949 y TLI=0.946).

Tabla 31

*Estadísticos e índices de bondad de ajuste del modelo de rasgos correlacionados*

Estadístico	Valor
N	906
Estimador	ULS
Chi-cuadrado	4172.789
Grados de libertad	1147
CFI ( $\geq 0.96$ )	0.949
TLI ( $\geq 0.95$ )	0.946
RMSEA ( $\leq 0.06$ )	0.054
Intervalo de confianza (90 %)	0.052 - 0.056
Valor- <i>p</i> RMSEA	0.000
SRMR ( $\leq 0.08$ )	0.060

Con base en la información obtenida, se puede afirmar que el modelo propuesto (Figura 6) muestra una alta correspondencia con los datos bajo estudio. En consecuencia, se puede sostener que la escala identifica los mismos componentes de la convivencia escolar considerados en su versión original. Estos hallazgos se suman al conjunto de evidencias de confiabilidad y validez relacionadas con la estructura interna de la ECE-A acumuladas en el presente estudio.

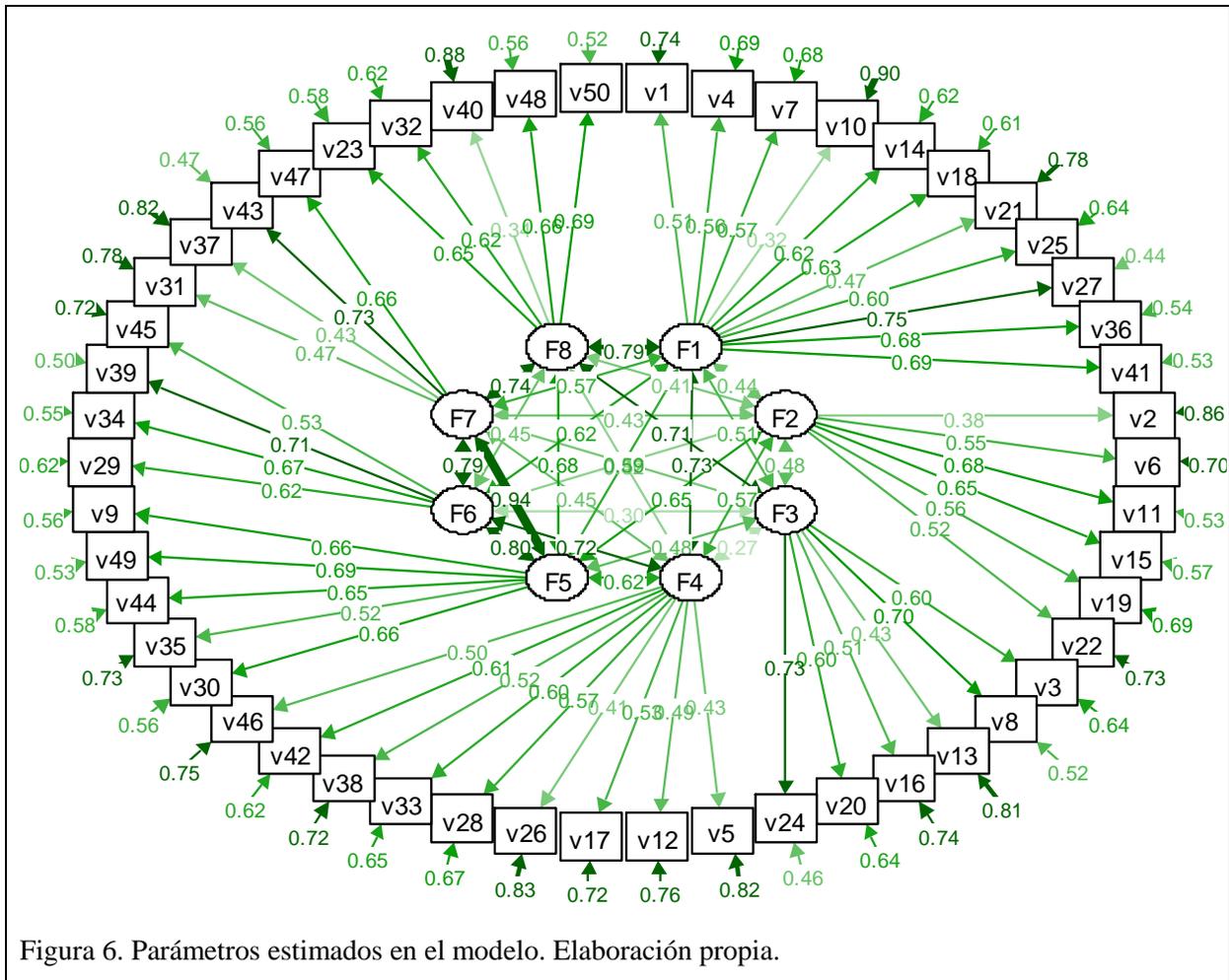
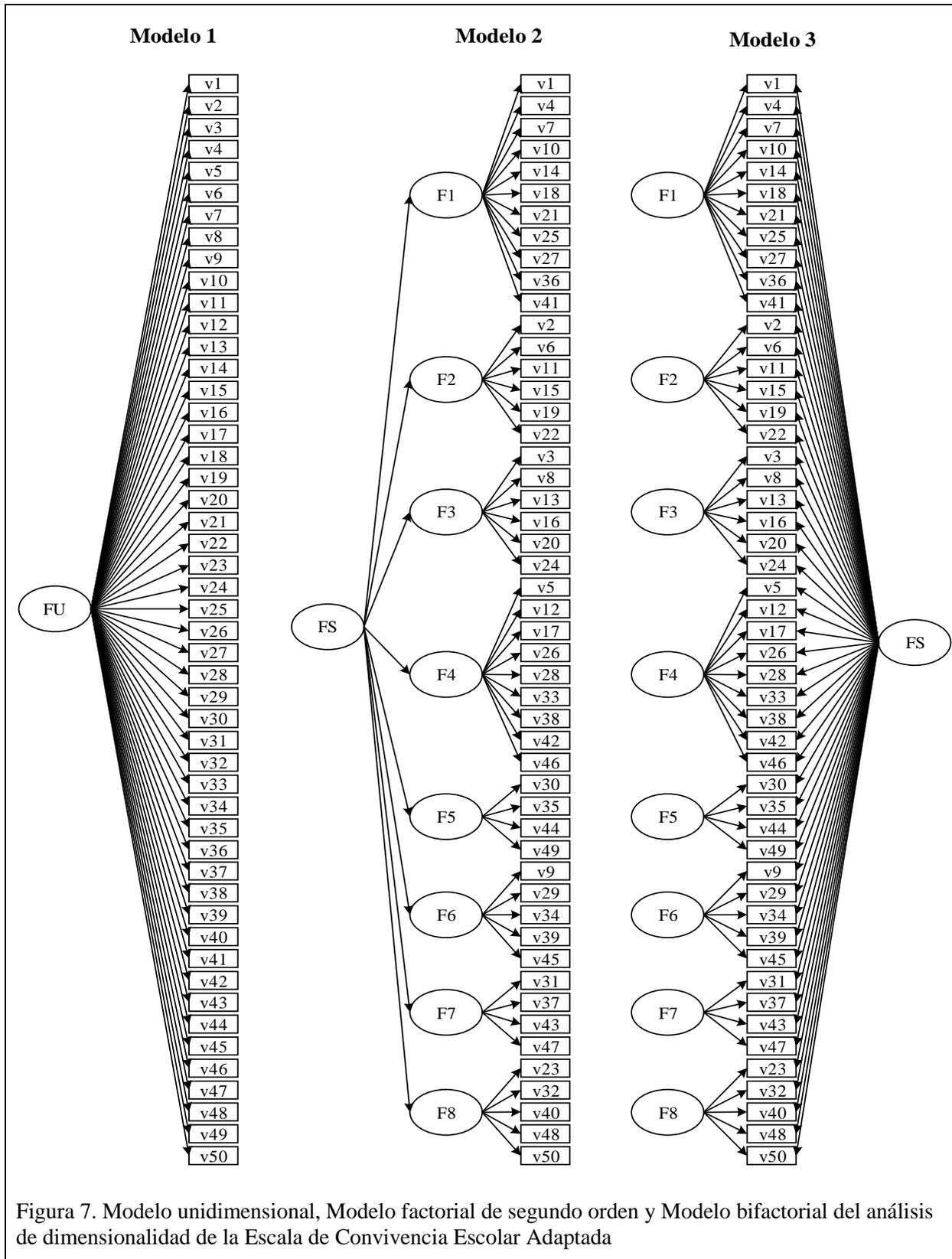


Figura 6. Parámetros estimados en el modelo. Elaboración propia.

**Análisis 2: Análisis de dimensionalidad.** Para estudiar la dimensionalidad de la ECE-A, se propuso una estrategia analítica en donde se compararon los niveles de ajuste, cargas factoriales e índices de consistencia interna de 3 modelos de medida (Figura 7), los cuales aspiran a explicar de manera más precisa la estructura de la escala. Así, se exploró el ajuste del modelo unidimensional (Modelo 1), del modelo factorial de segundo orden (Modelo 2) y del modelo bifactor (Modelo 3), basados en la matriz de correlación polícórica (Anexo I). Para la valoración del ajuste de los modelos se utilizaron los indicadores de ajuste absoluto (SRMR y RMSEA) y los indicadores de ajuste comparativo (TLI y CFI), y se tomaron como referencia los valores descritos en la Tabla 32.



Así, el modelo unidimensional (Modelo 1) registró un  $X^2$  (1175, N=906)=8390.782,  $p<.05$  y pobres niveles de ajuste comparativo y absoluto (Tabla 32) lo que sugiere que el modelo propuesto no ajusta a las respuestas de los sujetos en cuestión. Sin embargo, tal como se observa en la Tabla 33, las cargas factoriales de los ítems fueron positivas y estadísticamente significativas, siendo el ítem 13 aquel que obtuvo la carga factorial más baja (.244) y el ítem 9 el que obtuvo la más alta (.661).

Tabla 32  
*Estadísticos de ajuste e índices de proximidad asociados con los modelos propuestos*

	N	$X^2$	gl	CFI	TLI	RMSEA (90 % de confianza)	SRMR
Modelo 1	906	8390.782	1175	0.878	0.873	0.082 (0.081 - 0.084)*	0.085
Modelo 2	906	5968.502	1167	0.919	0.915	0.067 (0.066 - 0.069)*	0.072
Modelo 3	906	5257.810	1125	0.930	0.924	0.064(0.062 - 0.065)*	0.068

Nota: N=Número de sujetos;  $X^2$ =Valor de chi-cuadrada; gl=Grados de libertad.

Por su parte, el modelo factorial de segundo orden (Modelo 2) presentó bajos índices de ajuste comparativo (CFI=.919, TLI=.915) y niveles aceptables en los índices de ajuste absoluto (RMSEA=0.67, SRMR=0.072). Además, registró un  $X^2$  (1167, N=906)=5968.502,  $p<.05$ , lo que sugiere que el modelo presentó un buen ajuste a los datos (Tabla 32).

Las cargas de los factores F1, F2, F3, F4, F5, F6 F7 y F8 en el factor de segundo orden fueron de .824, .615, .584, .732, .877, .779, .814 y .810, respectivamente, todos con una  $p<.05$ , lo que sugiere un factor global relativamente fuerte. La tabla 33 presenta las cargas factoriales con la transformación Schmid-Leiman. Finalmente, al igual que el modelo 1, todas las cargas factoriales registradas fueron positivas y mayores a .30. Por último, el modelo bifactorial (Modelo 3), conformado por un factor general y ocho factores de primer orden presentó un  $X^2$  (1125, N=906)=5257.810,  $p<.05$  y índices aceptables de ajuste absoluto (RMSEA=0.064, SRMR=0.068)

pero niveles bajos de ajuste comparativo (CFI=0.930, TLI=0.924). A pesar de la falta de ajuste del modelo, es importante notar que el modelo bifactorial mostró un ajuste más óptimo en comparación con el Modelo 1 y el Modelo 2.

Tabla 33  
Cargas factoriales estimadas asociadas con los modelos propuestos

Ítem	Modelo A	Modelo B										Modelo C							
	FU	FS	F1	F2	F3	F4	F5	F6	F7	F8	FG	F1	F2	F3	F4	F5	F6	F7	F8
v1	.455	.419	.374								.420	.288							
v4	.497	.459	.409								.462	.301							
v7	.506	.471	.419								.457	.391							
v10	.298	.272	.242								.288	.122							
v14	.552	.505	.450								.528	.238							
v18	.557	.517	.460								.510	.395							
v21	.424	.391	.348								.396	.240							
v25	.523	.489	.436								.467	.445							
v27	.661	.617	.550								.607	.481							
v36	.607	.564	.503								.565	.386							
v41	.612	.566	.504								.582	.310							
v2	.259	.239		.360							.243	.279							
v6	.342	.333		.501							.313	.559							
v11	.433	.416		.626							.406	.586							
v15	.433	.407		.613							.412	.466							
v19	.386	.348		.524							.379	.256							
v22	.322	.312		.469							.290	.515							
v3	.382	.350			.564						.356		.451						
v8	.452	.408			.657						.432		.427						
v13	.244	.243			.391						.202		.549						
v16	.311	.299			.481						.272		.564						
v20	.371	.346			.557						.339		.524						
v24	.473	.430			.693						.451		.476						
v5	.349	.319				.368					.332			.225					
v12	.363	.354				.408					.317			.541					
v17	.422	.393				.453					.405			.279					
v26	.289	.290				.334					.231			.617					
v28	.467	.419				.483					.476			.084					
v33	.443	.427				.493					.400			.574					
v38	.418	.391				.451					.397			.313					
v42	.496	.457				.527					.491			.214					
v46	.374	.360				.415					.332			.509					
v30	.567	.573					.417				.566				.386				
v35	.445	.451					.328				.439				.356				
v44	.558	.567					.413				.554				.471				
v49	.588	.598					.436				.583				.546				
v9	.607	.600					.437				.640				-.084				
v29	.486	.484						.494			.486					.367			
v34	.527	.526						.537			.526					.443			
v39	.556	.556						.568			.556					.480			
v45	.413	.407						.416			.410					.288			
v31	.385	.386							.355		.396							.100	
v37	.337	.342							.315		.338							.243	
v43	.573	.587							.540		.580							.927	
v47	.541	.547							.503		.553							.203	
v23	.540	.530								.494	.543								.225
v32	.506	.501								.467	.505								.317
v40	.256	.257								.239	.227								.477
v48	.548	.544								.506	.550								.304
v50	.561	.556								.518	.551								.524
<b>Ω</b>	<b>.929</b>	<b>.922</b>	<b>.712</b>	<b>.687</b>	<b>.733</b>	<b>.681</b>	<b>.498</b>	<b>.576</b>	<b>.476</b>	<b>.555</b>	<b>.922</b>	<b>.570</b>	<b>.600</b>	<b>.662</b>	<b>.599</b>	<b>.400</b>	<b>.422</b>	<b>.397</b>	<b>.445</b>

Nota: v1=ítem 1; v2=ítem2; ..., v50=ítem50; FU= Factor unidimensional; FS=Factor de segundo orden;F1= Gestión interpersonal positiva; F2= Victimización; F3= Disruptividad; F4= Red social de iguales; F5= Agresión; F6= Ajuste normativo; F7= Indisciplina; F8= Desidia docente; FG=Factor General; Ω=Índice Omega

En relación con la confiabilidad/precisión, en la parte inferior de la tabla 33 se presentan los coeficientes de consistencia interna asociados a cada modelo. Así, el modelo unidimensional registró un  $\Omega=.929$ , el modelo factorial de segundo orden alcanzó índices del orden de .922 para el total de la escala y de .712, .687, .733, .681, .498, .576, .476 y .555 para las 8 subescalas (o factores), respectivamente. En contraste, en lo que respecta al Modelo 3 (modelo bifactorial), las estimaciones de los coeficientes de confiabilidad para los factores fueron sustancialmente inferiores: .570, .600, .662, .599, .400, .422, .397 y .445 para las subescalas F1, F2, F3, F4, F5, F6, F7 y F8, respectivamente; mientras que el total de la escala registró un  $\Omega=.922$ .

**Análisis 3: Análisis del funcionamiento diferencial de los ítems.** Previo al análisis del funcionamiento diferencial de los ítems se calcularon sus estadísticos descriptivos en función del sexo (tabla 34) y nivel educativo (tabla 35) de los estudiantes. De manera general, en cuanto al sexo de los estudiantes se observan diferencias absolutas inferiores a 1.00 en todos los ítems, siendo el ítem 49 (*He insultado a algún compañero(a).*) aquel que presentó la mayor diferencia entre mujeres y hombres (0.72 y 1.14, respectivamente); y el ítem 27 (*Mis profesores(as) me ayudan cuando lo necesito.*) el de la menor diferencia (2.97 y 2.98, respectivamente).

Tabla 34

*Estadísticos descriptivos de los ítems de la Escala de Convivencia Escolar Adaptada en función del sexo de los estudiantes*

	Femenino				Masculino			
	N	Media	DE	Mediana	N	Media	DE	Mediana
v1	447	3.31	.95	4	456	3.27	.93	4
v2	447	.47	.92	0	456	.40	.85	0
v3	448	1.64	1.28	2	455	1.47	1.22	2
v4	448	2.83	1.07	3	454	2.80	1.06	3
v5	448	2.73	1.01	3	455	2.85	.98	3
v6	447	.47	.96	0	456	.78	1.09	0
v7	448	3.04	1.07	3	455	3.06	1.05	3
v8	447	1.52	1.28	2	453	1.70	1.31	2
v9	446	3.49	.83	4	455	3.47	.77	4
v10	446	3.48	1.07	4	456	3.36	1.12	4
v11	448	1.25	1.20	1	456	1.53	1.24	1
v12	448	2.93	1.16	3	456	2.67	1.10	3
v13	444	2.57	1.24	3	453	2.41	1.33	2
v14	448	2.99	1.02	3	456	3.09	.99	3
v15	448	.30	.83	0	455	.33	.85	0
v16	447	2.34	1.11	2	456	2.33	1.11	2
v17	447	2.89	1.07	3	453	2.85	1.01	3
v18	448	3.02	1.16	3	455	2.99	1.13	3
v19	445	.93	1.17	0	454	1.06	1.22	1
v20	448	1.93	1.28	2	455	1.81	1.27	2
v21	448	2.98	1.34	4	454	2.97	1.35	4
v22	448	.96	1.19	0	455	.85	1.16	0
v23	448	.79	1.23	0	454	.84	1.24	0
v24	448	1.35	1.23	1	454	1.43	1.22	1
v25	447	2.62	1.26	3	453	2.78	1.22	3
v26	447	2.51	1.20	3	454	2.25	1.18	2
v27	447	2.97	1.18	3	456	2.98	1.15	3
v28	448	2.87	1.21	3	455	2.68	1.19	3
v29	447	3.36	.98	4	453	3.12	1.08	3
v30	448	.50	.90	0	453	.89	.99	1
v31	446	1.00	1.31	0	451	1.21	1.38	1
v32	444	1.19	1.27	1	454	1.11	1.24	1
v33	445	3.00	1.01	3	455	3.04	1.02	3
v34	446	3.10	1.14	4	455	2.98	1.13	3
v35	445	.84	1.12	0	455	.97	1.10	1
v36	446	3.17	1.18	4	455	3.22	1.10	4
v37	446	.78	1.07	0	452	1.15	1.14	1
v38	445	3.16	1.16	4	454	3.11	1.15	4
v39	446	3.07	1.06	3	453	3.06	1.03	3
v40	446	2.36	1.37	2	456	2.20	1.33	2
v41	446	2.91	1.23	3	456	2.97	1.16	3
v42	446	3.09	1.05	3	456	2.89	1.05	3
v43	446	.46	.92	0	456	.62	.99	0
v44	446	.35	.94	0	456	.54	1.00	0
v45	445	2.99	1.18	3	453	2.75	1.26	3
v46	443	3.33	1.11	4	455	3.36	1.10	4
v47	443	1.37	1.18	1	454	1.54	1.25	2
v48	444	1.44	1.22	1	456	1.46	1.16	1
v49	443	.72	1.01	0	456	1.14	1.15	1
v50	444	1.72	1.43	2	455	1.85	1.43	2

Asimismo, se registraron diferencias en la mayoría de todos los ítems en función del nivel educativo de los estudiantes. Sin embargo, estas diferencias fueron más fuertes en comparación con las registradas anteriormente. Así, en el ítem 50 (*Hay alumnos(as) que les caen mal a los profesores(as).*) se registró la diferencia más alta entre el nivel primaria y secundaria con valores de 0.90 y 2.26, respectivamente; mientras que el ítem 46 (*Siento que tengo amigos(as).*) y el ítem 12 (*Mis compañeros(as) me ayudan cuando lo necesito.*), no presentaron diferencias entre ningún grupo, registrando medias de 3.34 y 2.80, respectivamente.

Tabla 35

*Estadísticos descriptivos de los ítems de la Escala de Convivencia Escolar Adaptada en función del nivel educativo de los estudiantes*

	Primaria				Secundaria			
	N	Media	DE	Mediana	N	Media	DE	Mediana
v1	321	3.46	.90	4	585	3.20	.95	3
v2	320	.46	.93	0	586	.42	.86	0
v3	321	1.03	1.15	1	585	1.85	1.21	2
v4	320	3.23	1.10	4	585	2.58	.97	3
v5	321	2.92	.99	3	585	2.72	.99	3
v6	321	.74	1.12	0	585	.56	.99	0
v7	321	3.45	.92	4	585	2.84	1.06	3
v8	318	.98	1.13	1	585	1.95	1.25	2
v9	320	3.72	.65	4	584	3.35	.84	4
v10	320	3.44	1.00	4	585	3.40	1.15	4
v11	321	1.28	1.22	1	586	1.45	1.22	1
v12	321	2.80	1.19	3	586	2.80	1.10	3
v13	319	2.03	1.36	2	581	2.73	1.18	3
v14	321	3.60	.79	4	586	2.73	.98	3
v15	321	.29	.81	0	585	.33	.85	0
v16	320	1.98	1.08	2	586	2.53	1.08	2
v17	320	3.11	1.00	3	583	2.74	1.04	3
v18	321	3.39	1.05	4	585	2.80	1.14	3
v19	320	.79	1.05	0	582	1.10	1.26	1
v20	321	1.45	1.31	1	585	2.10	1.19	2
v21	321	3.21	1.39	4	584	2.85	1.30	3
v22	321	1.02	1.22	1	585	.84	1.15	0
v23	321	.43	1.03	0	584	1.03	1.29	0
v24	321	.82	1.09	0	584	1.70	1.19	2
v25	320	3.23	1.13	4	583	2.41	1.21	2
v26	321	2.33	1.24	2	583	2.41	1.17	2
v27	320	3.42	.99	4	586	2.73	1.18	3
v28	321	2.87	1.24	3	585	2.72	1.18	3
v29	320	3.37	1.04	4	583	3.17	1.04	4
v30	320	.55	.90	0	584	.78	.99	0
v31	320	.92	1.39	0	580	1.21	1.32	1
v32	319	.61	1.11	0	582	1.44	1.23	2
v33	321	3.12	1.04	3	582	2.97	.99	3
v34	321	3.24	1.07	4	583	2.92	1.15	3
v35	320	.88	1.15	0	583	.92	1.09	1
v36	321	3.57	.91	4	583	2.99	1.20	3
v37	320	1.04	1.05	1	581	.93	1.15	0
v38	321	3.43	1.04	4	581	2.97	1.18	3
v39	318	3.34	.94	4	584	2.91	1.07	3
v40	321	1.77	1.34	2	584	2.56	1.28	3
v41	321	3.29	1.04	4	584	2.76	1.23	3
v42	321	3.19	1.00	4	584	2.88	1.07	3
v43	321	.30	.72	0	584	.67	1.04	0
v44	321	.36	.87	0	584	.49	1.02	0
v45	318	3.03	1.18	3	583	2.78	1.24	3
v46	319	3.34	1.18	4	582	3.34	1.07	4
v47	320	1.08	1.14	1	580	1.66	1.21	2
v48	321	.89	1.10	0	582	1.76	1.13	2
v49	321	.66	.96	0	581	1.07	1.15	1
v50	320	.90	1.19	0	582	2.26	1.31	2

De acuerdo con los resultados del análisis de dimensionalidad, se estimaron las puntuaciones de los estudiantes con base en un modelo multidimensional (Modelo 3, en el análisis de dimensionalidad) y bajo el paradigma de la Teoría de la Respuesta al Ítem (Drasgow, Levine y Williams, 1985), a través del método Esperado *a posteriori* (EAP). El ajuste del modelo multidimensional y la estimación de las puntuaciones se realizó en el paquete *mirt* (Chalmers, 2012; Chalmers, 2014) del programa R versión 3.1.1. (ver sintaxis en el Anexo H).

Para estudiar el funcionamiento diferencial de los ítems (DIF) se utilizó el enfoque de regresión logística propuesto por Swaminathan y Rogers (1990) y Zumbo (1999) para detectar DIF uniforme y no-uniforme así como su magnitud.

Tabla 36

*Modelos de regresión logística utilizados para detectar DIF*

	Variable dependiente	Variabes independientes
Modelo 1	Respuesta al ítem	Habilidad
Modelo 2	Respuesta al ítem	Habilidad + Grupo
Modelo 3	Respuesta al ítem	Habilidad + Grupo + Habilidad*Grupo

Nota: Elaboración propia basada en “A handbook on the theory and methods of differential item functioning (DIF)” por B. Zumbo; 1999; *Ottawa: National Defense Headquarters*.

En la Tabla 36 se presentan los modelos de regresión logística utilizados para detectar el tipo y magnitud de DIF de cada uno de los ítems. El Modelo 1 asume que la respuesta de los sujetos es explicada por el nivel de habilidad (puntuación estimada previamente) en toda la escala. El Modelo 2 supone que la respuesta de los sujetos está condicionada por el nivel de habilidad y el grupo de pertenencia de los sujetos (hombre, mujer o nivel primaria y nivel secundaria). Por último, el Modelo 3 considera que la respuesta de los sujetos a un ítem está condicionada por la habilidad, el grupo de pertenencia y por el producto de la interacción de la habilidad y el grupo de pertenencia.

Así, el DIF uniforme puede detectarse a través de la prueba de significatividad estadística de las diferencias de los valores del Modelo 1 y 2, y el DIF no-uniforme de la comparación del Modelo 2 y 3. Asimismo, las diferencias de los coeficientes de determinación ( $R^2$ ) del Modelo 1 y 2, y del Modelo 2 y 3 permiten conocer la magnitud del DIF uniforme y no-uniforme, respectivamente.

En la Tabla 37 se presentan los resultados del DIF para las variables sexo y nivel educativo. Como se puede observar, 19 de 50 ítems presentaron un funcionamiento diferencial en cuanto al sexo de los estudiantes (16 de tipo uniforme y 3 de tipo no uniforme) y 34 de 50 ítems presentaron un funcionamiento diferencial en cuanto al nivel educativo (31 de tipo uniforme y 3 de tipo no uniforme). Sin embargo, como lo señalan Li y Stout (1996) el método utilizado tiene un buen poder de detección pero conlleva a la presencia de error Tipo I (considerar un ítem con DIF cuando en realidad éste no lo tiene). En consecuencia, se consideró la magnitud del DIF para valorar si un ítem presentaba un funcionamiento diferencial con un alto efecto para alguno de los grupos.

Siguiendo dicho procedimiento, se encontró que la magnitud de los ítems identificados con DIF (uniforme o no-uniforme) no es mayor a 0.13 tanto para la variable sexo como para la variable nivel educativo, lo que sugiere que el efecto de los ítems detectados con DIF es insignificante (Zumbo, 1999), y no representa un efecto significativo en las probabilidades de adscribirse [o no] a cada ítem.

Tabla 37

*Análisis del funcionamiento diferencial de los ítems en función del sexo y nivel educativo de los estudiantes*

Ítems	Sexo		Nivel escolar	
	Tipo	Magnitud	Tipo	Magnitud
1. Mis profesores(as) se llevan bien entre ellos.			Uniforme	0.0053
3. Dentro de la escuela hay problemas de vandalismo.	Uniforme	0.0027	Uniforme	0.0272
4. Hay buenas relaciones entre los profesores(as) y los alumnos(as).			Uniforme	0.0430
5. Los alumnos(as) nos llevamos bien entre nosotros.	Uniforme	0.0027		
6. Algún compañero(a) me ha golpeado.	Uniforme	0.0166	Uniforme	0.0137
7. Los padres de familia se llevan bien con los profesores(as).			Uniforme	0.0312
8. Hay peleas en las que la gente se golpea.			Uniforme	0.0378
9. Aprendo en mi escuela.			Uniforme	0.0225
10. Mis padres acuden a la escuela cuando son llamados por mis profesores.	Uniforme	0.0037		
11. Algún compañero(a) me ha insultado.	Uniforme	0.0042		
12. Mis compañeros(as) me ayudan cuando lo necesito.	Uniforme	0.0051		
13. Hay alumnos(as) que no dejan dar clase.	No uniforme	0.0022	Uniforme	0.0175
14. Los profesores(as) son respetados.			Uniforme	0.0870
15. Me siento amenazado(a) en la escuela.	No uniforme	0.0071		
16. Hay alumnos(as) que no respetan las reglas.	No uniforme	0.0045	Uniforme	0.0170
17. Me uno a las actividades que realizan los demás.			Uniforme	0.0060
18. Los profesores(as) dan ejemplo de buenas relaciones.			Uniforme	0.0233
20. Hay alumnos(as) que siempre se meten en peleas.			No uniforme	0.0050
21. Los profesores(as) evitan que nos burlemos unos de otros.			Uniforme	0.0084
22. Me he sentido excluido, aislado o rechazado por compañeros(as).			Uniforme	0.0087
23. Los profesores(as) sólo explican para los más listos(as) de la clase.			Uniforme	0.0235
24. Algunos alumnos(as) destruyen el material y las instalaciones.			Uniforme	0.0388
25. Los profesores(as) nos ayudan a resolver nuestros problemas.	Uniforme	0.0036	Uniforme	0.0328
26. Mis compañeros(as) se interesan por mí.	Uniforme	0.0039		
27. Mis profesores(as) me ayudan cuando lo necesito.			Uniforme	0.0269
29. Dejo trabajar a los demás sin molestarlos.	Uniforme	0.0072		
30. He golpeado a algún compañero(a).	Uniforme	0.0288	Uniforme	0.0045
31. Sólo cumplo las reglas que me convienen.			Uniforme	0.0035
32. Las reglas de los profesores(as) son injustas.			Uniforme	0.0423
36. Mis padres se llevan bien con mis profesores(as).			Uniforme	0.0209
37. Me han castigado en la escuela.	Uniforme	0.0130	Uniforme	0.0085
38. Me gusta trabajar en equipo.			Uniforme	0.0119
39. Cumplo las reglas de mi escuela y de mi grupo.			Uniforme	0.0083
40. Hay profesores(as) que castigan siempre a los mismos alumnos(as).			Uniforme	0.0209
41. Mis profesores(as) valoran mi trabajo y me animan a mejorar.			Uniforme	0.0114
42. Ayudo a mis compañeros(as) cuando me lo piden.	Uniforme	0.0042		
43. Interrumpo la clase porque me aburro.	Uniforme	0.0066	Uniforme	0.0115
44. He amenazado a algún compañero.	Uniforme	0.0163		
45. Respeto la opinión de los demás aunque no esté de acuerdo.	Uniforme	0.0028		
46. Siento que tengo amigos(as).			No uniforme	0.0059
47. Me aburro en la escuela.			Uniforme	0.0111
48. Los profesores(as) ponen actividades aburridas.			Uniforme	0.0421
49. He insultado a algún compañero(a).	Uniforme	0.0167	No uniforme	0.0042
50. Hay alumnos(as) que les caen mal a los profesores(as).			Uniforme	0.0925

## Capítulo 6. Discusión

La presente investigación tuvo como objetivo principal desarrollar e implementar una metodología para la adaptación de instrumentos de medición en contextos educativos, atendiendo puntualmente las directrices, estándares y recomendaciones emitidas por distintos organismos especializados en el diseño, desarrollo, adaptación y validación de instrumentos de medición psicológica y educativa; y a través de la implementación de dicha metodología a la adaptación de la Escala de Convivencia Escolar (Ortega-Ruiz et al., 2013) al contexto de Baja California. En lo particular, se articularon un conjunto de consideraciones metodológicas que permitieran alcanzar dicho objetivo y se documentaron evidencias de validez, confiabilidad, equivalencia y sesgo asociadas con el proceso de adaptación de la escala en cuestión.

La metodología propuesta se conformó de seis etapas. En la primera de éstas, la etapa de traducción, se transformaron las instrucciones y términos presentes en la versión original, a fin de adecuarlos al contexto de esta entidad. El ejercicio contó con la participación y acompañamiento de una especialista elegida en apego a un perfil determinado con antelación.

La instrumentación de esta etapa permitió identificar a un número considerable de ítems ( $k=14$ ) que requirieron de alguna modificación, lo que representa un hallazgo importante que alerta sobre la necesidad de emplear procedimientos para la traducción y adaptación de instrumentos en el contexto iberoamericano, aun cuando el idioma fuente y el idioma objetivos sean los mismos.

Si bien se cumplió con el objetivo propuesto por esta etapa, es importante subrayar que para futuras réplicas de sus procedimientos debería incluirse a más de un traductor como

responsable de dicha tarea. Aun considerando que los criterios propuestos como parte del perfil del especialista fueron los adecuados, dicho ejercicio podría haberse enriquecido con la participación de otros especialistas con otros perfiles y formados en otras áreas de especialización (lingüistas, traductores profesionales, etc.). A este respecto se ha sugerido que la integración de un equipo multidisciplinario debe considerar especialistas que permitan evaluar la precisión de la traducción; examinar los aspectos estructurales, funcionales y semánticos de la traducción; valorar el apego a las convenciones de escritura técnica de ítems; e identificar aquellos contenidos con sesgo cultural (INEE, 2011).

Por su parte, la etapa de validación lingüística permitió conocer las valoraciones de un grupo de expertos en cuanto a la equivalencia lingüística de los ítems, específicamente en cuanto a aspectos gramaticales y semánticos. En lo particular, se considera un acierto haber incluido a un sociolingüista, un traductor y un especialista en convivencia escolar, como integrantes del grupo de especialistas en quienes descansó la presente actividad. Sus valoraciones permitieron acumular elementos de naturaleza cuantitativa y cualitativa con respecto al nivel de equivalencia lingüística de cada ítem, así como identificar a un conjunto de ítems ( $k=15$ ) susceptibles de ajustes o modificaciones.

Los resultados de la etapa anterior refuerzan el empleo de procedimientos que incluyan las valoraciones de expertos en los procesos de adaptación de todo instrumento, ya que permiten identificar con mayor precisión diferencias culturales presentes en el lenguaje. Sin embargo, y sin menospreciar las aportaciones que puedan realizar los especialistas a título personal, es deseable contar con una sesión a manera de panel, en el que los jueces contrasten sus valoraciones, intercambien puntos de vista y establezcan acuerdos. Para la réplica de la presente metodología se

recomienda realizar una sesión de panel con expertos, preferentemente presencial, donde se expongan y contrasten sus argumentos, lo que permitiría identificar errores de difícil detección cuando se trabaja de manera individual, siendo posible llegar a un consenso en la valoración de cada ítem (INEE, 2011).

La etapa de validación de contenido se implementó con el propósito de evaluar la congruencia, relevancia, claridad y grado de suficiencia de los ítems que conformaban la versión original de la ECE. Al igual que las etapas anteriores, conocer y contrastar las valoraciones que los jueces hacen sobre estos aspectos resulta imprescindible en todo proceso de adaptación y validación de instrumentos.

La presente etapa permite hacer precisiones a los acuerdos establecidos en las etapas anteriores. Tal fue el caso del ítem 17 (en su forma original: *Me uno a las actividades que realizan los demás, en su forma original*) cuya redacción fue modificada en la etapa de traducción, y valorado como equivalente en términos lingüísticos en la etapa de validación lingüística, cuando en la etapa de validación de contenido los jueces acordaron que el fraseo original del ítem resultaba de mayor pertinencia para la población a la que se aplicaría dicha escala (estudiantes de primaria y secundaria).

En suma, los procedimientos utilizados en esta etapa redondean la evidencia de validez de contenido de la ECE acumulada en las etapas anteriores, ya que considera los resultados previos como insumos, se sustenta en las valoraciones de especialistas en medición, en la disciplina (convivencia escolar) y en los ámbitos donde se realizará la aplicación (agentes educativos del sistema educativo estatal), y permite el establecimiento de acuerdos entre jueces. Un aspecto deseable, más no imprescindible, sería el contar con la participación de los autores de la versión

original del instrumento a fin de realizar precisiones y aportaciones que enriquecieran al proceso de adaptación. De acuerdo a Byrne et al. (2009), la colaboración entre los autores del test original y los adaptadores ayuda a generar mejores instrumentos para la evaluación intercultural.

La cuarta etapa del proceso implicó la aplicación piloto de los ítems que derivaron del procedimiento anterior. Esta prueba permitió valorar el funcionamiento inicial de la escala y de sus ítems, estimar el tiempo en que los alumnos tardan en responder a la escala, registrar aquellas incidencias que pudieran comprometer la aplicación, así como capturar todas aquellas expresiones asociadas a problemas de interpretación y comprensión de las instrucciones, opciones de respuesta y fraseo de los ítems.

Así, fue posible identificar que algunos estudiantes no comprendieron el ítem 50 (*Hay alumnos(as) a los que los profesores(as) les tienen tirria*), valorado de manera óptima en las etapas anteriores. Si bien podría subestimarse el presente hallazgo, éste permite ilustrar la contribución que tiene esta etapa al proceso en su conjunto pues representa el último filtro antes de la aplicación a mayor escala, además de que arroja información que hace las veces de insumo para la planeación y organización de dicha aplicación.

La quinta etapa del proceso consistió en la aplicación del instrumento a la población objetivo, en este caso, estudiantes de quinto y sexto de primaria y de primero a tercero de secundaria. Se siguieron y documentaron puntualmente los procedimientos comprometidos en la selección de la muestra, la capacitación a aplicadores, la aplicación de la ECE y en la captura y elaboración de la base de datos, respetando las recomendaciones, estándares y directrices a los que se ha hecho alusión en este estudio (AERA, APA y NCME, 2014; ITC, 2010). Si bien la muestra de estudiantes que participaron en esta aplicación no es representativa de la población escolar de

Baja California, es importante recordar que la selección de la muestra buscaba reproducir las características de la muestra del estudio original en el que participaron estudiantes españoles, tal y como se sugiere en la literatura especializada.

Ahora bien, en lo relativo a los análisis de equivalencia y sesgo comprometidos en la sexta etapa del proceso de adaptación, se documentaron ampliamente las evidencias de validez, confiabilidad, niveles de equivalencia y de presencia-ausencia de sesgo, producto del empleo de un conjunto de técnicas estadísticas indicadas para este propósito. Así, los resultados del análisis de la equivalencia de constructo confirmó la estructura teórico-hipotética de la versión original de la ECE en el conjunto de datos obtenidos por la ECE-A, indicando el ajuste de los datos al modelo donde las ocho dimensiones que subyacen a la convivencia escolar están siendo identificadas y medidas en la versión adaptada. Las ocho dimensiones registraron índices de consistencia interna aceptables.

Cabe mencionar que estos análisis se apoyaron en el uso de la matriz de correlaciones policóricas y el método de estimación de mínimos cuadrados no ponderados, así como en el uso de las cargas factoriales para calcular los índices de consistencia interna con mayor precisión, respetando en todo momento los supuestos que se desprenden de la naturaleza y niveles de medida de las variables involucradas. Autores tales como Gignac et al., (2007), Reise et al., (2010), y Hancock y Mueller (2001) han realizado pronunciamientos importantes en esta materia, destacando la necesidad de incorporar dichos procedimientos en los esquemas analíticos que subyacen a los estudios orientados al análisis de la estructura interna de los instrumentos.

Por su parte, en cuanto a los análisis de dimensionalidad, el modelo de medida *bifactor* resultó ser el que mejor ajusta a los datos y el que permite estudiar un rasgo común entre todos los

ítems que conforman el constructo *convivencia escolar*, así como a los subconjuntos de ítems (dimensiones) que integran la escala, lo que posibilita el uso de las puntuaciones de las dimensiones de manera independiente.

En la actualidad, existe en la literatura especializada un amplio reconocimiento a la importancia de realizar estudios de dimensionalidad como parte del proceso de desarrollo, adaptación y/o validación de un test (Reise et al., 2010; Reise, 2012). Sin embargo, en el contexto nacional, esta práctica aún no se ha visto generalizada, por lo que su inclusión en esta metodología resulta un elemento que contribuye a la promoción de este tipo de prácticas en el país.

Finalmente se analizó el funcionamiento diferencial de los ítems (DIF) teniendo como variables de criterio, por un lado, el sexo, y por otro, el nivel educativo de los estudiantes, utilizando la puntuación estimada de los sujetos desde un modelo *bifactor*, en el marco de la teoría de la respuesta al ítem. Lo anterior permitió contar con una puntuación estimada de los sujetos más cercana a su “puntuación verdadera”, aislándose el efecto de la multidimensionalidad de la escala, tal y como lo sugieren Zumbo (1999) y Fukohara (2009).

A este respecto, si bien se identificaron un número importante de ítems que presentan un funcionamiento diferencial (19 ítems en función del sexo, y 34 ítems en función del nivel educativo), la magnitud de estas diferencias, de acuerdo con Zumbo (1999), es considerada como insignificante. Es decir, no representan un impacto significativo en las probabilidades que tienen los sujetos de adscribirse [o no] a cada uno de los ítems.

Como puede apreciarse, y tal como lo sostienen Byrne et al. (2009), las actuales prácticas de medición demandan que los profesionales de la evaluación que participan en estudios

interculturales tengan un profundo entendimiento de los distintos efectos que el idioma y la cultura pueden tener en el diseño de un test, así como conocimiento de las herramientas necesarias para evaluar estos efectos tanto a nivel de ítem como a nivel de escala. Además, necesitan comprender la importancia de vincular procedimientos de jueceo y métodos empíricos en la adaptación de instrumentos para uso intercultural.

A este respecto, el presente estudio permitió sugerir un conjunto de consideraciones metodológicas asociadas con la traducción y adaptación de instrumentos de medición psicológica y educativa. Este estudio pasa a formar parte del conjunto de investigaciones realizadas en nuestro país (aún reducido) que reconocen la necesidad de utilizar metodologías, procedimientos y esquemas analíticos que proporcionen equidad en las evaluaciones y justicia a los examinados, con reconocimiento a la contribución que hacen las diferencias culturales en la explicación de los resultados educativos (Byrne et al., 2009; Solano et al., 2009; INEE, 2011).

### Referencias

- Abad, F. J., Olea, J., Ponsoda, V., y García, C. (2011). *Medición de ciencias sociales y de la salud*. Madrid, España: Síntesis.
- Álvarez-García, D., Núñez, J. C., Rodríguez, C., Álvarez, L., y Dobarro, A. (2011). Propiedades psicométricas del Cuestionario de Violencia Escolar - Revisado (CUVE-R). *Revista de Psicodidáctica*, 16(1), 59–83.
- American Education Research Association, American Psychological Association, y National Council on Measurement in Education. (1985). *Standards for Educational and Psychological Testing 1985*. Washington, DC: American Education Research Association.
- American Education Research Association, American Psychological Association, y National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing 1999*. Washington, DC: American Education Research Association.
- American Education Research Association, American Psychological Association, y National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing 2014*. Washington, DC: American Education Research Association.
- APA. (2013). *Going International: A Practical Guide for Psychologists. Using Translated/Adapted Measurement Scales Internationally*. American Psychological Association.
- Bazdresch, M. (2009). La vida cotidiana escolar en la formación valoral: un caso. *Revista Iberoamericana Sobre Calidad, Eficacia y Cambio En Educación*, 7(2), 50–71.

Borsboom, D., Mellenbergh, G. J., y van Heerden, J. (2004). The Concept of Validity. *Psychological Review*, *111*(4), 1061–1071. doi:10.1037/0033-295X.111.4.1061

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, *20*(4), 872–882.

Byrne, B. M., Oakland, T., Leong, F. T. L., van de Vijver, F., Hambleton, R. K., Cheung, F. M., y Bartram, D. (2009). A critical analysis of cross-cultural research and testing practices: Implications for improved education and training in psychology. *Training and Education in Professional Psychology*, *3*(2), 94–105. doi:10.1037/a0014516

Byrne, B. M., Stewart, S. M., y Lee, P. W. (2004). Validating the Beck depression inventory-II for Hong Kong Community adolescents. *International Journal of Testing*, *4*(3), 199–216.

Byrne, B. M., y Watkins, D. (2003). The Issue Of Measurement Invariance Revisited. *Journal of Cross-Cultural Psychology*, *34*(2), 155–175. doi:10.1177/0022022102250225

Casagrande, J. (1954). The Ends of Translation. *International Journal of American Linguistics*, *20*(4), 335–340.

Caso, J., Díaz, C., y Chaparro, A. (2012). Caracterización da violencia escolar en escuelas secundarias de Baja California. *Revista Mexicana de Psicología*.

Caso, J., Díaz, C., y Chaparro, A. (2013). Aplicación de un procedimiento para la optimización de la medida de la convivencia escolar. *Revista Iberoamericana de Evaluación Educativa*. Recuperado de <https://repositorio.uam.es/handle/10486/661805>

Chalmers, P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, 48(6), 1-29. Recuperado de <http://www.jstatsoft.org/v48/i06/>.

Chalmers, P. (2014). Package “mirt”. Recuperado de <http://cran.r-project.org/web/packages/mirt/mirt.pdf>

Choi, S. W., Gibbons, L. E., y Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, 39(8), 1.

Cronbach, L. (1988). Five perspectives on validation argument. In *Test validity* (pp. 3–17). Hillsdale, NJ: Erlbaum.

Cook, L., Schmitt-Cascallar, A., y Brown, C. (2005). Adapting Achievement and Aptitude Tests: A Review of Methodological Issues. En Hambleton, R. K., Mereda, P., y Spielberger, C. (2005). *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (1st ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Delors, J. (1996). *La educación encierra un tesoro: informe para la UNESCO de la Comisión Internacional sobre la Educación para el Siglo Veintiuno*. UNESCO. Recuperado de <http://dide.minedu.gob.pe/xmlui/handle/123456789/1847>

Drasgow, F. (1984). Scrutinizing psychological tests: Measurement equivalence and equivalent relations with external variables are the central issues. *Psychological Bulletin*, 95(1), 134–135.

Drasgow, F., Levine, M. V., y Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *Journal of Mathematical and Statistical Psychology*, 38, 67-86.

Elosua, P. (2003). Sobre la validez de los tests. *Psicothema*, 15(2), 315–321.

Elosua, P. (2012). Test publicados en España: usos, costumbres y asignaturas pendientes. *Papeles Del Psicólogo*, 33(1), 12–21.

Elosua, P. y López, A. (1999). Funcionamiento diferencial de los ítems y sesgo en la adaptación de dos pruebas verbales. *Psicológica*, 20(1), 23–40.

Evers, A., Muñiz, J., Hagemester, C., HstmæLingen, A., Lindley, P., Sjöberg, A., y Bartram, D. (2013). Assessing the quality of tests: revision of the EFPA review model. *Psicothema*, 25(3), 283–291.

Fierro, C. (2013). Convivencia inclusiva y democrática. Una perspectiva para gestionar la seguridad escolar. *Sinética*, 40. Recuperado de [http://www.sinectica.iteso.mx/?seccion=articulo&lang=es&id=561\\_convivencia\\_inclusiva\\_y\\_democratica\\_una\\_perspectiva\\_para\\_gestionar\\_la\\_seguridad\\_escolar](http://www.sinectica.iteso.mx/?seccion=articulo&lang=es&id=561_convivencia_inclusiva_y_democratica_una_perspectiva_para_gestionar_la_seguridad_escolar)

Fierro, C., y Caso, J. (2013). Presentación. *Revista Iberoamericana de Evaluación Educativa*, 6(2), 7–12.

- Fukuhara, H. (2009). A Differential Item Functioning Model for Testlet-based Items Using a Bi-factor Multidimensional Item Response Theory Model: A Bayesian Approach. *Electronic Theses, Treatises and Dissertations*, 4384.
- Gadermann, A. M., Guhn, M., y Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research y Evaluation*, 17(3), 1–13.
- Geisinger, K. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, 6(4), 304–312.
- Gignac, G. E., Palmer, B. R., y Stough, C. (2007). A confirmatory factor analytic investigation of the TAS–20: Corroboration of a five-factor model and suggestions for improvement. *Journal of Personality Assessment*, 89(3), 247–257.
- Hambleton, R. K. (1996). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, 10, 229–240.
- Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment*, 17(3), 164–172.
- Hambleton, R. K. (2004). Theory, methods, and practices in testing for the 21st century. *Psicothema*, 16(4), 696–701.

Hambleton, R. K. (2005). Issues, Designs, and Technical Guidelines for Adapting Tests Into Multiple Languages and Cultures. En Hambleton, R. K., Mereda, P., y Spielberger, C. (2005). *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (1st ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hambleton, R. K., Mereda, P., y Spielberger, C. (2005). *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. New Jersey, London: Lawrence Erlbaum Associates.

Hambleton, R. K., Yu, J., y Slater, S. C. (1999). Field test of the ITC Guidelines for adapting educational and psychological tests. *European Journal of Psychological Assessment*, 15(3), 270–276.

Hancock, G. R., y Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In *Structural equation modeling: Present and future—A Festschrift in honor of Karl Joreskog* (pp. 195–216). Lincolnwood, IL: Scientific Software International.

He, J., y van de Vijver, F. (2012). Bias and Equivalence in Cross-Cultural Research. *Online Readings in Psychology and Culture*, 2(2). doi:10.9707/2307-0919.1111

Hulin, C., y Mayer, L. (1986). Psychometric equivalence of a translation of the Job Descriptive Index into Hebrew. *Journal of Applied Psychology*, 71(1), 83.

Instituto Nacional para la Evaluación de la Educación. (2011). *La Teoría del Error de Traducción de Pruebas y las evaluaciones internacionales de TIMSS y PISA* (Primera.). México, D.F.: Instituto Nacional para la Evaluación de la Educación.

International Test Commission. (2010). International Test Commission Guidelines for Translating and Adapting Tests. Recuperado de <http://www.intestcom.org>

Klerk, G. (2008). Cross-cultural testing. In *Online Readings in Testing and Assessment*. International Test Commission. Recuperado de <http://www.intestcom.org/Publications/ORTA.php>

Kline, R. B. (2010). *Principles and Practice of Structural Equation Modeling* (3rd edition.). New York: The Guilford Press.

Li H. y Stout W. (1996). A New Procedure for Detection of Crossing DIF. *Psychometrika*, 61(4), 647-677.

Luo, H. (2011). *Some Aspects on Confirmatory Factor Analysis of Ordinal Variables and Generating Non-normal Data*. Upssala Universitet, Upssala, Suecia. Recuperado de <http://www.diva-portal.org/smash/record.jsf?pid=diva2:405108>

Mereda, P. (2005). Cross-Cultural Adaptation of Educational and Psychological Testing. En Hambleton, R. K., Mereda, P., y Spielberger, C. (2005). *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (1st ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Messick, S. (1995). Validity of psychological assessment. *Educational Testing Service*, 50(9), 741-749.

- Miller, K., Mont, D., Maitland, A., Altman, B., y Madans, J. (2010). Results of a cross-national structured cognitive interviewing protocol to test measures of disability. *Quality y Quantity*, 45(4), 801–815. doi:10.1007/s11135-010-9370-4
- Muñiz, J., Elosua, P., y Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*, 25(2), 151–157.
- Muñiz, J., y Fernández-Hermida, J. R. (2010). La opinión de los psicólogos españoles sobre el uso de los tests. *Papeles Del Psicólogo*, 31(1), 108–121.
- Muñiz, J., Fernández-Hermida, J. R., Fonseca-Pedrero, E., Campillo-Álvarez, Á., y Peña-Suárez, E. (2011). Evaluación de tests editados en España. *Papeles Del Psicólogo*, 32(2), 113–128.
- Muñiz, J., y Hambleton, R. K. (1996). Directrices para la traducción y adaptación de los tests. *Papeles Del Psicólogo*, (66), 63–70.
- Organization for Economic Co-operation and Development. (2010). Translation and adaptation guidelines for PISA 2012. Recuperado de <http://www.oecd.org/pisa/pisaproducts/49273486.pdf>
- Ortega-Ruiz, R., Del Rey, R., y Casas, J. A. (2013). La Convivencia Escolar: clave en la predicción del Bullying. *Revista Iberoamericana de Evaluación Educativa*, 6(2), 91–102.
- Ortega-Ruiz, R., Del Rey, R., y Sánchez, V. (2012). *Nuevas dimensiones de la convivencia escolar y juvenil. Ciberconducta y relaciones en la Red: Ciberconvivencia*. Madrid, España: Observatorio Estatal de la Convivencia Escolar.

Orue, I., y Calvete, E. (2012). La justificación de la violencia como mediador de la relación entre la exposición a la violencia y la conducta agresiva en infancia. *Psicothema*, 24(1), 42–47.

Steven P. Reise (2012) The Rediscovery of Bifactor Measurement Models, *Multivariate Behavioral Research*, 47:5, 667-696, DOI: 10.1080/00273171.2012.715555

Reise, S. P., Moore, T. M., y Haviland, M. G. (2010). Bifactor Models and Rotations: Exploring the Extent to which Multidimensional Data Yield Univocal Scale Scores. *Journal of Personality Assessment*, 92(6), 544–559. doi:10.1080/00223891.2010.496477

Revelle, W. (2014). psych: Procedures for psychological, psychometric, and personality research. *R Package Version, 1019*. Recuperado de <http://cran.univ-lyon1.fr/web/packages/psych/>

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36.

Rosseel, Y. (2014). Package “lavaan.” Recuperado de <http://cran.md.tsukuba.ac.jp/web/packages/lavaan/lavaan.pdf>

Sireci, S., Patsula, L., y Hambleton, R. K. (2005). Statistical Methods for Identifying Flaws in the Test Adaptation Process. En Hambleton, R. K., Mereda, P., y Spielberger, C. (2005). *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (1st ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Swaminathan H., y Rogers H. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational Measurement*, 27, 361-370.

Solano-Flores, G., Backhoff, E., y Contreras-Niño, L. Á. (2009). Theory of Test Translation Error. *International Journal of Testing*, 9(2), 78–91. doi:10.1080/15305050902880835

Tanzer, N. K. (1995). Cross-cultural bias in Likert-type inventories: Perfect matching factor structures and still biased? *European Journal of Psychological Assessment*, 11(3), 194–201.

Tanzer, N. K. (2005). Developing Tests for Use in Multiple Languages and Cultures A Plea for Simultaneous Development. En Hambleton, R. K., Mereda, P., y Spielberger, C. (2005). *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (1st ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Tanzer, N. K., y Sim, C. Q. (1999). Adapting instruments for use in multiple languages and cultures: A review of the ITC Guidelines for Test Adaptations. *European Journal of Psychological Assessment*, 15(3), 258–269.

Van de Vijver, F., y Hambleton, R. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 1(2), 89–99.

Van de Vijver, F., y Poortinga, Y. H. (2002). Structural Equivalence in Multilevel Research. *Journal of Cross-Cultural Psychology*, 33(2), 141–156. doi:10.1177/0022022102033002002

Van de Vijver, F., y Watkins, D. (2006). Assessing Similarity of Meaning at the Individual and Country Level. *European Journal of Psychological Assessment*, 22(2), 69–77. doi:10.1027/1015-5759.22.2.69

Van de Vijver, F., y Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *European Review of Applied Psychology*, 54(2), 119–135.  
doi:10.1016/j.erap.2003.12.004

Zumbo, B. D. (1999). A handbook on the theory and methods of differential item functioning (DIF). *Ottawa: National Defense Headquarters*. Recuperado de <http://faculty.educ.ubc.ca/zumbo/DIF/handbook.pdf>

**Anexos**

**Anexo A. Procedimiento general para de la adaptación de la Escala de Convivencia Escolar**

Fase	Etapas	Procedimiento	Participantes	Materiales	Productos
Empírico-racional	1. Traducción	Traducción	#1 Especialista en elaboración de test #1 Traductor	#Versión original de la escala	#Reporte de Modificaciones 1
		Modificación 1		#Reporte de Modificaciones 1	#Versión 1 ECE-A
	2. Validación lingüística	Validación lingüística	#1 Traductor #1 Sociolingüista #1 Especialista en convivencia Escolar	#Objetivos y estructura de la escala #Versión original de la escala #Versión 1 ECE-A	#Reporte de Modificaciones 2
		Modificación 2		#Reporte de Modificaciones 2	#Versión 2 ECE-A
	3. Validación de contenido	Validación de contenido	#1 Especialista en Convivencia Escolar #1 Especialista en elaboración de test #1 Director de primaria #1 Director de secundaria #1 Docente de primaria #1 Docente de secundaria	#Objetivos y estructura de la escala #Versión original de la escala #Versión 2 ECE-A	#Reporte de Modificaciones 3
		Modificación 3		#Reporte de modificaciones 3	#Versión 3 ECE-A
	4. Pilotaje	Muestreo		#Características de la muestra española	#Muestra para pilotaje
		Aplicación		#Versión 3 ECE-A	#Instrumentos respondidos
		Captura de información	#21 estudiantes de quinto grado de primaria	#Formato de reporte de incidencias #Instrumentos respondidos #Incidencias de aplicación	#Incidencias de aplicación #Base de datos de pilotaje #Base de datos de incidencias
		Análisis de información		#Base de datos de pilotaje	#Resultados de análisis de pilotaje
		Modificación 4		#Resultados de análisis de pilotaje. #Base de datos de incidencias.	#Reporte de modificaciones 3 #Versión 3 ECE-A (Final)
	5. Aplicación	Muestreo		#Características de la muestra de España	#Muestra para aplicación en Baja California
Guía de aplicación				#Guía de aplicación	
Capacitación de aplicadores		#12 estudiantes de la licenciatura en Psicología y Ciencias de la Educación	#Guía de aplicación	#Aplicadores capacitados	
Aplicación		#12 Aplicadores capacitados	#Versión 3 ECE-A #Reportes de incidencia	#Instrumentos respondidos #Informe de reportes de incidencia	
Captura de información			#Instrumentos respondidos #Informe de reportes de incidencia	#Base de datos de la ECE-A	
Análisis de sesgo					
6. Análisis de equivalencia y sesgo	Análisis iniciales				
	Equivalencia de constructo	906 estudiantes de primaria y secundaria	#Base de datos de la ECE-A	#Reporte de resultados	
	Análisis de sesgo				

**Anexo B. Traducción**

Sección de la hoja electrónica para la tarea de traducción

	A	B	C	D	E	F	G	H	I
1		<b>Original</b>	<b>Traducción</b>	<b>Observaciones</b>					
2	1	Los profesores/as se llevan bien entre ellos							
3	2	He tenido miedo de venir a la escuela							
4	3	Dentro del centro hay problemas de vandalismo							
5	4	Hay buenas relaciones entre los profesores/as y alumnos/as							
6	5	Los alumnos/as nos llevamos bien							
7	6	Algún compañero/a me ha golpeado							
8	7	Los padres se llevan bien con los profesores/as							
9	8	Hay peleas en las que la gente se pega							
10	9	Aprendo							
11	10	Las familias del alumnado se implican en las actividades							
12	11	Algún compañero/a me ha insultado							
13	12	Mis compañeros/as me ayudan cuando lo necesito							
14	13	Hay niños/as que no dejan dar clase							
15	14	Los profesores/as son respetados							
16	15	Me he sentido amenazado/a							
17	16	Hay alumnos/as que no respetan las normas							
18	17	Me uno a las actividades que realizan los demás							
19	18	Los profesores/as son ejemplo de buenas relaciones							
20	19	Me han robado							
21	20	Hay niños/as que siempre están metidos en peleas							
22	21	Los profesores/as evitan que nos burlemos unos de otros							
23	22	Me he sentido excluido, aislado o rechazado por compañeros/as							

**Anexo C. Validación lingüística**

Sección del formato de captura de valoraciones

A	B	C	D	E	F	G	H	J
#	Ítems de la Escala de Convivencia Escolar	Ítems traducidos de la Escala de Convivencia Escolar	Gramática	Semántica	Equivalencia	Observaciones o modificaciones propuestas	Seguimiento	
1								
2	1 Los profesores/as se llevan bien entre ellos	Los profesores(as) se llevan bien entre ellos						
3	2 He tenido miedo de venir a la escuela	He tenido miedo de venir a la escuela						
4	3 Dentro del centro hay problemas de vandalismo	Dentro de la escuela hay problemas de vandalismo						
5	4 Hay buenas relaciones entre los profesores/as y alumnos/as	Hay buenas relaciones entre los profesores(as) y los alumnos(as)						
6	5 Los alumnos/as nos llevamos bien	Los alumnos(as) nos llevamos bien entre nosotros						
7	6 Algún compañero/a me ha golpeado	Algún compañero(a) me ha golpeado						
8	7 Los padres se llevan bien con los profesores/as	Los padres de familia se llevan bien con los profesores(as)						
9	8 Hay peleas en las que la gente se pega	Hay peleas en las que la gente se golpea						
10	9 Aprendo	Aprendo						
11	10 Las familias del alumnado se implican en las actividades	Las familias de los alumnos(as) participan en las actividades escolares						
12	11 Algún compañero/a me ha insultado	Algún compañero(a) me ha insultado						
13	12 Mis compañeros/as me ayudan cuando lo necesito	Mis compañeros(as) me ayudan cuando lo necesito						
14	13 Hay niños/as que no dejan dar clase	Hay alumnos(as) que no permiten dar clase						
15	14 Ítem5 Los profesores/as son respetados	Los profesores(as) son respetados						
16	15 Me he sentido amenazado/a	Me he sentido amenazado(a)						
17	16 Hay alumnos/as que no respetan las normas	Hay alumnos(as) que no respetan las normas						
18	17 Me uno a las actividades que realizan los demás	Me integro en las actividades que realizan los demás						
19	18 Los profesores/as son ejemplo de buenas relaciones	Los profesores(as) son ejemplo de buenas relaciones						
20	19 Me han robado	Me han robado						
21	20 Hay niños/as que siempre están metidos en peleas	Hay alumnos(as) que siempre están en peleas						
22	21 Los profesores/as evitan que nos burlemos unos de otros	Los profesores(as) evitan que nos burlemos unos de otros						
23	22 Me he sentido excluido, aislado o rechazado por compañeros/as	Me he sentido excluido, aislado o rechazado por compañeros(as)						
24	23 Los profesores/as sólo explican para los listos/as de la clase	Los profesores(as) sólo explican para los listos(as) de la clase						
25	24 Algunos alumnos/as destruyen el material y las instalaciones	Algunos alumnos(as) destruyen el material y las instalaciones						
26	25 Los profesores/as nos ayudan a resolver nuestros problemas	Los profesores(as) nos ayudan a resolver nuestros problemas						
27	26 Mis compañeros/as se interesan por mí	Mis compañeros(as) se interesan por mí						
28	27 Mis profesores/as me ayudan cuando lo necesito	Mis profesores(as) me ayudan cuando lo necesito						

## Manual para los jueces en la etapa de validación lingüística

**Equivalencia lingüística de la traducción de la Escala de Convivencia Escolar****Antecedentes**

La convivencia escolar es un tema que ha ido adquiriendo relevancia tanto en el campo de investigación educativa como en las agendas de los propios sistemas educativos y de los medios de comunicación, formando parte de los intereses y preocupaciones de docentes, directivos, padres de familia y diversos sectores de la población (Fierro y Caso, 2013; Ortega-Ruiz, Del Rey y Casas, 2013). Existen distintas aproximaciones teóricas asociadas con el estudio de la convivencia escolar lo que ha contribuido a carecer de una delimitación conceptual clara del constructo, dificultando tanto su **operacionalización** como el desarrollo de instrumentos que permitan su medición (Caso, Díaz y Chaparro, 2013; Ortega-Ruiz, Del Rey y Casa, 2013; Fierro y Caso, 2013). Contar con instrumentos de medición que sirvan a este propósito permitirían además establecer las bases para el diseño e implementación de intervenciones específicas orientadas a mejorar la convivencia en los centros escolares (Caso, Díaz y Chaparro, 2013).

En este marco, *el Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California (México) y el Laboratorio de Estudios sobre Convivencia y Prevención de la Violencia de la Universidad de Córdoba (España), establecieron un acuerdo de colaboración en el que se comprometió* la adaptación de la Escala de Convivencia Escolar (ECE), utilizada en población española, a fin de caracterizar la convivencia escolar en estudiantes de educación primaria y secundaria en Baja California. En lo particular, el presente acuerdo compromete la adaptación de la ECE (Ortega-Ruiz, Del Rey y Casas, 2013), instrumento de autoinforme que valora distintas dimensiones de la convivencia escolar (gestión interpersonal positiva, victimización, relaciones entre iguales, agresión, ajuste normativo, indisciplina y desidia docente), a través de la percepción de los estudiantes. La versión original del instrumento se conforma por 50 ítems tipo Likert, con cinco opciones de respuesta, y explora la frecuencia en que se presentan determinadas situaciones en el contexto escolar.

Dadas las diversas consideraciones que realizan en materia de **equivalencia lingüística** organismos tales como la APA y la ITC, mismos que establecen que de requerirse que un instrumento desarrollado en un idioma y país determinado se aplique en otro idioma o en países con el mismo idioma pero con diferencias interculturales distintas, se propone realizar la adaptación de la ECE en población mexicana.

Realizar una adaptación válida involucra más procesos que la simple traducción de un idioma a otro. Se deben considerar los **aspectos psicológicos, psicométricos, culturales y lingüísticos** involucrados; así como todas las idiosincrasias específicas de un idioma o cultura particular que pueda causar sesgo en la evaluación (Tanzer & Sim, 1999). El fin último de la adaptación es lograr la equivalencia de constructo, así como la equivalencia lingüística entre las escalas de medición original y objetivo (APA, 2013).

Así, con el **propósito** de asegurar la equivalencia lingüística de los ítems que conforman la Escala de Convivencia Escolar (ECE), cuya posterior aplicación permita la caracterización de una muestra de estudiantes de escuelas primarias (quinto y sexto grado) y secundarias (primero, segundo y tercer año) en Baja California, se propuso el presente estudio.

### Metodología

En lo general, la **metodología** propuesta para atender este propósito se conforma por seis etapas (I) traducción; (II) validación lingüística; (III) validación de contenido; (IV) prueba piloto; (V) aplicación del instrumento; y (VI) análisis de sesgo. La descripción a cada una de estas etapas se presenta en la siguiente tabla:

Tabla 1.

*Etapas de la metodología propuesta para la adaptación de la Escala de Convivencia Escolar (Ortega-Ruiz, Del Rey y Casas, 2013).*

<b>Etapas</b>	<b>Descripción</b>
I. Traducción	Transformación de términos o secciones del ítem, en términos adecuados y entendidos en el contexto de Baja California.
II. Validación lingüística	Procedimiento para asegurar la calidad de la traducción de los ítems de la Escala de convivencia Escolar. A través de juicios de expertos se evalúa la equivalencia gramatical y semántica de la traducción de la traducción de cada ítem.
III. Validación de contenido	Procedimiento empleado para asegurar que la calidad de la escala no sea afectada negativamente por la traducción. A través de juicios de expertos se evalúa la claridad, pertinencia y suficiencia de cada ítem.
IV. Prueba piloto	Consiste en una aplicación a una muestra pequeña de estudiantes, con el objetivo de conocer el funcionamiento inicial de la escala y los ítems, y realizar las modificaciones necesarias.
V. Aplicación	Aplicación de la escala a una muestra de estudiantes más amplia y comparable con la muestra utilizada en España para validar la Escala de Convivencia Escolar.
VI. Análisis de sesgo y equivalencia	Empleo de un conjunto de técnicas estadísticas que permitan conocer el grado en el cual los ítems y la escala traducida son equivalentes con la escala original, en términos de medición del constructo y de las puntuaciones obtenidas.

Recientemente se atendió de manera satisfactoria la transformación de términos o secciones de los ítems en su versión original, en términos adecuados y entendidos en el contexto de Baja California (Etapa I).

A fin de proceder con la **validación lingüística de los ítems** que aseguren la calidad de la traducción de los ítems y la equivalencia gramatical y semántica de cada ítem, se propone el siguiente procedimiento en el que se requerirá del juicio de expertos.

**Descripción de la tarea**

En la siguiente tarea se contará con la participación de tres especialistas en traducción, sociolingüismo y convivencia escolar, quienes fungirán como jueces.

Cada uno de ellos evaluará la equivalencia lingüística de la traducción de los ítems, considerando los siguientes rasgos:

- Gramática. La traducción del ítem no presenta errores gramaticales y la sintaxis es simple y común para la población destinataria.
- Semántica. Las ideas y el significado transferidos al ítem traducido son iguales a los del ítem en el idioma fuente.
- Equivalencia. El ítem traducido y el ítem en el idioma fuente son equivalentes en términos gramaticales y semánticos.

**Cómo evaluar.**

Junto con el presente documento, se anexa un archivo de Excel titulado *Archivo de evaluación.xlsx* En las primeras dos columnas de este archivo se presentan los ítems de la versión original de la Escala de Convivencia Escolar, y los ítems correspondientes a la versión traducida.

Al costado derecho de estas dos columnas, se encuentran otras cuatro tituladas (1) Gramática, (2) Semántica, (3) Equivalencia y (4) Observaciones o modificaciones propuestas. Finalmente, se encuentra una columna iluminada en rojo que hace las veces de indicador para asegurar que no quede ningún ítem sin evaluar.

Para proceder con la evaluación, los jueces deberán seguir el siguiente procedimiento:

1. Para evaluar determinado ítem, posicionarse en la celda correspondiente al rasgo a evaluar (Gramática, Semántica o Equivalencia), tal y como se muestra en la siguiente imagen:

A	B	C	D	E	F	G	H
#	Ítems de la Escala de Convivencia Escolar	Ítems traducidos de la Escala de Convivencia Escolar	Gramática	Semántica	Equivalencia	Observaciones o modificaciones propuestas	Seguimiento
1							
2	1 Los profesores/as se llevan bien entre ellos	Los profesores(as) se llevan bien entre ellos					
3	2 He tenido miedo de venir a la escuela	He tenido miedo de venir a la escuela					
4	3 Dentro del centro hay problemas de vandalismo	Dentro de la escuela hay problemas de vandalismo					
5	4 Hay buenas relaciones entre los profesores/as y alumnos/as	Hay buenas relaciones entre los profesores(as) y los alumnos(as)					
6	5 Los alumnos/as nos llevamos bien	Los alumnos(as) nos llevamos bien entre nosotros					
7	6 Algún compañero/a me ha golpeado	Algún compañero(a) me ha golpeado					
8	7 Los padres se llevan bien con los profesores/as	Los padres de familia se llevan bien con los profesores(as)					
9	8 Hay peleas en las que la gente se pega	Hay peleas en las que la gente se golpea					

Nota: cuando se emiten las valoraciones correspondientes, la columna de *Seguimiento* cambia su color a verde

#	Ítems de la Escala de Convivencia Escolar	Ítems traducidos de la Escala de Convivencia Escolar	Gramática	Semántica	Equivalencia	Observaciones o modificaciones propuestas	Seguimiento
1	Los profesores/as se llevan bien entre ellos	Los profesores(as) se llevan bien entre ellos	1	0	1		
2	He tenido miedo de venir a la escuela	He tenido miedo de venir a la escuela	1	1	1		
3	Dentro del centro hay problemas de vandalismo	Dentro de la escuela hay problemas de vandalismo	1	1	0		
4	Hay buenas relaciones entre los profesores/as y alumnos/as	Hay buenas relaciones entre los profesores(as) y los alumnos(as)					
5	Los alumnos/as nos llevamos bien	Los alumnos(as) nos llevamos bien entre nosotros					

2. Se asignará “1” si se considera que el ítem cumple de manera satisfactoria con el rasgo a evaluar (Gramática, Semántica o Equivalencia), “-1” si considera que no cumple de manera satisfactoria con el rasgo a evaluar, y “0” si presenta dudas con respecto al grado de cumplimiento del rasgo a evaluar.

2.1. En caso de que se asigne “-1” o “0”, se deberá emitir alguna recomendación o propuesta de modificación a la traducción del ítem, misma que habrá de incluirse en las celdas correspondientes a la columna *Observaciones o modificaciones propuestas*.

Nota: Si usted asigna un valor diferente a -1, 0, 1, el sistema se detendrá y desplegará el siguiente mensaje:



3. El fondo del encabezado del archivo cambiará de color blanco a color verde, cuando usted haya concluido el presente ejercicio.

#	Ítems de la Escala de Convivencia Escolar	Ítems traducidos de la Escala de Convivencia Escolar	Gramática	Semántica	Equivalencia	Observaciones o modificaciones propuestas	Seguimiento
1	Los profesores/as se llevan bien entre ellos	Los profesores(as) se llevan bien entre ellos	1	0	1		
2	He tenido miedo de venir a la escuela	He tenido miedo de venir a la escuela	1	1	1		
3	Dentro del centro hay problemas de vandalismo	Dentro de la escuela hay problemas de vandalismo	1	1	0		
4	Hay buenas relaciones entre los profesores/as y alumnos/as	Hay buenas relaciones entre los profesores(as) y los alumnos(as)	1	1	1		
5	Los alumnos/as nos llevamos bien	Los alumnos(as) nos llevamos bien entre nosotros	1	1	1		
6	Algún compañero/a me ha golpeado	Algún compañero(a) me ha golpeado	1	1	1		
7	Los padres se llevan bien con los profesores/as	Los padres de familia se llevan bien con los profesores(as)	1	1	1		
8	Hay peleas en las que la gente se pega	Hay peleas en las que la gente se golpea	1	1	1		
9	Aprendo	Aprendo	1	1	1		

4. Al finalizar, es de suma importancia proceda a “guardar” los cambios realizados en el archivo, tal y como se hace de manera tradicional. Posteriormente, envíe el documento a la dirección de correo electrónico: [carlos8diaz@gmail.com](mailto:carlos8diaz@gmail.com)

**Anexo D. Validación de contenido**

Material para los jueces participantes en la etapa de validación de contenido

**UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA**  
INSTITUTO DE INVESTIGACIÓN Y DESARROLLO EDUCATIVO



## Validación de contenido de la adaptación de la Escala de Convivencia Escolar

---

Ensenada, Baja California

1 de abril de 2014

### Validación del contenido de la adaptación de la Escala de Convivencia Escolar

El presente proceso de validación tiene el objetivo de recolectar evidencias de la congruencia, relevancia, claridad y suficiencia de los ítems que componen cada una de las dimensiones de la Escala de Convivencia Escolar (ECE). Con base en la evaluación que se derive de este ejercicio, se podrá establecer la validez de la adaptación, al contexto de Baja California, de la ECE.

#### Instrucciones

Marque el número que mejor refleje su opinión en relación con los criterios de congruencia, relevancia y claridad de cada ítem, así como la suficiencia de los ítems considerados para cada dimensión; de acuerdo a la información que se muestra en la tabla 1. Asimismo, emita las sugerencias u observaciones necesarias en el espacio respectivo a cada ítem.

Tabla 1.

*Criterios de validación del contenido de la adaptación de la Escala de Convivencia Escolar.*

<b>Criterio</b>	<b>Descripción</b>	<b>Escala</b>
Congruencia	Grado en que el ítem es consecuente con la dimensión correspondientes	(1) No congruente (2) Poco congruente (3) Congruente (4) Muy congruente
Relevancia	Grado en que la información contenida en el ítem es importante para evaluar la dimensión correspondiente.	(1) No relevante (2) Poco relevante (3) Relevante (4) Muy relevante
Claridad	Grado en que el ítem comunica de manera clara y directa la información, es decir que no admite más de una interpretación posible	(1) No claro (2) Poco claro (3) Claro (4) Muy claro
Suficiencia	Grado en que se incluyen los ítems necesarios para evaluar cada dimensión	(1) No suficiente (2) Poco suficiente (3) Suficiente (4) Muy suficiente

J -D1

**Dimensión 1:** Gestión interpersonal positiva

**Descripción:** Mide la forma en la que los profesores gestionan las relaciones interpersonales en el centro educativo, aludiendo a la conducta social ejemplar, la buena comunicación con el alumnado y sus familias y a la atención de las dificultades que éstos pueden encontrar

Ítems	Congruencia	Relevancia	Claridad	Observaciones
01. Los profesores(as) se llevan bien entre ellos.	1 2 3 4	1 2 3 4	1 2 3 4	
02. Hay buenas relaciones entre los profesores(as) y los alumnos(as).	1 2 3 4	1 2 3 4	1 2 3 4	
03. Los padres de familia se llevan bien con los profesores(as).	1 2 3 4	1 2 3 4	1 2 3 4	
04. Las familias de los alumnos(as) se comprometen con las actividades escolares.	1 2 3 4	1 2 3 4	1 2 3 4	
05. Los profesores(as) son respetados.	1 2 3 4	1 2 3 4	1 2 3 4	
06. Los profesores(as) dan ejemplo de buenas relaciones.	1 2 3 4	1 2 3 4	1 2 3 4	
07. Los profesores(as) evitan que nos burlemos unos de otros.	1 2 3 4	1 2 3 4	1 2 3 4	
08. Los profesores(as) nos ayudan a resolver nuestros problemas.	1 2 3 4	1 2 3 4	1 2 3 4	
09. Mis profesores(as) me ayudan cuando se lo pido.	1 2 3 4	1 2 3 4	1 2 3 4	
10. Mis padres se llevan bien con mis profesores(as).	1 2 3 4	1 2 3 4	1 2 3 4	
11. Los profesores(as) valoran mi trabajo y me animan a mejorar.	1 2 3 4	1 2 3 4	1 2 3 4	

Suficiencia	Observaciones
(1) No suficiente.	
(2) Poco suficiente.	
(3) Suficiente.	
(4) Muy suficiente.	

J -D2

**Dimensión 2:** Victimización

**Descripción:** Aborda la percepción de ser maltratado, atacado, golpeado

Ítems	Congruencia	Relevancia	Claridad	Observaciones
01. He tenido miedo de venir a la escuela.	1 2 3 4	1 2 3 4	1 2 3 4	
02. Algún compañero(a) me ha golpeado.	1 2 3 4	1 2 3 4	1 2 3 4	
03. Algún compañero(a) me ha insultado.	1 2 3 4	1 2 3 4	1 2 3 4	
04. Me he sentido amenazado(a).	1 2 3 4	1 2 3 4	1 2 3 4	
05. Me han robado.	1 2 3 4	1 2 3 4	1 2 3 4	
06. Me he sentido excluido, aislado o rechazado por compañeros(as).	1 2 3 4	1 2 3 4	1 2 3 4	

Suficiencia	Observaciones
(1) No suficiente	
(2) Poco suficiente	
(3) Suficiente	
(4) Muy suficiente	

J\_-D3

**Dimensión 3:** Disruptividad

**Descripción:** Se analizan los aspectos relativos al vandalismo, los destrozos o impedir el normal desarrollo de las clases.

Ítems	Congruencia	Relevancia	Claridad	Observaciones
01. Dentro de la escuela hay problemas de vandalismo.	1 2 3 4	1 2 3 4	1 2 3 4	
02. Hay peleas en las que la gente se golpea.	1 2 3 4	1 2 3 4	1 2 3 4	
03. Hay alumnos(as) que no dejan dar clase.	1 2 3 4	1 2 3 4	1 2 3 4	
04. Algunos alumnos(as) destruyen el material y las instalaciones.	1 2 3 4	1 2 3 4	1 2 3 4	
05. Hay alumnos(as) que siempre se meten en peleas.	1 2 3 4	1 2 3 4	1 2 3 4	
07. Dejo trabajar a los demás sin molestarlos.	1 2 3 4	1 2 3 4	1 2 3 4	

Suficiencia	Observaciones
(1) No suficiente	
(2) Poco suficiente	
(3) Suficiente	
(4) Muy suficiente	

J -D4

**Dimensión 4:** Red social de iguales

**Descripción:** Incluye elementos de las buenas relaciones entre los iguales.

Ítems	Congruencia	Relevancia	Claridad	Observaciones
01. Los alumnos (as) nos llevamos bien entre nosotros.	1 2 3 4	1 2 3 4	1 2 3 4	
02. Mis compañeros(as) me ayudan cuando lo necesito.	1 2 3 4	1 2 3 4	1 2 3 4	
03. Participo en las actividades que realizan los demás.	1 2 3 4	1 2 3 4	1 2 3 4	
04. Mis compañeros(as) se interesan por mí.	1 2 3 4	1 2 3 4	1 2 3 4	
05. Le caigo bien a mis compañeros(as).	1 2 3 4	1 2 3 4	1 2 3 4	
06. Me gusta trabajar en equipo.	1 2 3 4	1 2 3 4	1 2 3 4	
07. Ayudo a mis compañeros(as) cuando me lo piden.	1 2 3 4	1 2 3 4	1 2 3 4	
08. Siento que tengo amigos(as).	1 2 3 4	1 2 3 4	1 2 3 4	
09. Digo y defiendo mis opiniones sin dañar o molestar a los demás.	1 2 3 4	1 2 3 4	1 2 3 4	

Suficiencia	Observaciones
(1) No suficiente	
(2) Poco suficiente	
(3) Suficiente	
(4) Muy suficiente	

J -D5

**Dimensión 5:** Agresión

**Descripción:** Comportamientos de agresión de diversa naturaleza como física, verbal, social, etc.

Ítems	Congruencia	Relevancia	Claridad	Observaciones
01. He insultado a algún compañero(a).	1 2 3 4	1 2 3 4	1 2 3 4	
02. He golpeado a algún compañero(a).	1 2 3 4	1 2 3 4	1 2 3 4	
03. He amenazado a otra persona.	1 2 3 4	1 2 3 4	1 2 3 4	
04. He excluido o rechazado a algún compañero(a).	1 2 3 4	1 2 3 4	1 2 3 4	

Suficiencia	Observaciones
(1) No suficiente	
(2) Poco suficiente	
(3) Suficiente	
(4) Muy suficiente	

J -D6

**Dimensión 6:** Ajuste normativo

**Descripción:** Valora el ajuste a las normas sociales como pedir permiso para hablar, o respetar a los compañeros

Ítems	Congruencia	Relevancia	Claridad	Observaciones
01. Cumplo las normas.	1 2 3 4	1 2 3 4	1 2 3 4	
02. Sólo cumplo las normas que me convienen.	1 2 3 4	1 2 3 4	1 2 3 4	
03. Pido la palabra y espero mi turno para hablar.	1 2 3 4	1 2 3 4	1 2 3 4	
04. Respeto la opinión de los demás aunque no la comparto.	1 2 3 4	1 2 3 4	1 2 3 4	
05. Las normas de los profesores(as) son injustas.	1 2 3 4	1 2 3 4	1 2 3 4	

Suficiencia	Observaciones
(1) No suficiente	
(2) Poco suficiente	
(3) Suficiente	
(4) Muy suficiente	

J\_-D7

**Dimensión 7:** Indisciplina

**Descripción:** Aspectos relativos al incumplimiento de las normas, castigos, interrupciones de clase o aburrimiento

Ítems	Congruencia	Relevancia	Claridad	Observaciones
01. Me han castigado en la escuela.	1 2 3 4	1 2 3 4	1 2 3 4	
02. Me aburro en la escuela.	1 2 3 4	1 2 3 4	1 2 3 4	
03. Interrumpo la clase porque me aburro.	1 2 3 4	1 2 3 4	1 2 3 4	
04. Hay alumnos(as) que no respetan las normas.	1 2 3 4	1 2 3 4	1 2 3 4	

Suficiencia	Observaciones
(1) No suficiente	
(2) Poco suficiente	
(3) Suficiente	
(4) Muy suficiente	

J -D8

**Dimensión 8:** Desidia docente

**Descripción:** Refleja comportamientos propios de la desidia o desinterés docente o su incompetencia profesional, observada por el alumnado, en definitiva, la baja percepción del alumnado relativa al interés y el apoyo del profesorado.

Ítems	Congruencia	Relevancia	Claridad	Observaciones
01. Hay alumnos(as) a los que los profesores(as) les tienen tirria.	1 2 3 4	1 2 3 4	1 2 3 4	
02. Los profesores(as) sólo explican para los más listos(as) de la clase.	1 2 3 4	1 2 3 4	1 2 3 4	
03. Hay profesores(as) que castigan siempre a los mismos alumnos(as).	1 2 3 4	1 2 3 4	1 2 3 4	
04. Aprendo en mi escuela.	1 2 3 4	1 2 3 4	1 2 3 4	
05. Los profesores(as) ponen actividades aburridas.	1 2 3 4	1 2 3 4	1 2 3 4	

Suficiencia	Observaciones
(1) No suficiente	
(2) Poco suficiente	
(3) Suficiente	
(4) Muy suficiente	

Presentación de la sesión de la etapa de validación de contenido

26/12/2014



## Validación de contenido de la adaptación de la Escala de Convivencia Escolar

Carlos David Díaz López



Maestría en Ciencias Educativas

## Orden del día

- Presentación
- Antecedentes
- Validación de contenido
- Receso
- Sesión plenaria

## Antecedentes

La convivencia escolar ha ido adquiriendo relevancia tanto en el campo de investigación educativa como en las agendas de los propios sistemas educativos y de los medios de comunicación, formando parte de los intereses y preocupaciones de docentes, directivos, padres de familia y diversos sectores de la población (Fierro y Caso, 2013; Ortega-Ruiz, Del Rey y Casas, 2013).

Existen distintas aproximaciones teóricas asociadas con el estudio de la convivencia escolar lo que ha contribuido a carecer de una delimitación conceptual clara del constructo, dificultando tanto su **operacionalización** como el **desarrollo de instrumentos** que permitan su medición.

## Antecedentes

### Acuerdo de colaboración



Laboratorio de Estudios sobre Convivencia y Prevención de la Violencia.  
Universidad de Córdoba (España)



INSTITUTO DE INVESTIGACIÓN Y DESARROLLO EDUCATIVO

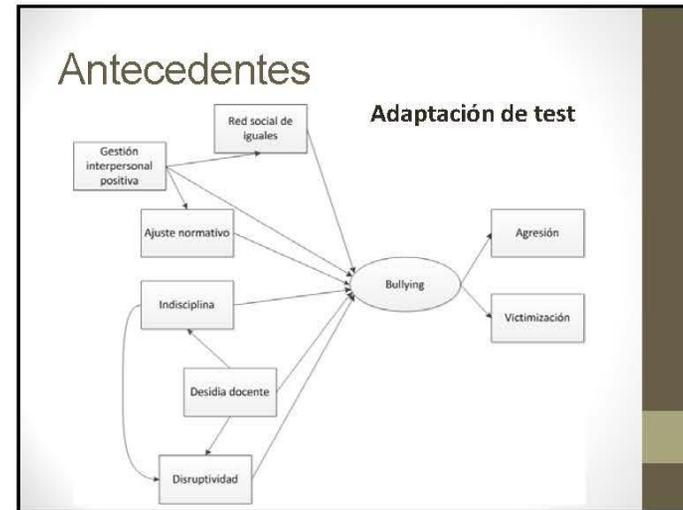
### Adaptación de la Escala de Convivencia Escolar

26/12/2014

## Antecedentes

### Escala de Convivencia Escolar

Dimensión	Número de ítems	
Gestión interpersonal positiva	11	*A través de la percepción de los <b>estudiantes de primaria y secundaria</b>  *5 opciones de respuesta 0=Nunca 1=Casi nunca 2=Algunas veces 3=Casi siempre 4=Siempre
Victimización	6	
Disruptividad	6	
Red social de iguales	9	
Agresión	4	
Ajuste normativo	5	
Indisciplina	4	
Desidia docente	5	
<b>Escala de Convivencia Escolar</b>	<b>50</b>	



## Antecedentes

### Adaptación de test

Una adaptación válida involucra más procesos que la simple traducción de un idioma a otro. Se deben considerar los **aspectos psicológicos, psicométricos, culturales y lingüísticos** involucrados; así como todas las idiosincrasias específicas de un idioma o cultura particular que pueda causar sesgo en la evaluación (Tanzer y Sim, 1999).

El fin último de la adaptación es lograr la equivalencia de constructo y de contenido, así como la equivalencia lingüística entre las escalas de medición original y objetivo (APA, 2013).

## Antecedentes

### Etapas del proyecto

Etapas	Descripción
I. Traducción	Transformación de términos o secciones del ítem, en términos adecuados y entendidos en el contexto de Baja California.
II. Validación lingüística	Procedimiento para asegurar la calidad de la traducción de los ítems de la Escala de convivencia Escolar. A través de juicios de expertos se evalúa la equivalencia gramatical y semántica de la traducción de la traducción de cada ítem.
III. Validación de contenido	Procedimiento empleado para asegurar que la calidad de la escala no sea afectada negativamente por la traducción. A través de juicios de expertos se evalúa la claridad, pertinencia y suficiencia de cada ítem.
IV. Prueba piloto del instrumento	Consiste en una aplicación a una muestra pequeña de estudiantes, con el objetivo de conocer el funcionamiento inicial de la escala y los ítems, y realizar las modificaciones necesarias.
V. Aplicación	Aplicación de la escala a una muestra de estudiantes más amplia y comparable con la muestra utilizada en España para validar la Escala de Convivencia Escolar.
VI. Análisis de sesgo y equivalencia	Empleo de un conjunto de técnicas estadísticas que permitan conocer el grado en el cual los ítems y la escala traducida son equivalentes con la escala original, en términos de medición del constructo y de las puntuaciones obtenidas.

26/12/2014

Descripción de la tarea

## VALIDACIÓN DE CONTENIDO

## Validación de contenido

### Criterios de evaluación

Criterio	Descripción	Escala
Congruencia	Grado en que el ítem es consecuente con la dimensión correspondientes	(1) No congruente (2) Poco congruente (3) Congruente (4) Muy congruente
Relevancia	Grado en que la información contenida en el ítem es importante para evaluar la dimensión correspondiente.	(1) No relevante (2) Poco relevante (3) Relevante (4) Muy relevante
Claridad	Grado en que el ítem comunica de manera clara y directa la información, es decir que no admite más de una interpretación posible	(1) No claro (2) Poco claro (3) Claro (4) Muy claro
Suficiencia	Grado en que se incluyen los ítems necesarios para evaluar cada dimensión.	(1) No suficiente (2) Poco suficiente (3) Suficiente (4) Muy suficiente

## Validación de contenido

Dimensión 1: Gestión interpersonal positiva

**Descripción:** Mide la forma en la que los profesores gestionan las relaciones interpersonales en el centro educativo, aludiendo a la conducta social ejemplar, la buena comunicación con el alumnado y sus familias y a la atención de las dificultades que estos pueden encontrar

Ítem	Congruencia	Relevancia	Claridad	Observaciones
01. Los profesores(as) se llevan bien entre ellos.	1 2 3 4	1 2 3 4	1 2 3 4	
02. Hay buenas relaciones entre los profesores(as) y los alumnos(as).	1 2 3 4	1 2 3 4	1 2 3 4	
03. Los padres de familia se llevan bien con los profesores(as).	1 2 3 4	1 2 3 4	1 2 3 4	
04. Las familias de los alumnos(as) se comprometen con las actividades escolares.	1 2 3 4	1 2 3 4	1 2 3 4	
05. Los profesores(as) son respetados.	1 2 3 4	1 2 3 4	1 2 3 4	
06. Los profesores(as) dan ejemplo de buenas relaciones.	1 2 3 4	1 2 3 4	1 2 3 4	
07. Los profesores(as) evitan que nos hagamos unos de otros.	1 2 3 4	1 2 3 4	1 2 3 4	
08. Los profesores(as) nos ayudan a resolver nuestros problemas.	1 2 3 4	1 2 3 4	1 2 3 4	
09. Mis profesores(as) me ayudan cuando se lo pido.	1 2 3 4	1 2 3 4	1 2 3 4	
10. Mis padres se llevan bien con mis profesores(as).	1 2 3 4	1 2 3 4	1 2 3 4	
11. Los profesores(as) valoran su trabajo y se animan a mejorar.	1 2 3 4	1 2 3 4	1 2 3 4	

(1) No suficiente  
 (2) Poco suficiente  
 (3) Suficiente  
 (4) Muy suficiente

## Validación de contenido

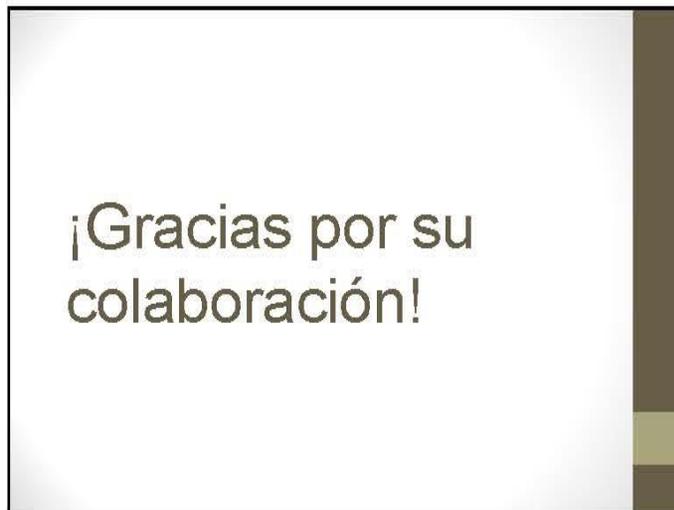
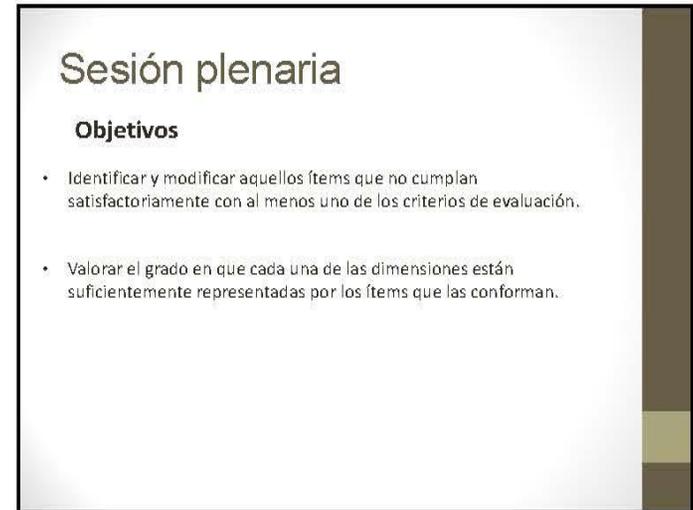
### Escala de Convivencia Escolar

Dimensión	Número de ítems
Gestión interpersonal positiva	11
Victimización	6
Disruptividad	6
Red social de iguales	9
Agresión	4
Ajuste normativo	5
Indisciplina	4
Desidia docente	5
<b>Escala de Convivencia Escolar</b>	<b>50</b>

\*A través de la percepción de los estudiantes de primaria y secundaria

\*5 opciones de respuesta  
 0= Nunca  
 1= Casi nunca  
 2= Algunas veces  
 3= Casi siempre  
 4= Siempre

26/12/2014



Sección de la hoja de evaluación de contenido

	A	B	C	D	E	F	G	H	I	J	
1					JUEZ 1						
2	Dimensión	Descriptor	ítem	Item_E	Congruencia	Relevancia	Claridad	Suficiencia	Observaciones	Congruencia R	
3	1. Gestión interpersonal positiva	Mide la forma en la que los profesores gestionan las relaciones interpersonales en el centro educativo, aludiendo a la conducta social ejemplar, la buena comunicación con el alumnado y sus familias y a la atención de las dificultades que éstos pueden encontrar	i1	Los profesores(as) se llevan bien entre ellos							
4			i2	Hay buenas relaciones entre los profesores(as) y los alumnos(as)							
5			i3	Los padres de familia se llevan bien con los profesores(as)							
6			i4	Las familias de los alumnos(as) se comprometen con las actividades escolares							
7			i5	Los profesores(as) son respetados							
8			i6	Los profesores(as) dan ejemplo de buenas relaciones							
9			i7	Los profesores(as) evitan que nos burlemos unos de otros							
10			i8	Los profesores(as) nos ayudan a resolver nuestros problemas							
11			i9	Mis profesores(as) me ayudan cuando se lo pido							
12			i10	Mis padres se llevan bien con mis profesores(as)							
13					i11	Los profesores(as) valoran mi trabajo y me animan a mejorar					
14	2. Victimización	Aborda la percepción de ser maltratado, atacado, golpeado	i1	He tenido miedo de venir a la escuela							
15			i2	Algún compañero(a) me ha golpeado							
16			i3	Algún compañero(a) me ha insultado							
17			i4	Me he sentido amenazado(a)							
18			i5	Me han robado							
19			i6	Me he sentido excluido, aislado o rechazado por compañeros(as)							
20	3. Disruptividad	Se analizan los aspectos relativos al vandalismo, los destrozos o impedir el normal desarrollo de las clases	i1	Dentro de la escuela hay problemas de vandalismo							
21			i2	Hay peleas en las que la gente se golpea							
22			i3	Hay alumnos(as) que no dejan dar clase							
23			i4	Algunos alumnos(as) destruyen el material y las instalaciones							
24			i5	Hay alumnos(as) que siempre se meten en peleas							
25			i6	Dejo trabajar a los demás sin molestarlos							
26			i1	Los alumnos(as) nos llevamos bien entre nosotros							
27			i2	Mis compañeros(as) me ayudan cuando lo necesito							

**Anexo E. Prueba piloto**

Hoja de respuesta

### HOJA DE RESPUESTA ESCALA DE CONVIVENCIA ESCOLAR

**INSTRUCCIONES PARA EL ALUMNO**

1. Marca tu respuesta llenando completamente el círculo que corresponda.
2. Usa lápiz 2 ó 2 ½. No uses pluma o marcador.
3. No maltrates, dobles o engrapes la hoja.
4. En caso de error, borra completa y limpiamente.
5. Contesta la totalidad de las preguntas.



Ejemplo: A B ● D

**FOLIO**

**Datos generales**

**Edad**

**Sexo**

Femenino  Masculino

**Datos de la escuela**

**Nivel primaria**

NOMBRE DE LA ESCUELA:

**Grado**

Quinto  Sexto

**Grupo**

(A) (B) (C)

**Turno al que asistes**

Matutino  Vespertino

**Nivel secundaria**

NOMBRE DE LA ESCUELA:

**Grado**

Primero  Segundo  Tercero

**Grupo**

(A) (B) (C) (D) (E)

**Turno al que asistes**

Matutino  Vespertino

**Tipo de secundaria al que asistes**

General  Privada  Técnica  Telesecundaria

**Escala de Convivencia Escolar**

A) Nunca

B) Casi nunca

C) Algunas veces

D) Casi siempre

E) Siempre

1 (A) (B) (C) (D) (E)

2 (A) (B) (C) (D) (E)

3 (A) (B) (C) (D) (E)

4 (A) (B) (C) (D) (E)

5 (A) (B) (C) (D) (E)

6 (A) (B) (C) (D) (E)

7 (A) (B) (C) (D) (E)

8 (A) (B) (C) (D) (E)

9 (A) (B) (C) (D) (E)

10 (A) (B) (C) (D) (E)

11 (A) (B) (C) (D) (E)

12 (A) (B) (C) (D) (E)

13 (A) (B) (C) (D) (E)

14 (A) (B) (C) (D) (E)

15 (A) (B) (C) (D) (E)

16 (A) (B) (C) (D) (E)

17 (A) (B) (C) (D) (E)

18 (A) (B) (C) (D) (E)

19 (A) (B) (C) (D) (E)

20 (A) (B) (C) (D) (E)

21 (A) (B) (C) (D) (E)

22 (A) (B) (C) (D) (E)

23 (A) (B) (C) (D) (E)

24 (A) (B) (C) (D) (E)

25 (A) (B) (C) (D) (E)

26 (A) (B) (C) (D) (E)

27 (A) (B) (C) (D) (E)

28 (A) (B) (C) (D) (E)

29 (A) (B) (C) (D) (E)

30 (A) (B) (C) (D) (E)

31 (A) (B) (C) (D) (E)

32 (A) (B) (C) (D) (E)

33 (A) (B) (C) (D) (E)

34 (A) (B) (C) (D) (E)

35 (A) (B) (C) (D) (E)

36 (A) (B) (C) (D) (E)

37 (A) (B) (C) (D) (E)

38 (A) (B) (C) (D) (E)

39 (A) (B) (C) (D) (E)

40 (A) (B) (C) (D) (E)

41 (A) (B) (C) (D) (E)

42 (A) (B) (C) (D) (E)

43 (A) (B) (C) (D) (E)

44 (A) (B) (C) (D) (E)

45 (A) (B) (C) (D) (E)

46 (A) (B) (C) (D) (E)

47 (A) (B) (C) (D) (E)

48 (A) (B) (C) (D) (E)

49 (A) (B) (C) (D) (E)

50 (A) (B) (C) (D) (E)

**Anexo F. Aplicación**

Cuadernillo de aplicación

---

# **Escala de Convivencia Escolar**

---

## Escala de Convivencia Escolar

### Instrucciones:

A continuación se presentan una serie de preguntas acerca de situaciones que ocurren dentro de tu escuela. En este cuestionario **no existen respuestas buenas ni malas**, sólo nos interesa que contestes de manera sincera. **Te pedimos que no anotes tu nombre**, para que nadie pueda conocer tus respuestas.

Por favor, **no hagas anotaciones** en este cuadernillo, pues será utilizado después por otros compañeros.

Para contestar el cuestionario, deberás utilizar sólo la *Hoja de respuestas*, donde señalarás la frecuencia en la que sucede en tu escuela rellenando completamente el círculo que corresponda, siguiendo esta clave:

A = Nunca

B = Casi nunca

C = Algunas veces

D = Casi siempre

E = Siempre

Por ejemplo:

1. En mi escuela se hacen simulacros de sismo.

1	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
---	-----------------------	-----------------------	----------------------------------	-----------------------	-----------------------

Por favor, no dejes ninguna pregunta sin responder.

## Escala de Convivencia Escolar

---

1. Mis profesores(as) se llevan bien entre ellos.
2. He tenido miedo de venir a la escuela.
3. Dentro de la escuela hay problemas de vandalismo.
4. Hay buenas relaciones entre los profesores(as) y los alumnos(as).
5. Los alumnos(as) nos llevamos bien entre nosotros.
6. Algún compañero(a) me ha golpeado.
7. Los padres de familia se llevan bien con los profesores(as).
8. Hay peleas en las que la gente se golpea.
9. Aprendo en mi escuela.
10. Mis padres acuden a la escuela cuando son llamados por mis profesores.
11. Algún compañero(a) me ha insultado.
12. Mis compañeros(as) me ayudan cuando lo necesito.
13. Hay alumnos(as) que no dejan dar clase.
14. Los profesores(as) son respetados.
15. Me siento amenazado(a) en la escuela.

---

-Utiliza la hoja de respuesta, no hagas anotaciones en este cuadernillo.-

### **Escala de Convivencia Escolar**

---

16. Hay alumnos(as) que no respetan las reglas.
17. Me uno a las actividades que realizan los demás.
18. Los profesores(as) dan ejemplo de buenas relaciones.
19. Me han robado en la escuela.
20. Hay alumnos(as) que siempre se meten en peleas.
21. Los profesores(as) evitan que nos burlemos unos de otros.
22. Me he sentido excluido, aislado o rechazado por compañeros(as).
23. Los profesores(as) sólo explican para los más listos(as) de la clase.
24. Algunos alumnos(as) destruyen el material y las instalaciones.
25. Los profesores(as) nos ayudan a resolver nuestros problemas.
26. Mis compañeros(as) se interesan por mí.
27. Mis profesores(as) me ayudan cuando lo necesito.
28. Digo y defiendo mis opiniones sin dañar o molestar a los demás.
29. Dejo trabajar a los demás sin molestarlos.
30. He golpeado a algún compañero(a).

---

-Utiliza la hoja de respuesta, no hagas anotaciones en este cuadernillo.-

## Escala de Convivencia Escolar

---

31. Sólo cumplo las reglas que me convienen.
32. Las reglas de los profesores(as) son injustas.
33. Le caigo bien a mis compañeros(as).
34. Pido la palabra y espero mi turno para hablar.
35. He excluido o rechazado a algún compañero(a).
36. Mis padres se llevan bien con mis profesores(as).
37. Me han castigado en la escuela.
38. Me gusta trabajar en equipo.
39. Cumplo las reglas de mi escuela y de mi grupo.
40. Hay profesores(as) que castigan siempre a los mismos alumnos(as).
41. Mis profesores(as) valoran mi trabajo y me animan a mejorar.
42. Ayudo a mis compañeros(as) cuando me lo piden.
43. Interrumpo la clase porque me aburro.
44. He amenazado a algún compañero.
45. Respeto la opinión de los demás aunque no esté de acuerdo.

---

-Utiliza la hoja de respuesta, no hagas anotaciones en este cuadernillo.-

## Escala de Convivencia Escolar

---

46. Siento que tengo amigos(as).
47. Me aburro en la escuela.
48. Los profesores(as) ponen actividades aburridas.
49. He insultado a algún compañero(a).
50. Hay alumnos(as) que les caen mal a los profesores(as).

**¡GRACIAS POR TU PARTICIPACIÓN!**

---

-Utiliza la hoja de respuesta, no hagas anotaciones en este cuadernillo.-

Reporte de incidencias



**Reporte de incidencias**  
**Escala de Convivencia Escolar**



Nombre de la escuela:	
Clave:	Grupo:
Nombre del aplicador:	
Teléfono:	Correo electrónico:

Número de cuadernillos recibidos:	Número de exámenes cancelados:
Número de cuadernillos devueltos:	Número de aplicaciones efectivas:
Fecha de aplicación:	Horario(s) de aplicación:
Hora de inicio de la aplicación:	Hora en que concluyó el primero de los examinados:
Hora en que concluyó el último de los examinados:	

Reporte de incidencias:
(1)
(2)
(3)
(4)
(5)
(6)
(7)

(8)	
(9)	
(10)	
(11)	
(12)	
(13)	
(14)	
(15)	
(16)	
(17)	
(18)	
(19)	
(20)	
(21)	
Firma del aplicador:	





**Anexo G. Frecuencias y porcentajes de las respuestas a los ítems**

1. Mis profesores(as) se llevan bien entre ellos.		
	Frecuencia	Porcentaje
Nunca	19	2.1
Casi nunca	21	2.3
Algunas veces	130	14.3
Casi siempre	243	26.8
Siempre	493	54.4
Total	906	100.0

2. He tenido miedo de venir a la escuela.		
	Frecuencia	Porcentaje
Nunca	688	75.9
Casi nunca	88	9.7
Algunas veces	105	11.6
Casi siempre	4	.4
Siempre	21	2.3
Total	906	100.0

3. Dentro de la escuela hay problemas de vandalismo.		
	Frecuencia	Porcentaje
Nunca	247	27.3
Casi nunca	179	19.8
Algunas veces	284	31.3
Casi siempre	119	13.1
Siempre	77	8.5
Total	906	100.0

4. Hay buenas relaciones entre los profesores(as) y los alumnos(as).		
	Frecuencia	Porcentaje
Nunca	40	4.4
Casi nunca	44	4.9
Algunas veces	236	26.1
Casi siempre	309	34.1
Siempre	276	30.5
Total	905	100.0

5. Los alumnos(as) nos llevamos bien entre nosotros.		
	Frecuencia	Porcentaje
Nunca	23	2.5
Casi nunca	60	6.6
Algunas veces	243	26.8
Casi siempre	335	37.0
Siempre	245	27.0
Total	906	100.0

6. Algún compañero(a) me ha golpeado.		
	Frecuencia	Porcentaje
Nunca	605	66.8
Casi nunca	127	14.0
Algunas veces	114	12.6
Casi siempre	30	3.3
Siempre	30	3.3
Total	906	100.0

7. Los padres de familia se llevan bien con los profesores(as).		
	Frecuencia	Porcentaje
Nunca	24	2.6
Casi nunca	48	5.3
Algunas veces	193	21.3
Casi siempre	229	25.3
Siempre	412	45.5
Total	906	100.0

8. Hay peleas en las que la gente se golpea.		
	Frecuencia	Porcentaje
Nunca	244	27.0
Casi nunca	174	19.3
Algunas veces	272	30.1
Casi siempre	117	13.0
Siempre	96	10.6
Total	903	100.0

9. Aprendo en mi escuela.		
	Frecuencia	Porcentaje
Nunca	9	1.0
Casi nunca	15	1.7
Algunas veces	76	8.4
Casi siempre	235	26.0
Siempre	569	62.9
Total	904	100.0

10. Mis padres acuden a la escuela cuando son llamados por mis profesores.		
	Frecuencia	Porcentaje
Nunca	52	5.7
Casi nunca	19	2.1
Algunas veces	67	7.4
Casi siempre	130	14.4
Siempre	637	70.4
Total	905	100.0

11. Algún compañero(a) me ha insultado.		
	Frecuencia	Porcentaje
Nunca	284	31.3
Casi nunca	208	22.9
Algunas veces	260	28.7
Casi siempre	89	9.8
Siempre	66	7.3
Total	907	100.0

12. Mis compañeros(as) me ayudan cuando lo necesito.		
	Frecuencia	Porcentaje
Nunca	37	4.1
Casi nunca	72	7.9
Algunas veces	250	27.6
Casi siempre	223	24.6
Siempre	325	35.8
Total	907	100.0

13. Hay alumnos(as) que no dejan dar clase.		
	Frecuencia	Porcentaje
Nunca	94	10.4
Casi nunca	95	10.6
Algunas veces	247	27.4
Casi siempre	207	23.0
Siempre	257	28.6
Total	900	100.0

14. Los profesores(as) son respetados.		
	Frecuencia	Porcentaje
Nunca	23	2.5
Casi nunca	34	3.7
Algunas veces	203	22.4
Casi siempre	269	29.7
Siempre	378	41.7
Total	907	100.0

15. Me siento amenazado(a) en la escuela.		
	Frecuencia	Porcentaje
Nunca	759	83.8
Casi nunca	69	7.6
Algunas veces	41	4.5
Casi siempre	15	1.7
Siempre	22	2.4
Total	906	100.0

16. Hay alumnos(as) que no respetan las reglas.		
	Frecuencia	Porcentaje
Nunca	55	6.1
Casi nunca	125	13.8
Algunas veces	355	39.2
Casi siempre	205	22.6
Siempre	166	18.3
Total	906	100.0

17. Me uno a las actividades que realizan los demás.		
	Frecuencia	Porcentaje
Nunca	22	2.4
Casi nunca	57	6.3
Algunas veces	250	27.7
Casi siempre	260	28.8
Siempre	314	34.8
Total	903	100.0

18. Los profesores(as) dan ejemplo de buenas relaciones.		
	Frecuencia	Porcentaje
Nunca	38	4.2
Casi nunca	67	7.4
Algunas veces	160	17.7
Casi siempre	225	24.8
Siempre	416	45.9
Total	906	100.0

19. Me han robado en la escuela.		
	Frecuencia	Porcentaje
Nunca	451	50.0
Casi nunca	161	17.8
Algunas veces	187	20.7
Casi siempre	54	6.0
Siempre	49	5.4
Total	902	100.0

20. Hay alumnos(as) que siempre se meten en peleas.		
	Frecuencia	Porcentaje
Nunca	155	17.1
Casi nunca	205	22.6
Algunas veces	277	30.6
Casi siempre	139	15.3
Siempre	130	14.3
Total	906	100.0

21. Los profesores(as) evitan que nos burlemos unos de otros.

	Frecuencia	Porcentaje
Nunca	86	9.5
Casi nunca	63	7.0
Algunas veces	123	13.6
Casi siempre	147	16.2
Siempre	486	53.7
Total	905	100.0

22. Me he sentido excluido, aislado o rechazado por compañeros(as).

	Frecuencia	Porcentaje
Nunca	487	53.8
Casi nunca	158	17.4
Algunas veces	166	18.3
Casi siempre	49	5.4
Siempre	46	5.1
Total	906	100.0

23. Los profesores(as) sólo explican para los más listos(as) de la clase.

	Frecuencia	Porcentaje
Nunca	566	62.5
Casi nunca	107	11.8
Algunas veces	119	13.1
Casi siempre	57	6.3
Siempre	56	6.2
Total	905	100.0

24. Algunos alumnos(as) destruyen el material y las instalaciones.

	Frecuencia	Porcentaje
Nunca	287	31.7
Casi nunca	208	23.0
Algunas veces	247	27.3
Casi siempre	101	11.2
Siempre	62	6.9
Total	905	100.0

---

25. Los profesores(as) nos ayudan a resolver nuestros problemas.

---

	Frecuencia	Porcentaje
Nunca	59	6.5
Casi nunca	99	11.0
Algunas veces	225	24.9
Casi siempre	191	21.2
Siempre	329	36.4
Total	903	100.0

---



---

26. Mis compañeros(as) se interesan por mí.

---

	Frecuencia	Porcentaje
Nunca	78	8.6
Casi nunca	111	12.3
Algunas veces	299	33.1
Casi siempre	220	24.3
Siempre	196	21.7
Total	904	100.0

---



---

27. Mis profesores(as) me ayudan cuando lo necesito.

---

	Frecuencia	Porcentaje
Nunca	38	4.2
Casi nunca	68	7.5
Algunas veces	191	21.1
Casi siempre	190	21.0
Siempre	419	46.2
Total	906	100.0

---



---

28. Digo y definiendo mis opiniones sin dañar o molestar a los demás.

---

	Frecuencia	Porcentaje
Nunca	59	6.5
Casi nunca	66	7.3
Algunas veces	223	24.6
Casi siempre	228	25.2
Siempre	330	36.4
Total	906	100.0

---

29. Dejo trabajar a los demás sin molestarlos.		
	Frecuencia	Porcentaje
Nunca	31	3.4
Casi nunca	31	3.4
Algunas veces	128	14.2
Casi siempre	212	23.5
Siempre	501	55.5
Total	903	100.0

30. He golpeado a algún compañero(a).		
	Frecuencia	Porcentaje
Nunca	515	57.0
Casi nunca	208	23.0
Algunas veces	138	15.3
Casi siempre	23	2.5
Siempre	20	2.2
Total	904	100.0

31. Sólo cumplo las reglas que me convienen.		
	Frecuencia	Porcentaje
Nunca	447	49.7
Casi nunca	154	17.1
Algunas veces	147	16.3
Casi siempre	63	7.0
Siempre	89	9.9
Total	900	100.0

32. Las reglas de los profesores(as) son injustas.		
	Frecuencia	Porcentaje
Nunca	401	44.5
Casi nunca	153	17.0
Algunas veces	224	24.9
Casi siempre	59	6.5
Siempre	64	7.1
Total	901	100.0

33. Le caigo bien a mis compañeros(as).		
	Frecuencia	Porcentaje
Nunca	29	3.2
Casi nunca	39	4.3
Algunas veces	162	17.9
Casi siempre	326	36.1
Siempre	347	38.4
Total	903	100.0

34. Pido la palabra y espero mi turno para hablar.		
	Frecuencia	Porcentaje
Nunca	40	4.4
Casi nunca	51	5.6
Algunas veces	172	19.0
Casi siempre	214	23.7
Siempre	427	47.2
Total	904	100.0

35. He excluido o rechazado a algún compañero(a).		
	Frecuencia	Porcentaje
Nunca	443	49.1
Casi nunca	224	24.8
Algunas veces	152	16.8
Casi siempre	46	5.1
Siempre	38	4.2
Total	903	100.0

36. Mis padres se llevan bien con mis profesores(as).		
	Frecuencia	Porcentaje
Nunca	35	3.9
Casi nunca	57	6.3
Algunas veces	133	14.7
Casi siempre	146	16.2
Siempre	533	59.0
Total	904	100.0

37. Me han castigado en la escuela.		
	Frecuencia	Porcentaje
Nunca	424	47.1
Casi nunca	205	22.8
Algunas veces	182	20.2
Casi siempre	58	6.4
Siempre	32	3.6
Total	901	100.0

38. Me gusta trabajar en equipo.		
	Frecuencia	Porcentaje
Nunca	37	4.1
Casi nunca	58	6.4
Algunas veces	152	16.9
Casi siempre	156	17.3
Siempre	499	55.3
Total	902	100.0

39. Cumpló las reglas de mi escuela y de mi grupo.		
	Frecuencia	Porcentaje
Nunca	24	2.7
Casi nunca	49	5.4
Algunas veces	175	19.4
Casi siempre	253	28.0
Siempre	401	44.5
Total	902	100.0

40. Hay profesores(as) que castigan siempre a los mismos alumnos(as).		
	Frecuencia	Porcentaje
Nunca	135	14.9
Casi nunca	114	12.6
Algunas veces	240	26.5
Casi siempre	197	21.8
Siempre	219	24.2
Total	905	100.0

41. Mis profesores(as) valoran mi trabajo y me animan a mejorar.		
	Frecuencia	Porcentaje
Nunca	47	5.2
Casi nunca	70	7.7
Algunas veces	180	19.9
Casi siempre	196	21.7
Siempre	412	45.5
Total	905	100.0

42. Ayudo a mis compañeros(as) cuando me lo piden.		
	Frecuencia	Porcentaje
Nunca	31	3.4
Casi nunca	42	4.6
Algunas veces	194	21.4
Casi siempre	275	30.4
Siempre	363	40.1
Total	905	100.0

43. Interrumpo la clase porque me aburro.		
	Frecuencia	Porcentaje
Nunca	625	69.1
Casi nunca	140	15.5
Algunas veces	92	10.2
Casi siempre	27	3.0
Siempre	21	2.3
Total	905	100.0

44. He amenazado a algún compañero.		
	Frecuencia	Porcentaje
Nunca	702	77.6
Casi nunca	88	9.7
Algunas veces	61	6.7
Casi siempre	23	2.5
Siempre	31	3.4
Total	905	100.0

45. Respeto la opinión de los demás aunque no esté de acuerdo.		
	Frecuencia	Porcentaje
Nunca	59	6.5
Casi nunca	61	6.8
Algunas veces	203	22.5
Casi siempre	194	21.5
Siempre	384	42.6
Total	901	100.0

46. Siento que tengo amigos(as).		
	Frecuencia	Porcentaje
Nunca	41	4.6
Casi nunca	39	4.3
Algunas veces	92	10.2
Casi siempre	127	14.1
Siempre	602	66.8
Total	901	100.0

47. Me aburro en la escuela.		
	Frecuencia	Porcentaje
Nunca	257	28.6
Casi nunca	205	22.8
Algunas veces	280	31.1
Casi siempre	88	9.8
Siempre	70	7.8
Total	900	100.0

48. Los profesores(as) ponen actividades aburridas.		
	Frecuencia	Porcentaje
Nunca	250	27.7
Casi nunca	215	23.8
Algunas veces	285	31.6
Casi siempre	90	10.0
Siempre	63	7.0
Total	903	100.0

---

---

49. He insultado a algún compañero(a).		
	Frecuencia	Porcentaje
Nunca	442	49.0
Casi nunca	192	21.3
Algunas veces	192	21.3
Casi siempre	43	4.8
Siempre	33	3.7
Total	902	100.0

---

---

---

---

50. Hay alumnos(as) que les caen mal a los profesores(as).		
	Frecuencia	Porcentaje
Nunca	247	27.4
Casi nunca	145	16.1
Algunas veces	226	25.1
Casi siempre	128	14.2
Siempre	156	17.3
Total	902	100.0

---

---

**Anexo H. Sintaxis de los análisis de equivalencia y sesgo**

```
#Instalar paquetes Rcmdr, psych, lavaan, semPlot, mirt y lordif
#Importar base de datos en R
library(Rcmdr)

#importar base desde Rcommander con nombre "b.ece"

##ANÁLISIS INICIALES
#Cálculo del alfa ordinal y la Ptbis corregida de los ítems por escala
y por dimensión

#Cargar paquete psych
library(psych)

#Generar conjuntos de datos de la escala y de cada dimensión
attach(b.ece)
ece <- b.ece [paste("v", 1:50, sep= "")]
f1 <- b.ece [paste("v", c(1,4,7,10,14,18,21,25,27,36,41), sep= "")]
f2 <- b.ece [paste("v", c(2,6,11,15,19,22), sep= "")]
f3 <- b.ece [paste("v", c(3,8,13,16,20,24), sep= "")]
f4 <- b.ece [paste("v", c(5,12,17,26,28,33,38,42,46), sep= "")]
f5 <- b.ece [paste("v", c(30,35,44,49), sep= "")]
f6 <- b.ece [paste("v", c(9,29,34,39,45), sep= "")]
f7 <- b.ece [paste("v", c(31,37,43,47), sep= "")]
f8 <- b.ece [paste("v", c(23,32,40,48,50), sep= "")]

#Calcular matrices policóricas de cada conjunto de datos
m.ece <- polychoric(ece)
f1.m <- polychoric(f1)
f2.m <- polychoric(f2)
f3.m <- polychoric(f3)
f4.m <- polychoric(f4)
f5.m <- polychoric(f5)
f6.m <- polychoric(f6)
f7.m <- polychoric(f7)
f8.m <- polychoric(f8)

#calcular alfas y ptbis sobre las matrices correspondientes
alpha(m.ece$rho)
alpha(f1.m$rho)
alpha(f2.m$rho)
alpha(f3.m$rho)
alpha(f4.m$rho)
alpha(f5.m$rho)
alpha(f6.m$rho)
alpha(f7.m$rho)
alpha(f8.m$rho)
```

```
##ANÁLISIS 1: EQUIVALENCIA DE CONSTRUCTO
library(lavaan)
library(semPlot)

#Especificación del modelo
m1 <- 'F1 =~ v1 + v4 + v7 + v10 + v14 + v18 + v21 + v25 + v27 + v36 +
v41
      F2 =~ v2 + v6 + v11 + v15 + v19 + v22
      F3 =~ v3 + v8 + v13 + v16 + v20 + v24
      F4 =~ v5 + v12 + v17 + v26 + v28 + v33 + v38 + v42 + v46
      F5 =~ v30 + v35 + v44 + v49 + v9\nF6 =~ v29 + v34 + v39 + v45
      F7 =~ v31 + v37 + v43 + v47
      F8 =~ v23 + v32 + v40 + v48 + v50'

#Estimación del modelo
cfa.1 <- cfa(m1, sample.cov=m.ece$rho, sample.nobs=906,
estimator="ULS", std.lv=TRUE)

#Estudio del ajuste
summary(cfa.1, fit.measures=TRUE, standardized=TRUE)

#Graficar modelo
library(semPlot)
semPaths(cfa.1, layout="circle", style="LISREL", sizeMan=4,
sizeMan2=3, sizeLat=4, sizeLat2=3)
semPaths(cfa.1, what="stand", layout="circle", style="LISREL",
sizeMan=4, sizeMan2=3, sizeLat=4, sizeLat2=3)
```

```
##ANÁLISIS 2: DIMENSIONALIDAD
```

```
#MODELO UNIDIMENSIONALID
```

```
m.a <- 'FU =~ v1 + v4 + v7 + v10 + v14 + v18 + v21 + v25 + v27 + v36 +  
v41 + v2 + v6 + v11 + v15 + v19 + v22 + v3 + v8 + v13 + v16 + v20 +  
v24 + v5 + v12 + v17 + v26 + v28 + v33 + v38 + v42 + v46 + v30 + v35 +  
v44 + v49 + v9 + v29 + v34 + v39 + v45 + v31 + v37 + v43 + v47 + v23 +  
v32 + v40 + v48 + v50'
```

```
#MODELO FACTORIAL DE SEGUNDO ORDEN
```

```
m.b <- 'F1 =~ v1 + v4 + v7 + v10 + v14 + v18 + v21 + v25 + v27 + v36 +  
v41
```

```
      F2 =~ v2 + v6 + v11 + v15 + v19 + v22
```

```
      F3 =~ v3 + v8 + v13 + v16 + v20 + v24
```

```
      F4 =~ v5 + v12 + v17 + v26 + v28 + v33 + v38 + v42 + v46
```

```
      F5 =~ v30 + v35 + v44 + v49 + v9
```

```
      F6 =~ v29 + v34 + v39 + v45
```

```
      F7 =~ v31 + v37 + v43 + v47
```

```
      F8 =~ v23 + v32 + v40 + v48 + v50
```

```
      FS =~ F1 + F2 + F3 + F4 + F5 + F6 + F7 + F8
```

```
      F1 ~~ 0*F2
```

```
      F1 ~~ 0*F3
```

```
      F1 ~~ 0*F4
```

```
      F1 ~~ 0*F5
```

```
      F1 ~~ 0*F6
```

```
      F1 ~~ 0*F7
```

```
      F1 ~~ 0*F8
```

```
      F2 ~~ 0*F3
```

```
      F2 ~~ 0*F4
```

```
      F2 ~~ 0*F5
```

```
      F2 ~~ 0*F6
```

```
      F2 ~~ 0*F7
```

```
      F2 ~~ 0*F8
```

```
      F3 ~~ 0*F4
```

```
      F3 ~~ 0*F5
```

```
      F3 ~~ 0*F6
```

```
      F3 ~~ 0*F7
```

```
      F3 ~~ 0*F8
```

```
      F4 ~~ 0*F5
```

```
      F4 ~~ 0*F6
```

```
      F4 ~~ 0*F7
```

```
      F4 ~~ 0*F8
```

```
      F5 ~~ 0*F6
```

```
      F5 ~~ 0*F7
```

```
      F5 ~~ 0*F8
```

```
      F6 ~~ 0*F7
```

```
      F6 ~~ 0*F8
```

```
      F7 ~~ 0*F8'
```

```
#MODELO BIFACTORIAL
```

```
m.c <- 'F1 =~ v1 + v4 + v7 + v10 + v14 + v18 + v21 + v25 + v27 + v36 +
v41
```

```
      F2 =~ v2 + v6 + v11 + v15 + v19 + v22
```

```
      F3 =~ v3 + v8 + v13 + v16 + v20 + v24
```

```
      F4 =~ v5 + v12 + v17 + v26 + v28 + v33 + v38 + v42 + v46
```

```
      F5 =~ v30 + v35 + v44 + v49 + v9
```

```
      F6 =~ v29 + v34 + v39 + v45
```

```
      F7 =~ v31 + v37 + v43 + v47
```

```
      F8 =~ v23 + v32 + v40 + v48 + v50
```

```
      FG =~ v1 + v4 + v7 + v10 + v14 + v18 + v21 + v25 + v27 + v36 +
v41 + v2 + v6 + v11 + v15 + v19 + v22 + v3 + v8 + v13 + v16 + v20 +
v24 + v5 + v12 + v17 + v26 + v28 + v33 + v38 + v42 + v46 + v30 + v35 +
v44 + v49 + v9 + v29 + v34 + v39 + v45 + v31 + v37 + v43 + v47 + v23 +
v32 + v40 + v48 + v50
```

```
      FG ~~ 0*F1
```

```
      FG ~~ 0*F2
```

```
      FG ~~ 0*F3
```

```
      FG ~~ 0*F4
```

```
      FG ~~ 0*F5
```

```
      FG ~~ 0*F6
```

```
      FG ~~ 0*F7
```

```
      FG ~~ 0*F8
```

```
      F1 ~~ 0*F2
```

```
      F1 ~~ 0*F3
```

```
      F1 ~~ 0*F4
```

```
      F1 ~~ 0*F5
```

```
      F1 ~~ 0*F6
```

```
      F1 ~~ 0*F7
```

```
      F1 ~~ 0*F8
```

```
      F2 ~~ 0*F3
```

```
      F2 ~~ 0*F4
```

```
      F2 ~~ 0*F5
```

```
      F2 ~~ 0*F6
```

```
      F2 ~~ 0*F7
```

```
      F2 ~~ 0*F8
```

```
      F3 ~~ 0*F4
```

```
      F3 ~~ 0*F5
```

```
      F3 ~~ 0*F6
```

```
      F3 ~~ 0*F7
```

```
      F3 ~~ 0*F8
```

```
      F4 ~~ 0*F5
```

```
      F4 ~~ 0*F6
```

```
      F4 ~~ 0*F7
```

```
      F4 ~~ 0*F8
```

```
      F5 ~~ 0*F6
```

```
      F5 ~~ 0*F7
```

```
      F5 ~~ 0*F8
```

```
      F6 ~~ 0*F7
```

```
      F6 ~~ 0*F8
```

```
      F7 ~~ 0*F8'
```

```
#Estimación de modelo 1, 2 Y 3
cfa.a <- cfa(m.a, sample.cov=m.ece$rho, sample.nobs=906,
estimator="ULS", std.lv=TRUE)
cfa.b <- cfa(m.b, sample.cov=m.ece$rho, sample.nobs=906,
estimator="ULS", std.lv=TRUE)
cfa.c <- cfa(m.c, sample.cov=m.ece$rho, sample.nobs=906,
estimator="ULS", std.lv=TRUE)

#Explorar resultados de manera general
summary(cfa.a, fit.measures=TRUE)
summary(cfa.b, fit.measures=TRUE, standardized=TRUE)
summary(cfa.c, fit.measures=TRUE)

#Explorar parámetros
parameterEstimates(cfa.a, ci=FALSE, standardized=TRUE)
parameterEstimates(cfa.b, ci=FALSE, standardized=TRUE)
parameterEstimates(cfa.c, ci=FALSE, standardized=TRUE)

#Explorar más índices de ajuste
fitMeasures(cfa.a)
fitMeasures(cfa.b)
fitMeasures(cfa.c)
```

```
##ANÁLISIS 3: DIFFERENTIAL ITEM FUNCTIONING

#Estudio de AJUSTE DE MODELO BIFACTOR
library(mirt)

#Modelo BIFACTOR
m.bfact<-
c(1,2,3,1,4,2,1,3,6,1,2,4,3,1,2,3,4,1,2,3,1,2,8,3,1,4,1,4,6,5,7,8,4,6,
5,1,7,4,6,8,1,4,7,5,6,4,7,8,5,8)
bfact<-bfactor(ece, m.bfact, SE=TRUE, itemtype="graded")
summary(bfact)

##Generar parámetros de ítems
coef.1<-coef(bfact)

##Generar Zh de los sujetos
fit.1 <- personfit(bfact)

#Exportar la habilidad de los sujetos la base de datos e importarla
("ece.dif")
library(Rcmdr)

#Calculo del DIF
library(lordif)
attach(ece.dif)

#Definir conjunto de datos, variables de agrupación y de habilidad de
los sujetos
dif.ece <- ece.dif[paste("v", 1:50, sep= "")]
sx <- ece.dif$sexo
nv <- ece.dif$nivel
fit <- ece.dif$fit

#Definir ítems para analizar
itm <- c(1:50)

#DIF
dif.sx <- rundif(itm,dif.ece,fit,sx,criterion="CHISQR", alpha=0.01,
beta.change=0.1, pseudo.R2="McFadden", R2.change=0.02)
dif.nv <- rundif(itm,dif.ece,fit,nv,criterion="CHISQR", alpha=0.01,
beta.change=0.1, pseudo.R2="McFadden", R2.change=0.02)
```

**Anexo I. Análisis de normalidad y matriz de correlaciones policóricas**

Ítem	N	Media	Simetría		Curtosis	
			Estadístico	Std. Error	Estadístico	Std. Error
1. Mis profesores(as) se llevan bien entre ellos.	906	3.29	-1.384	.081	1.670	.162
2. He tenido miedo de venir a la escuela.	906	.43	2.207	.081	4.662	.162
3. Dentro de la escuela hay problemas de vandalismo.	906	1.56	.302	.081	-.863	.162
4. Hay buenas relaciones entre los profesores(as) y los alumnos(as).	905	2.81	-.762	.081	.190	.162
5. Los alumnos(as) nos llevamos bien entre nosotros.	906	2.79	-.602	.081	-.019	.162
6. Algún compañero(a) me ha golpeado.	906	.62	1.683	.081	2.066	.162
7. Los padres de familia se llevan bien con los profesores(as).	906	3.06	-.924	.081	.132	.162
8. Hay peleas en las que la gente se golpea.	903	1.61	.300	.081	-.931	.163
9. Aprendo en mi escuela.	904	3.48	-1.754	.081	3.335	.162
10. Mis padres acuden a la escuela cuando son llamados por mis profesores.	905	3.42	-2.014	.081	3.151	.162
11. Algún compañero(a) me ha insultado.	907	1.39	.505	.081	-.650	.162
12. Mis compañeros(as) me ayudan cuando lo necesito.	907	2.80	-.612	.081	-.428	.162
13. Hay alumnos(as) que no dejan dar clase.	900	2.49	-.452	.082	-.803	.163
14. Los profesores(as) son respetados.	907	3.04	-.898	.081	.332	.162
15. Me siento amenazado(a) en la escuela.	906	.31	3.022	.081	8.887	.162
16. Hay alumnos(as) que no respetan las reglas.	906	2.33	-.145	.081	-.547	.162
17. Me uno a las actividades que realizan los demás.	903	2.87	-.601	.081	-.290	.163
18. Los profesores(as) dan ejemplo de buenas relaciones.	906	3.01	-.987	.081	.082	.162
19. Me han robado en la escuela.	902	.99	.988	.081	-.018	.163
20. Hay alumnos(as) que siempre se meten en peleas.	906	1.87	.161	.081	-.936	.162
21. Los profesores(as) evitan que nos burlemos unos de otros.	905	2.98	-1.071	.081	-.162	.162
22. Me he sentido excluido, aislado o rechazado por compañeros(as).	906	.91	1.130	.081	.300	.162
23. Los profesores(as) sólo explican para los más listos(as) de la clase.	905	.82	1.333	.081	.549	.162
24. Algunos alumnos(as) destruyen el material y las instalaciones.	905	1.38	.491	.081	-.713	.162
25. Los profesores(as) nos ayudan a resolver nuestros problemas.	903	2.70	-.573	.081	-.712	.163
26. Mis compañeros(as) se interesan por mí.	904	2.38	-.310	.081	-.682	.162
27. Mis profesores(as) me ayudan cuando lo necesito.	906	2.98	-.883	.081	-.180	.162
28. Digo y defiendo mis opiniones sin dañar o molestar a los demás.	906	2.78	-.726	.081	-.334	.162
29. Dejo trabajar a los demás sin molestarlos.	903	3.24	-1.408	.081	1.416	.163
30. He golpeado a algún compañero(a).	904	.70	1.389	.081	1.502	.162
31. Sólo cumplo las reglas que me convienen.	900	1.10	.952	.082	-.364	.163
32. Las reglas de los profesores(as) son injustas.	901	1.15	.785	.081	-.419	.163
33. Le caigo bien a mis compañeros(as).	903	3.02	-1.039	.081	.769	.163
34. Pido la palabra y espero mi turno para hablar.	904	3.04	-1.035	.081	.252	.162
35. He excluido o rechazado a algún compañero(a).	903	.91	1.150	.081	.570	.163
36. Mis padres se llevan bien con mis profesores(as).	904	3.20	-1.284	.081	.626	.162
37. Me han castigado en la escuela.	901	.97	.955	.081	.053	.163

Ítem	N	Media	Estadístico	Simetria		Curtosis	
				Std. Error	Estadístico	Std. Error	Std. Error
38. Me gusta trabajar en equipo.	902	3.13	-1.156	.081	.327	.163	.163
39. Cumplo las reglas de mi escuela y de mi grupo.	902	3.06	-.971	.081	.292	.163	.163
40. Hay profesores(as) que castigan siempre a los mismos alumnos(as).	905	2.28	-.287	.081	-1.049	.162	.162
41. Mis profesores(as) valoran mi trabajo y me animan a mejorar.	905	2.95	-.900	.081	-.179	.162	.162
42. Ayudo a mis compañeros(as) cuando me lo piden.	905	2.99	-.926	.081	.335	.162	.162
43. Interrumpo la clase porque me aburro.	905	.54	1.880	.081	3.004	.162	.162
44. He amenazado a algún compañero.	905	.45	2.358	.081	4.850	.162	.162
45. Respeto la opinión de los demás aunque no esté de acuerdo.	901	2.87	-.833	.081	-.269	.163	.163
46. Siento que tengo amigos(as).	901	3.34	-1.693	.081	1.896	.163	.163
47. Me aburro en la escuela.	900	1.45	.441	.082	-.655	.163	.163
48. Los profesores(as) ponen actividades aburridas.	903	1.45	.430	.081	-.615	.163	.163
49. He insultado a algún compañero(a).	902	.93	1.011	.081	.233	.163	.163
50. Hay alumnos(as) que les caen mal a los profesores(as).	902	1.78	.184	.081	-1.242	.163	.163

**Matriz de correlación policórica**

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18	v19	v20
v1	1																			
v2	.108	1																		
v3	.139	.114	1																	
v4	.337	.183	.202	1																
v5	.220	.103	.177	.225	1															
v6	.084	.181	.090	.138	.201	1														
v7	.325	.158	.187	.328	.198	.146	1													
v8	.172	.056	.437	.311	.140	.216	.287	1												
v9	.317	.235	.174	.337	.211	.189	.334	.248	1											
v10	.117	.068	.067	.158	.120	.140	.201	.033	.223	1										
v11	.164	.225	.233	.127	.258	.486	.152	.255	.134	.106	1									
v12	.197	.152	.035	.148	.275	.165	.124	.027	.257	.153	.218	1								
v13	.011	.046	.283	.189	.089	.050	.129	.252	.097	-.104	.154	.075	1							
v14	.293	.152	.335	.366	.292	.051	.353	.355	.324	.176	.206	.187	.274	1						
v15	.237	.314	.185	.104	.170	.392	.195	.210	.262	.153	.410	.206	.012	.270	1					
v16	.075	-.010	.320	.182	.174	.086	.145	.338	.077	-.062	.214	.091	.443	.296	.110	1				
v17	.221	.136	.094	.205	.254	.151	.219	.113	.304	.269	.095	.242	.030	.291	.205	.048	1			
v18	.365	.162	.221	.387	.232	.148	.406	.228	.375	.209	.159	.202	.109	.332	.269	.160	.308	1		
v19	.187	.129	.218	.184	.081	.206	.145	.233	.195	.022	.352	.114	.216	.185	.308	.098	.072	.227	1	
v20	.154	.035	.334	.220	.151	.128	.172	.397	.131	.049	.263	.083	.375	.287	.210	.357	.104	.163	.308	1
v21	.197	.169	.225	.238	.139	.072	.186	.121	.287	.142	.115	.237	.134	.274	.252	.102	.149	.316	.157	.109
v22	.093	.253	.068	.060	.206	.396	.119	.058	.151	.050	.409	.262	.051	.103	.330	.179	.135	.136	.237	.130
v23	.277	.168	.262	.326	.141	.142	.260	.287	.356	.065	.181	.124	.198	.336	.339	.214	.104	.375	.260	.229
v24	.208	.084	.398	.269	.162	.173	.288	.388	.226	.054	.295	.066	.336	.383	.232	.383	.110	.243	.311	.417
v25	.315	.120	.234	.339	.186	-.028	.322	.180	.336	.131	.141	.209	.135	.380	.103	.166	.274	.386	.134	.214
v26	.132	.182	-.022	.091	.164	.167	.085	.005	.149	.167	.184	.493	-.014	.056	.179	.021	.260	.214	.078	-.013
v27	.376	.131	.201	.393	.193	.104	.412	.264	.434	.222	.151	.276	.163	.428	.244	.161	.372	.487	.163	.238
v28	.235	.071	.094	.184	.068	.170	.226	.118	.315	.292	.201	.178	-.001	.174	.167	.037	.246	.276	.192	.109
v29	.176	.140	.154	.190	.106	.216	.190	.194	.365	.171	.210	.135	.048	.221	.134	.097	.213	.183	.189	.145
v30	.182	.078	.178	.200	.142	.415	.123	.253	.280	.188	.333	.172	.141	.217	.248	.166	.137	.154	.340	.221
v31	.048	.089	.099	.197	.102	.167	.122	.169	.251	.085	.168	.116	.067	.159	.066	.133	.117	.222	.165	.118
v32	.213	.081	.210	.296	.103	.058	.266	.274	.352	.106	.103	.057	.168	.310	.181	.187	.188	.327	.213	.202
v33	.306	.214	.102	.173	.305	.193	.158	.091	.273	.149	.284	.341	-.036	.222	.242	.079	.323	.207	.148	.043
v34	.181	.030	.158	.204	.155	.128	.268	.162	.320	.212	.138	.204	.066	.238	.082	.090	.273	.215	.196	.066
v35	.079	.109	.150	.121	.094	.285	.169	.151	.315	.180	.293	.061	.120	.168	.192	.054	.154	.130	.179	.158
v36	.356	.112	.192	.347	.132	.110	.593	.246	.357	.287	.165	.157	.132	.344	.198	.121	.251	.414	.129	.146
v37	.163	.145	.051	.083	.006	.234	.125	.159	.202	.124	.197	.080	-.018	.051	.189	-.008	.114	.095	.153	.064
v38	.204	.021	.127	.166	.202	.015	.214	.103	.234	.109	.106	.340	.126	.283	.154	.162	.340	.185	.078	.143
v39	.168	.061	.179	.179	.190	.120	.223	.183	.397	.195	.189	.173	.027	.258	.176	.104	.232	.290	.162	.116
v40	.083	.093	.174	.230	.102	.025	.221	.237	.059	-.018	.097	.051	.345	.311	.088	.239	.060	.106	.131	.290
v41	.279	.135	.145	.350	.218	.152	.343	.190	.374	.214	.233	.229	-.003	.371	.262	.042	.326	.424	.178	.144
v42	.267	.137	.142	.202	.225	.166	.225	.164	.256	.200	.145	.326	.051	.216	.179	.141	.250	.283	.144	.070
v43	.186	.099	.172	.229	.053	.186	.232	.278	.351	.125	.156	.094	.051	.188	.211	.187	.180	.193	.173	.200
v44	.224	.070	.172	.168	.210	.299	.173	.289	.301	.197	.300	.171	.024	.147	.299	.110	.097	.235	.254	.179
v45	.232	.124	.046	.159	.139	.152	.172	.053	.204	.235	.180	.191	-.046	.178	.119	.041	.183	.256	.100	.018
v46	.270	.209	.064	.180	.169	.150	.151	.030	.290	.163	.219	.338	-.085	.089	.281	.078	.240	.206	.096	-.036
v47	.196	.151	.124	.236	.126	.056	.261	.227	.431	.174	.157	.101	.143	.238	.135	.137	.226	.246	.213	.170
v48	.232	.130	.237	.311	.121	.030	.302	.308	.350	.046	.122	.112	.178	.344	.132	.247	.202	.269	.094	.218
v49	.169	.033	.173	.190	.168	.313	.189	.287	.333	.144	.388	.172	.105	.213	.207	.151	.188	.218	.252	.232
v50	.224	.114	.338	.369	.155	.136	.295	.354	.306	.006	.216	.105	.348	.433	.203	.295	.165	.332	.247	.324

	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	v35	v36	v37	v38	v39	v40	
v21	1																				
v22	.083	1																			
v23	.277	.166	1																		
v24	.141	.170	.289	1																	
v25	.366	.122	.388	.213	1																
v26	.179	.298	.061	.053	.195	1															
v27	.374	.163	.420	.239	.597	.250	1														
v28	.243	.145	.150	.130	.184	.211	.378	1													
v29	.145	.083	.155	.158	.222	.138	.252	.387	1												
v30	.181	.232	.252	.296	.163	.162	.276	.354	.395	1											
v31	.208	.116	.285	.175	.169	.076	.182	.179	.265	.352	1										
v32	.265	.055	.410	.270	.322	-.001	.382	.156	.186	.270	.323	1									
v33	.197	.408	.096	.095	.202	.433	.306	.241	.232	.220	.119	.099	1								
v34	.172	.095	.179	.151	.221	.157	.335	.415	.420	.371	.251	.235	.235	1							
v35	.108	.231	.258	.267	.106	.072	.158	.211	.298	.321	.196	.166	.166	.292	1						
v36	.298	.151	.332	.198	.356	.159	.492	.304	.271	.237	.172	.309	.285	.373	.256	1					
v37	.094	.114	.136	.069	-.005	.015	.139	.208	.267	.291	.132	.191	.077	.265	.275	.220	1				
v38	.191	.143	.160	.174	.259	.198	.318	.157	.194	.171	.104	.102	.355	.340	.122	.320	.017	1			
v39	.157	.038	.202	.186	.230	.099	.355	.361	.432	.422	.291	.313	.295	.515	.250	.369	.275	.291	1		
v40	.130	.034	.213	.265	.231	-.004	.224	-.017	-.022	.080	.099	.224	.007	.015	.081	.149	.016	.023	.001	1	
v41	.327	.187	.402	.172	.436	.249	.531	.366	.261	.254	.206	.331	.347	.326	.203	.462	.146	.289	.362	.129	
v42	.208	.139	.196	.215	.225	.237	.357	.276	.289	.226	.172	.149	.289	.352	.212	.308	.161	.294	.357	.023	
v43	.158	.077	.277	.229	.255	.011	.268	.240	.466	.485	.320	.363	.120	.415	.374	.358	.425	.172	.419	.101	
v44	.256	.208	.372	.202	.147	.125	.252	.256	.310	.489	.263	.338	.236	.263	.469	.256	.446	.159	.297	.074	
v45	.199	.060	.171	.119	.187	.195	.230	.306	.325	.274	.109	.086	.280	.324	.314	.256	.134	.242	.369	-.046	
v46	.142	.401	.119	.039	.155	.390	.227	.214	.106	.202	.040	.049	.528	.204	.150	.268	.059	.275	.200	-.053	
v47	.189	.079	.250	.250	.249	.061	.315	.232	.245	.330	.282	.307	.137	.295	.297	.353	.231	.232	.382	.191	
v48	.195	.059	.347	.294	.374	.078	.376	.141	.217	.289	.212	.446	.102	.280	.226	.316	.208	.226	.308	.258	
v49	.098	.150	.262	.310	.191	.101	.254	.278	.359	.580	.267	.257	.223	.374	.456	.283	.340	.256	.407	.044	
v50	.253	.147	.407	.410	.364	.015	.379	.178	.139	.318	.221	.413	.109	.224	.211	.322	.146	.181	.185	.416	

	v41	v42	v43	v44	v45	v46	v47	v48	v49	v50
v41	1									
v42	.315	1								
v43	.304	.306	1							
v44	.198	.340	.487	1						
v45	.269	.326	.320	.265	1					
v46	.333	.234	.134	.231	.216	1				
v47	.306	.303	.505	.398	.180	.141	1			
v48	.360	.247	.398	.330	.157	.096	.555	1		
v49	.285	.315	.477	.544	.337	.244	.394	.337	1	
v50	.361	.165	.363	.266	.116	.082	.288	.439	.362	1