



UNIVERSIDAD CATÓLICA ANDRÉS BELLO
FACULTAD DE HUMANIDADES Y EDUCACIÓN
ESCUELA DE PSICOLOGÍA

COMPARACIÓN DE LA TEORÍA DE RESPUESTA AL ÍTEM Y DE LA TEORÍA CLÁSICA DE LOS TESTS EN FUNCIÓN DEL ANÁLISIS DE RESULTADOS DE UNA ESCALA DE DEPRESIÓN

Trabajo de Investigación presentado por:

Héctor ACOSTA

a la

Escuela de Psicología

Como un requisito parcial para obtener el título de

Licenciado en Psicología

Profesor Guía:

José Gregorio DE LLANO

Caracas, 22 de Julio de 2011

A mis padres y hermanos

AGRADECIMIENTOS

En primer lugar quiero agradecerle a mi madre por su amor y apoyo incondicional. Porque sin ella simplemente nada de lo que habría alcanzado hasta ahora sería posible. Gracias por ser quien eres, por ser mi guía y mi ejemplo simplemente gracias por ser la mejor madre del mundo.

Agradezco a mi padre por siempre estar allí de manera incondicional y brindarme su amor, cariño y confianza. Te agradezco por siempre tener palabras de apoyo y por tu compañía. Gracias por ser un gran amigo y aun mejor padre.

Agradezco a mi hermano Carlos por su apoyo y cariño incondicional. Por enseñarme a darle sentido a mi vida, ya que como dice Nietzsche “sin música la vida no tendría sentido” y tu hermano me mostraste la buena música.

A mi hermano Iván por ser un gran ejemplo de dedicación, trabajo y entrega. Por su apoyo y cariño incondicional. Y por siempre tener palabras que me orienten y me guíen cuando más los necesito.

A mi tutor, José Gregorio de Llano, por su tiempo y apoyo incondicional y por que sin el difícilmente este trabajo pudo haber sido realidad. Gran profesor y tutor pero aún mejor persona. De verdad muchísimas gracias.

A mis compadres Alberto, Rogelio por ser mis amigos y mis confidentes a lo largo de estos años y por siempre brindarme un apoyo incondicional y estar presentes cuando los he necesitado.

A Mari y a Daniela M. por ser personas que siempre me han brindado su apoyo y cariño incondicional, y por ser grandes amigas en este largo recorrido.

A Daniela C. porque simplemente sin ti hubiera sido difícil terminar este trabajo. Gracias por tu tiempo y apoyo incondicional. De verdad muchísimas gracias.

Gracias a todas aquella personas que han estado a lo largo de mi carrera y que han contribuido con su presencia en hacer de este camino rocoso algo mucho más agradable. Entre ellas están: Adriana, “las frutas”, Luna, Laura E., Tepy, Dani, Caro y muchas otras que siempre estarán en mis recuerdos por haberme acompañado en este recorrido. A todos ustedes muchísimas gracias.

ÍNDICE DE CONTENIDO

RESUMEN	IX
INTRODUCCIÓN	1
MARCO TEÓRICO	3
MÉTODO	33
Objetivo General	33
Objetivos Específicos	33
Definición Conceptual y Operacional de la Variable	34
Constructo a medir: Depresión	34
Definición Operacional de Depresión	35
Diseño de Investigación	35
Muestra	38
Instrumentos y/o Aparatos	39
Prueba Piloto	41
Procedimiento	42
Análisis de Datos	44
Consideraciones Éticas	46
ANÁLISIS DE RESULTADOS	49
Características de la Muestra	49
Puntajes totales de la escala de depresión (TCT)	52

Análisis Factorial	55
Posición del ítem en el continuo (TCT)	56
Supuesto de invarianza (TCT)	58
Ajuste del modelo de Rasch	59
Puntajes en el parámetro θ (Rasch)	63
Parámetro b (Rasch)	64
Supuesto de Invarianza (Rasch)	68
Comparación de los puntajes de los sujetos obtenidos por TCT y Rasch	70
Comparación de los parámetros de los ítems	71
Comparación de los índices de confiabilidad	73
Comparación del cumplimiento del supuesto de invarianza de los parámetros de los ítems	74
DISCUSIÓN DE RESULTADOS	75
CONCLUSIONES	83
LIMITACIONES Y RECOMENDACIONES	86
REFERENCIAS BIBLIOGRÁFICAS	87
ANEXO A. Escala de Depresión	92
ANEXO B Tablas de resultados de Análisis Factorial de la Escala de Depresión	96
ANEXO C. Estadísticos descriptivos de los ítems.....	102
ANEXO D. Correlaciones de las posiciones de ítems entre las submuestras	106

ÍNDICE DE TABLAS

Tabla 1. Diferencias entre la TCT y la TRI	18
Tabla 2. Estadísticos descriptivos de la Edad de la muestra	49
Tabla 3. Número y porcentaje de estudiantes por escuela	50
Tabla 4. Estadísticos descriptivos para la escala total de Depresión	53
Tabla 5. Coeficiente Alfa de consistencia interna, casos válidos y excluidos para su cálculo	54
Tabla 6. Ítems con baja correlación con los puntajes totales del test	57
Tabla 7. Índice de dificultad de los ítems de acuerdo a la TCT	58
Tabla 8. Correlaciones entre las distintas submuestras en relación a la posición del ítem en el continuo (Índice de Dificultad)	59
Tabla 9. Tabla de indicadores de ajuste global en puntajes ponderados y estándar	60
Tabla 10. Tabla de indicadores de ajuste de los ítems en puntajes ponderados y estándar	61
Tabla 11. Ítems que no se ajustan al modelo de Rasch y cuál fue el criterio que violaron	62
Tabla 12. Estadísticos descriptivos del parámetro Depresión en escala logit	63
Tabla 13. Estadísticos descriptivos del índice de dificultad “b” en escala logit	63
Tabla 14. Media y desviación de cada uno de los ítems en relación al índice de dificultad “b”	65

Tabla 15. Correlaciones entre las 10 submuestras en cuanto al parámetro “b”	69
Tabla 16. Correlación entre los puntajes obtenidos por medio de la TCT y el parámetro de los sujetos de acuerdo al modelo de Rasch	70
Tabla 17. Posición de los ítems de acuerdo a la TCT transformados a puntajes Z	72
Tabla 18. Correlación entre los puntajes obtenidos por medio de la TCT y el del parámetro de los ítems de acuerdo al modelo de Rasch	72
Tabla 19. Índices de confiabilidad Alfa de Cronbach (TCT), y confiabilidad por ítem y por persona (Rasch)	74
Tabla 20. Estadísticos de las correlaciones entre las distintas submuestras en relación a la posición del ítem y al parámetro b	74
Tabla B1. Tabla de varianza explicada total	97
Tabla B2. Matriz de componentes rotados	99
Tabla C1. Estadísticos descriptivos de los ítems	102
Tabla D1 Correlaciones de la posiciones de los ítems entre las submuestras de acuerdo a la TCT	107
Tabla D2. Correlaciones de las posiciones de los ítems entre las submuestras de acuerdo al modelo de Rasch	108

ÍNDICE DE FIGURAS/GRÁFICOS

Figura 1. Escalamiento conjunto de ítems y personas	7
Figura 2. Curva Característica del Ítem	8
Figura 3. Formula de probabilidad según el modelo de uno, dos o tres Parámetros	10
Figura 4. Histograma de frecuencia de la edad de los sujetos de la muestra	50
Figura 5. Gráfico de Barras que muestra el porcentaje de sujetos por carrera	51
Figura 6. Gráfico de Barras que muestra el porcentaje de sujetos por semestre ..	52
Figura 7. Histograma de frecuencia de los puntajes obtenidos en la escala	53
Figura 8. Grafico P-P de normalidad de los puntajes totales de los sujetos en la escala	54
Figura 9. Gráfico de sedimentación de Cattell	56
Figura 10. Histograma de frecuencia de los puntajes de los sujetos en el parámetro θ	64
Figura 11. Histograma de frecuencia de los puntajes de los ítems en relación al parámetro θ	66
Figura 12. Mapas de ítems y de personas en relación al parámetro depresión	68
Figura 13. Gráfico de dispersión que muestra la asociación entre Puntajes totales de los sujetos en la TCT y los valores que obtuvieron en el parámetro θ TRI	71
Figura 14. Gráfico de dispersión que muestra la asociación entre los índices de dificultad de la TCT y b del modelo de Rasch	73

RESUMEN

En la presente investigación se comparó de manera empírica los resultados obtenidos por un grupo de sujetos en una escala de depresión por medio de dos teorías psicométricas distintas, la Teoría Clásica de los Tests (TCT) y el modelo de Rasch de un parámetro. Esta comparación se centra en tres indicadores principales y la comprobación de un supuesto: (a) la estimación de los resultados de las personas en la escala de depresión; (b) los parámetros de los ítems, teniendo como indicadores la posición del ítem (TCT) y el umbral del ítem que determina el parámetro "b" del modelo de Rasch ;(c) la calidad psicométrica del test, tomando como único indicador la confiabilidad; y (d) el supuesto de la invarianza de la dificultad de los ítems a través de las submuestras. Para realizar esta comparación empírica se trabajó con los resultados obtenidos por 500 estudiantes de la Universidad Católica Andrés Bello (UCAB) en una escala de depresión, dentro de los cuales se encontraban 250 hombres y 250 mujeres.

En los resultados obtenidos, se pudo observar que ambas teorías son bastante equivalentes en cuanto a la estimación del nivel depresión y en cuanto a la estimación de los parámetros de los ítems, obteniendo correlaciones mayores a 0,90. En cuanto a la estimación de la confiabilidad, también se obtuvieron resultados similares entre ambas teorías ya que de acuerdo a ambas teorías el instrumento resultó ser confiable. Y en cuanto a la comprobación del supuesto de invarianza, los parámetros de los ítems estimados por la TCT se mostraron más invariantes en relación a los obtenidos por el modelo de Rasch. A pesar de que se obtuvieron resultados equivalentes entre ambas teorías se puede concluir que el modelo de Rasch brinda una mayor profundidad de análisis tanto en la descripción de los resultados obtenidos por las personas como de los ítems; ganando de este modo mayor precisión predictiva en el proceso de medición.

Palabras clave: Comparación empírica, modelo de Rasch, Teoría Clásica de los Test, parámetros de los ítems y supuesto de invarianza.

INTRODUCCIÓN

A principios del siglo XX, luego de los primeros trabajos clásicos publicados por Spearman (1904, 1907 y 1913; cp. Muñiz, 2000) donde explica los principios de la Teoría Clásica de los Tests (TCT) se puede marcar el inicio de la Teoría de los Tests. Luego de más de medio siglo de predominio de esta metodología en la creación y análisis de test psicológicos, en 1968 Lord y Novick publican *Statistical Theories of Mental Test* (Teorías Estadísticas de los Test Mentales), donde sintetizan y critican todo lo hecho anteriormente por la TCT, lo que le abre paso a nuevas propuestas como la Teoría de Respuesta al Ítem (TRI), que surge cuando Rasch publica en 1960 sobre el modelo logístico de un parámetro, marcando el inicio de esta nueva teoría (Muñiz, 2000).

A lo largo de mucho tiempo, ambas teorías han sido comparadas en cuanto a las ventajas que ofrece una sobre otra, este tipo de comparaciones se ha llevado en dos niveles: a nivel teórico y a nivel empírico. A nivel teórico existen publicaciones donde se enumeran las diferencias de ambas teorías en cuanto a sus supuestos básicos como por ejemplo las de Muñiz (1997), Muñiz (2000) y Cortada (2004).

A nivel empírico existen investigaciones como las de Fan (1998), Nukhet (2002), Silvestre (2009), Magno (2009) y Hernández (2009), que contrastan ambas teorías según la metodología de evaluación psicométrica que cada una de ellas propone para el análisis de ítems y para estimar los resultados obtenidos por los sujetos.

En el presente estudio se pretende: comparar los resultados obtenidos por personas en una escala de depresión, comparar la posición de los ítems en el continuo latente, comparar la estimación de la confiabilidad del instrumento y la comprobación del supuesto de que la posición del ítem no varía en función de la muestra en la cual se aplique. Para lograr estos objetivos se utilizarán las técnicas estadísticas propuestas por los estudios antes mencionados, que permiten

contrastar de manera empírica la metodología que brinda la Teoría Clásica de los Tests (TCT) y el modelo de Rasch de un parámetro.

Dentro del contexto académico donde se va a elaborar este trabajo de grado, puntualmente la UCAB, sólo se encontró un trabajo de investigación que utilice o mencione el modelo de Rasch. De este modo, enmarcándose dentro del campo de la psicometría, mediante la elaboración de éste estudio se pretende obtener evidencia empírica que demuestre la utilidad de la metodología que ofrece el modelo de Rasch para el análisis de las propiedades psicométricas de los tests psicológicos, a través de su comparación empírica con la TCT.

La metodología propuesta por la TRI ha sido utilizada en múltiples ambientes académicos y profesionales, así como en estudios y publicaciones de las distintas ramas de la psicología a nivel mundial, pero poco aplicada en el contexto donde se enmarca el presente estudio. De acuerdo a los objetivos del presente estudio que se enmarcan dentro del campo psicométrico, existe también una contribución al campo de la psicología clínica, ya que desde la psicometría se aborda un constructo clínico, lo que permitiría generar aportes a nivel teórico y una contribución a la evaluación y diagnóstico de la depresión.

Durante la presente investigación se contó con el permiso y aprobación previa de las autoridades de la UCAB y de las escuelas de la misma para la administración de los instrumentos. De igual forma, se contó con el consentimiento informado de los participantes, se le informó a los mismos del procedimiento de manera verbal y se aseguró y aplicó la confidencialidad de sus respuestas; todo ello en función del Código Deontológico de la Práctica de la Investigación en Psicología (Escuela de Psicología UCAB, 2002). Adicionalmente, en el aspecto ético y en consonancia con el mismo código; se preservaron los derechos de autor para todos los documentos de investigación consultados y reportados en esta investigación.

MARCO TEÓRICO

El origen de la Teoría de los Tests puede ubicarse en los primeros trabajos clásicos realizados por Spearman (1904, 1907, 1913; cp. Muñiz, 2000) en los que se establecen los fundamentos de la Teoría Clásica de los Tests (TCT). El objetivo central Spearman era encontrar un modelo estadístico que fundamentase adecuadamente las puntuaciones en los tests y permitiera la estimación de los errores de medida asociados con el proceso de medición. El modelo de Spearman destaca por su simplicidad matemática y su importancia para la psicología. Este modelo asume que la puntuación empírica de un sujeto en un test consta de dos componentes que se suman; el primero es la puntuación verdadera del sujeto (V) y el segundo es el puntaje de error (e) que va unido a cualquier proceso de medición (Muñiz, 2000).

De acuerdo a Muñiz (2000) el modelo de la TCT de Spearman plantea que cuando se le administra a una persona un test no hay manera de saber su puntuación verdadera, lo único que tenemos es su puntuación empírica; es decir los puntos que este obtiene en el test y su puntuación verdadera se estimarán con base en los supuestos de la TCT. Dichos supuestos plantean que el rasgo medido por un test de ejecución puede ser representado como un continuo latente, siendo en esta escala donde el individuo toma cierta posición, según sea su desempeño.

Por otra parte, Magnusson (1969) muestra que el desempeño de un individuo puede variar de una medición a otra (puntaje de error). En la TCT el puntaje de error es entendido como la diferencia entre el puntaje observado y el puntaje verdadero; y tal diferencia se puede deber a factores de azar o a la sensibilidad del instrumento a las fuentes de varianza de error (por ejemplo: adivinación, subjetividad de los evaluadores, presencia del evaluador, etc.). Este puntaje de error tiene tres propiedades básicas: (a) la media de los errores es igual a cero, (b) la correlación entre los puntajes de error en diferentes

aplicaciones es cero y (c) la correlación entre los puntajes de error y los puntajes verdaderos es cero.

Lord y Novick en 1968 (cp. Muñiz, 2000) publican *Statistical Theories of Mental Test* (Teorías Estadísticas de los Test Mentales), donde sintetizan y critican la mayoría de los aspectos planteados anteriormente por la TCT; a los cuales la misma no podía dar respuesta, abriendo de este modo nuevas perspectivas como la propuesta que hoy se conoce como Teoría de Respuesta al Ítem (TRI), que tiene como inicio 1960 cuando Rasch publica sobre su modelo logístico de un parámetro (Muñiz, 2000).

Por su parte, Hernández (2009) y Muñiz (1997) proponen que la TRI no contradice ni los supuestos ni los principios fundamentales de la TCT, sino que hace propuestas adicionales que permitirán responder cuestiones que la TCT no podía. El nombre de Teoría de Respuesta al Ítem proviene de la idea de que este enfoque se basa en las propiedades de los ítems más que en la del test global.

Muñiz (1998) plantea lo siguiente:

Si el eje angular de la TCT es asumir que la puntuación empírica viene dada por la puntuación verdadera más un error aleatorio ($X = V + e$), la TRI va a hacer una asunción ciertamente más restrictiva, a saber, que existe una relación matemática o función que conecta la competencia de los sujetos con la probabilidad de que éstos respondan correctamente a los ítems. En otras palabras, que dada la competencia de una persona en la variable medida, conocemos la probabilidad que tiene de acertar el ítem (p. 11).

De acuerdo con Muñiz (1997, 1998) y Gómez e Hidalgo (2003) los objetivos generales de la TRI son, en primer lugar, proporcionar mediciones de variables psicológicas que sean invariantes respecto a los instrumentos de medida utilizados; y segundo lugar, disponer de instrumentos de medidas cuyas propiedades sean invariantes e independientes respecto a las personas evaluadas, lo que implica que el índice de dificultad de un ítem sea independiente

de la muestra en la cual es aplicado, siendo entonces una propiedad invariante del instrumento.

En este sentido, Silvestre (2009) plantea que los modelos de TRI producen estadísticos de los ítems independientes de la muestra a la cual se les administre y que los estadísticos de las personas en los constructos medidos son independientes del grupo de ítems que estas respondan. Así mismo, Lord (1980, cp. Stage, 2003) y Lamprianou (2007) resaltan que la invarianza de los parámetros de los ítems en diferentes grupos de personas es una de las características más importantes de la TRI. La invarianza de los parámetros es crucial para determinar la generalidad del instrumento a través de diferentes poblaciones o condiciones de medición.

La invarianza de los parámetros posibilita que independientemente de que el modelo se ajuste o no a los datos, las características del test no dependerán de la muestra y las medidas de los sujetos no dependerán del test utilizado. En este sentido, Prieto y Delgado (2003) proponen que la medida sólo puede ser considerada válida y generalizable si no depende de las condiciones específicas por la que es obtenida. De igual manera, la diferencia entre dos ítems no debe depender de las personas específicas que se utilicen para cuantificarla.

Para los instrumentos de medida la invarianza es uno de los requisitos indispensables ya que implica que las propiedades de los ítems no varían a través de subgrupos de sujetos a los que se les administre el test ya que los puntajes de los sujetos no son función de variables demográficas o características particulares de cada grupo. La invarianza también supone que las características de los ítems no están siendo afectadas por la inclusión o exclusión de otros ítems al test. El cumplimiento de este supuesto permite que los sujetos sean comparados independientemente de la muestra de ítems que se haya utilizado para evaluarlos (Gómez e Hidalgo, 2003).

Otra característica importante es que en el modelo de Rasch las estimaciones, tanto de las características de los ítems como de los individuos, se

obtienen en la misma escala de medición, una escala logarítmica que va de $-\infty$ a $+\infty$, lo que facilita su comparación, mientras que en la TCT las estimaciones se obtienen en diversas escalas (Asun y Zuñiga, 2008).

Bond y Fox (2001) plantean que es muy importante partir de la noción básica de medición en el modelo de Rasch, según la cual una medida debe referirse a un constructo unidimensional y creciente. En este sentido, toda variable que intentemos medir debe ser susceptible de ser comprendida dentro de este esquema en el que va aumentando progresivamente sobre una única dimensión. Este orden progresivo del constructo debe ser elaborado teóricamente y es lo que en principio permite diferenciar a personas en los extremos del constructo que queremos medir. La propuesta básica de la medición es que podamos diferenciar la cantidad de rasgo que tiene cada persona. Así, en pruebas de habilidad se puede hablar de mayor o menor capacidad. En pruebas orientadas a medir variables de índole personal y social también existe esta progresión, pero no de habilidad, sino de nivel en el constructo que es revelado por la conducta descrita en el ítem.

Basados en la lógica del orden, los softwares que utilizan el modelo de Rasch ejecutan una transformación logarítmica sobre los resultados de los sujetos y de los ítems para convertirlos de una escala ordinal a una escala de intervalo. Esta transformación representa la capacidad o habilidad del sujeto y la dificultad de los ítems detectados en la muestra. De este modo, tanto las capacidades de las personas como las propiedades de los ítems se representan en las posiciones a lo largo del continuo de una escala de intervalo (Bond y Fox, 2001).

El modelo de Rasch se refiere al caso en que se supone que una dimensión subyacente simple genera un conjunto de respuestas observadas a un conjunto de ítems. Según este modelo la probabilidad de responder a un determinado ítem de un conjunto es una función de dos componentes: los parámetros del ítem y un parámetro para la persona, que refleja dónde está localizada la misma en el atributo subyacente (Richaud, 2006).

Se puede observar como ejemplo en la figura 1 extraída de Bond y Fox (2001), un gráfico llamado Mapa de Ítems y Personas. En el eje vertical (y) se muestra el recorrido del continuo latente del parámetro a medir θ . Este continuo latente oscila de $-\infty$ a $+\infty$, implicando que a mayor puntaje mayor cantidad de parámetro. En el centro del grafico se encuentra una línea que divide a los ítems y a las personas. Del lado izquierdo se encuentran representados por las "x" las personas y su posición relativa en relación al constructo representado en el eje "y". Del lado derecho se muestran los ítems que conforman el instrumento (representados por su número dentro del test) y su posición relativa con respecto al continuo latente. De este modo se puede observar la utilidad para el análisis ya que el que tanto los puntajes de los ítems como las personas son expresadas en una misma escala, por lo que se puede describir su comportamiento en una misma representación gráfica.

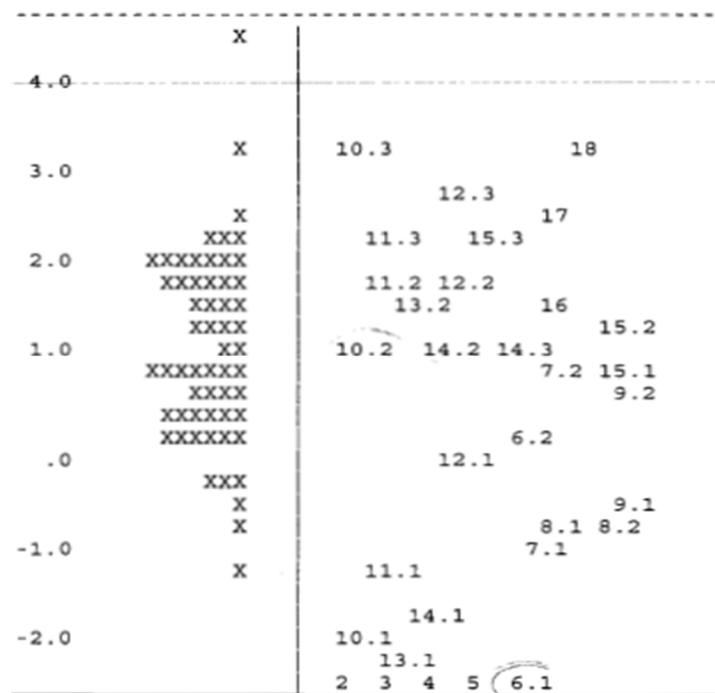


Figura 1. Escalamiento conjunto de ítems y personas. Cada x a la izquierda representa dos personas y a la derecha cada ítem es representado por su número en el instrumento. Tomado de Bond y Fox (2001). "Applying the Rasch model: Fundamental measurement in the Human Sciences" (Aplicando el modelo de Rasch: Fundamentos de la medición dentro de las ciencias humanas)

Según Muñiz (1997) la unidad conceptual fundamental de la TRI implica la asunción de una relación funcional entre los valores que miden los ítems (Rasgo Latente θ) y la probabilidad de acertar estos ($P(\theta)$), siendo representada dicha función por la Curva Característica del Ítem (CCI) que se muestra en la Figura 2.

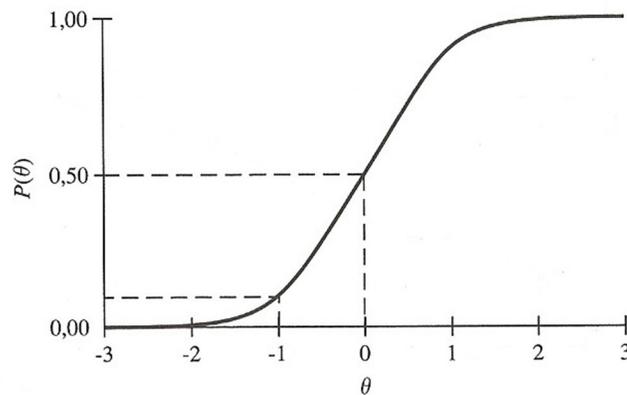


Figura 2. Curva Característica del Ítem

En el modelo de Rasch la forma en la que se relaciona la probabilidad de responder a un ítem x , está modelada como una función de la posición respectiva en el parámetro y en el ítem. En las pruebas que miden logros y habilidades la posición respectiva del sujeto se expresa en términos de habilidad y la posición del ítem se expresa en términos de dificultad del ítem. En pruebas que midan actitud o cualquier rasgo de la personalidad los términos nivel de habilidad y dificultad del ítem no son apropiados y se utiliza la noción de actitud hacia algo y posición del ítem en la escala del rasgo. Para ser neutral en cuanto a las áreas de aplicación de los términos utilizados, es pertinente hablar de la posición del sujeto en el continuo y de la posición del ítem en el continuo (Wilson, 2005).

Asimismo, cada ítem de la escala puede representarse gráficamente con una curva o función no lineal (en general, una función logística) que caracteriza las propiedades específicas del ítem (CCI). En la evaluación educacional, la dimensión latente subyacente (θ) como un conjunto de ítems se llama *habilidad*,

mientras que en la evaluación de la personalidad describe la forma con que las personas responden a los ítems de un inventario, las características de estos ítem y el lugar que cada individuo tiene en la dimensión subyacente de personalidad (Richaud,2006).

La probabilidad de acertar un ítem sólo depende de los valores de la variable medida por el ítem, esto implica que sujetos con distintas niveles en dicha variable tendrán probabilidades distintas de acertar los ítems. Para definir adecuadamente la CCI se debe explicar los parámetros que se observan en dicha curva.

En primer lugar, el parámetro “a” se denomina índice de discriminación, siendo su valor proporcional a la pendiente de la recta tangente de la CCI en el punto máximo de la pendiente de la curva, mientras mayor sea la pendiente mayor será el índice de discriminación. Por otra parte, el parámetro “b” se denomina índice de dificultad y es el valor del parámetro (θ) correspondiente al punto máximo de la pendiente CCI. Por último, el parámetro “c” representa la probabilidad de acertar el ítem al azar por parte de un sujeto. Así, la CCI queda definida cuando se especifican estos tres parámetros y se adopta una determinada función matemática para la curva (Muñiz, 1997).

Según Muñiz (1997) y Cortada (2004) existen variados modelos explicativos de la TRI como son los modelos logísticos de un parámetro o modelo de Rasch, modelo logístico de dos parámetros y el modelo logístico de tres parámetros. La elección del modelo a utilizar se da en función de las características del diseño. El modelo de un parámetro llamado índice de dificultad o umbral del ítem (Rash, 1960; cp. Richaud, 2006), describe la probabilidad de adherir a un determinado ítem y refleja el punto de inflexión de la curva en el cual la probabilidad de adherir al ítem o de rechazarlo es 0,5. El parámetro indica el punto de transición entre la probabilidad de rechazar el ítem como no autodescriptivo y la de adherir al ítem como autodescriptivo, con valores bajos que indican alta adhesión al parámetro o ítems no difíciles (Richaud, 2006).

El modelo de Rasch de un parámetro es el más popular debido a su sencillez ya que la respuesta a un ítem sólo depende de la competencia de la persona y de la dificultad del ítem, es decir, de θ y de b (Muñiz, 1997; Cortada, 2004).

Por otra parte, el modelo de dos parámetros plantea que la probabilidad de acertar el ítem está determinada por la capacidad del sujeto, la dificultad del ítem (b) y por el índice de discriminación del ítem (a). Y por último, el modelo de los tres parámetros incluye a la capacidad de responder el ítem al azar (parámetro c) como un tercer factor útil para estimar la probabilidad de respuesta correcta a un ítem, adicional al índice de discriminación (parámetro a) y al índice de dificultad (parámetro b). En la figura 3 se muestran las fórmulas para obtener la probabilidad de que un individuo responda de manera correcta un ítem según cada uno de los modelos que involucran a los tres parámetros logísticos (Muñiz, 1997).

Un parámetro:	$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$
Dos parámetros:	$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$
Tres parámetros:	$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$

Figura 3. Formula de probabilidad según el modelo de uno, dos o tres parámetros

Muñiz (1997) muestra otro de los supuestos planteados por la TRI que es la unidimensionalidad e independencia local. En este sentido, la CCI establece la relación funcional entre la probabilidad de acertar un ítem y los valores de θ , así pues si el modelo es correcto, la probabilidad de acertar un ítem únicamente dependerá de un factor que es el parámetro a medir θ . El supuesto de la independencia local implica que dado un valor determinado en el constructo latente (θ), la probabilidad de que un individuo responda correctamente a un ítem determinado es independiente de los demás ítems utilizados en el test (Sánchez, 2008).

Sánchez (2008), plantea que el supuesto de la unidimensionalidad implica que el conjunto de ítems utilizados en el test pretende cuantificar una única dimensión o variable latente. Si se cumple la unidimensionalidad, de ello se deriva matemáticamente que existe independencia local entre los ítems, lo que implica que para un individuo con determinado valor en la variable unidimensional su respuesta al ítem no viene influida por sus respuestas en los otros, ya que si esto ocurriera se caería en una contradicción, debido a que la variable unidimensional no daría cuenta de toda la varianza de los ítems, sino parte de esta varianza dependería de otros ítems (Muñiz, 1997).

Un buen proceso de medición en educación, psicología o en cualquiera de las ciencias sociales permitirá medir una habilidad a la vez, ya que si no se confundiría dos o más atributos en las medidas y puntajes que se deriven de las mismas. Por esto, cada uno de los ítems debe contribuir de manera importante para la investigación de los constructos y conceptos de manera unidimensional (Bond y Fox, 2001).

Por otra parte, Muñiz (1997) plantea que hay que comprobar en qué grado los resultados predichos con esos valores coinciden con los obtenidos, es decir, hay que comprobar el ajuste del modelo a los datos. En este sentido si los datos se ajustan a el modelo implica que los valores $P(\theta)$ pronosticadas por el modelo no difieren estadísticamente de los obtenidos empíricamente.

De acuerdo con Bond y Fox (2001), no hay duda que los modelos matemáticos no se encuentran en el mundo real. Estos modelos describen una idealización intangible, idealizaciones que realizan todas las ciencias. Los datos empíricos muestran lo imperfecto que es el mundo. En el modelo de Rasch se utilizan los estadísticos de ajuste para ayudarnos a detectar las discrepancias entre el mismo (idealización intangible) y los datos recolectados en la práctica. En el análisis de Rasch el proceso de estimación tiene dos fases principales: la calibración de la dificultad de los ítems y de la estimación de las habilidades de los sujetos, y la comprobación del ajuste de los resultados al modelo, en el sentido que la interacción entre las habilidades y las dificultades den cuenta de toda la

varianza y por lo tanto el resultado sea un instrumento unidimensional que modele apropiadamente un constructo unidimensional y progresivo. Lo que hace de este último un proceso indispensable en la metodología de análisis que te brindan los modelos de TRI.

En lo que respecta a la confiabilidad la TRI presenta no sólo índices de confiabilidad centrados en el instrumento, por lo que además de brindar un índice de confiabilidad del set de ítems que conformen un instrumento, también brinda un índice de confiabilidad de las personas (Bond y Fox, 2001).

Este índice de confiabilidad centrado en las personas indica la replicabilidad del orden de las mismas con respecto al continuo latente θ si le damos a esta muestra de personas otra serie de ítems que midan el mismo constructo. Lo que se trabaja en términos de confiabilidad de las personas es que los ítems que miden los distintos constructos sean el objetivo de la habilidad de nuestra muestra. La confiabilidad de las personas no requiere sólo habilidades bien medidas por un grupo de ítems, sino también una alta gama de niveles de la habilidad que estén presentes lo largo de la muestra, lo que demuestra una jerarquía entre las personas y la diferencia de éstas en relación a los constructos. La confiabilidad de los ítems determina la replicabilidad de los mismos a lo largo del continuo latente del constructo con muestras distintas de sujetos con niveles comparables en el constructo (Bond y Fox ,2001).

La TRI ya tiene un desarrollo histórico importante y se pueden enumerar algunas investigaciones donde se ha utilizado dicha metodología para la elaboración de instrumentos psicométricos, para la evaluación de la calidad psicométrica de test psicológicos, y de los parámetros obtenidos por los resultados de los sujetos y los ítems de los test. Ejemplo de esto es el estudio realizado por Cadavid, Delgado y Prieto (2007) que tuvieron como objetivo realizar una escala de depresión considerando las ventajas que puede ofrecer el modelo de Rasch en cuanto a la escala de intervalo del modelo, la independencia de los parámetros ítems y sujetos y las estimaciones de confiabilidad y validez tanto para los sujetos como para el grupo. Para ello utilizaron una muestra de 318 sujetos con edades

entre 18 y 24 años. En sus conclusiones plantean que el modelo de Rasch les ha permitido evaluar, comparar y escalar conjuntamente ítems de cuestionarios tradicionales. Concretamente, encontraron que el cuestionario diseñado en el estudio posee propiedades psicométricas apropiadas de confiabilidad (Ítems= 0,97 y Personas= 0,82) y validez, ajustándose al modelo de Rasch y presentando un promedio en la posición de los ítems superior a la posición presentada por los sujetos. Esto significa que el cuestionario es útil para medir niveles de depresión clínica.

Por su parte, Ferrando y Chico (2000) presentaron como propósito demostrar que la TRI es también útil para analizar y escalar un test de personalidad de longitud alta (más de 20 ítems) y proporcionar información de interés acerca de las propiedades de los ítems. Se utilizó la escala de deseabilidad social o SDS en sus siglas en inglés. En este estudio se demostró que es posible utilizar la TRI en el análisis de un cuestionario de personalidad relativamente largo y obtener de éste información tanto en lo que se refiere al comportamiento de los ítems individuales como al del test total. En esta adaptación de la SDS se observó que posee propiedades psicométricas aceptables ya que sus puntuaciones miden esencialmente una sola dimensión y la mayoría de sus ítems son escalables de acuerdo al modelo logístico de dos parámetros. Por otra parte, su nivel es adecuado para el grupo normativo utilizado y, de acuerdo con su curva de información, parece ser utilizable en un amplio rango de niveles en el rasgo a medir.

Otro ejemplo de la utilidad y aplicabilidad de la TRI es el estudio realizado por Marco, González y Gómez (2000), el objetivo general de su estudio era comparar el Análisis Factorial Confirmatorio (AFC) y la TRI como procedimientos alternativos para evaluar la equivalencia psicométrica en el contexto de la traducción de instrumentos de medida, donde se plantearon que tanto el AFC como la TRI son metodologías alternativas para abordar el Funcionamiento Diferencial de los Ítems (FDI) evaluando ambos la equivalencia psicométrica cuando un instrumento es traducido a otro idioma, esperando que si la metodología de TRI es

efectiva, el instrumento no varié en su estructura psicométrica. Los investigadores concluyeron que tanto el AFC con estructura de medias latentes como la TRI, ofrecen resultados similares en el análisis del FDI dentro del contexto de la traducción de instrumentos de medida.

Prieto y Delgado (2003), realizaron el análisis de un test en matemáticas por medio del modelo de Rasch, lo que permitió ilustrar las ventajas del mismo en la evaluación colectiva y en el diagnóstico individual, así como presentar las principales técnicas empleadas en el proceso. Plantean en principio la importancia de los ajustes de los datos al modelo, en el caso de que los datos no se ajustaran al modelo podía deberse a la multidimensionalidad o sesgo en los ítems, a la falta de precisión en los enunciados o en las opciones, a la poca motivación al responder, a los errores al anotar las respuestas, etc. En el caso de este estudio los datos se ajustaron al modelo de un parámetro lo que es de vital importancia para el análisis de los mismos. En este estudio se expusieron ventajas del modelo de Rasch sobre el modelo de la TCT, una de estas ventajas es la propiedad de medición conjunta de los parámetros del ítem y de las personas presentes en la misma escala y la posibilidad de establecer un diagnóstico individual de los sujetos en relación a la habilidad o variable medida.

Los estudios antes presentados son algunos ejemplos de la utilidad y viabilidad práctica que tiene la metodología de la TRI para la creación y análisis de instrumentos psicométricos. Ahora bien, a través del presente estudio se pretende demostrar la utilidad práctica y la viabilidad que puede tener el modelo de Rasch, y para demostrarlo se compara con la TCT en el contexto del estudio, entendiendo que éste es un estudio que pretende exponer la teoría pero que esta ya tiene una tradición dentro del campo de la psicometría (Muñiz, 2000).

Si bien la TRI data de una época un poco más reciente a la TCT, ya es una teoría que tiene casi cincuenta años desde la publicación de Rasch, lo que ya la hace una teoría de tradición dentro de la psicometría y dentro de la teoría de los test psicológicos. Muestra de esto es la presencia de la TRI en libros de texto icónicos para el estudio de la Teoría de los Test, como lo son Anastasi y Urbina

(1998), Nunnally y Berstein (1995), además en libros editados por la American Psychological Association (Sociedad Americana de Psicología) (2007) sobre el entendimiento de estadísticas multivariadas, donde le dedican un capítulo completo a la TRI. Así mismo se pueden señalar los múltiples estudios publicados en Journals de gran prestigio en el área de la psicología, donde utilizan la TRI, como en el Journal of Personality and Social Psychology (Revista de personalidad y psicología social) y el Journal de la American Psychological Association (Revista de la Sociedad Americana de Psicología). De este modo es innegable la tradición que tiene esta teoría dentro de la historia de la teoría de los test, además de su aplicación práctica.

Ahora bien; ya que se ha ilustrado la propuesta de análisis que brinda la TRI como metodología de análisis psicométrico de los test, es oportuno contrastarla en cuanto a sus supuestos teóricos con la TCT, en aras de conocer si la misma brinda o no aportes y, en caso positivo, de comprender cuáles son las ventajas o desventajas en relación al análisis clásico de los test psicológicos.

En este sentido, Cortada (2004) plantea que la principal diferencia entre la TCT y la TRI es que la relación entre el valor esperado y el rasgo en la TCT es de tipo lineal ($X = V + e$), mientras que en los diversos modelos de la TRI las relaciones pueden ser funciones de tipo exponencial, tales como los modelos de Poisson, de la ojiva normal, del error binomial, el modelo de Rasch o los modelos logísticos de dos y tres parámetros. Mientras la TCT incorpora una serie de supuestos no posibles de comprobar, la TRI ofrece la posibilidad de contar con modelos cuyos supuestos sí pueden ser juzgados empíricamente, con lo que podemos formarnos un juicio respecto de la validez de las estimaciones obtenidas (Asún y Zuñiga, 2008).

Muñiz (1997, 2010) estima que la primera diferencia entre ambas metodologías puede parecer un poco obvia ya que el énfasis de la TCT se centra en la construcción total del test, mientras que en la TRI el énfasis se encuentra en la construcción de los ítems. Resalta que las asunciones de la TCT son más generales en cuanto al nivel de explicación y la TRI es mucho más específica,

pero a su vez más robusta al momento de someter a prueba sus supuestos, lo que implica que sacrifica generalidad para ganar precisión predictiva en el proceso de medición.

Adicionalmente, Silvestre (2009) y Stage (2003) exponen que otra de las diferencias ya antes mencionadas es que en la TCT no parece existir a nivel teórico invarianza de las propiedades de los test entre las muestras, ya que el test depende de la muestra en la cual se aplica y la TRI garantiza invarianza de los ítems entre las muestras. Por ejemplo, en la TCT el índice de dificultad de un ítem está determinado por la proporción de personas que responden de manera correcta el ítem, lo que implica que para cada muestra el ítem va a tener un índice de dificultad distinto. Por otra parte, en la TRI los puntajes de los sujetos son invariantes en relación al instrumento utilizado a diferencia de la TCT, donde los puntajes de los sujetos dependen del instrumento por el cual están siendo medidos. Como medida paralela de acuerdo a la TCT se crean varios instrumentos que midan el mismo constructo lo cual afecta la economía de ítems.

Para Muñiz (1997), en la TCT los puntajes de los sujetos se encuentran entre cero y la puntuación máxima que obtenga el sujeto en el test (o alguna transformación de estos puntajes), y en la TRI los puntajes se ubican desde $-\infty$ a $+\infty$. También es relevante la medición conjunta, que de acuerdo con Prieto y Delgado (2003) implica que los parámetros de las personas (θ) y de los ítems (parámetros a , b o c) se expresan en las mismas unidades de medida y en consecuencia se localizan en el mismo continuo.

Esta característica coloca al modelo de Rasch y a los modelos de TRI en general en ventaja en relación a la TCT. Esta propiedad dota al modelo de Rasch de un carácter más realista que el de la TCT, ya que no es razonable mantener el supuesto de la invarianza de los ítems puesto que es evidente que no todos los ítems miden la misma cantidad del constructo. Otra gran ventaja que brinda esta propiedad es que permite analizar las interacciones entre las personas y los ítems en relación al constructo latente (θ). De este modo, la interpretación de las puntuaciones no se fundamenta necesariamente en normas de grupo como ocurre

en la TCT, sino en la identificación de los ítems, lo implica que las personas tienen distintas probabilidades (altas o bajas) de responder. Vale destacar que esta característica dota a los modelos de TRI de una gran riqueza diagnóstica (Prieto y Delgado, 2003).

Para Muñiz (1997), la descripción de los ítems en la TCT se da por medio del índice de dificultad y del índice de discriminación de los ítems, en cambio en la TRI la descripción de los ítems se realiza por los parámetros a , b y c (Stage, 2003), aunque puede determinarse el aporte de cada uno de los ítems por medio del análisis de cualquiera de los tres parámetros, ya que existen análisis de ajuste del modelo estándar en la TRI.

El tamaño muestral ideal necesario para poder realizar un análisis del test en el TCT podría funcionar con un número que oscile entre 200 y 500 sujetos, y en la TRI se recomienda que sean más de 500 sujetos, lo que coloca a la TCT en ventaja en relación a la TRI ya que es más práctico elaborar instrumentos con menor cantidad de sujetos (Muñiz, 1997). En la tabla 1 se muestran de las diferencias propuestas por Muñiz (2010) en base a los supuestos teóricos de ambas metodologías.

Tabla 1

Diferencias entre la TCT y la TRI

Aspectos	Teoría Clásica	Teoría de Respuesta a los Ítems
Modelo	Lineal	No Lineal
Asunciones	Débiles (fáciles de cumplir por los datos)	Fuertes (dificiles de cumplir por los datos)
Invarianza de las mediciones	No	Si
Invarianza de las propiedades del test	No	Si
Escala de las puntuaciones	Entre cero y la puntuación máxima en el test	Entre $-\infty$ y $+\infty$
Énfasis	Test	Ítem
Relación Ítem-Test	Sin especificar	Curva Característica del Ítem
Descripción de los ítems	Índices de Dificultad y de Discriminación	Parámetros a, b, c
Errores de medida	Error típico de medida común para toda la muestra	Función de Información (varía según el nivel de aptitud)
Tamaño Muestral	Puede funcionar bien con muestras entre 200 y 500 sujetos aproximadamente	Se recomiendan más de 500 sujetos, aunque depende del modelo

Nota. Tabla comparativa presentada por Muñiz (2010) en su trabajo titulado "Las teorías de los tests: Teoría Clásica de los Test y Teoría de respuesta al ítem"

Un segundo nivel de comparación de ambas metodologías, además de las proposiciones teóricas que plantea cada una, es el nivel empírico de comparación. En base a criterios objetivos y cuantitativos se realizarán comparaciones del funcionamiento de cada una de las metodologías y se demostrar así qué tan equivalentes o distantes son en los resultados que presentan.

Ejemplo de la comparación empírica de ambas teorías se observa en el trabajo propuesto por Fan (1998) donde realiza una comparación empírica de los

estadísticos de los ítems y los puntajes obtenidos por los sujetos por medio de la metodología de la TCT y por la TRI. Comparó ambas teorías en cuanto a: (a) el nivel de habilidad de los sujetos estimado por ambas metodologías, (b) la dificultad de los ítems, (c) la discriminación de los ítems, (d) cuan invariante es la dificultad de los ítems medidos en distintas muestras y (e) qué tan invariante es la discriminación de los ítems en ambas teorías a través de muestras distintas de sujetos. Obtuvo en primer lugar que no existían diferencias significativas entre la estimación del nivel de habilidad que poseen los sujetos y en cuanto a la dificultad de los ítems, ya que los puntajes de las correlaciones eran superiores a 0,90; lo que implica que ambas teorías son bastante equivalentes en este aspecto.

Luego de que el modelo de Rasch mostró ser más eficiente a la hora de medir el índice de dificultad del ítem a través de distintas muestras, lo que contribuye a la utilidad de la invarianza de los ítems a través de las muestras del modelo de Rasch, se mostraron discrepancias en cuanto al índice de discriminación de los ítems con correlaciones de hasta de 0,19; donde Fan (1998) concluyó que se iban a obtener resultados distintos dependiendo de cuál metodología se aplicara y de si se toma en cuenta el modelo de dos o tres parámetros. Por último los resultados estimados por medio de la TRI se mostraron invariantes entre muestras distintas comparando el parámetro a y c en cuanto al índice de discriminación (Fan, 1998).

Vale señalar que en su estudio Fan (1998) plantea que el índice de dificultad del ítem de la TCT siendo este la proporción de personas que respondieron de manera correcta el ítem, se encuentra en una escala de tipo ordinal y el índice de discriminación del ítem representado por la correlación punto-biserial se encuentra en puntajes de correlación, por lo que para la comparación con los parámetros a y b de la TRI, que se encuentran en una escala de intervalo, transformó ambos índices a puntajes Z de Fisher; asumiendo que tanto los índices de dificultad como los de discriminación se distribuyen de manera normal, lo que facilitará la comparación tomando como criterio la equivalencia de

los índices entre los puntajes estimados por las metodologías brindadas por ambas teorías.

En la mayoría de los estudios, las correlaciones entre la estimación de la TRI y la TCT son altas, pero una de las ventajas principales de la metodología de análisis de la TRI es el que las respuestas observadas no están relacionadas de manera lineal con el rasgo latente que está siendo medido (Revelle, Condon y Wilt, 2010).

Por su parte, Nukhet (2002) realizó una comparación empírica sobre el test y los ítems, centrada en las metodologías de ambas teorías, a través los resultados obtenidos por 1118 estudiantes de 6to grado de 11 escuelas primarias ubicadas en Ancara, Turquía. El test que se les administró a los niños fue el de matrices progresivas de RAVEN. Uno de los propósitos de este estudio era comparar el índice de dificultad obtenido por medio de ambas teorías, correlacionándolos y además comparando los índices de discriminación obtenidos con el objetivo de observar qué tan equivalentes eran dichos índices. Un segundo propósito era comprobar qué tan invariantes eran los índices de dificultad, el parámetro “b”, el índice de discriminación y el parámetro “a” a través de las muestras.

Nukhet (2002), encontró un alto grado de asociación entre los índices de dificultad de los ítems obtenidos por la TCT y la TRI de un parámetro (Rasch), de dos parámetros (a y b) y la teoría de los tres parámetros (a, b y c) con correlaciones que oscilaban desde 0,89 hasta 0,98. Sin embargo, se encontró un menor nivel de asociación en los índices de discriminación de los ítems, ya que no se observó una asociación alta entre el índice de discriminación de la TCT y la teoría de los tres parámetros de la TRI. Estos resultados nos pueden indicar que puede existir un mayor nivel de equivalencia entre los índices de dificultad de ambas metodologías en relación con el índice de discriminación. En relación a la invarianza de los índices de dificultad, el autor encontró que son equivalentes ambas metodologías en cuanto a un nivel alto de invarianza con correlaciones superiores en ambas metodologías de 0,83 hasta 0,92.

En otro estudio, Silvestre (2009) comparó empíricamente las teorías en cuanto a los parámetros de los ítems y los resultados de los sujetos en un test de conocimiento en ciencias biológicas. Los objetivos de la investigación eran determinar cuan equivalente era el nivel de dificultad de los ítems entre la metodología de la TCT y de la TRI, qué tan equivalentes son la TRI y la TCT en cuanto a las medidas de consistencia interna, si los ítems mostraban unidimensionalidad y qué tan equivalente es el funcionamiento diferencial de los ítems comparándolos entre el género y nivel de preparación basados en la TCT y en la TRI.

Silvestre (2009), obtuvo diferencias significativas en los resultados obtenidos por las dos metodologías en cuanto a la determinación de la dificultad de los ítems y los puntajes de los sujetos, siendo que resultó más fácil clasificar a los ítems en cuanto a su dificultad por medio de la metodología que brinda la TRI; ya que existieron diferencias significativas entre la dificultad de los ítems y los puntajes de los sujetos comparándolos en cada una de las metodologías, y habiendo más disparidad en los parámetros obtenidos por la TCT dependiendo de su género y del nivel de preparación de los sujetos. Otros resultados señalan que existió un grado de estabilidad a la hora de determinar la confiabilidad del test por medio de ambas metodologías ya que para la TCT el coeficiente de consistencia interna fue de 0,72; y para la TRI la confiabilidad de los ítems fue de 0,98 y de las personas fue de 0,70. Para calcular la dificultad del ítem y la habilidad de las personas según la TRI, el autor utilizó el software WINSTEPS, en donde se representan tanto el parámetro del ítem como el puntaje de los sujetos con media de 0, oscilando los puntajes entre -3 y 3 en una escala logarítmica.

Por su parte, Asún y Zuñiga (2008) plantearon un estudio cuyos objetivos eran: comparar el escalamiento de los sujetos, la estimación de parámetros de los ítems y la calidad psicométrica (confiabilidad y validez); entre la TCT y la TRI. Además de asociar la actitud con las variables originales obtuvieron una similitud de los parámetros de los sujetos y de los ítems calibrados por medio de la TCT y la TRI; y de este modo lograron exponer las ventajas de la TRI sobre la TCT en la

medición de actitudes sociales. Se obtuvo además que la TRI resultó útil tanto en la estimación del error con que se midió a los entrevistados como en la descripción de las características psicométricas del instrumento aplicado, permitiendo formular sugerencias para su tratamiento. Sin embargo, no modificaron significativamente el escalamiento de los individuos ni cambiaron el grado de asociación entre la actitud y las variables adicionales ya que se encontraron semejanzas en las estimaciones por parte de ambas metodologías. Se debe señalar que la metodología de TRI utilizada fue la de modelos politómicos; modelos alternativos a los modelos dicotómicos (respuesta correcta e incorrecta) de uno, dos y tres parámetros, por lo que no profundizaremos en estos métodos alternativos de TRI, ya que para efectos de esta investigación se utilizará uno de los modelos dicotómicos.

En un estudio relacionado, Magno (2009) buscaba demostrar las diferencias entre la TRI y la TCT basándose en la estimación de dificultad del ítem, en los valores de consistencia interna, en la variación en los resultados en el test y en los errores de medida; utilizando dos modelos de un test de química. Los resultados muestran que: la dificultad de los ítems por medida por la TRI no mostró cambios a través de las muestras en comparación con la TCT donde sí se observaron cambios a través de las mismas, lo cual indica mayor nivel de invarianza en la TRI que en la TCT. Además, los índices de dificultad de los ítems se mostraron más estables en los dos modelos del test de química con la TRI en comparación con la TCT; y los coeficientes de consistencia interna de la TRI se mostraron más estables a través de los dos test en comparación por los propuestos por la TCT. Por último, por medio de la TRI se obtuvieron significativamente menos errores de medida en los puntajes obtenidos por los sujetos en comparación con la TCT. Estos resultados evidencian marcadas diferencias entre la TRI y la TCT en cuanto a las muestras, las pruebas, los parámetros de los ítems y los coeficientes de confiabilidad de ambas metodologías. Así, estos resultados pueden ser de utilidad para los investigadores para determinar qué tipo de aproximación se debería utilizar.

Hernández (2009), buscaba comparar el índice de dificultad de los ítems y el nivel de discriminación de los ítems obtenido por medio de las dos teorías (TRI y TCT) en un test de habilidad mental. Para realizar dicha comparación, correlacionó el índice de dificultad del ítem (TCT) con el parámetro “b” de la TRI, y se correlacionó las correlaciones punto-biserial con el parámetro “a”. El autor calculó el coeficiente de determinación de ambas correlaciones. En los resultados se observó un gran nivel de equivalencia entre ambas metodologías en lo que respecta a los índices de dificultad del test y a los índices de discriminación. Tomando en cuenta los tres parámetros de la TRI se observó un mayor nivel de asociación entre el parámetro “a” y el índice de discriminación propuesto por la TCT.

Por su parte, Morales (2009) presentó un estudio que tenía como objetivo evaluar las propiedades psicométricas de un instrumento que mide logros en matemáticas por medio de la TCT y la TRI; obteniendo los indicadores de calidad psicométrica (confiabilidad y validez) por ambas metodologías así como el funcionamiento diferencial de cada ítem. En el estudio participaron 80 estudiantes (34 de matemáticas y 46 de inglés). A nivel de los estimadores de confiabilidad se obtuvo por medio de la TCT el coeficiente alfa de Cronbach de 0,77 y el de Guttman de 0,72. La estimación de la confiabilidad por persona fue de 0,77 y por ítem fue de 0,95 (TRI). Basados en los resultados de la estimación de la calidad del test el autor plantea que aunque existan marcadas diferencias teóricas a nivel de resultados, las mismas no se ven evidenciadas ya que se obtienen resultados similares entre ambas.

Un ejemplo de la exploración empírica del supuesto de invarianza se encuentra en el estudio presentado por Adedoyin, Nenty y Chilisa (2008), quienes se plantearon como objetivo encontrar si la dificultad de los ítems se muestra invariante a través de dos muestras distintas y encontrar si la dificultad de los ítems var[ía de una muestra a otra dependiendo del tamaño de la misma, utilizando la metodología de la TCT y la TRI. Para ello trabajaron con un

instrumento que medía el nivel conocimientos en matemáticas, el cual aplicaron a jóvenes estudiantes de una escuela secundaria.

Dentro de sus hipótesis los autores se plantearon que no se observarían diferencias en la dificultad de los ítems en distintas muestras ni en la TCT como ni en la TRI, pero que sí existiría una diferencia en la estimación de la dificultad del ítem en muestras de distintos tamaños; estimando los índices por la TCT, manteniéndose estables los índices en la TRI sin importar la variación del tamaño de la muestra. Utilizaron 155 muestras independientes de un total de 35262 estudiantes de escuelas secundarias en Botswana. Obtuvieron como resultado que la TCT no cumplió los criterios de invarianza entre muestras distintas de personas y en muestras con distinto tamaño, ya que se observaron diferencias significativas en el contraste de los puntajes por medio de un ANOVA. Adicionalmente, el índice de dificultad calculado por la TRI se mostró invariante tras el contraste de los puntajes por medio del análisis de varianza, al comparar distintas muestras de personas y muestras con distintos tamaños. Lo anterior sirve como indicador de que el índice de dificultad del ítem calculado por la TRI es independiente de la muestra a diferencia de la dificultad del ítem estimada por la TCT, que parece depender de la muestra en que se mide.

Progar y Socan (2008), realizaron otra comparación empírica de la TCT y la TRI. En este estudio sus objetivos de investigación eran tres principalmente. En primer lugar querían verificar qué tan comparables son los parámetros estimados de los ítems y de las personas por medio de ambas metodologías. En segundo lugar, qué tan invariantes son los parámetros de los ítems a través de distintos subgrupos. En tercer lugar, qué tan invariantes son los parámetros de los ítems y de las personas a través de diferentes grupos de ítems. Para comparar ambas metodologías se utilizó un instrumento de 76 ítems, 39 medían conocimiento en matemáticas y 37 medían conocimiento en ciencias. La muestra utilizada fue de 3913 estudiantes de 4to y 5to grado de 6 países europeos (Hungría, Letonia, Holanda, Noruega, Escocia y Eslovenia), de los cuales 937 eran estudiantes

femeninos de 4to grado, 1021 estudiantes masculinos de 4to grado, 963 estudiantes femenino de 5to grado y 992 estudiantes masculinos de 5to grado.

Para comparar los parámetros de las personas, Progar y Socan (2008) correlacionaron la proporción de respuestas correctas (TCT) y los valores en θ (TRI) tanto para el grupo de ítems de matemática como para el de ciencia. Para comparar los parámetros de los ítems se correlacionaron el índice de dificultad y de discriminación de la TCT con los parámetros “a” y “b” de la TRI. Para comparar qué tan invariantes son los parámetros de los ítems se correlacionaron éstos a través del grado (4to y 5to grado), género (masculino, femenino) y entre los países antes mencionados. Además, se compararon las medias de los parámetros (con t correlacionadas y ANOVAS). Obtuvieron correlaciones altas para los puntajes en ambas pruebas ($r_{\text{matematica}}= 0,984$ y $r_{\text{ciencias}}= 0,990$). Por otro lado, para los índices de dificultad ($r= 0,972$ y $r= 0,922$) y discriminación ($r= 0,935$ y $r= 0,831$) obtuvieron puntajes igual de elevados.

Los mismos autores, para comprobar la invarianza de los parámetros de los ítems obtuvieron correlaciones altas mayores a 0,90 en ambas metodologías en cuanto al grado de los estudiantes y al género; y en lo que respecta a los países se encontraron correlaciones superiores a 0,75 tanto para las pruebas de ciencia como las de matemáticas en lo que respecta al índice de dificultad. Los investigadores concluyeron que tanto los parámetros de los sujetos como los de los ítems son altamente comparables ya que las correlaciones obtenidas son muy altas. En cuanto a la comprobación de la invarianza se puede concluir que existe alto de nivel de invarianza a través de los grupos en relación al índice de dificultad de los ítems y en los índices de discriminación.; pero también se puede concluir en el estudio que los parámetros de los ítems de la TRI se muestran más invariantes a través de las muestras cuando existe un ajuste perfecto de los datos al modelo teórico.

Tanto los trabajos de Fan (1998), Nukhet (2002), Silvestre (2009), Asún y Zuñiga (2008), Magno (2009), Hernández (2009), Morales (2009), Pregar y Socan

(2008) y Adedoyin et al. (2008); constituyen ejemplos de cómo se pueden comparar ambas teorías no sólo a nivel teórico sino también a nivel empírico.

En la presente investigación se pretende comparar empíricamente los indicadores psicométricos arrojados por la TRI más específicamente por medio del modelo de Rasch de un parámetro y la TCT. Para ello es necesaria la aplicación de un instrumento, que en el presente estudio es una escala de depresión utilizada por Cadavid, et al. (2007), para demostrar la utilidad en el ámbito clínico de la aplicación de escalas realizadas por medio de la metodología de Rasch; ya que esta metodología permite determinar el comportamiento, el nivel de intensidad o la magnitud del parámetro que mide cada ítem y en este caso más específico; permite determinar el nivel de depresión que es medido, en particular, por cada ítem.

Anastasi y Urbina (1998), plantean que los tests de personalidad son instrumentos que se utilizan para medir características emocionales, motivacionales, interpersonales y de actitudes; rasgos que se consideran distintos a las habilidades. El número de test de personalidad es muy numeroso y en su mayoría están conformados por un número importante de los inventarios de personalidad y las aplicaciones de las técnicas proyectivas que ayudan a evaluar y describir los rasgos de la personalidad. La principal ventaja del desarrollo de los inventarios, cuestionarios y escalas de personalidad sobre las pruebas proyectivas está en su simplicidad y en lo directo del método.

De esta forma, Hogan (2004) respalda esta afirmación ya que las pruebas de personalidad objetivas se pueden corregir de manera impersonal; en cambio en las pruebas proyectivas al existir una elaboración en las respuestas por parte de los sujetos debe existir un profesional con un criterio para codificar las respuestas.

Existe una diversidad de uso de las pruebas de personalidad objetivas en el ramo de la psicología forense, la neuropsicología, el asesoramiento psicológico, la selección de personal, la investigación de la personalidad y la psicología clínica. En el campo clínico, las pruebas se utilizan para ofrecer una medición

estandarizada de los rasgos y las características de la personalidad sirviendo de este modo como herramientas de apoyo para la evaluación de estos pacientes (Hogan, 2004).

El mismo autor, presenta una clasificación de las pruebas objetivas de la personalidad en cuanto a la normalidad (normal o anormal) y en cuanto al campo dominio o campo a evaluar (general o específico). Para la psicología clínica, los inventarios o cuestionarios de la personalidad están diseñados para medir características anormales, psicopatológicas y problemas de las personas. En el contexto clínico la utilidad de poder medir estas características y rasgos de la personalidad estriba en que son métodos alternativos a la observación clínica ya que son más simples para evaluar el estado actual de una persona en cuanto a la presencia de cualquier tipo de sintomatología; por ejemplo en el caso de la depresión, la ansiedad, el riesgo suicida, estrés, etc. Las pruebas de amplio dominio ofrecen extensa y diversa información sobre los rasgos generales de la personalidad del individuo, mientras que las pruebas de dominio específico son muy útiles cuando se tiene un objetivo más restringido que se quiere evaluar en el paciente; ya que requieren de menor tiempo para descartar la presencia y el nivel funcional del rasgo.

En este sentido, Cadavid et al. (2007) plantean que los autoinformes de depresión son uno de los medios más utilizados para detectar la sintomatología depresiva. En su mayoría son cuestionarios elaborados por medio de la Teoría Clásica de los Test que se sustentan en observaciones psiquiátricas y teorías psicopatológicas dejando de lado metodologías psicométricas más eficientes para su evaluación y análisis. En este sentido, estos autores decidieron realizar un cuestionario que integrara distintas fuentes empleadas para la evaluación de la depresión. El mismo incluye ítems inspirados en los enunciados del Inventario de Beck para la depresión (BDI), del Cuestionario de ansiedad estado - rasgo (STAI) y de la Escala de afectividad positiva y negativa (PANAS).

Cada test psicológico tiene una teoría que lo subyace y que guía su construcción y lo condicionan de acuerdo con los avances teóricos y estadísticos

de cada momento. El motivo de esto es que los tests psicológicos son instrumentos de medida mediante los cuales los psicólogos infieren y toman decisiones sobre aspectos importantes de las personas en los campos de la psicología clínica, el asesoramiento psicológico, la selección de personal, etc. Vale acotar que son las teorías psicométricas las que van a permitir la estimación de las propiedades de los tests para que se garantice una adecuada toma de decisiones. Sin embargo, sin las teorías que subyacen a los test, no se podrían estimar las calidades psicométricas de los mismos (si es confiable o válido), lo que es indispensable para su utilización (Muñiz, 2010).

En este sentido, es necesario mencionar de manera breve cuales son las teorías que subyacen las distintas escalas que componen el instrumento utilizado en la presente investigación: BDI-II, PANAS y STAI. Pero esta exposición no se debe hacer sin antes hacer una breve reseña del rasgo que intenta medir el instrumento, que es el de depresión y más específicamente medir si se está en presencia o no de un episodio depresivo mayor; para luego observar como las distintas teorías abordan el constructo y determinan la forma en que será medido.

En este sentido, Krauss y Halgin (2009) plantean que la depresión se enmarca dentro de los trastornos del estado del ánimo o del humor y que estos trastornos implican una perturbación en el estado emocional de una persona. Las características de los trastornos anímicos están centradas en: la gravedad del episodio (si este es leve, moderado o grave), la reincidencia de los síntomas, los indicadores que pueden reflejar la naturaleza de síntomas destacados; y, en el caso específico de la mujer, si ha experimentado el episodio anímico luego de haber dado a luz (episodio posparto).

Por otra parte, el estado de ánimo deprimido es una de las condiciones de malestar psicológico más frecuentes en los seres humanos y que potencialmente conlleva a un peligro mayor de mortalidad debido a las conductas suicidas que se pueden llevar a cabo (Belloch, Sandín y Ramos, 1995). La patología fundamental de los trastornos afectivos es el humor; el estado emocional interno de una persona, y no el afecto; que es la expresión externa del contenido emocional

presente. Los pacientes que sufren únicamente episodios depresivos se dice que padecen un trastorno depresivo mayor, llamado también trastorno unipolar (Kaplan, Sadock y Grebb, 2005).

Para entender los trastornos anímicos, para Krauss y Halgin (2009) es importante comprender el concepto de episodio, que para estos autores se entiende como un período delimitado de tiempo en el cual se evidencian determinados síntomas de un trastorno. En este sentido, el trastorno depresivo mayor está comprendido por episodios agudos, mientras que los episodios depresivos mayores se componen de síntomas depresivos. Por otra parte, la distimia o trastorno distímico se observa cuando la persona presenta una depresión más crónica pero menos grave.

Según el DSM-IV (1995) la característica esencial de un episodio depresivo mayor es que durante un período de al menos dos semanas la persona presenta un estado de ánimo deprimido o una pérdida de interés o placer en casi todas las actividades que realiza. El sujeto también experimenta al menos otros cuatro síntomas de una lista que incluye: cambios de apetito o peso, del sueño y de la actividad psicomotora; falta de energía; sentimientos de infravaloración o culpa; dificultad para pensar, concentrarse o tomar decisiones, y pensamientos recurrentes de muerte o ideación, planes o intentos suicidas. Para indicar la existencia de un episodio depresivo mayor, al menos uno de los síntomas debe ser de nueva presentación o haber empeorado claramente si se compara con el estado del sujeto antes del episodio, lo que es llamado estado premórbido. Los síntomas han de mantenerse la mayor parte del día, casi cada día, durante al menos dos semanas consecutivas. Además, el episodio debe acompañarse de un malestar clínico significativo y que afecte las distintas áreas de funcionamiento tanto social, como laboral o de otras áreas importantes de la actividad del individuo.

Por definición, de acuerdo con el DSM-IV (1995) un episodio depresivo mayor no se debe a los efectos fisiológicos directos de las drogas (por ejemplo en el contexto de una intoxicación por alcohol o la abstinencia de cocaína), a los

efectos secundarios de los medicamentos (por ejemplo corticoides) o a la exposición a tóxicos. Así mismo, el episodio no es debido a los efectos fisiológicos directos de ninguna enfermedad médica (por ejemplo hipotiroidismo). Además, si los síntomas empiezan antes de transcurrir dos meses de la pérdida de un ser querido y no persisten más allá de estos dos meses, generalmente se consideran resultado de un duelo, a menos que estén asociados a un deterioro funcional importante o incluyan preocupaciones patológicas de inutilidad, ideación suicida, síntomas psicóticos o enlentecimiento psicomotor.

Una vez explicado el rasgo que se está intentando medir por el instrumento, es momento de hacer una breve descripción de las teorías que los subyacen a cada uno los test que sirven de base para el instrumento que se utiliza en la presente investigación.

En primer lugar, el inventario de Beck para la depresión (BDI-II) es un autoinforme que evalúa la presencia e intensidad de los síntomas depresivos (Bravo, Echeburúa y Aizpiri, 2007). El BDI-II se basa en la teoría que plantea Beck sobre la depresión. La premisa básica de este autor es que en los trastornos depresivos se da una distorsión o sesgo sistemático a nivel de procesamiento de la información, que viene dado como consecuencia de cualquier acontecimiento que suponga pérdida o fracaso; además plantean que es frecuente que aparezcan sentimientos de tristeza y de abatimiento (Belloch et al., 1995)

De acuerdo con los mismos autores, la teoría de Beck plantea que en las personas depresivas aparece un sesgo o distorsión en el procesamiento acerca de los acontecimientos que las afectan. De esta forma, los individuos depresivos parecen valorar excesivamente esos sucesos negativos, los consideran globales, frecuentes e irreversibles, mostrando lo que se conoce como la tríada cognitiva negativa; lo que se traduce en una visión negativa del yo, del mundo y del futuro. Esta tríada cognitiva es la causa de una amplificación de la persona de los sentimientos de tristeza e indefensión, bien sea en su duración, en su frecuencia o en su intensidad, de forma que interfieren con la capacidad adaptativa del individuo. Este procesamiento cognitivo distorsionado o sesgado que aparece en

la depresión conduce, según Beck, al resto de los síntomas afectivos, conductuales, motivacionales y fisiológicos que caracterizan al trastorno.

El STAI, al igual que el inventario de Beck, forma parte de las seis pruebas psicológicas más utilizadas por los psicólogos de 6 países europeos (Bravo et al., 2008). La ansiedad como estado fue definida según Spielberger (1966; cp. Lorenzo, Grau, Fumero, Vizcaíno, Martín y Prado, 2003) como la condición o estado emocional que es transitorio y se caracteriza por sentimientos tanto de tensión como de aprensión subjetivas, que son percibidas a nivel consciente; y por un aumento en la actividad a nivel del sistema nervioso autónomo; además plantean que tal estado puede variar en intensidad y fluctuar a través del tiempo. Por otra parte, la ansiedad como rasgo se define como las diferencias que se dan entre los individuos que son relativamente estables, a nivel de la propensión y de la ansiedad, esto es, las diferencias entre las personas en su tendencia a responder a situaciones percibidas como amenazantes con elevaciones de la intensidad de la ansiedad-estado (Spielberger, 1966; cp. Lorenzo et al. 2003)

Por último, la escala de afectividad positiva y negativa (PANAS) surge de un estudio efectuado por Watson y Tellegen (1985 cp. Sandín, Chorot, Lostao, Joiner, Santed y Valiente, 1999) sobre un alto número de investigaciones que tenían como objetivo investigar sobre el afecto. Sugieren estos autores que los dos factores principales que aparecen en forma consistente a lo largo de los estudios y que configuran la estructura del afecto son: afecto positivo y el afecto negativo.

De acuerdo con Sandín, et al. (1999):

El afecto positivo representa el punto hasta el cual la persona se siente entusiasta, activa, alerta, con energía y con participación gratificante. El afecto negativo representa una dimensión general de distrés subjetivo y participación desagradable que incluye una variedad de estados aversivos como disgusto, ira, culpa, miedo y nerviosismo. Se asocia el afecto positivo con la extraversión y el negativo con el neurotismo y al distrés somatopsíquico (p. 38)

Como ya se dijo, en el ámbito académico de la Universidad Católica Andrés Bello no existe un trabajo en la base de datos que utilice o mencione el modelo de Rasch. De este modo, la elaboración de éste estudio aportará mayor evidencia empírica sobre la comprensión y utilidad de la metodología de la TRI en el análisis psicométrico de los test dentro del contexto donde se realiza.

Se tiene entonces que los objetivos del presente estudio son: (a) comparar los resultados de los individuos producidos por la TCT y el modelo de Rasch; en un test que pretende medir el nivel de depresión en estudiantes universitarios, (b) comparar las propiedades de los ítems; tomando como único indicador la posición del ítem en el continuo, estimado en la TCT por el índice de dificultad del ítem y a través del modelo de Rasch por el parámetro b que determina la posición del ítem en el continuo latente, (c) comparar la confiabilidad del instrumento estimada por Rasch y la TCT y (d) comprobar el supuesto de invarianza de la posición de los ítems por medio de la comparación de distintas sub-muestras de la población total encuestada.

Para comparar ambas teorías en base a estos criterios empíricos se utilizarán los resultados obtenidos en el cuestionario de depresión propuesto por Cadavid et al. (2007) en una muestra de estudiantes de la UCAB.

MÉTODO

Objetivo General

El objetivo del presente estudio es comparar la metodología de análisis propuesta por la Teoría Clásica de los Test y el modelo de Rasch a través del análisis de los resultados obtenidos por estudiantes universitarios en una escala de depresión.

Objetivos Específicos:

Específicamente se pretende comparar empíricamente 4 criterios:

1. El nivel de depresión que obtienen los estudiantes universitarios, estimado por la TCT y por la TRI por medio del modelo de Rasch de un parámetro.
2. Las propiedades de los ítems, tomando como único indicador la posición del ítem en el continuo, estimado en la TCT por la proporción de sujetos que respondieron de manera afirmativa el ítem y a través del modelo de Rasch por el parámetro b que determina el umbral del ítem.
3. La calidad psicométrica del test:

Confiabilidad, estimado por la TCT por medio del coeficiente de consistencia interna alfa de Cronbach y por el modelo de Rasch por a través del método de confiabilidad por ítem y por personas.

4. Comprobar el supuesto de invarianza de la posición que ocupan los ítems en el continuo, por medio de la comparación de los resultados de la dificultad de distintas sub-muestras de la población total encuestada.

Definición Conceptual y Operacional de la Variable

Constructo a medir: Depresión

Según el DSM – IV (1995) la característica esencial de un episodio depresivo mayor es que durante un período de al menos 2 semanas la persona presenta un estado de ánimo deprimido o una pérdida de interés o placer en casi todas las actividades. El sujeto también experimenta al menos otros cuatro síntomas de una lista que incluye cambios de apetito o peso, del sueño y de la actividad psicomotora; falta de energía; sentimientos de infravaloración o culpa; dificultad para pensar, concentrarse o tomar decisiones, y pensamientos recurrentes de muerte o ideación, planes o intentos suicidas. Para indicar la existencia de un episodio depresivo mayor, al menos uno de los síntomas debe ser de nueva presentación o haber empeorado claramente si se compara con el estado del sujeto antes del episodio, lo que es llamado estado premórbido. Los síntomas han de mantenerse la mayor parte del día, casi cada día, durante al menos 2 semanas consecutivas. El episodio debe acompañarse de un malestar clínico significativo y que afecte las distintas áreas de funcionamiento tanto social, laboral o de otras áreas importantes de la actividad del individuo.

Por definición, un episodio depresivo mayor no es debido a los efectos fisiológicos directos de las drogas (por ejemplo: en el contexto de una intoxicación por alcohol o la abstinencia de cocaína), a los efectos secundarios de los medicamentos (por ejemplo: corticoides) o a la exposición a tóxicos. Asimismo, el episodio no es debido a los efectos fisiológicos directos de ninguna enfermedad médica (por ejemplo: hipotiroidismo). Además, si los síntomas empiezan antes de transcurrir 2 meses de la pérdida de un ser querido y no persisten más allá de estos 2 meses, generalmente se consideran resultado de un duelo (v. pág. 700), a menos que estén asociados a un deterioro funcional importante o incluyan preocupaciones mórbidas de inutilidad, ideación suicida, síntomas psicóticos o enlentecimiento psicomotor.

Definición Operacional de Depresión

Debido a que los objetivos de la presente investigación es comparar la TCT y la TRI, más específicamente el modelo de Rasch, es necesario definir operacionalmente el constructo Depresión de acuerdo a cada una de las metodologías de análisis. En este sentido la Depresión será medida de acuerdo a la TCT los puntajes máximos que pueden obtener los sujetos en este test pueden oscilar desde 0 a 87 puntos, valorando con un punto cada respuesta afirmativa a los ítems. Así a mayor puntaje en la escala mayor será el nivel de depresión en el que se encuentra el sujeto.

Por su parte en la metodología de Rasch los puntajes de los sujetos en Depresión Mayor pueden oscilar de $-\infty$ a $+\infty$, lo que implica que a mayor puntaje obtenga el sujeto dentro del continuo latente será proporcional al grado de Depresión que presenta el sujeto.

Diseño de Investigación

La presente investigación es de tipo no experimental, transversal, descriptivo y psicométrica. Ya que se compararon los resultados obtenidos por los sujetos en una escala de depresión, se evaluaron los parámetros de los ítems, la confiabilidad del instrumento y comprobar la invarianza de la dificultad de los ítems entre diferentes submuestras, obtenidos estos indicadores y comparados los mismos por medio de la metodología que brinda la TCT y la TRI más específicamente el modelos de Rasch de un parámetro. De este modo se puede clasificar como un tipo de investigación no experimental o ex post facto, debido a que implicó una búsqueda empírica y sistemática donde el investigador no posee control directo sobre las variables independientes, ya que el hecho y sus manifestaciones ya han ocurrido o las mismas variables son intrínsecamente no manipulables, como su grado de Depresión (Kerlinger y Lee, 2002).

En el presente estudio se analizaron y compararon los resultados obtenidos por los sujetos por medio del modelo de Rasch y la TCT, lo que indica que también es un estudio de tipo descriptivo y transversal, ya que según lo planteado por Hernández, Fernández y Baptista (1991) los estudios descriptivos transversales son aquellos en los que se mide una o más variables en un punto determinado en el tiempo y posteriormente se realizan análisis descriptivo del cómo se comportan las mismas.

En este caso se pretendió comparar los resultados obtenidos por un grupo de sujetos en una escala de depresión por medio de dos metodologías psicométricas distintas, la TCT y el modelo de Rasch. Esta comparación se centra en tres indicadores principales y la comprobación de un supuesto: (a) los resultados de los sujetos en el test, (b) los parámetros de los ítems, (c) la calidad psicométrica del test y (d) el supuesto de la invarianza de la dificultad de los ítems a través de las submuestras

Para comprobar el primer objetivo de investigación se empleó el mismo procedimiento utilizado por Fan (1998) en su estudio, donde se correlacionaron los puntajes en la escala de Depresión estimados por el modelo de Rasch con los puntajes obtenidos por la TCT, tomando en cuenta que a mayor sea el monto de asociación más equivalentes van a ser las estimaciones por medio de ambas metodologías.

Previo la comparación de los índices de dificultad de los ítems fue necesaria la transformación de dicho índice estimado por la TCT que se encuentra representado en puntajes de probabilidad, a una escala de intervalo, para la comparación con el parámetro b de la TRI que se encuentran ya en una escala de intervalo. Así que se deben transformar ambos índices a puntajes Z de Fisher asumiendo que el índice de dificultad se distribuye de manera normal, lo que facilitará la comparación de los índices entre los puntajes estimados por las metodologías brindadas por ambas teorías, replicando la metodología utilizada por Fan (1998).

Se tomó en cuenta la investigación de Hernández (2009) para la comprobación del segundo objetivo de investigación se comparará los parámetros de los ítems en cuanto a la posición del ítem. Se correlacionó la posición del ítem representada por el parámetro dificultad del modelo de Rasch con el índice que representa la posición del ítem en la TCT.

Para la comprobación del tercer objetivo de investigación se observó si existieron diferencias significativas en la confiabilidad obtenida por los métodos de α de Cronbach y de división por mitades (TCT) en relación a los métodos de confiabilidad por persona y confiabilidad por ítem propuesto por Rasch, del mismo modo que fue realizado por Silvestre (2009) en su estudio donde tomo en cuenta como criterio base para que el instrumento sea confiable un monto de .60 independiente del método que se utilice para estimarlo.

En cuanto a la comprobación del supuesto de invarianza de la dificultad de los ítems se empleó el método utilizado por Fan (1998) donde se le aplicaron el mismo grupo de ítems a muestras aleatorias. En este caso se transformaron los puntajes estándar Z de Fisher para que puedan estar en un nivel de medida de intervalo y de este modo poder correlacionarse. Luego se correlacionaron los resultados obtenidos por las submuestras en relación al índice de dificultad del ítem (TCT) y estas correlaciones se promediaron. Luego se realizó el mismo proceso con los resultados obtenidos por los sujetos en el parámetro b del modelo de Rasch lo que implica que se correlacionaron los resultados obtenidos en las submuestras en relación al parámetro b y luego se promediaron para compararse con el promedio de las correlaciones del índice de dificultad.

Por las características de este estudio donde se buscó evaluar los resultados obtenidos por los sujetos en la prueba, evaluar los parámetros de los ítems, además de la confiabilidad del instrumento comparando dos metodologías de trabajo a nivel psicométrico como lo son el modelo de Rasch y la TCT, se puede considerar éste un estudio de tipo psicométrico.

Muestra

El tipo de muestreo que se utilizó en el presente estudio fue intencional o también llamado accidental de acuerdo con Kerlinger y Lee (2002), ya que se seleccionaron las aulas de clases que se encuentran en el campus de la UCAB a conveniencia y una vez que se ingresaba a el aula se solicitaba el permiso del profesor encargado para que permitiera administrar el cuestionario a los estudiantes en su hora de clases.

En total la muestra estuvo compuesta por 500 estudiantes de la UCAB dentro de los cuales se encontraban 250 hombres y 250 mujeres. Presentaron una edad promedio de 20 años oscilando desde los 17 hasta los 44 años.

La muestra estuvo conformada por estudiantes de las carreras de Psicología, Derecho, Educación, Comunicación Social, Ingeniería, Ciencias Sociales, Contaduría, Economía, Administración y Letras de la UCAB. Participaron en el estudio estudiantes que cursaban desde el semestre 1 hasta el semestre 10, teniendo una participación mayor de estudiantes que cursan el 4to semestre de su carrera de pregrado universitaria. Es importante aclarar que las personas que participaron en la muestra no son una población clínicamente diagnosticada.

En el análisis de resultados se pueden observar un análisis más detallado de las características demográficas de la muestra en relación a la edad, las escuelas que participaron en el estudio y el nivel de la carrera que se encontraban cursando los estudiantes al momento de responder el cuestionario.

Para comparar el supuesto de invarianza de los parámetros de los ítems a través de ambas metodologías (TCT y TRI) se seleccionaron 10 sub-muestras totalmente aleatorizadas del total de 500 personas. Cada sub-muestra estaba compuesta por un total de 30 personas.

Instrumentos y/o Aparatos

El instrumento utilizado fue una escala de Depresión elaborada por Cadavid, et al. (2007) utilizada con el objeto de evaluar indicadores de depresión en la población joven española para mejorar la calidad técnica de medición del constructo (Véase Anexo A). La escala consiste en preguntas cerradas y dicotómicas, relacionadas con un episodio depresivo mayor, afectos negativos y positivos. Incluye ítems inspirados en los enunciados del Inventario de Beck para la depresión (BDI), del Cuestionario de ansiedad estado - rasgo (STAI), de la Escala de afectividad positiva y negativa (PANAS), así como de criterios del Manual Diagnóstico y Estadístico de los Trastornos Mentales (DSM-IV) ya que existen síntomas del episodio depresivo mayor que no eran tomados en cuenta por los otros inventarios, cuestionarios y escalas.

En primer lugar el inventario de Beck para la depresión (BDI-II) es un auto informe que evalúa la presencia e intensidad de los síntomas depresivos (Bravo, Echeburúa y Aizpiri ,2007). Este fue un instrumento creado con la intención de que cubriera todos los criterios diagnósticos sintomáticos de los trastornos depresivos por el DSM-IV (Sanz, Perdigon y Vazquez, 2003). Es un instrumento de autoreporte que sirve para medir la gravedad de la depresión en adultos y adolescentes. La prueba contiene 21 ítems y los datos de Los indicadores de confiabilidad son de .92 en pacientes externos, de .93 en población universitaria y una confiabilidad test-retest 0,93. Además parece que cumplía con los criterios de validez convergente y divergente (Hogan (2004).

Las preguntas derivadas del Inventario de Beck para la Depresión (BDI) se formularon en estilo afirmativo, abarcando ambos géneros. Se utilizaron los enunciados que representaban el nivel más alto de perturbación psicoafectiva; también se incluyeron los cambios del BDI-II. Varios de los enunciados que provienen de este test fueron divididos en dos oraciones separadas, al considerar que una misma pregunta exploraba dos aspectos diferentes. Ejemplo de esto es la afirmación; *Me siento tan triste o tan desgraciado que no puedo soportarlo*, se dividió en dos afirmaciones independientes, así: *Me siento tan triste que no puedo*

soportarlo y Me siento una persona tan desgraciada que no puedo soportarlo (Cadavid, et al.2007).

En lo que respecta al cuestionario de ansiedad y rasgo (STAI) se valora el estado emocional transitorio de atención tensa y aprensión, así como de hiperactividad del Sistema Nervioso Autónomo. Se valora la ansiedad como rasgo como una predisposición estable a percibir las situaciones como amenazantes para la persona. El cuestionario está compuesto por 20 ítems. La versión española de esta prueba presenta unos índices de consistencia interna entre 0,84 y 0,87.

La escala de afectividad positiva y negativa (PANAS) es un cuestionario compuesto por 20 ítems de los cuales 10 representan a afectos positivos claramente marcados y los otros 10 afectos negativos marcados. El objetivo de sus creadores fue obtener indicadores lo más puros posibles del afecto. 0,89 y 0,87.

El inventario de Beck para la depresión (BDI) es un auto informe que evalúa la presencia e intensidad de los síntomas depresivos (Bravo, et al. 2008). Este fue un instrumento creado con la intención de que cubriera todos los criterios diagnósticos sintomáticos de los trastornos depresivos por el DSM-IV (Sanz, Perdigón y Vázquez). En el cuestionario de ansiedad y rasgo (STAI) se valora el estado emocional transitorio de atención tensa y aprensión, así como de hiperactividad del Sistema Nervioso Autónomo. La escala de afectividad positiva y negativa (PANAS) es un cuestionario compuesto por 20 ítems de los cuales 10 representan a afectos positivos claramente marcados y los otros 10 afectos negativos marcados.

La escala al estar elaborada por medio del modelo de Rasch, los ítems que están presentes en ella presentan niveles diferentes y progresivos del constructo depresión. Está constituida por 62 ítems de los cuales 34 son afirmaciones que indagan la presencia de pensamientos, sentimientos y conductas relacionadas con la depresión patológica. Además 28 términos que implican palabras aisladas que hacen referencias estados afectivos positivos y negativos (Cadavid, et al. 2007)

Como se mencionó anteriormente de acuerdo al modelo de la TCT los puntajes máximos que pueden obtener los sujetos en este test pueden oscilar desde 0 a 62 puntos, valorando con un punto cada respuesta afirmativa a los ítems. Así a mayor puntaje en la escala mayor será el nivel de depresión en el que se encuentra el sujeto. Y según el modelos de Rasch los puntajes de los sujetos pueden oscilar de $-\infty$ a $+\infty$, implica que a mayor puntaje obtenga el sujeto dentro del continuo latente será proporcional al nivel de Depresión que presenta el sujeto.

Para el procesamiento de los datos bajo la propuesta del modelo Rasch, se utilizó el software Winsteps. El Winsteps es un software basado en Windows que sirve como aplicación para realizar los análisis que respectan a el modelo de Rasch, más específicamente en pruebas académicas, encuestas de actitud y distintos instrumentos de medida (Linacre, 2011). En lo que respecta al análisis exploratorio de datos y el procesamiento de los mismos bajo la metodología de la TCT se utilizó el software estadístico SPSS 17.

Prueba Piloto

Stone (2003) plantea una alternativa a la aplicación tradicional de pruebas piloto que se suele utilizar tomando una muestra de sujetos para calibrar los instrumentos. Plantea que para poder librar al instrumento de la mayor cantidad de errores posibles se debe seguir el siguiente procedimiento. En primer lugar el investigador debe leer los ítems en voz alta y grabarlo, luego escuchar que tal suenan los ítems y hacer las modificaciones necesarias. En segundo lugar, cuando esté satisfecho con la forma en como suenan debe pedirle a otro sujeto que los lea y los responda en voz alta. Como acto seguido debe pedirle que le exprese su razonamiento al responder cada ítem, para ello debe incentivarlo a vocalizar cada detalle y opción del ítem y cuál fue el razonamiento específico que tomó para llegar a la respuesta, se debe motivar a que piense en voz alta. Se debe tomar una grabación de lo expresado por el sujeto para que luego sea considerado por el investigador. Stone (2003) expone que seguir este procedimiento de probar el instrumento con una persona podrá ayudar a cubrir la mayoría de los errores que tiene la composición del instrumento en cuanto a el

tipeo, las palabras o frases confusas y el formato. De este modo en el presente trabajo se pretende seguir este procedimiento para poder calibrar el instrumento antes de su aplicación a la muestra definitiva.

Procedimiento

1. Se seleccionó la muestra de manera intencional de la población de estudiantes de la UCAB.
2. En primer lugar realizó la aplicación del cuestionario dentro de las aulas de la UCAB. Se seleccionó las aulas a conveniencia y se le solicitó permiso a los profesores para poder ingresar al aula y administrarles el cuestionario a los alumnos. Se procedió a explicarles la necesidad de su colaboración para la realización de la investigación y se les garantizó la confidencialidad individual de los datos.
3. Se transcribieron la respuesta de los sujetos en las encuestas en una base de datos en los programas SPSS y WINSTEPS.
4. Se calcularon los resultados obtenidos por los estudiantes tanto por medio de la TCT como por medio del modelo de Rasch.
5. Se realizó un análisis exploratorio de la variable medida donde se determinó el comportamiento de la variable en cuanto a normalidad, medidas de tendencia central, dispersión y forma.
6. Se analizaron si todos los datos se ajustan al modelo de Rasch para así luego poder estimar los parámetros por medio de esta metodología.
7. Se analizaron los resultados obtenidos en la escala de Depresión por medio de la metodología de la TCT, en cuanto los resultados obtenidos por los sujetos en la escala, evaluar los parámetros de los ítems (posición del ítem).
8. Se calculó la confiabilidad del test bajo la TCT a través de coeficiente de consistencia interna alfa de Cronbach.

9. Se analizó los resultados obtenidos en la escala de depresión por medio del modelo de Rasch en cuanto los resultados obtenidos por los sujetos en la prueba, evaluar la cantidad de parámetro medido por el ítem, además de la calidad psicométrica por medio de la confiabilidad por persona y por ítem.
10. Se realizó la comparación de los puntajes obtenidos por los sujetos en la prueba analizando la correlación entre el puntaje de los sujetos obtenidos por los puntajes logarítmicos derivados del modelo de Rasch y los obtenidos por la TCT en la suma de los puntajes de los sujetos en la escala.
11. Se realizó la comparación por medio de una correlación producto momento entre los indicadores de los parámetros de los ítems de la TCT y el parámetro b del modelo de Rasch.
12. Se calcularon los coeficientes de confiabilidad y de validez propuestos por la TCT y los propuestos por el modelo de Rasch.
13. Para comprobar el supuesto de invarianza de los parámetros de los ítems se seleccionaron 10 sub-muestras totalmente aleatorizadas utilizando la aplicación de selección de casos aleatorizados del SPSS. Cada sub-muestra tenía una $n=30$ personas.
14. Para comprobar el supuesto de invarianza de los parámetros de los ítems a través de las sub-muestras se correlacionaron los parámetros de cada sub-muestra y luego se promediaron los coeficientes de correlación. Este proceso se llevó a cabo tanto para los índices de dificultad obtenidos para la TCT, como para el nivel de rasgo medido por el ítem de acuerdo al modelo de Rasch.
15. Se analizó los resultados pertinentes a los cálculos realizados para ver cómo se comportan los sujetos con respecto a la variable dependiente.
16. Se realizó la discusión de resultados y la conclusión que se derive de los mismos.

Análisis de los Datos

Por medio del análisis derivado del programa WINSTEPS de los datos se obtuvo la posición del ítem (parámetro b), la confiabilidad por personas y por ítems. Donde los puntajes serán representados por medio de la transformación Logit de los parámetros.

Luego para analizar los datos por medio de la TCT se utilizó el programa SPSS 17, donde se obtuvo el puntaje total del sujeto en el test por medio de la sumatoria de las respuestas correctas. Se obtuvo el índice que determina la posición del ítem. Por medio de este programa también se calculó el coeficiente de consistencia interna alfa de Cronbach para estimar la confiabilidad.

Para realizar la comparación de los resultados estimados por ambas teorías, en el presente estudio se siguió con la metodología de Fan (1998) planteó que el índice que determina la posición del ítem de la TCT para que pueda ser comparado con el parámetro b de la TRI que se encuentran es una escala de intervalo, se deben transformar ambos índices a puntajes Z de Fisher asumiendo que tanto los índice de dificultad se distribuyen de manera normal, lo que facilitó la comparación de los índices entre los puntajes estimados por las metodologías brindadas por ambas teorías.

De este modo los pasos que se siguieron para realizar el análisis de datos son:

1. Se cargó la base de datos tanto en el programa SPSS 17 como en el programa Winsteps.
2. Se calculó los datos obtenidos por medio de la TCT y del modelo de Rasch.
3. Se realizó un análisis exploratorio de los datos para determinar el comportamiento de la variable medida, tomando en cuenta que este factor también puede influir en la calidad psicométrica del test. De este modo también se calculara un Análisis Factorial para comprobar

que la escala es explicada por un único factor que es el que está midiendo.

4. Se calculó los puntajes de los sujetos por medio de la TCT y del modelo de Rasch respectivamente.
5. Se calculó la posición de los ítems por medio de ambas teorías utilizando el Winsteps y SPSS respectivamente.
6. Se calculó el índice de ajuste del modelo por medio del programa Winsteps.
7. Se calculó el coeficiente de confiabilidad Alfa de Cronbach utilizando el SPSS, además de calcular la confiabilidad por persona y por ítem utilizando el Winsteps.
8. Se realizó el re-escalamiento de los puntajes obtenidos por los sujetos y del índice de dificultad obtenido por la TCT, a puntajes estándares Z utilizando como herramienta el SPSS.
9. Se realizó una comparación de los resultados estimados por ambas teorías por medio de una correlación producto momento de Pearson, utilizando como herramienta el software SPSS.
10. Se realizó una comparación de los índices de dificultad de los ítems estimados por ambas metodologías por medio de una correlación producto momento de Pearson.
11. Se comprobó el supuesto de invarianza de los parámetros de los ítems a través de las sub-muestras correlacionando los parámetros de cada sub-muestra y luego se promediaron los coeficientes de correlación. Este proceso se llevó a cabo tanto para los índices de dificultad obtenidos para la TCT, como para el nivel de rasgo medido por el ítem de acuerdo al modelo de Rasch.

Consideraciones Éticas

Para la presente investigación se consideran diversos aspectos éticos lo que implica que previamente a la administración de los instrumentos, se contó con el permiso y aprobación escrita de las autoridades de la UCAB y de las escuelas dentro de las cuales se administró el instrumento, en concordancia con el Código Deontológico de la Práctica de la Investigación en Psicología (Escuela de Psicología UCAB, 2002).

De la misma forma, previo al momento de la aplicación, se debe contar con el consentimiento informado de los participantes. Este fue obtenido en respuesta a una invitación clara y detallada del procedimiento realizada de manera verbal a todos los sujetos, todo ello en concordancia con el Código Deontológico de la Práctica de la Investigación en Psicología (Escuela de Psicología UCAB, 2002).

Por otra parte, en concordancia con el mismo código; en su sección de Principios generales el investigador se dirigió en todo momento con total respeto y cordialidad hacia los sujetos. Además, en resguardo de los participantes se les informó sobre las características fundamentales de recolección de información.

Por otra parte, de acuerdo al Artículo 57 del Código de Ética del Psicólogo en Venezuela (Federación de Psicólogos de Venezuela, 1981), se resguardará la integridad física y mental de la persona al asegurarse que expresen su voluntad de aceptar o rechazar su condición de sujeto de experimentación y de suspender la experiencia en cualquier momento. Se les señaló claramente que se trataba de una colaboración voluntaria, por lo que tenían derecho, en cualquier momento, de negarse a participar en la misma.

Aunado a eso, en concordancia con el Artículo 59 del Código de Ética del Psicólogo en Venezuela (Federación de Psicólogos de Venezuela, 1981) y con el Código Deontológico de la Práctica de la Investigación en Psicología (Escuela de Psicología UCAB, 2002), se les aclaró a los sujetos el objetivo general de la investigación expresándoles que los resultados obtenidos tendrán un uso de investigación.

Adicionalmente, en consonancia con el artículo 60 del mismo código; en su sección de Principios Generales, en el presente trabajo se preserva el anonimato y la confidencialidad de las respuestas emitidas por cada uno de los sujetos en el presente estudio; y se explicó de forma verbal, clara y explícita, que la información proporcionada por ellos está protegida, y sólo será utilizada para la investigación. De forma similar, se protegen de forma activa los derechos individuales al hacer uso confidencial y exclusivo para efectos de la investigación de la información recabada, como especifica el Código Deontológico de la Práctica de la Investigación en Psicología (Escuela de Psicología UCAB, 2002).

Del mismo modo, se asegura la transparencia del proceso de investigación, se siguieron los estándares metodológicos de la investigación psicológica y se analiza a cabalidad la información recopilada. Asimismo, se reportan e interpretan los resultados en función de los datos recabados y en función de los lineamientos teóricos pertinentes. Todo lo anterior en función de lo especificado en el Código Deontológico de la Práctica de la Investigación en Psicología (Escuela de Psicología UCAB, 2002).

Se cuida además la protección de los derechos de autor, afirmando que todos los documentos de investigación consultados son reportados como exclusivos de los autores que realizaron la escritura y producción de dichas investigaciones y que se reconocen a través de citas específicas, el origen de las ideas y materiales de otros autores que fueron empleados en la redacción de la presente investigación en consonancia con el Artículo 73 del código de ética de psicólogos de Venezuela y con el Código de Ética Deontológico de la Práctica de la Investigación en Psicología (Escuela de Psicología UCAB, 2002).

Adicionalmente, de acuerdo al Artículo 63 del mismo código, se publicarán y comunicarán los resultados de la presente investigación y de acuerdo al Artículo 69 del mismo y del Código Deontológico de la Práctica de la Investigación en Psicología (Escuela de Psicología UCAB, 2002); en su sección de Análisis y reporte de resultados, el investigador asumirá máxima responsabilidad en divulgación de información psicológica.

Finalmente, en concordancia con el Artículo 70 y del Código Deontológico de la Práctica de la Investigación en Psicología (Escuela de Psicología, 2002); en su sección de Autoría y propiedad intelectual, se adjudicará justo reconocimiento a la labor de todos los que tomaron parte en el presente proyecto de investigación, reconociéndose públicamente la colaboración recibida y dándoseles la mención adecuada.

ANALISIS DE RESULTADOS

Características de la muestra

En la presente investigación se encuestó a 500 estudiantes de UCAB de los cuales estaban conformados por 250 (50%) estudiantes femeninos y 250 (50%) masculinos.

Como se muestra en la tabla 2 las edades oscilaron desde los 17 (Mínimo) hasta los 44 años (Máximo), la media fue de 20,13 años, con un coeficiente de variación de 11,27% lo que implica que la muestra es muy homogénea en relación a la media de la edad. Presenta una asimetría de 3,016 que implica que la mayoría de los valores se encuentran agrupados hacia los valores inferiores de edad, esto parece estar siendo afectado por un dato extremo de un estudiante que tiene 44 años de edad que parece influir en el monto de la asimetría. Por último se obtuvo una curtosis de 24,93 colocando a la muestra como muy leptocúrtica lo que implica que la mayoría de los datos se agrupan alrededor de 20,13 (Media). En la figura 4 se muestra el comportamiento de la variable edad.

Tabla 2

Estadísticos descriptivos de la Edad de la muestra

Media	20,13
Desviación estándar	2,27
Coeficiente de variación	11,27%
Asimetría	3,016
Curtosis	24,93
Mínimo	17
Máximo	44

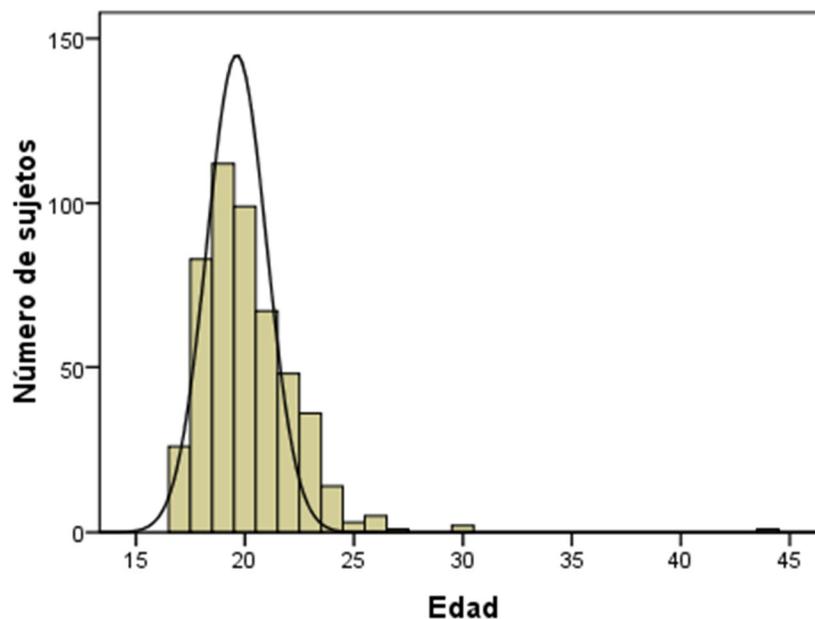


Figura 4. Histograma de frecuencia de la edad de los sujetos de la muestra

La muestra estuvo conformada por estudiantes de las carreras de Psicología, Derecho, Educación, Comunicación Social, Ingeniería, Ciencias Sociales, Contaduría, Economía, Administración y Letras. En la tabla 3 se muestra el número de estudiantes que participaron en la investigación y el porcentaje de estudiantes de las distintas escuelas que conformaron la muestra total, estos porcentajes se ven representados en el gráfico de barras de la figura 5.

Tabla 3

Número y porcentaje de estudiantes por escuela

Carrera	Frecuencia	Porcentaje
Psicología	111	22,2%
Derecho	30	6,0%
Educación	20	4,0%
Comunicación Social	38	7,6%
Ingeniería	163	32,6%
Ciencias Sociales	34	6,8%
Contaduría	37	7,4%
Economía	14	2,8%
Administración	50	10,0%
Letras	2	0,4%
Casos perdidos	1	0,2%
Total	500	100%

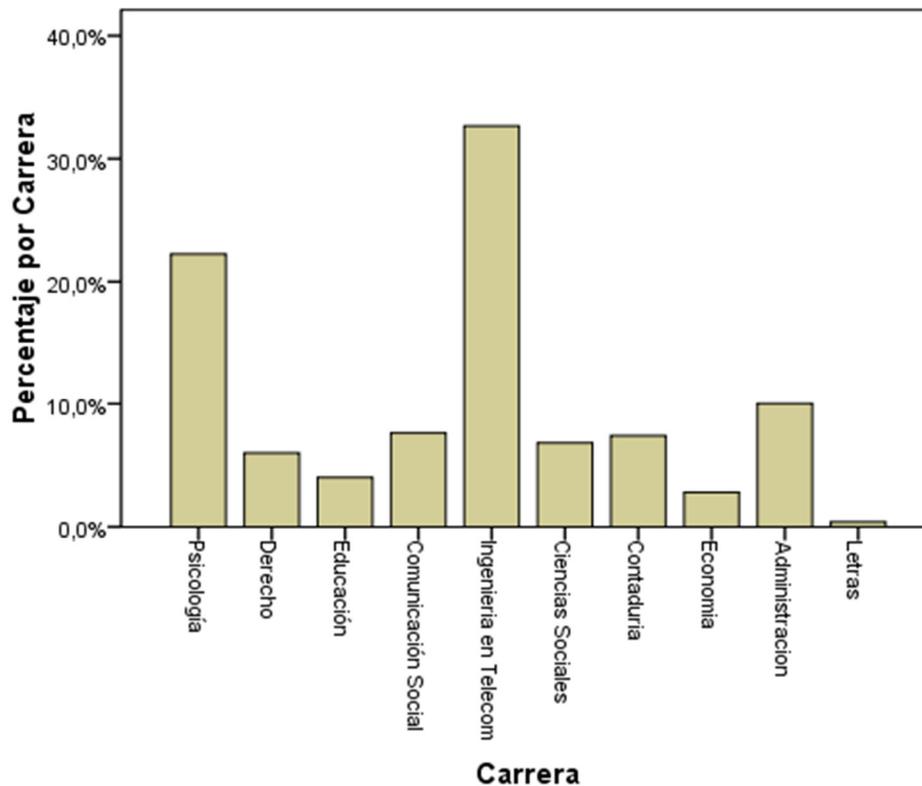


Figura 5. Gráfico de Barras que muestra el porcentaje de sujetos por carrera

En cuanto al semestre que cursaban los estudiantes que participaron en el estudio se observa en la figura 6 que existen estudiantes de todos los periodos de las carreras pero existe una mayor presencia de estudiantes del 4 semestre (32%), del segundo semestre (26%) y del sexto semestre (10,8%), el resto de los estudiantes se repartieron en el resto de los semestres ninguno de ellos teniendo un porcentaje mayor al 10 % de la muestra. Es importante aclarar que las carreras de régimen anual se consideraron con su semestre equivalente al periodo en que fueron administradas las escalas a la muestra y como la administración de las escalas se llevó a cabo en el segundo semestre del presente periodo académico los estudiantes que cursaban primer año se consideró en su equivalente que sería segundo semestre, los de segundo año en cuarto semestre, los de tercer año en sexto semestre, los de cuarto año en octavo semestre y los de quinto año en decimo semestre.

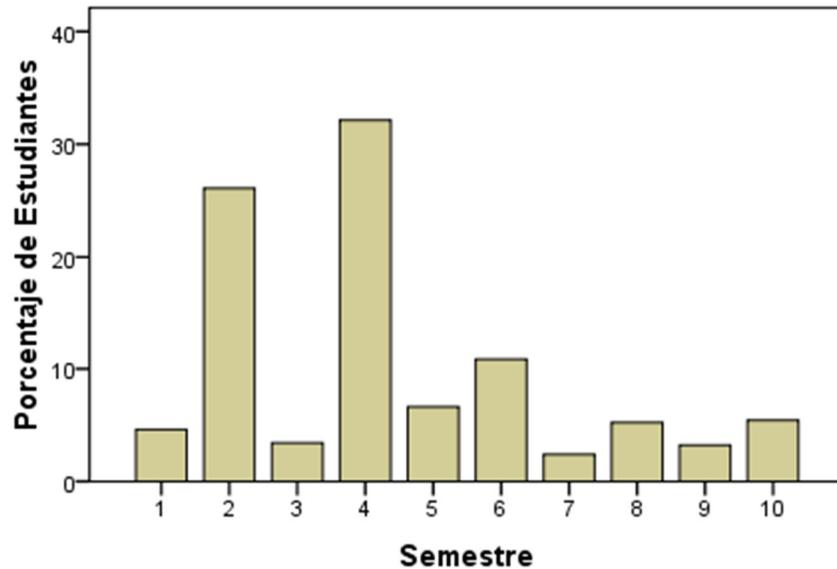


Figura 6. Gráfico de Barras que muestra el porcentaje de sujetos por semestre

Puntajes totales de la escala de depresión (TCT)

Los puntajes totales obtenidos por los sujetos en la escala se formaron a partir de la sumatoria de los ítems dicotómicos, siendo 1 cuando los sujetos consideraban como verdadera la afirmación y 0 cuando consideraban falsa la afirmación. En este sentido el puntaje máximo obtenido por la sumatoria de los ítems es de 62 que implica un nivel alto de depresión y sería la puntuación mínima 0 implicando niveles bajos del constructo depresión.

Los resultados obtenidos se muestran en la tabla 4 donde se observa que la media de los puntajes obtenidos fue de 14,22, con una desviación estándar de 10,48 y un coeficiente de variación de 73,69% lo que indica una distribución muy heterogénea de los datos, una asimetría de 0,902 que nos muestra que la mayoría de los datos se agrupa hacia los valores inferiores. Una curtosis de 0,146 implicando que la distribución es levemente leptocúrtica. Los puntajes de los sujetos oscilaron de un valor mínimo del recorrido de 0 y aun valor máximo de 50.

En la figura 7 se muestra la distribución de los puntajes por medio de un histograma.

Tabla 4

Estadísticos descriptivos para la escala total de Depresión

Media	14,22
Desviación estándar	10,48
Coefficiente de variación	73,69%,
Asimetría	0,902
Curtosis	0,146
Mínimo	0
Máximo	50

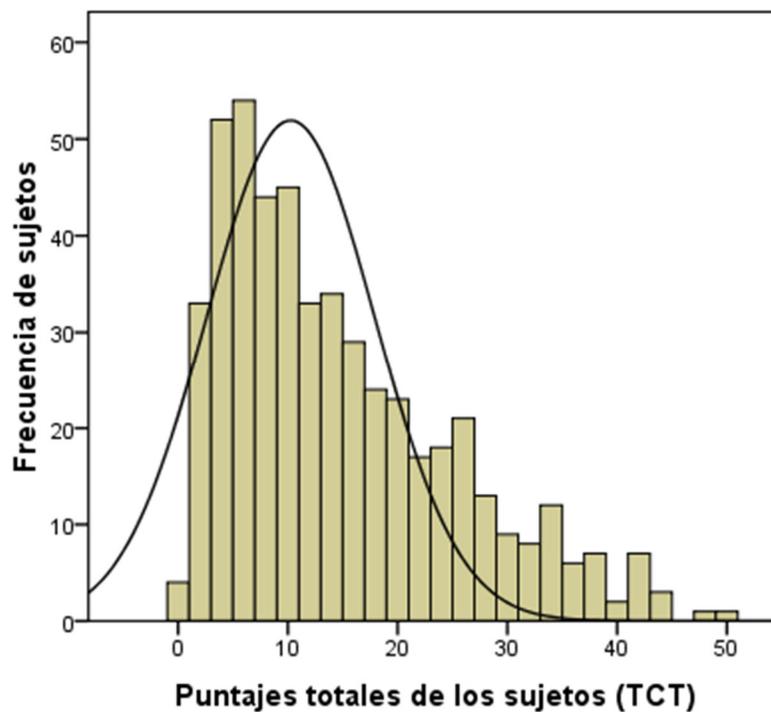


Figura 7. Histograma de frecuencia de los puntajes obtenidos en la escala

En cuanto a la normalidad en los puntajes, se observa que la distribución no se ajusta a la curva normal tomando en cuenta que se obtuvo en la prueba de

normalidad Kolmogorov-Smirnov igual a 2,698 (Sig.=0,00) y en el gráfico de normalidad P-P presente en la figura 8.

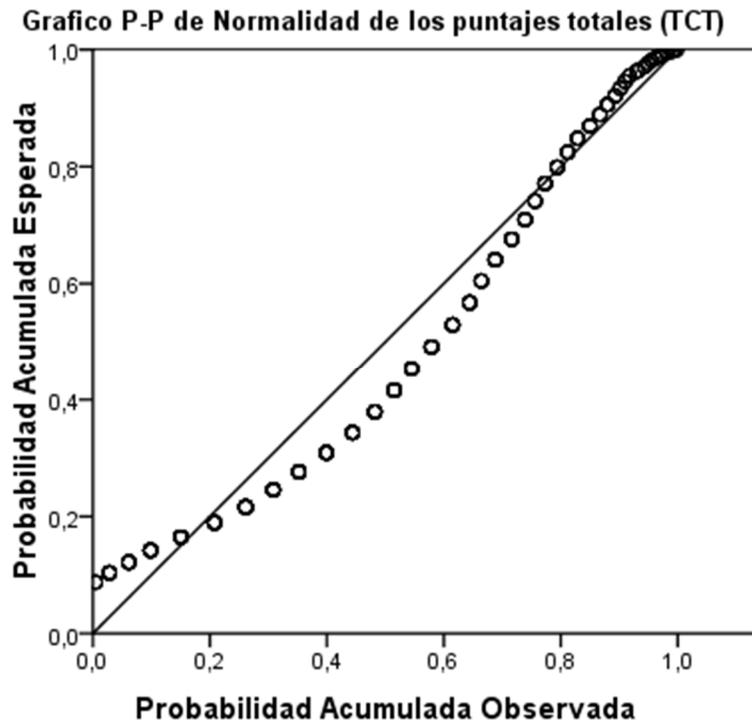


Figura 8. Grafico P-P de normalidad de los puntajes totales de los sujetos en la escala

Para el cálculo de la confiabilidad basándonos en la metodología de TCT se utilizó como indicador el coeficiente de consistencia interna alfa de Crombach y se obtuvo como resultado un coeficiente alfa de 0,927 lo que implica que la escala tiene un nivel alto de consistencia interna, lo que la hace confiable. Los resultados se observan en la tabla 5.

Tabla 5

Coeficiente alfa de consistencia interna, casos válidos y excluidos para su cálculo

Alfa de Cronbach	0,927
Casos Validos	436
Casos Excluidos	64
Casos Totales	500

Nota. Los casos excluidos son aquellos sujetos que dejaron de responder algún ítem.

Análisis factorial

Para observar el comportamiento de la escala en cuanto a su estructura se realizó un análisis de componentes principales utilizando el método de rotación ortogonal Varimax para comprobar la estructura factorial que subyace al test.

En primer lugar se obtuvo un coeficiente Kaiser-Mayer-Olking (KMO) de 0,889 lo que implica que existe una buena adecuación de la muestra para poder realizar el análisis factorial. En cuanto a la prueba de Esfericidad de Barlett arrojó un Chi-cuadrado de 10086,869 con una significancia de 0,00. Los resultados tanto en el KMO como en la prueba de esfericidad de Barlett indican que la muestra es adecuada para realizar el análisis factorial, y que las correlaciones entre los ítems y los factores son significativas por lo que se puede continuar con el análisis de la estructura factorial del instrumento.

Tomando en cuenta el criterio de normalización de Kaiser (1960) donde se deben tomar como factores todos aquellos componentes que superen el valor crítico de 1,5 autovalores, en los resultados se observa que existen 8 componentes que superan este puntaje como se puede observar en el Anexo B.1. Pero el factor 1 explica un 21,16% de la varianza que es la que supera a la suma de los siguientes 6 componentes que arroja el análisis.

Si también tomamos en cuenta la prueba de Sedimentación de Cattell (1966) que expresa la cantidad de autovalores (Eje vertical y) explicados por cada factor (Eje horizontal x), que se observa en la figura 9, se puede observar claramente como existe un factor principal que explica la mayoría de la varianza del test de acuerdo a los resultados obtenidos. Además podemos observar en la matriz rotada (Anexo B.2) que la mayor cantidad de ítems cargan en el factor 1. De acuerdo a estos criterios se puede determinar que la estructura del test es unifactorial.

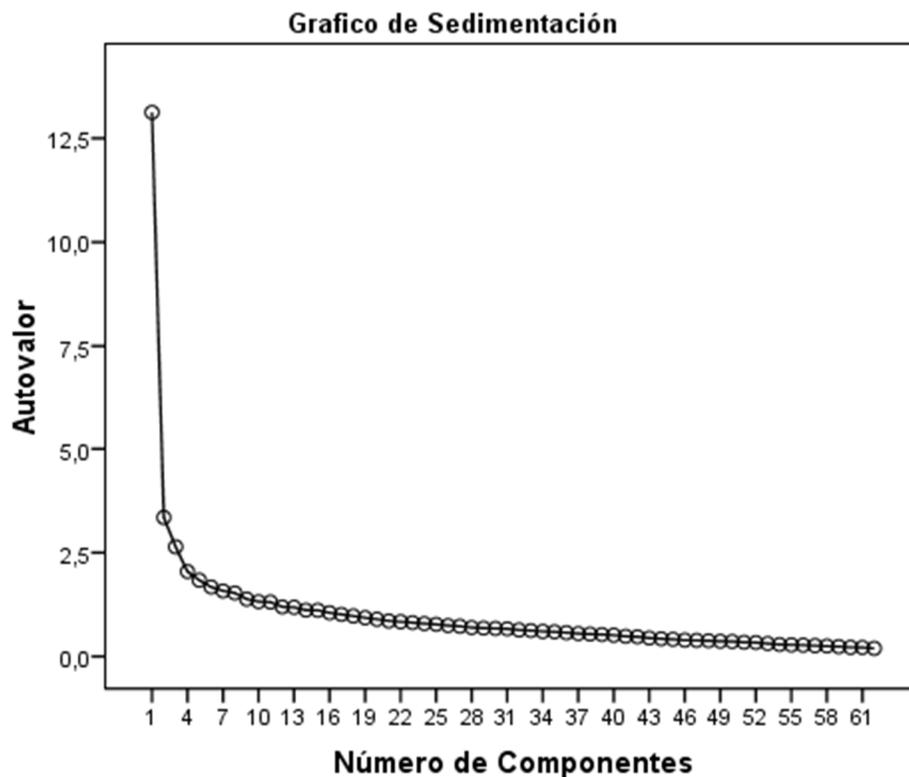


Figura 9. Gráfico de sedimentación de Cattell

Posición del ítem en el continuo TCT

Los ítems de la escala se puntuaron de manera dicotómica donde se adjudicó el valor de 1 cuando los sujetos respondían de manera afirmativa al ítem y 0 cuando lo hacían de manera negativa. En el Anexo C se muestra la tabla que indica el comportamiento de los ítems en cuanto a su media, varianza, desviación estándar, correlación con el puntaje total en el test y la variación del monto de la confiabilidad si el ítem es eliminado. En este último indicador se observa que todos los ítems parecen aportar a la confiabilidad ya que el monto de la misma no aumenta cuando el reactivo es eliminado, por lo que se puede determinar que todos los ítems contribuyen a la confiabilidad del instrumento que como se indicó anteriormente es de $\alpha=0,927$, pero debido a la alta cantidad de ítems y el nivel de confiabilidad que posee el instrumento no parece ser un indicador lo

suficientemente sensible para observar el aporte del ítem a la calidad del instrumento.

Ahora bien si tomamos en cuenta la correlación del ítem con el puntaje total del test observamos en la tabla 6 que existen 7 ítems que correlacionan de manera muy baja con el instrumento, lo que indica que éstos ítems tienen un aporte muy pobre al comportamiento del instrumento.

Tabla 6

Ítems con baja correlación con los puntajes totales del test

Ítems		Correlación ítem-test
2	Estoy a dieta para adelgazar	0,060
6	Las cosas que antes solían irritarme me importan poco	0,073
15	Me suicidaría si tuviese la oportunidad	0,101
26	Estoy durmiendo más que antes	0,011
30	He perdido totalmente el interés por el sexo	0,106
32	Estoy comiendo menos que antes	0,079
54	Sobreexcitado	-0,038

Nota. El criterio para definir los ítems con poco aporte fue $r_{xy} < 0,11$; $\alpha = 0,01$, para $n = 500$

Anastasi y Urbina (1998) y Magnusson (1969) en relación al indicador de la posición del ítem en el continuo llamado también índice de dificultad del ítem en la TCT, se define como el porcentaje o proporción de personas que contestan correctamente el ítem. Y mientras más fácil sea el reactivo o menos cantidad del constructo mida mayor será la proporción de personas que lo respondan de manera afirmativa. En la tabla 7 se muestran los índices de dificultad, expresados en porcentaje, obtenidos para los ítems que conforman la escala.

Tabla 7

Índice de dificultad de los ítems de acuerdo a la TCT

Ítem	Índice de Dificultad						
1	7,8%	17	11,2%	33	7,4%	49	30%
2	23,6%	18	6,2%	34	24,7%	50	48,8%
3	25,5%	19	9%	35	28,6%	51	24,7%
4	8,8%	20	9,4%	36	18,2%	52	26,3%
5	10,2%	21	57,1%	37	24,8%	53	31,7%
6	46,4%	22	10,2%	38	30,9%	54	12,5%
7	41,2%	23	15,4%	39	38,7%	55	28,5%
8	22,4%	24	0,8%	40	32,9%	56	20,1%
9	3,2%	25	2%	41	10,5%	57	16%
10	12,8%	26	16,2%	42	17,6%	58	44,9%
11	25,7%	27	16%	43	26,1%	59	14,5%
12	4,4%	28	11,4%	44	41,7%	60	53,2%
13	52,8%	29	8,6%	45	48,6%	61	40%
14	3,2%	30	2,8%	46	42,6%	62	21,9%
15	1,2%	31	7%	47	65,7%		
16	25,4%	32	21,2%	48	45,1%		

Supuesto de invarianza (TCT)

En la tabla 8 se muestran los resultados de las correlaciones 45 correlaciones obtenidas al relacionar las 10 submuestras aleatorias que fueron seleccionadas. Se puede observar que las correlaciones oscilan desde 0,829 hasta 0,988, teniendo como media 0,882 y desviación de 0,039. Los 62 ítems fueron incluidos en la totalidad de las correlaciones. En el anexo D.1 se muestran las correlaciones totales y los casos válidos para cada una de ellas.

Tabla 8

Correlaciones entre las distintas submuestras en relación a la posición del ítem en el continuo (Índice de Dificultad)

Muestras Correlacionadas	r_{xy}	Muestras Correlacionadas	r_{xy}
1-2	,987	3-10	,834
1-3	,844	4-5	,863
1-4	,855	4-6	,893
1-5	,853	4-7	,844
1-6	,970	4-8	,950
1-7	,870	4-9	,929
1-8	,876	4-10	,866
1-9	,900	5-6	,886
1-10	,869	5-7	,841
2-3	,843	5-8	,887
2-4	,843	5-9	,867
2-5	,850	5-10	,988
2-6	,960	6-7	,869
2-7	,855	6-8	,904
2-8	,850	6-9	,899
2-9	,896	6-10	,902
2-10	,860	7-8	,904
3-4	,851	7-9	,871
3-5	,840	7-10	,829
3-6	,869	8-9	,891
3-7	,923	8-10	,896
3-8	,876	9-10	,851
3-9	,877	Media	,882

Ajuste del modelo de Rasch

Para estimar la probabilidad de poder analizar este instrumento por medio del modelo de Rasch, es necesario previamente verificar si los datos se ajustan a este modelo (De Llano, 2005).

Bond y Fox (2007), Prieto y Delgado (2003) y Muñiz (1997) plantean como criterios viables para medir el ajuste de los datos al modelo el análisis de la suma de los residuales estandarizados (Outfit) y la suma de los residuales ponderados (Infit) así como las puntuaciones medias, donde se establece que si los valores Infit u Outfit superan el valor de 1,3 para muestras menores de 500 sujetos implica que no existe ajuste del modelo ni de los ítems ni de la muestra. Además de acuerdo a Bond y Fox (2007) si los puntajes Infit y Outfit transformados a puntuaciones t normalizadas superan los valores de -2 y +2 (con un alfa de 0,05) ni las personas ni los ítems se ajustarían el modelo.

Tabla 9

Tabla de indicadores de ajuste global en puntajes ponderados y estándar

	Indicadores de Ajuste Global			
	Media	Desviación	Media t	Desviación t
Infit (Ítems)	0,99	0,18	-0,10	2,9
Outfit (Ítems)	1	0,61	-0,20	3
Infit (Personas)	1	0,18	0	1,1
Outfit (Personas)	0,99	0,70	0	1

De acuerdo a los resultados obtenidos que se muestran en la tabla 9 en general existe un ajuste global al modelo tanto para los ítems como para las personas ya que la media de las sumatorias Infit y Outfit no superan el valor criterio de 1,3, ni superan el valor crítico de puntajes t mayores a -2 y +2.

En la tabla 10 se pueden observar los índices de ajuste individuales de cada uno de los ítems, para verificar cuales son los que se ajustan o no al modelo de Rasch. En la tabla 11 se muestra cuáles fueron los 30 ítems que no se ajustaron a el modelo y cuál fue el criterio de desajuste.

Tabla 10

Tabla de indicadores de ajuste de los ítems en puntajes ponderados y estándar

Ítem	Infit	Outfit	Infit t	Outfit t	Ítem	Infit	Outfit	Infit t	Outfit t
1	.91	.65	-.7	-1.2	32	1.38*	1.72*	5.2^	3.9^
2	1.48*	2.24*	6.7^	6.6^	33	.84	.42	-1.2	-2.3^
3	.99	.85	-.2	-1.2	34	.98	.85	-.2	-1.1
4	1.06	.88	.6	-.3	35	.90	.74	-1.8	-2.4^
5	.98	.89	-.1	-.3	36	.88	.65	-1.7	-2.2^
6	1.53*	1.89*	9.9^	9.3^	37	1.01	.98	.2	-.1
7	1.46*	1.85*	8.7^	8.4^	38	.87	.77	-2.4^	-2.2^
8	1.02	.93	.4	-.4	39	.85	.77	-3.2^	-2.8^
9	.90	.41	-.4	-1.6	40	.92	.93	-1.6	-.7
10	.94	.82	-.6	-.7	41	1.01	1.41*	.1	1.4
11	1.04	.99	.7	.0	42	.80	.57	-2.8^	-2.7^
12	.91	.47	-.4	-1.6	43	.96	.89	-.7	-.8
13	1.07	1.12	1.6	1.5	44	.83	.75	-3.7^	-3.3^
14	.85	.32	-.7	-2.1^	45	.91	.85	-2.0	-2.1^
15	.98	.69	.1	-.5	46	1.14	1.29	3.0^	3.3^
16	.92	.78	-1.3	-1.7	47	.93	.92	-1.4	-.8
17	1.18	1.49*	1.7	1.8	48	.89	.80	-2.5^	-2.8^
18	.92	.52	-.5	-1.6	49	.75	.61	-4.9^	-4.0^
19	.87	.74	-1.1	-.9	50	1.01	.99	.2	-.1
20	1.03	1.10	.3	.4	51	.91	.82	-1.6	-1.3
21	1.27	1.53*	5.5^	5.7^	52	.75	.57	-4.5^	-3.9^
22	.86	.51	-1.3	-2.1	53	.72	.58	-5.8^	-4.6^
23	.86	.62	-1.8	-2.1	54	1.44*	4.11*	4.1^	7.5^
24	1.01	.49	.2	-1.0	55	.82	.65	-3.3^	-3.3^
25	.94	2.37*	-.1	2.2^	56	.90	.78	-1.5	-1.4
26	1.44*	2.63*	4.9^	5.9^	57	.83	.59	-2.2^	-2.4^
27	.97	.76	-.3	-1.2	58	1.00	1.07	.0	.9
28	1.02	.97	.2	.0	59	.79	.46	-2.6^	-3.1^
29	.84	.86	-1.4	-.4	60	.90	.89	-2.2^	-1.4
30	1.04	1.23	.2	.6	61	1.15	1.17	3.1^	1.9
31	1.03	.88	.3	-.3	62	1.03	.91	.5	-.6

Nota. Serán señalizados con un “” los que superen el criterio de desajuste infit u Outfit >1,3 y con el símbolo “^” los que superen el valor t > -2 o +2.*

Tabla 11

Ítems que no se ajustan al modelo de Rasch y cuál fue el criterio que violaron

Ítems que no se Ajustan al modelo		Supera el valor Crítico
2	Estoy a dieta para adelgazar	Infit, Outfit y t
6	Las cosas que antes solían irritarme me importan poco	Infit, Outfit y t
7	Estoy comiendo más que antes	Infit, Outfit y t
14	Me detesto	T
17	Antes era capaz de llorar, pero ahora no puedo, incluso aunque quiera	Outfit
21	Estoy durmiendo menos que antes	Infit, Outfit y t
25	Me siento una persona tan desgraciada que no puedo soportarlo	Infit, Outfit y t
26	Estoy durmiendo más que antes	Infit, Outfit y t
32	Estoy comiendo menos que antes	Infit, Outfit y t
33	Todo me sale mal, no tengo suerte con nada	T
35	Tranquilo/a	T
36	Oprimido	T
38	Entusiasmado	T
39	Confiado	T
41	Desasosegado	Outfit
42	Alegre	T
44	Satisfecho	T
45	Angustia	T
46	Inspirad	T
48	Relajado	T
49	Seguro	T
52	Cómodo	T
53	Animado	T
54	Sobreexcitado	Infit, Outfit y t
55	Motivado	T
57	Vacio	Infit, Outfit y t
59	Bien	T
60	Tenso	T
61	Agitado	T

Puntajes en el parámetro θ (Rasch)

Los puntajes de los sujetos en relación al parámetro depresión (parámetro θ) son representados en una escala logarítmica que oscila de $-\infty$ a $+\infty$, mientras mayor sea el monto que posean los sujetos del parámetro, en este caso depresión, mayor será su puntaje en la escala logarítmica ($+\infty$). Los resultados se muestran en la tabla 12 donde se observa que los puntajes en el parámetro oscilan desde -6,13 a 1,92, siendo la media de esta distribución -1,96 lo que implica que la mayoría de los sujetos de esta población poseen poca cantidad del parámetro depresión. La muestra es muy heterogénea (71%) y levemente coleada hacia los valores inferiores ($As=-0,13$). En la figura 10 se muestra el comportamiento de los resultados en un histograma.

Tabla 12

Estadísticos descriptivos del parámetro Depresión en escala logit.

Media	-1,96
Desviación estándar	1,4
Coefficiente de variación	71%
Asimetría	-0,13
Curtosis	-0.01
Mínimo	-6,13
Máximo	1,92

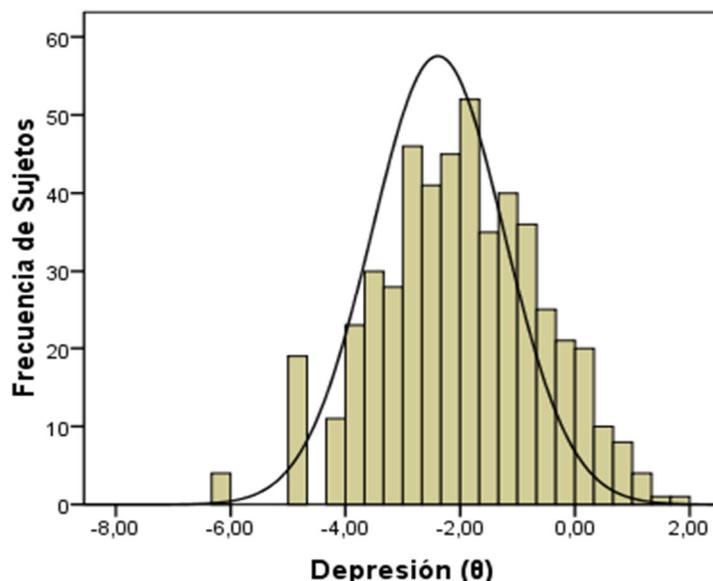


Figura 10. Histograma de frecuencia de los puntajes de los sujetos en el parámetro θ

Parámetro b (Rasch)

Los índices de dificultad se presentan en una escala Logit donde se observa cuáles son los ítems que miden mayor o menor cantidad del parámetro, en la tabla 13 se pueden observar los resultados descriptivos donde la media del índice de dificultad b es de -0,001 con un recorrido que va desde -2,84 a 3,75. Hay una leve asimetría positiva de los datos que implican que puede que haya mayor cantidad de ítems que puntúen por debajo de la media del parámetro. En la tabla 14 se observan los resultados detallados de cada uno de los ítems en cuanto a su media y desviación en relación a el parámetro depresión y en la figura 11 se ven representado el comportamiento de los ítems.

Tabla 13

Estadísticos descriptivos del índice de dificultad b en escala logit

Media	-0,001
Desviación estándar	1,42
Coefficiente de variación	0,07%
Asimetría	0,429
Curtosis	-0,019
Mínimo	-2,84
Máximo	3,75

Tabla 14

Media y desviación de cada uno de los ítems en relación al índice de dificultad b

Ítem	Dificultad (Logit)	Error de Estimación	Ítem	Dificultad (Logit)	Error de Estimación
1	1.21	.18	32*	-.19	.12
2*	-.36	.12	33*	1.29	.18
3	-.49	.12	34	-.44	.12
4	1.08	.17	35*	-.71	.11
5	.88	.16	36*	.07	.13
6*	-1.75	.11	37	-.45	.12
7*	-1.46	.11	38*	-.84	.11
8	-.29	.12	39*	-1.32	.11
9	2.28	.26	40	-.97	.11
10	.57	.15	41*	.87	.16
11	-.51	.12	42*	.11	.13
12	1.92	.23	43	-.53	.12
13	-2.10	.10	44*	-1.48	.11
14*	2.28	.26	45*	-1.86	.11
15	3.33	.42	46*	-1.53	.11
16	-.49	.12	47	-2.84	.11
17*	.76	.16	48*	-1.67	.11
18	1.51	.20	49*	-.78	.11
19	1.05	.17	50	-1.87	.11
20	.99	.17	51	-.44	.12
21*	-2.34	.11	52*	-.55	.12
22	.88	.16	53*	-.89	.11
23	.31	.14	54*	.62	.15
24	3.75	.51	55*	-.69	.12
25*	2.79	.33	56	-.09	.13
26*	.23	.14	57*	.26	.14
27	.25	.14	58	-1.66	.11
28	.73	.15	59*	.41	.14
29	1.10	.17	60*	-2.11	.11
30	2.41	.28	61*	-1.38	.11
31	1.36	.19	62	-.23	.12

Nota. Serán señalados con un “” los ítems que no se ajustan a el modelo de Rasch*

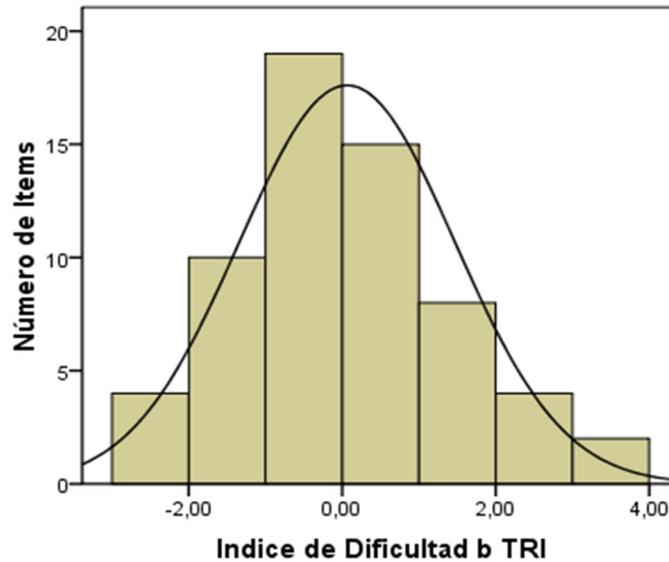


Figura 11. Histograma de frecuencia de los puntajes de los ítems en relación al parámetro θ

Una de las importantes ventajas del modelo de Rasch es que tanto el puntaje de los sujetos como el comportamiento de los ítems en relación al parámetro a medir, en este caso Depresión, se encuentran en la misma escala logit por lo que se puede observar en un mismo gráfico como se muestra en la figura 12 el comportamiento de los sujetos y de los ítems en relación a el parámetro para un mayor valor descriptivo.

Tomando en cuenta este gráfico se puede observar que la mayoría de las personas que participaron en la muestra poseen valores muy bajos en el parámetro Depresión lo que concuerda con lo esperado ya que la muestra es no clínica. De hecho se observa que hay un número importante de personas que no son medidas de manera adecuada por el instrumento ya que no hay ningún ítem que las pueda discriminar en relación a el parámetro, esto se puede observar en que los puntajes de los sujetos tienen un límite inferior de -6,13 y el de los ítems es de -2,84, lo que implica que todos los sujetos que se encuentren por debajo de este valor crítico no están siendo medidos de manera adecuada por alguno de los ítems que conforman la escala de depresión.

Por su parte los ítems parecen repartirse a lo largo de todo el parámetro (Min= -2,84 y Max= 3,75) con una media de -0,001, lo que nos muestra que existen ítems que midan baja, media y alta cantidad del parámetro. Pero también gracias a este gráfico se puede observar en la figura 12 que existen ítems que no parecen tener suficientes sujetos como para calibrarse completamente como por ejemplo suicidio (1), suicidio (2) y desgraciado. Lo que podría indicar que estos ítems no parecen ser adecuados para esta muestra pero sí podrían serlo para una muestra clínica o que podría existir un problema con la formulación de las preguntas.

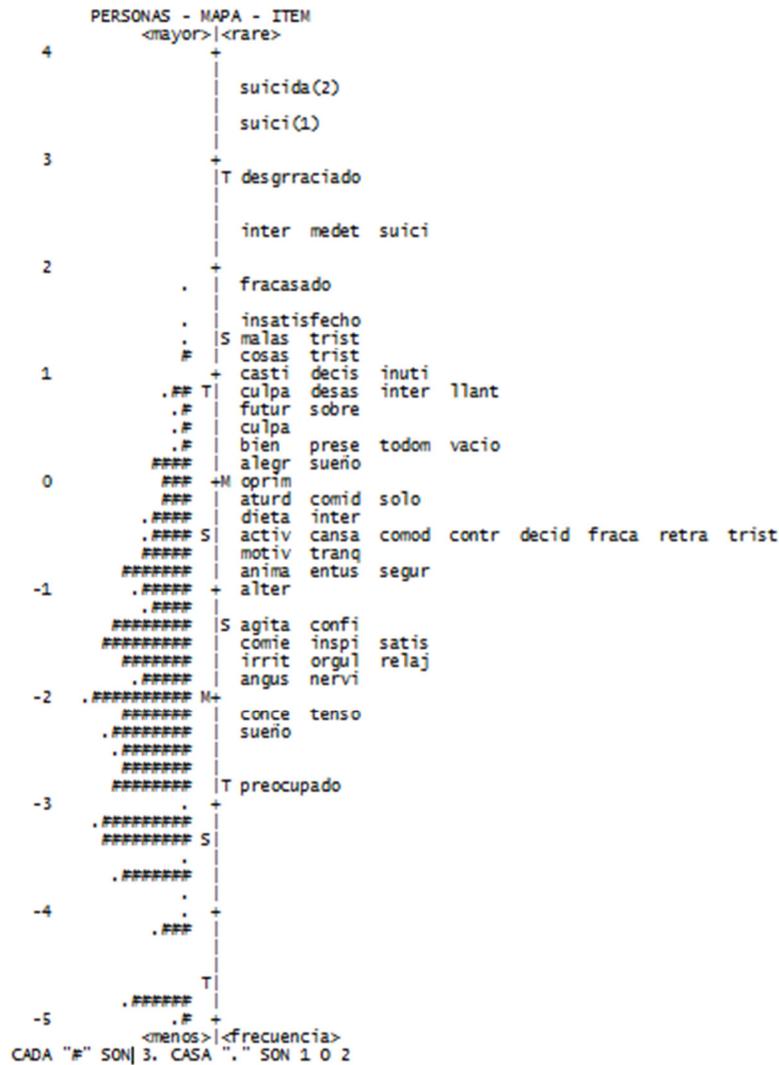


Figura 12. Mapas de ítems y de personas en relación al parámetro depresión. Del lado izquierdo se muestran los sujetos representados por # o ". Según la frecuencia. Del lado derecho se muestran los ítems representados por la etiqueta verbal correspondiente.

Supuesto de invarianza (Rasch)

En la tabla 15 se muestran los resultados de las correlaciones 45 correlaciones obtenidas al relacionar las 10 submuestras aleatorias que fueron seleccionadas. Se puede observar que las correlaciones oscilan desde 0,829 hasta 0,988, teniendo como media 0,826 y desviación de 0,061. Hay que destacar

que como se observa en el anexo D2 de las salidas de las correlaciones de los parámetros de los ítems, que existen valores ausentes ya que existieron ítems que no fueron incluidos ya que presentaban varianza 0 lo que lo excluye de la correlación y que puede incidir en los resultados obtenidos.

Tabla 15

Correlaciones entre las 10 submuestras en cuanto al parámetro b

Muestras Correlacionadas	r_{xy}	Muestras Correlacionadas	r_{xy}
1-2	,988	3-10	,802
1-3	,767	4-5	,820
1-4	,746	4-6	,807
1-5	,801	4-7	,756
1-6	,956	4-8	,925
1-7	,781	4-9	,879
1-8	,810	4-10	,814
1-9	,815	5-6	,846
1-10	,834	5-7	,809
2-3	,763	5-8	,883
2-4	,729	5-9	,799
2-5	,796	5-10	,980
2-6	,946	6-7	,788
2-7	,766	6-8	,845
2-8	,786	6-9	,811
2-9	,805	6-10	,867
2-10	,827	7-8	,872
3-4	,763	7-9	,821
3-5	,812	7-10	,795
3-6	,785	8-9	,815
3-7	,896	8-10	,892
3-8	,848	9-10	0,773
3-9	,789	Media	0,827

Comparación de los puntajes de los sujetos obtenidos por TCT y Rasch

Siguiendo el procedimiento utilizado por Fan (1998) de comparación de ambas metodologías se correlacionaron los puntajes obtenidos por los sujetos sumando cada uno de los ítems en la escala (TCT) y los puntajes de los sujetos en de acuerdo a el parámetro θ Depresión en escala logit. Como se puede observar en la tabla 16 el monto de la correlación es alta ($r_{xy}=0,952$) y significativo con un alfa de 0,05 y 0,01. En la figura 13 se muestra el grafico de dispersión donde se observa la tendencia y el grado de la relación entre ambos índices.

Tabla 16

Correlación entre los puntajes obtenidos por medio de la TCT y del parámetro de los sujetos de acuerdo al modelo de Rasch

		PuntajesTCT	PuntajesRasch
PuntajesTCT	Correlación de Pearson	1	,952**
	Sig. (2-colas)		,000
	N	500	500
PuntajesRasch	Correlación de Pearson	,952**	1
	Sig. (2-colas)	,000	
	N	500	500

*Nota. ** Correlación significativa a un alfa de 0.01 (2-colas).*

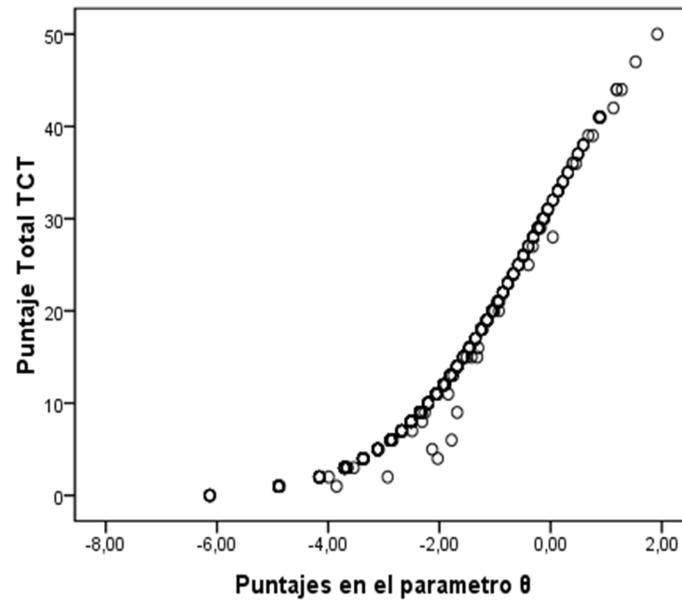


Figura 13. Gráfico de dispersión que muestra la asociación entre Puntajes totales de los sujetos en la TCT y los valores que obtuvieron en el parámetro θ TRI

Comparación de los parámetros de los ítems

El índice de dificultad de la TCT que es la proporción de personas que responden de manera afirmativa el ítem, se encuentra en una escala ordinal. Es necesario poder transformarlo a una escala continua de acuerdo a lo planteado por Fan (1998). Este autor propone que se deben transformar estos puntajes a valores estándar de la distribución Z de Fisher. En la tabla 17 se muestran estos puntajes de índice de dificultad transformados a los puntajes Z. Esta transformación permite poder comparar de modo correlacional los índices de dificultad de la TCT con el índice de dificultad b que brinda la metodología de Rasch.

En la tabla 18 se muestra que el monto de la correlación obtenida entre ambos puntajes es muy alta ($r_{xy}=0,949$) con un alfa de 0,05. En la figura 14 se muestra un dispersigrama donde se observa la tendencia y el grado de la relación entre ambos índices. Se observa que existe una correlación alta y positiva lo que

implica que los puntajes obtenidos por medio de ambas metodologías están muy relacionados.

Tabla17

Posición de los ítems de acuerdo a la TCT transformados a puntajes Z

Ítem	Índice de Dificultad						
1	0,96	17	0,75	33	0,98	49	-0,43
2	-0,03	18	1,06	34	-0,10	50	-1,60
3	-0,15	19	0,88	35	-0,34	51	-0,10
4	0,90	20	0,86	36	0,31	52	-0,20
5	0,81	21	-2,12	37	-0,10	53	-0,53
6	-1,45	22	0,81	38	-0,48	54	0,67
7	-1,12	23	0,48	39	-0,97	55	-0,33
8	0,05	24	1,40	40	-0,61	56	0,19
9	1,25	25	1,32	41	0,79	57	0,45
10	0,65	26	0,43	42	0,35	58	-1,36
11	-0,16	27	0,45	43	-0,18	59	0,54
12	1,17	28	0,73	44	-1,16	60	-1,87
13	-1,85	29	0,91	45	-1,59	61	-1,05
14	1,25	30	1,27	46	-1,21	62	0,08
15	1,37	31	1,01	47	-2,65		
16	-0,14	32	0,12	48	-1,37		

Tabla 18

Correlación entre los puntajes obtenidos por medio de la TCT y del parámetro de los ítems de acuerdo al modelo de Rasch

		Índice de Dificultad TCT	Parámetro b Rasch
Índice de Dificultad TCT	Correlación de Pearson	1	,949**
	Sig. (2-colas)		,000
	N	62	62
Índice de Dificultad b Rasch	Correlación de Pearson	,949**	1
	Sig. (2-colas)	,000	
	N	62	62

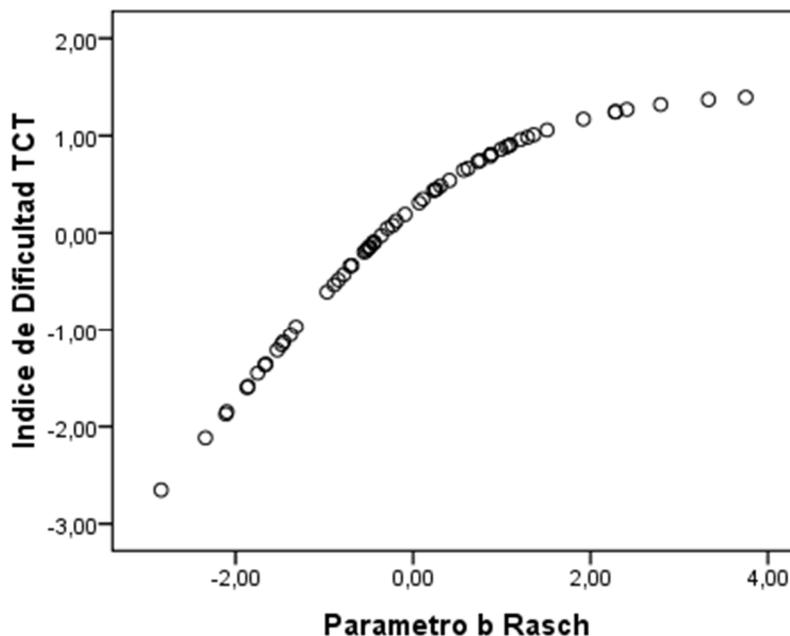


Figura 14. Gráfico de dispersión que muestra la asociación entre los índices de dificultad de la TCT y b del modelo de Rasch

Comparación de los índices de confiabilidad

Para comparar la medición de la calidad psicométrica del test se siguió la metodología expuesta por Silvestre (2009) donde por la TCT se tomó el coeficiente alfa de Cronbach y por el modelo de Rasch la confiabilidad por ítem y por persona. Si tomamos como criterio para que los valores sean confiables los índices de confiabilidad deben ser mayores a 0,60, como se puede observar en la tabla 19 tanto el alfa de Cronbach es de 0,927, la confiabilidad por persona del modelo de Rasch es de 0,88 y la confiabilidad por ítem es mayor a 0,98, lo que nos indica que de acuerdo a ambas metodologías de análisis el instrumento es confiable.

Tabla 19

Índices de confiabilidad Alfa de Cronbach (TCT) y confiabilidad por ítem y por persona (Rasch)

TCT	TRI	
Alfa de Cronbach	Confiabilidad por Persona	Confiabilidad por Ítem
0,927	0,88	0,98

Comparación del cumplimiento del supuesto de Invarianza de los parámetros de los ítems

Se puede apreciar según los resultados obtenidos que se muestran en la tabla 20 que la TCT parece mostrar mayor nivel de invarianza en relación a la TRI, ya que se observó una media mayor en las correlaciones de las posiciones de los ítems entre las distintas submuestras. Además se observa mayor dispersión en los puntajes de las correlaciones en la TRI ya que las correlaciones oscilan desde 0,729 hasta 0,988 (DE=0,061), en cambio el rango de correlaciones mostradas por la TCT es de menor recorrido ya que sus puntajes oscilan de 0,829 hasta 0,988 (DE=0,039).

Tabla 20

Estadísticos de las correlaciones entre las distintas submuestras en relación a la posición del ítem y al parámetro b

	TCT	TRI
Media	0,881	0,826
Desviación	0,039	0,061
Mínimo	0,829	0,729
Máximo	0,988	0,988

Ya que se tienen tanto la media como la desviación típica de las correlaciones se procedió a hacer un análisis adicional para comprobar si las diferencias entre las medias de estas correlaciones son significativas, por lo que se calculó una t de Student de grupos correlacionados lo que dio como resultado una $t=12,076$ con un nivel de significancia de 0,00, lo que implica que las diferencias entre las medias de ambas metodologías es estadísticamente significativa.

DISCUSIÓN DE RESULTADOS

Para poder analizar cabalmente éste instrumento por medio del modelo de Rasch, es necesario previamente verificar si los datos obtenidos se ajustan a este modelo (De Llano, 2005). Ósea hay que comprobar en qué grado los resultados los resultados predichos con esos valores coinciden con los obtenidos (Muñiz, 1997).

Tomando en cuenta los criterios de ajuste propuestos por Bond y Fox (2007), Prieto y Delgado (2003) y Muñiz (1997) en el presente estudio se obtuvo, como se muestra en la tabla 9, que en general existe un ajuste global al modelo tanto para los ítems como para las personas, lo que implica que los datos predichos por el modelo matemático no difieren estadísticamente de los puntajes obtenidos empíricamente en general de la escala total por lo que se puede analizar el instrumento de acuerdo al modelo de Rasch de un parámetro.

Como plantean Bond y Fox (2001) si existe ajuste de los resultados al modelo, la interacción entre el nivel del rasgo medido que posean los sujetos y el umbral de los ítems a lo largo del continuo del rasgo dan cuenta de toda la varianza explicada, y por lo tanto esto implica que el instrumento es unidimensional y progresivo. La expresión psicométrica y estadística de la depresión representada por el instrumento analizado está modelando adecuadamente el constructo depresión que a nivel teórico también es unidimensional y progresivo. El que la escala sea unidimensional también puede ser validado con los resultados obtenidos por medio del análisis factorial, que de acuerdo al criterio propuesto por Cattell (1966) presenta un factor que explica la mayor cantidad de varianza como se observa en la figura 9.

Ahora bien, si observamos los índices de ajuste individuales de cada uno de los ítems, para verificar cuales son los que se ajustan o no al modelo de Rasch. Se observa que de los 62 ítems existen 30 que no se ajustaron al modelo, lo que implica que existe una varianza extra que no es explicada por el constructo

depresión, por lo que se recomendaría realizar una selección definitiva de los ítems del instrumento que permitan estimar de manera adecuada el nivel de depresión que presenta un sujeto. Para esto se excluirían los ítems que no se ajustan, los cuales son señalados en la tabla 11. Si tomamos el criterio de exclusión de ítems propuesto por la TCT observando la correlación ítem-test podemos decir que existen siete ítems que se correlacionan de manera muy baja con el test por lo que no aportan al test gran utilidad. Estos resultados denotan que el modelo de Rasch es mucho más exigente con la calidad de la medida que la TCT.

A nivel descriptivo se puede observar que el modelo de Rasch tiene mayor profundidad de análisis ya que tiene la gran ventaja de tener la misma escala logit para los ítems y para las personas. Ya que gracias a esto se puede discriminar claramente los niveles de rasgo, como se observa en la figura 12. Se puede observar que la mayoría de las personas que participaron en la muestra poseen valores bajos en el parámetro Depresión lo que concuerda con lo esperado ya que la muestra es no clínica. De hecho se observa que hay un número importante de personas que no son medidas de manera adecuada por el instrumento ya que no hay ningún ítem que las pueda discriminar en relación a su nivel en el parámetro, esto se puede observar en que los puntajes de los sujetos tienen un límite inferior de -6,13 y el de los ítems es de -2,84, lo que implica que todos los sujetos que se encuentren por debajo de este valor crítico no están siendo medidos de manera adecuada por alguno de los ítems que conforman la escala de depresión, pero que puede estar determinado en gran medida por ser una muestra no clínica, mientras que el instrumento pareciera ser más apropiado para niveles clínicos.

Si bien el determinar que los resultados estimados por ambas metodologías en relación al constructo, el modelo de Rasch permite determinar cuáles sujetos no están siendo medidos de manera adecuada por los ítems, o que el instrumento no está siendo sensible para medir a esa población en relación al constructo depresión.

Otra ventaja que se muestra en la figura 12 es que existen ítems que no parecen tener suficientes sujetos como para calibrarse con precisión, lo que podría indicar que estos ítems no parecen ser adecuados para esta muestra pero sí podrían serlo para una muestra clínica o también podría existir un problema con la formulación de las preguntas.

Si se sigue el procedimiento de análisis de la TCT se puede identificar que los resultados en cuanto a los niveles de depresión tampoco representan a una muestra clínica, ya que la distribución de sus puntajes es claramente asimétrica, se pierde profundidad de análisis en cuanto a los ítems y las personas en relación a la depresión. Se observa también que a diferencia de la TCT en la que todos los ítems tienen el mismo peso a la hora de medir el constructo (pesa igual reportar problemas de sueño que una ideación suicida), el modelo de Rasch ofrece un potencial importante de desarrollo conceptual en el análisis de la ubicación de los ítems a lo largo del constructo medido de Depresión, lo que a fin de cuentas corresponde a los niveles de depresión que indican ciertos comportamientos y que pueden llegar a ser críticos al momento del diagnóstico. Tomando en cuenta estos criterios se observa la mayor profundidad de análisis que brinda el modelo de Rasch, lo que genera un gran aporte para la psicología clínica por la calidad del análisis de un constructo clínico que te puede brindar el modelo de Rasch.

El objetivo del presente estudio es comparar la metodología de análisis propuesta por la Teoría Clásica de los Test y el modelo de Rasch de manera empírica a través del análisis de los resultados obtenidos por estudiantes universitarios en una escala de depresión. El primer criterio de comparación fue el nivel de depresión que obtienen los estudiantes universitarios, estimado por la TCT y por la TRI por medio del modelo de Rasch de un parámetro

Hay que recordar que en la evaluación habilidades o aptitudes, la dimensión latente subyacente (θ) como un conjunto de ítem se llama habilidad, mientras que en la evaluación de la personalidad, describe la forma con que las personas responden a los ítem de un inventario, las características de estos ítem y el lugar

que cada individuo tiene en la dimensión subyacente de personalidad (Richaud, 2006).

En este sentido para obtener la dimensión latente de la personalidad, en este caso depresión, por medio de la TCT para Muñiz (1997) los puntajes de los sujetos se encuentran entre cero y la puntuación máxima que obtenga el sujeto en el test que en este caso es de 62 ya que se suman el total de respuesta afirmativas del sujeto para medir su nivel en el constructo. Y en el modelo de Rasch los puntajes se ubican desde -3 a +3 dentro de una escala logarítmica. En este modelo el valor del parámetro de cada persona (θ) es el resultado de la probabilidad de responder a un determinado ítem de un conjunto, siendo esto una función de dos componentes: los parámetros del ítem y un parámetro para la persona, que refleja dónde está localizada la misma en el atributo subyacente (Richaud, 2006).

En los resultados obtenidos se observa una alta similitud en los resultados obtenidos de los sujetos tanto en la TCT como en el modelo de Rasch, ya que se obtuvo un nivel de asociación bastante elevado, como se observa en la figura 13 del nivel que poseen las personas de depresión. Lo que demuestra un nivel de equivalencia entre ambas teorías al momento de estimar cual es el nivel que posee cada persona en la dimensión subyacente de la personalidad que se intenta medir por un instrumento.

Este resultado se corresponde con lo obtenido por Fan (1998) donde obtuvo que no existían diferencias significativas entre la estimación del nivel de habilidad que poseen los sujetos, según la metodología de análisis utilizada, ya que los puntajes de las correlaciones eran superiores a 0,90, concluyendo que ambas teorías son equivalentes en este aspecto. Lo mismo se obtuvo en otro estudio elaborado para comparar los parámetros de las personas, donde Progar y Socan (2008) correlacionaron los valores de unas personas en dos test de habilidades (matemática y ciencia) y obtuvieron correlaciones altas para los puntajes en ambas pruebas. En este sentido Revelle, et al. (2010) afirman que la mayoría de los estudios las correlaciones entre las estimación de la TRI y la TCT

son altas. Por su parte Silvestre (2009) si obtuvo diferencias significativas en la estimación de los puntajes de los sujetos por medio de ambas metodologías.

El segundo criterio de comparación fue la estimación propiedades de los ítems por medio de ambas metodologías, tomando como único indicador la posición del ítem en el continuo, estimado en la TCT por la proporción de sujetos que respondieron de manera afirmativa el ítem y a través del modelo de Rasch por el parámetro b que determina el umbral del ítem.

En los resultados obtenidos en el presente estudio existe un nivel de asociación bastante elevado de la estimación de los parámetros de los ítems por medio de ambas metodologías ($r_{xy} = 0,949$) como se observa en la figura 14. Estos resultados implican en primero lugar que ambas metodologías están siendo bastante equivalentes al momento de estimar las propiedades de los ítems que conforman el instrumento, en este caso específico están siendo equivalentes al momento de estimar el umbral del ítem.

Este resultado es consistente con lo obtenido por Fan (1998) que realizó una comparación empírica de los estadísticos de los ítems en cuanto a la dificultad de los mismos. Obtuvo altas correlaciones entre el modelo de un parámetro de Rasch y la TCT mayores a 0,95. Por su parte, Nukhet (2002) al comparar ambas metodologías en cuanto a el índice de dificultad también encontró un alto grado de asociación entre los índices obtenidos por la TCT y la TRI de un parámetro (Rasch), de dos parámetros (a y b) y la teoría de los tres parámetros (a , b y c) con correlaciones que oscilaban desde 0,89 hasta 0,98. También, Hernández (2009) observó un gran nivel de equivalencia entre ambas metodologías en lo que respecta a los índices de dificultad del test y a los índices de discriminación. Progar y Socan (2008) también realizaron otra comparación empírica de la TCT y la TRI. Para comparar los parámetros de los ítems se correlacionaron el índice de dificultad y de discriminación de la TCT con los parámetros a y b de la TRI. Y también obtuvieron correlaciones altas para los índices de dificultad ($r = 0,972$ y $r = 0,922$) y discriminación ($r = 0,935$ y $r = 0,831$).

Pero se obtuvieron resultados diferentes a los obtenidos por Silvestre (2009) que obtuvo diferencias significativas en los resultados obtenidos por las dos metodologías en cuanto a la determinación de la dificultad de los ítems.

El tercer criterio de comparación que se utilizó se refirió a la calidad psicométrica del test. Para esto se tomó, dentro de la TCT el coeficiente alfa de Cronbach y dentro del modelo de Rasch la confiabilidad por ítems y por personas. Se tomó un criterio de confiabilidad mayor a 0,60. Como se puede observar en la tabla 19 tanto el alfa de Cronbach es de 0,927, la confiabilidad por persona del modelo de Rasch es de 0,88 y la confiabilidad por ítem es mayor a 0,98, lo que además de indicar que de acuerdo a ambas metodologías de análisis el instrumento es confiable, resultan equivalentes los resultados que se pueden obtener a partir a ambas metodologías. Estos resultados son análogos a los obtenidos en los estudios de Silvestre (2009), Magno (2009) y Morales (2009), donde este último autor plantea basado en los resultados de la estimación de la calidad del test, que aunque existan marcadas diferencias teóricas entre los modelos, a nivel de resultados empíricos no se ven tales diferencias ya que se obtienen resultados similares entre ambas.

Pero el modelo de Rasch ofrece no solo índices de confiabilidad centrados en el instrumento, por lo que además de brindar un índice de confiabilidad del set de ítems que conformen un instrumento, también brinda un índice de confiabilidad de las personas que indica la replicabilidad del orden de las personas con respecto al continuo latente θ si le damos a esta muestra de personas otra serie de ítems que midan el mismo constructo (Bond y Fox ,2001). Esto implica que el modelo de Rasch ofrece otro índice más para el análisis del instrumento de medida lo que da mayor profundidad de análisis de los resultados que se pueden tener al analizar la calidad psicométrica de un instrumento.

El cuarto criterio de comparación empírica entre ambas metodologías fue el comparar cómo se comportan ambas metodologías en la comprobación del supuesto de invarianza. Se puede apreciar según los resultados obtenidos que la TCT parece mostrar mayor nivel de invarianza en relación a la TRI, ya que se

observó una media mayor en las correlaciones de las posiciones de los ítems entre las distintas submuestras. Estos resultados discrepan con los obtenidos por Fan (1998) donde obtuvo que ambas metodologías se mostraran equivalentes en cuanto al nivel de invarianza que muestra el índice de dificultad (TCT) y el parámetro b de Rasch.

Los resultados obtenidos también discrepan de lo encontrado por Adedoyin, et al. (2008) quienes se plantearon como objetivo encontrar si la dificultad de los ítems se muestra invariante a través de dos muestras distintas y también encontrar si la dificultad de los ítems varia de una muestra a otra dependiendo del tamaño de la muestra, utilizando la metodología de la TCT y la TRI. Y obtuvieron como resultado que la TCT no cumplió los criterios de invarianza entre muestras distintas de personas y en muestras con distinto tamaño ya que se observaron diferencias significativas en el contraste de los puntajes por medio de un ANOVA. Y el índice de dificultad calculado por la TRI se mostró invariante tras el contraste de los puntajes por medio del análisis de varianza, al comparar distintas muestras de personas y muestras con distintos tamaños.

Por su parte Progar y Socan (2008) concluyeron en cuanto la comprobación de la invarianza que existe alto de nivel de invarianza a través de los grupos en relación al índice de dificultad de los ítems y en los índices de discriminación, pero que en el estudio que los parámetros de los ítems de la TRI se muestran más invariantes a través de las muestras cuando existe un ajuste perfecto de los datos al modelo teórico.

Tomando en cuenta lo propuesto por Progar y Socan (2008) hay que destacar que en los resultados obtenidos no todos los ítems que conforman el instrumento tuvieron un ajuste perfecto a el modelo lo que se vio reflejado en que no todos los ítems en cada submuestras pudieran formar parte en el cálculo de la correlación tal como se muestra en el anexo D.2, por lo que puede que esto haya influido en que la TCT mostrara mayor nivel de invarianza que las propiedades del ítem estimadas por el modelo de Rasch.

Hay que destacar también que en los estudios de Progar y Socan (2008), Adedoyin, et al. (2008) y Fan (1998) utilizaron muestras superiores a las mil personas y en muchos casos sub-muestras que igualaban este número. En el presente estudio cada una de las 10 submuestras tenía un tamaño pequeño ($n= 30$) y que de acuerdo con Muñiz (1997) coloca en ventaja a la TCT en relación al modelo de Rasch cuando se trabaja con pocos sujetos, por lo que esto puede explicar los resultados favorables a la TCT en relación a la comprobación del supuesto de invarianza.

.CONCLUSIONES

De acuerdo a los resultados obtenidos se puede decir que en primer lugar que queda demostrado que el modelo de Rasch es un método totalmente válido para estimar la unidimensionalidad del instrumento que se está midiendo y brinda practicidad de análisis a la hora de evaluar los ítems.

Queda evidenciado que el modelo de Rasch es una metodología que permite evaluar constructos no solo de habilidad si no como en este caso medir rasgos de la personalidad. Lo que brinda nuevas posibilidades a la psicología clínica ya que cuenta con una metodología con mayor profundidad de análisis que le permite medir sus constructos como en este caso es el de la depresión.

Si bien de acuerdo a los resultados obtenidos y contrastados con otros estudios se puede concluir que la TCT y el modelo de Rasch resultan equivalentes en la estimación de los puntajes de las personas, el que en este último tanto los ítems como las personas se puedan analizar dentro de una misma escala, brinda una mayor profundidad en su análisis ya que se puede evaluar el comportamiento de los ítems en función de las personas y de las personas en función de los ítems que intentan medir su nivel de depresión. En el ámbito del análisis clínico es muy relevante poder ubicar comportamientos concretos de los pacientes dentro de diferentes niveles del rasgo medido, lo cual puede ser valioso tanto en la medición de estos rasgos, como en la valoración clínica.

Ambas metodologías de análisis parecen ser bastante equivalentes en relación a la estimación de la posición que ocupa cada ítem en relación al constructo que se está midiendo, ya que se encontraron un alto nivel de asociación entre el índice de dificultad y el parámetro b de Rasch y esto se ve secundado por los resultados obtenidos por Fan (1998), Nukhet (2002) y Progar y Socan (2008).

En lo que respecta a la calidad psicométrica del test se observa que también ambas metodologías parecen ser equivalentes de acuerdo a los resultados obtenidos, lo que también es consistente con los resultados obtenidos por Silvestre (2009), Magno (2009) y Morales (2009). Pero el modelo de Rasch brinda un índice de estimación de confiabilidad no centrado en la consistencia de los ítems del instrumento, sino centrado en la replicabilidad del puntaje de las personas medidas por un grupo diferente de ítems que midan el mismo constructo. Esto resulta en una ventaja del modelo de Rasch ya que, de nuevo brinda mayor profundidad en el análisis al contarse con otros indicadores.

En lo que respecta al nivel de invarianza que mostraron ambas metodologías a través de las distintas submuestras, tomando en cuenta los parámetros de los ítems, se puede concluir que con las muestras utilizadas la TCT resulta más invariante que el modelo de Rasch. Esto puede deberse a que, como lo plantea Muñiz (1997), la TCT tiene mejor rendimiento en muestras pequeñas lo que supone una ventaja en cuanto a la economía y practicidad por el número de sujetos que requiere para estimar de manera adecuada los parámetros de los ítems. Esto también puede validar lo propuesto por Progar y Socan (2008) que plantean que el modelo de Rasch puede presentar mayor nivel de invarianza mientras exista un ajuste del modelo por parte de todo los ítems, pero que en este caso este factor pudo incidir ya que a pesar de que existió un ajuste global del instrumento hubo un gran número de ítems no se ajustaron a el modelo.

En este sentido se puede concluir que, si bien tanto la metodología propuesta por la TCT y el modelo de Rasch son bastante equivalentes en cuanto a la estimación de parámetros de los ítems y a la medición del nivel de depresión que poseen las personas, el modelo de Rasch brinda mayor profundidad de análisis de los instrumentos, tanto por la escala única en que se encuentran los ítems y las personas, también porque brinda indicadores estadísticos que determinan el nivel en que el instrumento se ajusta al modelo matemático tomando en cuenta los indicadores empíricos obtenidos y por último brinda índices que sirven como indicadores de la replicabilidad de los puntajes de los sujetos.

Queda en evidencia que la ventaja de la TCT sobre el modelo de Rasch en lo que respecta a la parsimonia de sus supuestos y de acuerdo a los resultados obtenidos, la TCT posee un mejor rendimiento con muestras pequeñas lo que supone una gran ventaja práctica. En contraste, el modelo de Rasch brinda una mayor profundidad de análisis tanto en la descripción de los resultados obtenidos por las personas como de los ítems; ganando de este modo mayor precisión predictiva en el proceso de medición.

LIMITACIONES Y RECOMENDACIONES

Una de las principales limitaciones de este estudio fue que existieron muchos ítems que no se ajustaron a el modelo de Rasch por lo que limita su alcance en cuanto al nivel de generalidad.

Sería de gran utilidad poder comprobar cómo se comporta la invarianza de los parámetros de los ítems en muestras de distintos tamaños, tanto grandes como pequeñas, para así comprobar si el modelo de Rasch es en realidad poco útil en muestras pequeñas.

Sería de gran utilidad poder comparar los resultados obtenidos en este estudio si se re-aplica el instrumento en una muestra distinta de personas para así poder determinar el funcionamiento de los ítems a través de muestras distintas y que tan equivalentes son los parámetros que esta produce.

Una de las limitaciones del estudio fue que no participó muestra clínica para así poder ayudar a calibrar en mayor medida el instrumento. Particularmente con sujetos ubicados en niveles altos del constructo depresión.

Sería conveniente aplicar el instrumento en una muestra clínica para así validar de manera convergente los resultados que se han obtenido donde se muestra que en muestras no clínicas como criterio divergente los puntajes de las personas en relación a la variable depresión son bastante bajos.

REFERENCIAS BIBLIOGRAFICAS

Adedoyin, O. Nenty, H y Chilisa, B (2008) Investigating the invariance of ítem difficulty parameter estimates based on CTC y IRT. *Educational Research and Review, Vol. 3 (2), pp. 083-093, March, 2008*

American Psychological Association. (2007). *Divisions of the American Psychological Association*. Recuperado en Noviembre 25, 2009, de <http://www.apa.org/about/division.html>

American Psychiatryc Association. (1995): DSM-IV. Breviario. Criterios diagnósticos. Barcelona: Masson, S. A.

Anastasi, A., y Urbina, S. (1998). *Tests psicológicos*. México: Prentice-Hall.

Asún, R y Zuñiga, C. (2008). Ventajas de los Modelos Politómicos de Teoría de Respuesta al Ítem en la Medición de Actitudes Sociales. El Análisis de un Caso. *PSYKHE, Vol.17, N° 2, 103-115*

Belloch, A., Sandín, B., y Ramos, F. (1995). *Manual de psicopatología*. Volumen 2. Madrid: Mc Graw Hill.

Bond, G y Fox, C. (2001). *Applying the Rasch model: Fundamental measurement in the Human Sciences*. London: LEA

Bond, G y Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the Human Sciences* (2 ed). London: LEA

Bravo, R., Echeburúa, E. y Aizpiri, J. (2007). Características psicopatológicas y dimensiones de personalidad en los pacientes dependientes del alcohol: un estudio comparativo. *Adicciones, Vol.19 Núm. 4, pp.373-382*

Cadavid, N. Delgado y A. Prieto, G. (2007). Construcción de una escala de depresión con el modelo de Rasch. *Psicothema, Vol. 19, nº 3, pp. 515-521*

Cattell, R.B. (1966). The meaning and strategic use of factor analysis. En R.B Cattell. *Handbook of experimental psychology* (pp.174-243). Chicago: Rand and McNally

Cortada, N (2004). Teoría de Respuesta al Ítem: Supuestos Basicos. *Evaluar 2004, n ° 4 (setiembre)*

De Llano, J.G. (2005). *Desarrollo de un instrumento para evaluar la docencia en espacios virtuales de formación: Variable de desarrollo de comunidades virtuales* (Trabajo de Asenso). Universidad Católica Andrés Bello, Caracas, Venezuela.

Escuela de Psicología. (2002). *Contribuciones a la deontología de la investigación en psicología*. Caracas: Publicaciones UCAB.

Fan, X. (1998). Item response theory and classical test theory: an empirical comparison of their Item/person statistics. *Educational and Psychological Measurement, 58*, 357-385.

Federación de Psicólogos de Venezuela (1981). Código de Ética Profesional del Psicólogo en Venezuela. Recuperado en Mayo 28, 2011, de <http://es.scribd.com/doc/14448662/Codigo-de-Etica-del-Psicologo-en-Venezuela>

Ferrando, P y Chico, E. (2000). Adaptación y análisis psicométrico de la escala de deseabilidad social de Marlowe y Crowne. *Psicothema Vol. 12, n° 3, pp. 383-389*

Gómez, J. e Hidalgo, D. (2003). Desarrollos recientes en psicometría. *Avances en Medición, 1 (1), 17-36*

Hernández, Fernández y Baptista. (1991). *“Metodología de la investigación”*. México: McGraw-Hill

Hernandez, R (2009) Comparison of the Item Discrimination and Item Difficulty of the Quick-Mental Aptitude Test using CTT and IRT Methods. *Journal of Educational and Psychological Assessment April 2009, Vol. 1, Issue 1, pp. 12-18*

Hogan, T. (2004). *Pruebas Psicológicas: Una Introducción Práctica*. México: Manual Moderno.

Kaplan, H., Sadock, B., y Grebb J. (2005). *Sinopsis de Psiquiatría*. Buenos Aires: Panamericana.

Kaiser, H.F. (1960). The application of electronics computers to factor analysis. *Journal of Educational and Psychological Assessment*, 20, pp. 141-161.

Kerlinger, F., y Lee, H. (2002). *Investigación del comportamiento: metodos de investigación en ciencias sociales* (4 ed.). México: McGraw Hill.

Krauss, S. y Halgin, R. (2009). *Psicología de la anormalidad*. (5 ed.). México: McGraw-Hill

Lamprianou, I. (2007). An investigation into the test equating methods used during 2006, and the potential for strengthening their validity and reliability. Mayo 28, 2011 de http://www.ofqual.gov.uk/files/Test_equating_report_final.pdf.

Linacre, J. (2011). User Guide to WINSTEPS. Recuperado en 13 de febrero de 2011 de <http://www.winsteps.com/index.htm>

Lorenzo, A., Grau, J., Fumero, A., Vizcaíno, M., Martín, M., y Prado, F. (2003). Validación del Inventario de Ansiedad Rasgo-Estado para Niños en adolescentes cubanos. *Psicología y Salud*, Vol. 13, Núm. 2:pp. 203-214

Magno, C. (2009) Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data. *Journal of Educational and Psychological Assessment* April 2009, Vol. 1, Issue 1, pp. 1-11

Magnusson, D. (1969). *Teoría de los Tests*. México: Editorial Trillas.

Marco, I. Gonzalez, V y Gomez, J. (2000) Teoría de respuesta al ítem y análisis factorial confirmatorio: dos métodos para analizar la equivalencia psicométrica en la traducción de cuestionarios. *Psicothema*, Vol. 12, Supl. nº 2, pp. 540-544

Morales, R. (2009). Evaluation of Mathematics Achievement Test: A Comparison between CTT and IRT. *The International Journal of Educational and Psychological Assessment*. April 2009, Vol. 1, Issue 1, pp. 19-26

Muñiz, J. (1997). *Introducción a la Teoría de Respuesta a los Ítems*. Madrid: Pirámide.

Muñiz, J. (1998). La medición en lo psicológico. *Psicothema*, 1998. Vol. 10, nº 1, pp. 1-21.

Muñiz, J. (2000). *Teoría Clásica de los Test*. Madrid: Pirámide.

Muñiz, J. (2010). Las teorías de los tests: Teoría Clásica y Teoría de Respuesta a los Ítems. *Papeles del Psicólogo*, Vol. 31, Núm. 1, enero-abril, 2010, pp. 57-66

Nukhet, C. (2002) A Study of Raven Standard Progressive Matrices test's item measures under classic and item response models: An empirical comparison. *Ankara University, Journal of Faculty of Educational Science*, 35 (1-2), 71-79.

Nunnally, J. Berstein, I. (1995) *Teoría Psicométrica*. México: McGraw- Hill

Prieto, G. y Delgado, A. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15, 94-100.

Progar, S. y Socan, G. (2008). An empirical comparison of Item Response Theory and Classical Test Theory. *Horizons of Psychology*, 17, 3, pp. 5-24

Revelle, W., Condon, D. y Wilt, J. (2010). Methodological Advances in Differential Psychology. Recuperado Abril 20, 2011 de <http://www.personality-project.org/revelle/publications/HID.methods.final.pdf>.

Richaud, M. (2006). Desarrollos del análisis factorial para el estudio de ítems dicotómicos y ordinales. *Interdisciplinaria*, 22,2, pp. 237-251

Sánchez, M. (2008). *Introducción a la Teoría de Respuesta al Ítems, una herramienta para el análisis de variables latentes: aplicación a la medición de la*

calidad de vida de la infancia (Trabajo de investigación). Departamento de Economía Aplicada y Organización de Empresas Universidad de Extremadura.

Sandín, B., Chorot, P., Lostao, L., Joiner, T., Santed, M. y Valiente, R. (1999). Escala PANAS de afecto positivo y negativo: validación factorial y convergencia transcultural. *Psicothema*, vol. 11, nro. 1, pp. 37-51.

Sanz, J., Perdigón, L. A., y Vázquez, C. (2003). Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 2. Propiedades psicométricas en población general. *Clínica y Salud*, 14, 249-280.

Stage, C. (2003). Teoría Clásica de la Medición o Teoría de respuesta al Ítem: La experiencia sueca. *Estudios públicos n 90* pp. 185-217

Stone, M. (2003). Sstantive Scale Construction. *Journal of Applied Measurement*, 4 (3), pp. 282-297.

Silvestre, J (2009) Item Response Theory and Classical Test Theory: An Empirical Comparison of Item/Person Statistics in a Biological Science Test. *De La Salle College of Saint Benilde, Journal of Educational and Psychological Assessment April 2009, Vol. 1, Issue 1, pp. 19-31.*

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. NJ: Lawrence Erlbaum Associates.

ANEXO A
Escala de Depresión

Este cuestionario busca conocer cómo se ha sentido usted últimamente. Por favor, léalo con detenimiento. A continuación marque la afirmación que más se corresponda con su estado de ánimo en las dos últimas semanas.

Si tiene alguna duda, por favor comuníquese al encuestador para guiarle. Gracias por su colaboración.

Edad: _____ Sexo: F / M Estado civil: _____ Nivel educativo: _____

Ocupación: _____ Carrera: _____ Año/Semestre: _____

Recuerde: debe marcar aquellas afirmaciones que haya experimentado
en las dos últimas semanas

- | | SI | NO |
|---|--------------------------|--------------------------|
| 1. Me siento tan triste que no puedo soportarlo | <input type="checkbox"/> | <input type="checkbox"/> |
| 2. Estoy a dieta para adelgazar | <input type="checkbox"/> | <input type="checkbox"/> |
| 3. He fracasado en las cosas que me he propuesto | <input type="checkbox"/> | <input type="checkbox"/> |
| 4. Siento que estoy siendo castigado/a | <input type="checkbox"/> | <input type="checkbox"/> |
| 5. He perdido todo el interés por los demás | <input type="checkbox"/> | <input type="checkbox"/> |
| 6. Las cosas que antes solían irritarme, me importan poco | <input type="checkbox"/> | <input type="checkbox"/> |
| 7. Estoy comiendo más que antes | <input type="checkbox"/> | <input type="checkbox"/> |
| 8. He perdido el interés por las actividades que antes realizaba con placer | <input type="checkbox"/> | <input type="checkbox"/> |
| 9. He pensado en quitarme la vida | <input type="checkbox"/> | <input type="checkbox"/> |
| 10. Me culpo por todo lo malo que sucede | <input type="checkbox"/> | <input type="checkbox"/> |
| 11. Estoy demasiado cansado/a para hacer nada | <input type="checkbox"/> | <input type="checkbox"/> |
| 12. Me siento una persona totalmente fracasada | <input type="checkbox"/> | <input type="checkbox"/> |
| 13. Tengo problemas para concentrarme | <input type="checkbox"/> | <input type="checkbox"/> |
| 14. Me detesto | <input type="checkbox"/> | <input type="checkbox"/> |
| 15. Me suicidaría si tuviese la oportunidad | <input type="checkbox"/> | <input type="checkbox"/> |
| | SI | NO |
| 16. Me siento triste en situaciones que antes no me hacían sentirme así | <input type="checkbox"/> | <input type="checkbox"/> |
| 17. Antes era capaz de llorar, pero ahora no puedo, incluso aunque quiera | <input type="checkbox"/> | <input type="checkbox"/> |
| 18. Todo me produce insatisfacción | <input type="checkbox"/> | <input type="checkbox"/> |

- | | | |
|--|--------------------------|--------------------------|
| 19. Me siento inútil y me culpo por ello | <input type="checkbox"/> | <input type="checkbox"/> |
| 20. Ya me es imposible tomar decisiones | <input type="checkbox"/> | <input type="checkbox"/> |
| 21. Estoy durmiendo menos que antes | <input type="checkbox"/> | <input type="checkbox"/> |
| 22. Me siento culpable constantemente | <input type="checkbox"/> | <input type="checkbox"/> |
| 23. Presiento que todo me va a salir mal | <input type="checkbox"/> | <input type="checkbox"/> |
| 24. He intentado matarme | <input type="checkbox"/> | <input type="checkbox"/> |
| 25. Me siento una persona tan desgraciada que no puedo soportarlo | <input type="checkbox"/> | <input type="checkbox"/> |
| 26. Estoy durmiendo más que antes | <input type="checkbox"/> | <input type="checkbox"/> |
| 27. Todo me aburre | <input type="checkbox"/> | <input type="checkbox"/> |
| 28. Siento que el futuro es desesperanzador y las cosas no mejorarán | <input type="checkbox"/> | <input type="checkbox"/> |
| 29. No puedo dejar de pensar cosas malas sobre mí | <input type="checkbox"/> | <input type="checkbox"/> |
| 30. He perdido totalmente el interés por el sexo | <input type="checkbox"/> | <input type="checkbox"/> |
| 31. Me siento más triste por la mañana que a otras horas del día | <input type="checkbox"/> | <input type="checkbox"/> |
| 32. Estoy comiendo menos que antes | <input type="checkbox"/> | <input type="checkbox"/> |
| 33. Todo me sale mal, no tengo suerte con nada | <input type="checkbox"/> | <input type="checkbox"/> |
| 34. Tardo más que antes en las actividades cotidianas | <input type="checkbox"/> | <input type="checkbox"/> |

Indique cómo se ha sentido en las dos últimas semanas, en términos generales:

- | | | | | | |
|--------------------|--------------------------------|--------------------------------|-------------------|--------------------------------|--------------------------------|
| 35. Tranquilo/a | SI
<input type="checkbox"/> | NO
<input type="checkbox"/> | 42. Alegre | SI
<input type="checkbox"/> | NO
<input type="checkbox"/> |
| 36. Oprimido/a | <input type="checkbox"/> | <input type="checkbox"/> | 43. Contrariado/a | <input type="checkbox"/> | <input type="checkbox"/> |
| 37. Activo/a | <input type="checkbox"/> | <input type="checkbox"/> | 44. Satisfecho/a | <input type="checkbox"/> | <input type="checkbox"/> |
| 38. Entusiasmado/a | <input type="checkbox"/> | <input type="checkbox"/> | 45. Angustiado/a | <input type="checkbox"/> | <input type="checkbox"/> |
| 49. Confiado/a | <input type="checkbox"/> | <input type="checkbox"/> | 46. Inspirado/a | <input type="checkbox"/> | <input type="checkbox"/> |
| 40. Alterado/a | <input type="checkbox"/> | <input type="checkbox"/> | 47. Preocupado/a | <input type="checkbox"/> | <input type="checkbox"/> |
| 41. Desasosegado/a | SI
<input type="checkbox"/> | NO
<input type="checkbox"/> | 48. Relajado/a | SI
<input type="checkbox"/> | NO
<input type="checkbox"/> |
| 49. Seguro/a | <input type="checkbox"/> | <input type="checkbox"/> | 56. Aturdido/a | <input type="checkbox"/> | <input type="checkbox"/> |
| 50. Nervioso/a | <input type="checkbox"/> | <input type="checkbox"/> | 57. Vacío/a | <input type="checkbox"/> | <input type="checkbox"/> |
| 51. Decidido/a | <input type="checkbox"/> | <input type="checkbox"/> | 58. Orgullosa/a | <input type="checkbox"/> | <input type="checkbox"/> |

- | | | | | | |
|---------------------|--------------------------|--------------------------|---------------|--------------------------|--------------------------|
| 52. Cómodo/a | <input type="checkbox"/> | <input type="checkbox"/> | 59. Bien | <input type="checkbox"/> | <input type="checkbox"/> |
| 53. Animado/a | <input type="checkbox"/> | <input type="checkbox"/> | 60. Tenso/a | <input type="checkbox"/> | <input type="checkbox"/> |
| 54. Sobreexcitado/a | <input type="checkbox"/> | <input type="checkbox"/> | 61. Agitado/a | <input type="checkbox"/> | <input type="checkbox"/> |
| 55. Motivado/a | <input type="checkbox"/> | <input type="checkbox"/> | 62. Solo/a | <input type="checkbox"/> | <input type="checkbox"/> |

ANEXO B

Tablas de resultados de Análisis Factorial de la Escala de Depresión

Tabla B.1.

Tabla de varianza explicada total.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	13,124	21,168	21,168	13,124	21,168	21,168	6,656	10,736	10,736
2	3,348	5,401	26,569	3,348	5,401	26,569	5,540	8,935	19,671
3	2,641	4,259	30,828	2,641	4,259	30,828	4,246	6,849	26,520
4	2,045	3,298	34,126	2,045	3,298	34,126	3,527	5,689	32,209
5	1,836	2,962	37,088	1,836	2,962	37,088	2,510	4,048	36,257
6	1,672	2,697	39,784	1,672	2,697	39,784	1,859	2,998	39,256
7	1,578	2,545	42,329	1,578	2,545	42,329	1,759	2,838	42,094
8	1,527	2,464	44,793	1,527	2,464	44,793	1,673	2,699	44,793
9	1,381	2,227	47,019						
10	1,320	2,129	49,148						
11	1,308	2,110	51,258						
12	1,194	1,926	53,185						
13	1,183	1,909	55,093						
14	1,115	1,799	56,892						
15	1,112	1,794	58,686						
16	1,052	1,697	60,384						
17	1,013	1,634	62,018						
18	,974	1,571	63,589						
19	,932	1,504	65,093						
20	,897	1,446	66,539						
21	,853	1,375	67,915						
22	,836	1,349	69,264						
23	,815	1,315	70,578						
24	,792	1,277	71,855						
25	,772	1,246	73,101						
26	,743	1,199	74,299						
27	,728	1,174	75,473						
28	,695	1,120	76,593						

29	,685	1,104	77,697					
30	,673	1,086	78,783					
31	,666	1,074	79,857					
32	,638	1,030	80,887					
33	,622	1,004	81,890					
34	,602	,971	82,862					
35	,595	,959	83,821					
36	,572	,922	84,743					
37	,550	,887	85,630					
38	,539	,869	86,500					
39	,529	,853	87,352					
40	,513	,827	88,179					
41	,486	,784	88,964					
42	,475	,766	89,729					
43	,445	,717	90,447					
44	,425	,686	91,132					
45	,414	,667	91,799					
46	,393	,634	92,433					
47	,387	,624	93,057					
48	,378	,609	93,666					
49	,367	,592	94,258					
50	,359	,579	94,837					
51	,343	,552	95,390					
52	,334	,539	95,929					
53	,310	,501	96,429					
54	,288	,464	96,893					
55	,278	,449	97,342					
56	,272	,439	97,782					
57	,261	,420	98,202					
58	,249	,402	98,604					
59	,236	,381	98,985					
60	,218	,352	99,337					
61	,217	,350	99,687					
62	,194	,313	100,000					

Tabla B.2.

Matriz de Componentes Rotados.

	Component							
	1	2	3	4	5	6	7	8
Me siento tan triste que no puedo soportarlo	,154	,196	,286	,491				,138
Estoy a dieta para adelgazar							,554	
He fracasado en las cosas que me he propuesto	,236	,184	,464	,111	,105		-,145	
Siento que estoy siendo castigado/a			,347		,161	-,176		,349
He perdido todo el interes por los demás	,272			,512	,258			
Las cosas que antes solian irritarme me importan poco	-,145			,104	,536	,164		
Estoy comiendo más que antes		,227			,111		-,617	,125
He perdido el interés por las actividades que antes realizaba con placer	,255	,125	,101		,447	,212	-,155	
He pensado en quitarme la vida			,136	,592		,388	-,112	
Me culpo por todo lo malo que sucede	,140	,112	,548	,296				
Estoy demasiado cansado/a para hacer nada	,189	,143	,276		,470			
Me siento una persona totalmente fracasada			,642	,292	-,120			
Tengo problemas para concentrarme	,156	,223	,158		,456		-,104	
Me detesto			,333	,671				
Me suicidaria si tuviera la oportunidad				,436		-,123		-,106

Me siento tan triste en situaciones que antes no me hacían sentirme así	,176	,338	,303	,147	,226	,202	,141	
Antes era capaz de llorar, pero ahora no puedo, incluso aunque quiera		,126	-,134	,425	,232	,317		
Todo me produce insatisfacción	,234		,123	,553	,266			
Me siento inútil y me culpo por ello	,238		,470	,279	,116	,151	-,189	
Ya me es imposible tomar decisiones	,157		,248	,293	,140		,149	
Estoy durmiendo menos que antes		,250		,112	,362		,206	-,452
Me siento culpable constantemente	,118	,214	,592	,181		,230		
Presiento que todo me va a salir mal	,246	,139	,710		,167			
He intentado matarme				,138		,767		
Me siento una persona tan desgraciada que no puedo soportarlo			,232	,575	-,141			
Estoy durmiendo más que antes			,116	-,146			-,216	,558
Todo me aburre	,347		,159		,557			,206
Siento que el futuro es desesperanzador y las cosas no mejoraran			,502		,164		,143	
No puedo dejar de pensar cosas malas sobre mi	,149	,145	,449	,321		,374		,124
He perdido totalmente el interés por el sexo			,118	-,131	,136	,486	,151	
Me siento más triste por la mañana que a otras horas de día	,224				,225	,274	,143	,240
Estoy comiendo menos que antes		,112				,199	,683	-,228

Todo me sale mal, no tengo suerte con nada	,200		,653					,152
Tardo más que antes en las actividades cotidianas	,188	,253	,235		,516			-,122
Tranquilo/a	,252	,625	,204					,106
oprimido	,139	,476	,129	,267	,215	-,132	,232	,238
activo	,667				,199			
entusias	,719	,109	,155	,107				
confiado	,565	,361	,225					
alterado	,180	,604	,150		,115	,104		,171
desasose		,204	,276	,404	,137	-,272	,144	
alegre	,638	,226		,167	,108		,161	,351
contrari	,156	,439	,122	,152	,129	,146		
satisfec	,553	,398	,148					
angustia	,182	,643	,145					-,139
inspirad	,615							-,132
preocupa	,140	,650						-,187
relajado	,376	,620						
seguro	,626	,347	,245					
nervioso	,188	,608						-,161
decidido	,602	,120	,140	,159				
comodo	,503	,489	,195					
animado	,714	,259	,128	,130	,124	,121		
sobreexi	-,229			,103			-,101	,285
motivado	,681	,191		,133	,149			
aturdido	,155	,501		,197	,185		,145	,241
vacio	,367	,218	,128	,290	,168	,160	,162	,230
orgullos	,590	,172	,175					-,141
bien	,570	,225	,150	,184	,122			,310
tenso	,156	,714			,147			
agitado		,589				,132		,174
solo	,383		,113	,207	,134		,207	,138

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

ANEXO C

Estadísticos descriptivos de los ítems

Tabla C.1

Estadísticos descriptivos de los ítems

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Cronbach's Alpha if Item Deleted
Me siento tan triste que no puedo soportarlo	14,28	105,337	,414	,926
Estoy a dieta para adelgazar	14,11	107,002	,060	,928
He fracasado en las cosas que me he propuesto	14,10	103,491	,453	,925
Siento que estoy siendo castigado/a	14,27	105,935	,290	,926
He perdido todo el interes por los demás	14,26	105,364	,367	,926
Las cosas que antes solian irritarme me importan poco	13,89	106,717	,073	,929
Estoy comiendo más que antes	13,94	106,006	,145	,928
He perdido el interés por las actividades que antes realizaba con placer	14,14	104,299	,389	,926
He pensado en quitarme la vida	14,33	106,570	,308	,927
Me culpo por todo lo malo que sucede	14,23	104,599	,436	,926
Estoy demasiado cansado/a para hacer nada	14,09	103,789	,415	,926
Me siento una persona totalmente fracasada	14,31	106,253	,339	,926
Tengo problemas para concentrarme	13,82	103,652	,376	,926
Me detesto	14,33	106,525	,347	,926
Me suicidaria si tuviera la oportunidad	14,35	107,515	,101	,927

Me siento tan triste en situaciones que antes no me hacían sentirme así	14,11	102,984	,517	,925
Antes era capaz de llorar, pero ahora no puedo, incluso aunque quiera	14,25	106,268	,208	,927
Todo me produce insatisfacción	14,29	105,600	,402	,926
Me siento inútil y me culpo por ello	14,27	104,979	,461	,926
Ya me es imposible tomar decisiones	14,27	105,765	,320	,926
Estoy durmiendo menos que antes	13,80	105,032	,240	,927
Me siento culpable constantemente	14,26	104,679	,479	,925
Presiento que todo me va a salir mal	14,20	103,737	,518	,925
He intentado matarme	14,35	107,428	,145	,927
Me siento una persona tan desgraciada que no puedo soportarlo	14,34	106,983	,245	,927
Estoy durmiendo más que antes	14,19	107,496	,011	,928
Todo me aburre	14,20	104,538	,411	,926
Siento que el futuro es desesperanzador y las cosas no mejoraran	14,25	105,650	,310	,926
No puedo dejar de pensar cosas malas sobre mí	14,27	104,989	,459	,926
He perdido totalmente el interés por el sexo	14,33	107,354	,106	,927
Me siento más triste por la mañana que a otras horas de día	14,29	106,026	,308	,926
Estoy comiendo menos que antes	14,15	106,890	,079	,928

Todo me sale mal, no tengo suerte con nada	14,29	105,346	,445	,926
Tardo más que antes en las actividades cotidianas	14,11	103,709	,436	,926
Tranquilo/a	14,07	102,696	,526	,925
oprimido	14,17	103,505	,516	,925
activo	14,11	103,614	,445	,925
entusias	14,05	102,117	,575	,924
confiado	13,97	101,935	,563	,924
alterado	14,03	102,415	,533	,925
desasose	14,26	105,391	,362	,926
alegre	14,18	103,016	,586	,925
contrari	14,10	103,480	,453	,925
satisfec	13,94	101,421	,608	,924
angustia	13,86	102,232	,518	,925
inspirad	13,93	103,866	,357	,926
preocupa	13,70	103,188	,445	,925
relajado	13,90	101,940	,550	,925
seguro	14,06	101,443	,657	,924
nervioso	13,88	102,864	,455	,925
decidido	14,11	103,045	,512	,925
comodo	14,09	101,704	,655	,924
animado	14,05	101,147	,682	,924
sobreexi	14,23	107,875	-,038	,928
motivado	14,07	102,068	,597	,924
aturdido	14,15	103,392	,504	,925
vacio	14,21	103,732	,529	,925
orgullos	13,91	102,618	,482	,925
bien	14,21	103,354	,593	,925
tenso	13,84	102,302	,511	,925
agitado	13,96	103,950	,353	,926
solo	14,15	104,140	,416	,926

ANEXO D

Correlaciones de las posiciones de ítems entre las submuestras

Tabla D.1

Correlaciones de las posiciones de ítems entre las submuestras de acuerdo a la TCT.

		1	2	3	4	5	6	7	8	9	10
Muestra1	Pearson Correlation	1	,987**	,844**	,855**	,853**	,970**	,870**	,876**	,900**	,869**
	Sig. (2-tailed)		,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	62	62	62	62	62	62	62	62	62	62
Muestra2	Pearson Correlation	,987**	1	,843**	,843**	,850**	,960**	,855**	,850**	,896**	,860**
	Sig. (2-tailed)	,000		,000	,000	,000	,000	,000	,000	,000	,000
	N	62	62	62	62	62	62	62	62	62	62
Muestra3	Pearson Correlation	,844**	,843**	1	,851**	,840**	,869**	,923**	,876**	,877**	,834**
	Sig. (2-tailed)	,000	,000		,000	,000	,000	,000	,000	,000	,000
	N	62	62	62	62	62	62	62	62	62	62
Muestra4	Pearson Correlation	,855**	,843**	,851**	1	,863**	,893**	,844**	,950**	,929**	,866**
	Sig. (2-tailed)	,000	,000	,000		,000	,000	,000	,000	,000	,000
	N	62	62	62	62	62	62	62	62	62	62
Muestra5	Pearson Correlation	,853**	,850**	,840**	,863**	1	,886**	,841**	,887**	,867**	,988**
	Sig. (2-tailed)	,000	,000	,000	,000		,000	,000	,000	,000	,000
	N	62	62	62	62	62	62	62	62	62	62
Muestra6	Pearson Correlation	,970**	,960**	,869**	,893**	,886**	1	,869**	,904**	,899**	,902**
	Sig. (2-tailed)	,000	,000	,000	,000	,000		,000	,000	,000	,000
	N	62	62	62	62	62	62	62	62	62	62
Muestra7	Pearson Correlation	,870**	,855**	,923**	,844**	,841**	,869**	1	,904**	,871**	,829**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000		,000	,000	,000
	N	62	62	62	62	62	62	62	62	62	62
Muestra8	Pearson Correlation	,876**	,850**	,876**	,950**	,887**	,904**	,904**	1	,891**	,896**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000		,000	,000
	N	62	62	62	62	62	62	62	62	62	62
Muestra9	Pearson Correlation	,900**	,896**	,877**	,929**	,867**	,899**	,871**	,891**	1	,851**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000		,000
	N	62	62	62	62	62	62	62	62	62	62
Muestra10	Pearson Correlation	,869**	,860**	,834**	,866**	,988**	,902**	,829**	,896**	,851**	1
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	
	N	62	62	62	62	62	62	62	62	62	62

** . Correlation is significant at the 0.01 level (2-tailed).

Tabla D.2

Correlaciones de las posiciones de ítems entre las submuestras de acuerdo al modelo de Rasch

		1R	2R	3R	4R	5R	6R	7R	8R	9R	10R
Muestra1R	Pearson	1	,988	,767	,746	,801	,956	,781	,810	,815	,834
	Correlation										
	Sig. (2-tailed)		,000	,000	,000	,000	,000	,000	,000	,000	,000
N		60	60	57	49	58	59	57	59	52	60
Muestra2R	Pearson	,988	1	,763	,729	,796	,946	,766	,786	,805	,827
	Correlation										
	Sig. (2-tailed)	,000		,000	,000	,000	,000	,000	,000	,000	,000
N		60	60	57	49	58	59	57	59	52	60
Muestra3R	Pearson	,767	,763	1	,763	,812	,785	,896	,848	,789	,802
	Correlation										
	Sig. (2-tailed)	,000	,000		,000	,000	,000	,000	,000	,000	,000
N		57	57	58	49	58	57	56	58	52	58
Muestra4R	Pearson	,746	,729	,763	1	,820	,807	,756	,925	,879	,814
	Correlation										
	Sig. (2-tailed)	,000	,000	,000		,000	,000	,000	,000	,000	,000
N		49	49	49	49	49	49	49	49	48	49
Muestra5R	Pearson	,801	,796	,812	,820	1	,846	,809	,883	,799	,980
	Correlation										
	Sig. (2-tailed)	,000	,000	,000	,000		,000	,000	,000	,000	,000
N		58	58	58	49	60	58	57	59	52	60
Muestra6R	Pearson	,956	,946	,785	,807	,846	1	,788	,845	,811	,867
	Correlation										
	Sig. (2-tailed)	,000	,000	,000	,000	,000		,000	,000	,000	,000
N		59	59	57	49	58	59	57	59	52	59
Muestra7R	Pearson	,781	,766	,896	,756	,809	,788	1	,872	,821	,795
	Correlation										
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000		,000	,000	,000
N		57	57	56	49	57	57	58	58	52	58
Muestra8R	Pearson	,810	,786	,848	,925	,883	,845	,872	1	,815	,892
	Correlation										
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000		,000	,000
N		59	59	58	49	59	59	58	60	52	60
Muestra9R	Pearson	,815	,805	,789	,879	,799	,811	,821	,815	1	,773
	Correlation										
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000		,000
N		52	52	52	48	52	52	52	52	52	52
Muestra10R	Pearson	,834	,827	,802	,814	,980	,867	,795	,892	,773	1
	Correlation										
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	
N		60	60	58	49	60	59	58	60	52	62