



UNIVERSIDAD VERACRUZANA

Especialización en Estudios de Opinión



UNIVERSIDAD VERACRUZANA

CENTRO DE ESTUDIOS DE OPINIÓN Y ANÁLISIS

ESPECIALIZACIÓN EN ESTUDIOS DE OPINIÓN

PROYECTO DE INTERVENCIÓN

**VALIDEZ Y CONFIABILIDAD DE INSTRUMENTOS DE
ESTUDIOS DE OPINIÓN**

Presenta

Israel Huesca Domínguez

Director(es)

Dr. Sergio Hernández González
Dr. Claudio Rafael Castro López

Xalapa-Enríquez, Ver., México, agosto del 2016



AGRADECIMIENTOS

A Dios, por estar siempre conmigo.

A mi familia, por apoyarme siempre.

A mis directores y sinodales:

Dr. Sergio Hernández González, por dirigirme, brindarme su apoyo, guiarme y por confiar en mí para este proyecto.

Dr. Claudio Rafael Castro López, por dirigirme, apoyarme y prestarme la base de datos empleada en este proyecto.

Mtra. Nery Sofía Huerta Pacheco, por compartirme sus conocimientos y estar al pendiente de que el proyecto quedara lo mejor posible.

Dra. Laura Oliva Zarate, por su tiempo para revisar mi proyecto y ser parte de mi jurado.

A mis profesores de la especialidad por compartirme parte de sus conocimientos y experiencia para mi crecimiento profesional.

A mis amigos por animarme en cada momento y a todos aquellos que influyeron positivamente en mi proyecto.

Muchas gracias



ÍNDICE

I. INTRODUCCIÓN	1
I.1 Planteamiento del problema.....	2
I.2 Objetivo general y específicos	5
I.3 Justificación.....	5
II. MARCO TEÓRICO	7
II.1 Método de recolección de información: Cuestionario	7
II.2.1 Confiabilidad	8
II.2.1 Validez	9
II.3 Escalas de medición de las variables	10
II.4 Escala de Likert.....	12
II.5 Alpha de Cronbach	13
II.6 Análisis del Impacto	15
II.7 Teoría de Respuesta al Ítem (TRI).....	16
II.7.1 Curvas características de los ítems	17
II.7.2 Supuestos de la TRI	18
II.7.3 Modelo de respuesta graduada de Samejima	18
III. MARCO CONTEXTUAL	20
III.1 Herramientas para la confiabilidad y validez	20
III.2 Importancia de las encuestas en estudios de opinión	24
IV. METODOLOGÍA	27
IV.1 Descripción del instrumento.....	27
IV.2 Descripción de los datos.....	28
IV.3 Herramientas estadísticas	29



IV.3.1 Estadística univariante.....	30
IV.3.2 Estadística bivariante.....	30
IV.3.3 Estadística multivariante	30
V. RESULTADOS	32
CONCLUSIONES	53
REFERENCIAS	57
ANEXOS	61



I. INTRODUCCIÓN

Hoy en día una forma para recolectar información en temáticas de estudios de opinión, mercado, preferencias, entre otras, son las encuestas (instrumento de medición). Éstas pueden ser digitales, impresas, telefónicas, etc. Dentro de las encuestas existen preguntas (ítems) en diferentes escalas de medición, tal como la escala de Likert (Likert, 1932), que es muy utilizada dado que son fáciles de interpretar y proporciona información cuando existen características que son difíciles de medir, incluyendo opciones de respuesta con tendencia positiva, negativa y neutra.

Actualmente, la técnica mayormente utilizada para validar instrumentos de medición en el que el constructo contiene escala tipo Likert es el Alpha de Cronbach (Cronbach, 1951), pero esta herramienta tiende a sesgarse debido a la inadecuada estructura de la escala. No obstante, para hacer una mejor construcción y validación de una encuesta existen técnicas más específicas que proporcionan mayor información, entre ellas se encuentra el Impacto de los Ítems (Juniper, Streiner & King, 1997) y la Teoría de Respuesta al Ítem particularmente el modelo de respuesta graduada de Samejima (1969, 1972, 1997).

Dichas técnicas proporcionan características claves de los ítems que ayudan a conocer la validez de la escala empleada, a su vez muestran puntos estratégicos que se pueden considerar para un análisis posterior. El Impacto de los Ítems busca identificar ítems de mayor peso, es decir, identifica los ítems mayormente seleccionados dentro del cuestionario. (Allen & Locker, 2002, & Vicente-Galindo, 2011)

Por otro lado, la Teoría de Respuesta al Ítem analiza los ítems identificados como relevantes que proporcionan información basándose en el comportamiento de los mismos. Esta teoría se fundamenta en las correlaciones de pares de ítems, que subsecuentemente inducen a la creación de una variable latente derivada de la variabilidad representada en los valores propios. Concluyendo con la representación gráfica de curvas características o funciones de respuestas,



que representan la aportación de los ítems dentro del cuestionario y la frecuencia de respuesta de cada una de sus categorías. (Sánchez-Rivero, 2004)

Al emplear conjuntamente las técnicas: Alpha de Cronbach, Impacto de los Ítems y Teoría de Respuesta al Ítem, se espera contar con mayor información al momento de elegir los ítems definitivos de un instrumento, mejorando su validez, fiabilidad y consistencia.

Existen diversas investigaciones que utilizan varias técnicas para validar cuestionarios, entre ellas se encuentra la investigación publicada en el 2015 (Vicente-Galindo et al.) que hace uso de métodos multivariantes para identificar el comportamiento de los aspectos de calidad de vida relacionados con la salud de pacientes con osteoporosis, otro es la propuesta metodológica de validación de un cuestionario de características de un juego serio que utiliza la Teoría de Respuesta al Ítem y otras herramientas complementarias. (Huerta-Pacheco, 2014)

I.1 Planteamiento del problema

Los cuestionarios son métodos de recolección de información que comúnmente son utilizados en diversos estudios sociales, donde la validación y confiabilidad de dicho instrumento es necesario si se quiere tener una herramienta que permita encontrar verdaderamente lo que se pretenda medir y que pueda ser replicable para una población específica.

Para el cálculo del coeficiente de fiabilidad se debe tener en cuenta todas las características de un estimador muestral, como el caso de contar con una muestra de tamaño suficiente (representativa), ya que el estimador es sensible al tipo de la muestra con el que se calcula, esto es, a mayor tamaño de muestra, mayor es la precisión del estimador. La fiabilidad es muy sensible con respecto a los grupos en que se estima, ya que si el instrumento va dirigido a diferentes poblaciones, es recomendable calcular el coeficiente de fiabilidad para cada uno. En la práctica de publicación de investigaciones aunque el instrumento no sea psicométrico se debe exponer el coeficiente de fiabilidad, debido a la sensibilidad que tiene el coeficiente



respecto a las características de la muestra, por lo que es conveniente calcular la fiabilidad en cada investigación. (Martínez-Arias et. al, 2014)

Los investigadores no siempre hacen lo mencionado, ya que es algo muy común utilizar los resultados de otros estudios. La falta de fiabilidad de un instrumento repercute en el análisis de los datos de las investigaciones. Algunos efectos de la baja fiabilidad son que en estudios en comparación de grupos reduce la potencia estadística de los contrastes y aumenta la varianza del efecto, en estudios correlacionales aumentará la variabilidad de la distribución muestral. Además, no se toman decisiones únicamente con el coeficiente de fiabilidad de un instrumento, sino que es de gran importancia conocer la validez del mismo dado que se encuentran muy relacionados.

En la validez de un instrumento también es importante considerar los grupos en lo que se están realizando las mediciones, ya que se pueden encontrar resultados diferentes en los instrumentos cuando se utilizan grupos como el género, la edad, nivel de educación, ocupación, entre otros, es por ello que la heterogeneidad de la muestra es relevante para medir la validez. Otro factor que altera la validez de un instrumento es el tiempo, ya que las aptitudes, percepciones, opiniones, etc., cambian de un momento a otro. (Anastasi & Urbina, 1998)

Es común encontrar instrumentos que no son validados, esta falta de validación se debe a varios factores como el desconocimiento, tiempos establecidos por el estudio, entre otros; un ejemplo es que se presentan preguntas (ítems) donde se hace uso inadecuado de las escalas de medición (cualitativa y cuantitativa), conllevando a una errónea conclusión de los resultados presentados. Existen diferentes tipos de validación de cuestionarios tales como la validación de contenido, criterio y constructo, éste último es el más complejo ya que requiere de métodos estadísticos robustos.

Por otro lado, existen técnicas como el Alpha de Cronbach que brinda una certeza cuantitativa de ítems cualitativos, basándose en escalas ordinales como la de Likert. Permitiendo la



identificación de relaciones entre los mismos. No obstante, la susceptibilidad de la técnica tiende a sesgar el valor de consistencia cuando se hace uso incorrecto de la escala. Además, el Alpha de Cronbach demuestra un puntaje de consistencia interna del instrumento, pero no llega más a detalle con los ítems, solo hace un criterio de exclusión de los mismos para reajustar dicha consistencia.

La validez y la confiabilidad, son mediciones que no se asumen sino que se prueban. Existen varios factores que afectan a la validez y la confiabilidad, como el que un instrumento se realice de manera improvisada, que se haga de un día para otro o en el mismo momento sin tener conocimientos del tema, que se utilicen instrumentos que han sido validados en otros países pero no en el que se desea aplicar. Otro problema es que el instrumento no esté el lenguaje de los usuarios lo cual hace inadecuada su aplicación. También, algo que causa sesgo es la condición en la que se está aplicando como la iluminación, el ruido, el frío, etc. (Hernández-Sampieri et. al, 2010)

Algo muy habitual que se da en el procedimiento de realización de instrumentos es traducirlo de un idioma distinto a otro, en este caso el investigador tiene que considerar si vale la pena realizar ésta alternativa, ya que utilizarlo es la mejor manera de poder realizar comparaciones de diferentes poblaciones pero hay que tener en cuenta que los constructos no son universales sino que los ítems apropiados para medir tal constructo pueden ser diferentes en culturas y lenguas, por lo que a veces es preferible construir un nuevo instrumento. (Morales-Vallejo et. al, 2003)

Respecto a lo anterior es un error que al contar con una buena traducción de un instrumento se garantizará la validez y confiabilidad de los constructos. La validez y la confiabilidad no son traducibles por lo que al traducir y adaptar un instrumento se tiene que realizar el procedimiento de un instrumento nuevo. La fiabilidad y la validez no son propiedades de los instrumentos sino de los datos obtenidos con el instrumento, por lo que se deben calcular en cada investigación.



- **Preguntas de investigación**

- ¿La unión de técnicas estadísticas de la confiabilidad y validez de constructo para ítems en escala ordinal darán mejor fiabilidad en la realización de instrumentos de estudios de opinión?
- ¿Qué estructura del constructo es adecuada para medir las aptitudes y opiniones planteadas en los objetivos bajo estudio?
- ¿Con base a técnicas estadísticas utilizadas en la validación y confiabilidad de cuestionarios, se podrán encontrar preguntas de mayor relevancia para reducir la estructura inicial de un constructo?

I.2 Objetivo general y específicos

- **General**

- Implementar una metodología de confiabilidad y validación de constructo en el contexto de estudios de opinión para ítems con escala ordinal.

- **Específicos**

- Identificar relaciones entre los ítems, con el fin de observar la consistencia interna.
- Determinar características de los ítems que muestren el impacto de los mismos.
- Determinar el número de categorías de los ítems que mejoren la fiabilidad.
- Seleccionar y reducir los ítems con mayor representatividad dentro de los instrumentos de medición (cuestionarios).

I.3 Justificación

Cuando en un estudio se realiza el instrumento con un objetivo de medición, se recurre generalmente al Alpha de Cronbach (Cronbach, 1951) que es una medida de consistencia interna del cuestionario, ya que proporciona el grado de fiabilidad de los ítems, mostrando un



valor estadístico que representa la relación de los mismos, donde si es muy cercano a uno la fiabilidad del instrumento se considera buena.

Esta técnica por si sola es una herramienta de certeza estadística para los investigadores, de que pueden obtener datos confiables de una problemática, sin embargo desconocen que los ítems aún proporcionan más información y que en dado caso pueden contener un sesgo natural presentado por las características de los mismos.

Además, un instrumento de medición puede ser confiable, pero no necesariamente válido, es decir, si un constructo es consistente en sus respectivos resultados pero no mide lo que se pretende medir por consiguiente no es válido. De los tres tipos de validación existentes el de constructo es el más complejo debido a que requiere de métodos estadísticos avanzados de los cuales se necesita tener una formación sólida en éste ámbito para poder realizar las metodologías, es por ello que se describen de manera sencilla diferentes técnicas que se espera den mayor información de las características de los ítems dentro del cuestionario.

Para ello, después de conocer la consistencia interna del cuestionario, se debe realizar una exploración de diversas técnicas preliminares que dan mayor indicación de los ítems, tales como el Impacto de Ítems y la Teoría de Respuesta al Ítem para contar con mayor información al momento de elegir los ítems definitivos de un instrumento, mejorando su validez, fiabilidad y consistencia.

La mayor bibliografía de validación de cuestionarios se enmarca dentro de la psicometría, pero los métodos estadísticos empleados son multidisciplinarios es por ellos que este proyecto de investigación se enmarca en los Estudios de Opinión, como la valoración de una metodología a seguir para realizar estudios más concretos y detallados en el contenido de las herramientas de medición (encuestas), además de que puede ser replicable a diversos estudios de investigación donde el cuestionario es la herramienta esencial e indispensable.



II. MARCO TEÓRICO

En este capítulo se describen los conceptos básicos necesarios para el desarrollo de este proyecto, empezando desde la descripción del método más usual de recolección de datos (el cuestionario), los requisitos que deben cumplirse, la descripción de las escalas de medida de las variables del cuestionario, donde se hace énfasis a la escala de Likert. Posteriormente, se describen técnicas que ayudan a conocer la validez y confiabilidad de un instrumento de medición.

II.1 Método de recolección de información: Cuestionario

Muchos problemas de investigación requieren de recolección de datos de una población a través de entrevistas, encuestas u otro tipo de instrumentos. El más utilizado es la encuesta, se llama encuesta a “un conjunto de preguntas tipificadas dirigidas a una muestra representativa, para averiguar estados de opinión o diversas cuestiones de hecho” (Real Academia Española, Asociación de Academias de la Lengua Española, 2014).

El cuestionario es un instrumento de medición que contiene una serie de preguntas escritas sobre un tema específico del que se buscan las opiniones personales de los individuo, por otro lado la encuesta es una técnica de investigación no experimental orientada a tener información del estado real de alguna situación a través de interrogatorio de una muestra de individuos (Pérez-Tejada, 2008). El procedimiento de la encuesta, se refiere a la combinación de técnicas que no son específicamente de ninguna disciplina sino que es multidisciplinaria, ya que muchas áreas la utilizan como instrumento de medición. Un aspecto importante dentro del diseño de un cuestionario es la formulación y construcción de las escalas de medición de cada uno de los ítems que la conformaran.

Cabe mencionar que la encuesta se utiliza cuando la información que se desea recabar no se puede obtener de forma fácil, así como no tener los recursos económicos que conlleva la aplicación del estudio. Un punto importante en la aplicación de una encuesta es que las



personas puedan y quieran proporcionar la información requerida, otro punto es que existen casos en que las personas no respondan bien a los reactivos, debido a que se trate de una situación embarazosa que lo incrimine haciendo de esto la distorsión de la respuesta emitida. (Anastasi & Urbina, 1998)

En el cuestionario se deben traducir los objetivos de la investigación de tal modo que cada uno de los ítems contenga el mismo contexto para que así se pueda responder la hipótesis de la investigación y/o proyecto. Cuando es impresa, se tiene que cuidar que el encuestado proporcione la información verdadera y no exista distorsión en los resultados. El lenguaje utilizado dentro del cuestionario es de vital importancia, ya que el vocabulario y la sintaxis hacen que exista una comunicación completa y precisa. En resumen el lenguaje del cuestionario debe ser ajustado al lenguaje del usuario que responderá el instrumento.

II.2.1 Confiabilidad

Se refiere a la falta de distorsión de un instrumento de medición, de tal modo que si se aplica de manera repetida al mismo usuario, se puedan obtener resultados semejantes. Otras formas de decir confiabilidad son: estabilidad, fiabilidad, consistencia, reproductibilidad, predictibilidad o falta de distorsión. Cuando un instrumento es altamente confiable se dice que se está midiendo con precisión o consistencia. (Kerlinger & Lee, 2002)

En la Tabla 1 se muestran varias opciones para el cálculo de la fiabilidad, donde se describe: el tipo de caso que puede existir, la herramienta recomendable y que procedimiento debería seguirse. La elección del coeficiente de fiabilidad dependerá de la finalidad del instrumento, el más utilizado en evaluación psicológica es el de consistencia interna, el cual solo requiere que el instrumento se aplique una sola vez. Algunas herramientas estadísticas para conocer la fiabilidad de un instrumento a través de consistencia interna son procedimientos basados en la división del test en dos mitades y métodos basados en las covarianzas de los ítems como lo es el Alpha de Cronbach, el coeficiente de Kuder-Richardson, Lambda de Guttman, entre otros. (Martínez-Arias, Hernández-Lloreda, & Hernández-Lloreda, 2014)



Tabla 1. Diferentes formas de conocer la fiabilidad de un instrumento

FUENTE PRINCIPAL	TIPO DE COEFICIENTE DE FIABILIDAD	PROCEDIMIENTO DE ANÁLISIS DE DATOS
-Cambio en las puntuaciones a lo largo del tiempo. -Muestreo de tiempos.	Coeficiente de estabilidad temporal o test-retest.	Correlaciones entre las puntuaciones obtenidas en dos aplicaciones del instrumento.
-Cambios en los contenidos de las formas. -Muestreo de contenidos.	Coeficiente de equivalencia entre formas alternativas.	Correlaciones entre las puntuaciones de las dos formas.
-Muestreo de contenidos + Muestreo de tiempos.	Formas alternativas en aplicación diferida.	Correlaciones entre las puntuaciones de las dos formas.
-Contenido de los ítems. -Muestreo de contenidos y heterogeneidad de los ítems.	Coeficiente de consistencia interna.	Procedimientos basados en las covarianzas o correlaciones entre partes del test.
-Cambio en los calificadores. -Muestreo de calificadores.	Acuerdo entre calificadores.	Porcentajes de acuerdo o coeficientes de asociación.

Fuente: Martínez-Arias, M. R., Hernández-Lloreda, M. V., & Hernández-Lloreda, M. J. (2014). *Psicometría*. Madrid: Alianza Editorial.

II.2.1 Validez

Su definición se resume en responder a la pregunta: ¿estamos midiendo lo que creemos que estamos midiendo?, haciendo hincapié en lo que se está midiendo. Por ejemplo si un instrumento pretende medir inteligencia, debe medir inteligencia y no memoria. Los diferentes tipos de validez más importantes son las que creó un comité junto a la Asociación Psicológica Americana, la Asociación Americana de Investigación Educativa y el Consejo Nacional de Mediciones utilizadas en Educación, dichos tipos son: validez de contenido, validez de criterio y validez de constructo. (Kerlinger & Lee, 2002)



-
- **Validez de contenido:** Hace referencia a la representatividad del contenido de un instrumento de medición, se basa en el juicio, esto es que el propio investigador o en conjunto con otros se juzgue la representatividad de los ítems. Debe mostrar el dominio del contenido de modo que las mediciones representen las variables medidas. Responde a la pregunta: ¿la sustancia o contenido de esta medida es representativa del contenido del universo de contenido de la propiedad que se mide? Ejemplo: si se habla de operaciones aritméticas no solo se tendría que hablar de la resta sino también de la suma, multiplicación y división.
 - **Validez de criterio:** Consiste en comparar los resultados obtenidos con una u otras variables externas o criterios que midan las mismas características, con la finalidad de conocer y corroborar la concordancia de la medición. Muchas veces el criterio mismo es la mayor dificultad de la validación del criterio.
 - **Validez de constructo:** Es el aspecto más importante, ya que es más teórico, donde se busca comprobar relaciones hipotéticas. Tiene una perspectiva científica, se refiere a que tanto un instrumento mide un concepto teórico, haciendo énfasis en lo que se está midiendo y cómo opera para medirlo. Tiene tres pasos, primero especifica relación teórica entre los conceptos, luego se correlacionan los conceptos para analizarlos y por último se interpreta la evidencia empírica.

II.3 Escalas de medición de las variables

Cada una de las preguntas que están contenidas en un instrumento representa variables con las que se cuenta, las cuales existen diferentes formas de medirlas. Los tipos de variables existentes son las cualitativas y cuantitativas. A continuación se describen los diferentes tipos de variables con las que se puede contar como las escalas de medición que tienen (Stevens, 1946):



-
- **Variables cuantitativas:** son aquellas representadas numéricamente, donde contienen valores reales en las que se pueden realizar las diferentes operaciones aritméticas. Las escalas de este tipo de variables son las de intervalo y de razón.
 - **Escala de intervalo:** Adquiere valores numéricos continuos en un rango, sus valores representan magnitudes con distancias iguales entre ellos, además, el cero toma un valor verdadero como por ejemplo las escalas de temperatura (Centígrados y Fahrenheit).
 - **Escala de razón:** Contiene valores numéricos en el que no se encuentra dado en intervalo. El cero no tiene valor, es decir, significa ausencia de una característica. Ejemplos: Longitud, estatura, número de personas, cantidad de objetos, etc.
 - **Las variables cualitativas,** como su nombre lo dice, representan cualidades, características, atributos, categorías, etc., que no pueden ser medidas con valores numéricos. Existen diferentes formas de escala de medición, variables nominales y ordinales.
 - **Escala nominal:** Son categorías que para representarlas se utilizan etiquetas como símbolos (A, B, C, nombres, características, etc.) o números (1, 2, 3) del cual no tienen sentido matemático. No importa el orden que tengan, si se cambia el orden de las categorías, no altera su significado. Ejemplo: género, partidos políticos, tipos de revistas, etc.
 - **Escala ordinal:** Al igual que la escala nominal, son categorías cuyas etiquetas también son símbolos o números, pero tienen puntuaciones que reflejan orden en el significado; a mayor puntuación más alto será el valor de la propiedad que



se esté midiendo. Ejemplo: nivel socioeconómico, grado de estudios, escala de Likert, entre otras.

II.4 Escala de Likert

Escala comúnmente utilizada en las preguntas de un cuestionario, fue creada en 1932 que consiste en categorías que van de valores jerárquicos de menos a más o viceversa, el cual contiene afirmaciones o juicios con el fin de indicar acuerdos o desacuerdos, aprobación o desaprobación, grados de interés, de importancia, entre otros. Se recomienda que este tipo de escalas contenga al menos cinco categorías (Likert, 1932). La puntuación en la escala califica al objeto de actitud medida, el cual puede ser algún individuo, objeto, marca, concepto, símbolo, profesión, etc. Algunos ejemplos de dicha escala son los que se muestran en la Tabla 2. (Hernández-Sampieri et al., 2010)

Tabla 2. Ejemplos de las estructura de una escala de Likert

No.	ESCALA				
1	Muy de acuerdo	De acuerdo	Ni acuerdo ni en desacuerdo	En desacuerdo	Muy en desacuerdo
2	Totalmente de acuerdo	De acuerdo	Neutral	En desacuerdo	Totalmente en desacuerdo
3	Siempre	La mayoría de las veces si	Algunas veces sí, algunas veces no	La mayoría de las veces no	Nunca
4	Completamente verdadero	Verdadero	Ni falso, ni verdadero	Falso	Completamente falso
5	Definitivamente si	Probablemente si	Indeciso	Probablemente no	Definitivamente no

Fuente: Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2010). *Metodología de la investigación*. México D.F.: McGraw-Gill.



<p>*Con guiones:</p> <p><i>Muy de acuerdo</i> <i>Muy en desacuerdo</i></p> <p>_____</p>				
<p>*Con números:</p> <p><i>Muy de acuerdo</i> <i>Muy en desacuerdo</i></p> <p>_____ 5 _____ 4 _____ 3 _____ 2 _____ 1</p>				

Figura 1. Otras formas de realizar una escala

Fuente: Morales-Vallejo, P., Urosa-Sanz, B., & Blanco-Blanco, A. (2003). *Construcción de escalas de actitudes tipo Likert*. Madrid: La Muralla.

Es también habitual expresar las respuestas solo con los extremos de las categorías donde puede estar dividido por números, espacios, letras, etc., como se muestra en la Figura 1. El número de categorías de respuesta más común es cinco, pero puede ser mayor o menor, lo que sí es importante aclarar es que en general a mayor número de categorías en una escala suele haber mayor puntaje en el coeficiente de fiabilidad.

Respecto a si debe tener número de categorías par o impar, en muchos casos incluir una respuesta central facilita al individuo para mostrar indecisión, mientras que tener categorías de respuesta par también es un caso válido, ya que una ventaja es que se elimina la posibilidad de que el individuo elija la respuesta central y puede ser útil para el análisis. No existe una regla específica del número de categorías que se deban emplear, hay autores que prefiere categorías impares y otros pares. (Morales-Vallejo, Urosa-Sanz, & Blanco-Blanco, 2003)

II.5 Alpha de Cronbach

El Alpha de Cronbach (Cronbach, 1951) es una medida de consistencia interna del cuestionario que tiene la finalidad de evaluar la magnitud en que los ítems de un instrumento están correlacionados, ya que proporciona el grado de fiabilidad de un conjunto de ítems, mostrando



un valor estadístico que representa la relación de los mismos, donde si es muy cercano a uno la fiabilidad del instrumento es excelente. El Alpha de Cronbach puede ser calculado de dos formas:

- Mediante la varianza de los ítems y la varianza del puntaje total.

$$\alpha = \left[\frac{k}{k-1} \right] \left[1 - \frac{\sum_{i=1}^k S_i^2}{S_t^2} \right]$$

Dónde:

- S_i^2 representa la varianza del ítem i ,
- S_t^2 se refiere a la varianza de los valores totales observados, y
- k es el número total de ítems.

- Mediante la correlación de los ítems para el Alpha de Cronbach estandarizado.

$$\alpha_{est} = \frac{kp}{1 + p(k-1)}$$

Dónde:

- k es el número total de ítems y
- p es el promedio de las correlaciones lineales entre los ítems.

Teóricamente puede incluir valores negativos, pero por lo regular está dado en un intervalo que va de 0 a 1, todo depende del conjunto de datos con los que se cuente, así como la dispersión que tiene cada uno de los ítems.

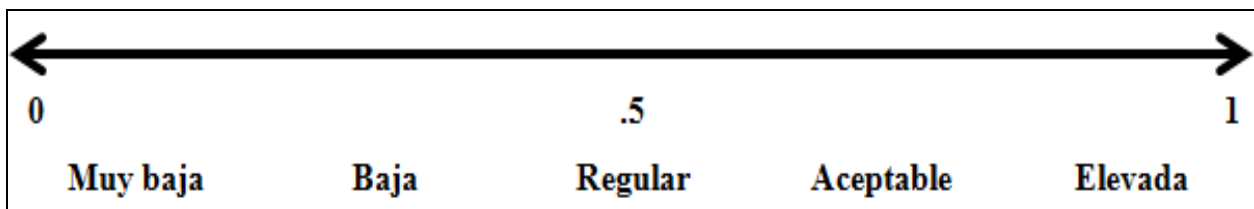


Figura 2. Interpretación del Alpha de Cronbach



Para la interpretación del Alpha de Cronbach, muchos autores han dado propuestas de alternativas pero no se ha llegado a un acuerdo general, únicamente se puede decir que a mayor valor de alpha, más elevada es la confiabilidad del instrumento, tal como se muestra en la Figura 2.

El valor de alpha se espera que sea alto, pero no muy cercado a la unidad ya que de ser el caso se considera que existe redundancia o duplicación, es decir, existen varios ítems en el instrumento que están midiendo lo mismo y por consiguiente tienen que ser eliminados. El valor de alpha es afectado por la cantidad de ítems con los que se cuenta, es por ello que a medida que incrementa el número de ítems, aumenta la varianza. Por otro lado, también se verá reflejado con el tamaño de la muestra, pues a mayor cantidad de usuarios que califiquen en la escala, mayor será la varianza esperada. (Celina-Oviedo & Campo-Arias, 2005)

II.6 Análisis del Impacto

El Análisis de Impacto es un método propuesto por Juniper y otros en 1997, que no posee demostración matemática, sin embargo, sirve para detectar sesgos de respuesta en los ítems que más fueron seleccionados por los usuarios dentro de un instrumento, este método tienen el requisito primordial que los ítems estén medidos en escala ordinal.

Supongamos que n cantidad de ítems están en escala de Likert de 5 categorías cada uno (0=Nunca, 1= Casi nunca, 2= a veces, 3= muchas veces y 4= siempre), los pasos a seguir para realizar el método son los siguientes:

- 1.-Se calcula la frecuencia de cada uno de los ítems.
- 2.- Se calcula la importancia de cada uno de los ítems del instrumento obteniendo el valor medio de cada ítem sin considerar la categoría “Nunca”.
- 3.- El impacto, siguiendo a Allen y Locker, 2002, se calcula multiplicando la frecuencia por la importancia.



Los pasos anteriores, pueden representarse en una tabla (Tabla 3) para facilitar su interpretación como la que se muestra a continuación, donde los ítems con mayor puntaje son los que mayor impacto tienen dentro del instrumento.

Tabla 3. Análisis del impacto para un conjunto de ítems

ÍTEM	FRECUENCIA	IMPORTANCIA	IMPACTO
1	Frecuencia 1	Importancia 1	Frecuencia 1 x Importancia 1
2
3
...
N	Frecuencia n	Importancia n	Frecuencia n x Importancia n

II.7 Teoría de Respuesta al Ítem (TRI)

La Teoría de Respuesta al Ítem es una forma de establecer relación de medidas observadas y el constructo con las puntuaciones de los ítems para que a través de los patrones de respuesta se pueda encontrar una estimación del valor del individuo en el constructo, de otra forma, establecen una relación matemática entre la respuesta a un ítem en específico y el nivel de aptitud o rasgo de un individuo. Para lo anterior existen diversos tipos de modelos que realizan esta funcionalidad, cada uno con sus propias características diferentes. (Martínez-Arias, Hernández-Lloreda, & Hernández-Lloreda, 2014)

La TRI es un método que sirve para conocer la funcionalidad de cada uno de los ítems, donde el interés principal es conocer si los usuarios están respondiendo correctamente los ítems (Anastasi & Urbina, 1998). Existen tres puntos importantes que se obtienen en la TRI la dificultad del ítem, la pseudoadivinación y la discriminación del ítem, que se definen:



- **Dificultad del ítem (α):** describe donde está situado el ítem en la escala de aptitud medida, esto quiere decir que representa la cantidad de aptitud que se necesita para que el usuario responda al ítem correctamente.
- **Discriminación del ítem (b):** refleja en punto en que se puede diferenciar si un usuario tiene una aptitud inferior o superior a la posición del ítem. Si un ítem no discrimina por consiguiente no es adecuado en el constructo.
- **Seudoadivinación:** representa la probabilidad de que un individuo conteste un ítem por adivinación o azar, aún con desconocimiento del propio individuo respecto al ítem.

II.7.1 Curvas características de los ítems

Si se grafica las probabilidades de que un usuario responda correctamente un ítem respecto a una función de aptitud, se obtendrá una curva característica, dicha curva se hacen para cada uno de los ítems y tienen una forma sigmoidea (Forma de S), ver figura 3. Esta curva no puede contener probabilidades menores que cero y mayores a uno, la probabilidad se aproxima a cero para niveles bajos de aptitud y se aproxima a la unidad para los niveles que son más altos.

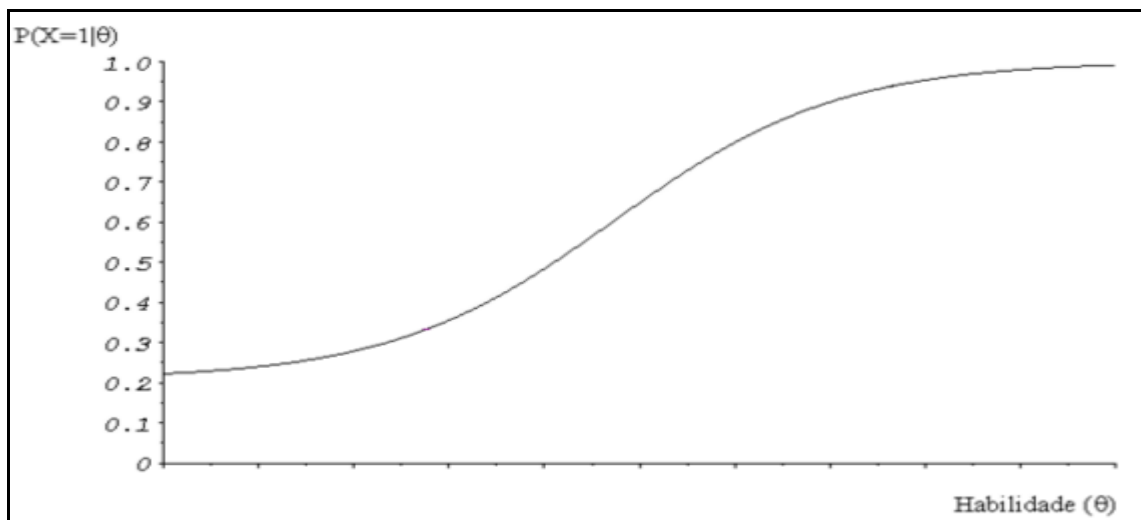


Figura 3. Ejemplo de una curva característica



II.7.2 Supuestos de la TRI

La Teoría de Respuesta al Ítem (TRI) tiene supuestos tales como la Unidimensionalidad e Independencia local (Martínez-Arias, 1995), las cuales se definen de la siguiente forma:

- **Unidimensionalidad:** Inicialmente se asume que hay un conjunto de rasgos que son responsables de la actuación de los usuarios dentro del instrumento que definen un espacio latente k dimensiones, pero se puede representar en dicho instrumento la posición de los usuarios y los ítems. En la Teoría de Respuesta al Ítem se supone que una única aptitud de medición es suficiente para poder representar los resultados de los usuarios y de los ítems. En resumen, quiere decir que un conjunto de ítems mide una sola actitud bajo estudio.
- **Independencia local:** Este supuesto significa que si se mantuviera constante las mediciones o aptitudes que explican al rendimiento del instrumento, entonces las respuestas de los examinados en un par de ítems son estadísticamente independientes, esto es, que no existe alguna relación entre las puntuaciones de los examinados a diferentes ítems del instrumento para el mismo nivel de aptitud medida.

II.7.3 Modelo de respuesta graduada de Samejima

Existen diferentes tipos de modelos que se ajustan al tipo de variable utilizada en la medición de los constructos dentro del instrumento, hay modelos para ítems medidos en escala dicotómica, nominal, ordinal, entre otros. El modelo utilizado para datos categóricos en escala ordinal es el de respuesta graduada (MRG) propuesto por Samejima (1969, 1972, 1997) para datos con más de 2 categorías ordenadas. El MRG es una generalización de los modelos de ítems en escala dicotómica a politómica pudiéndose aplicar para ítems con diferentes cantidades de categorías.

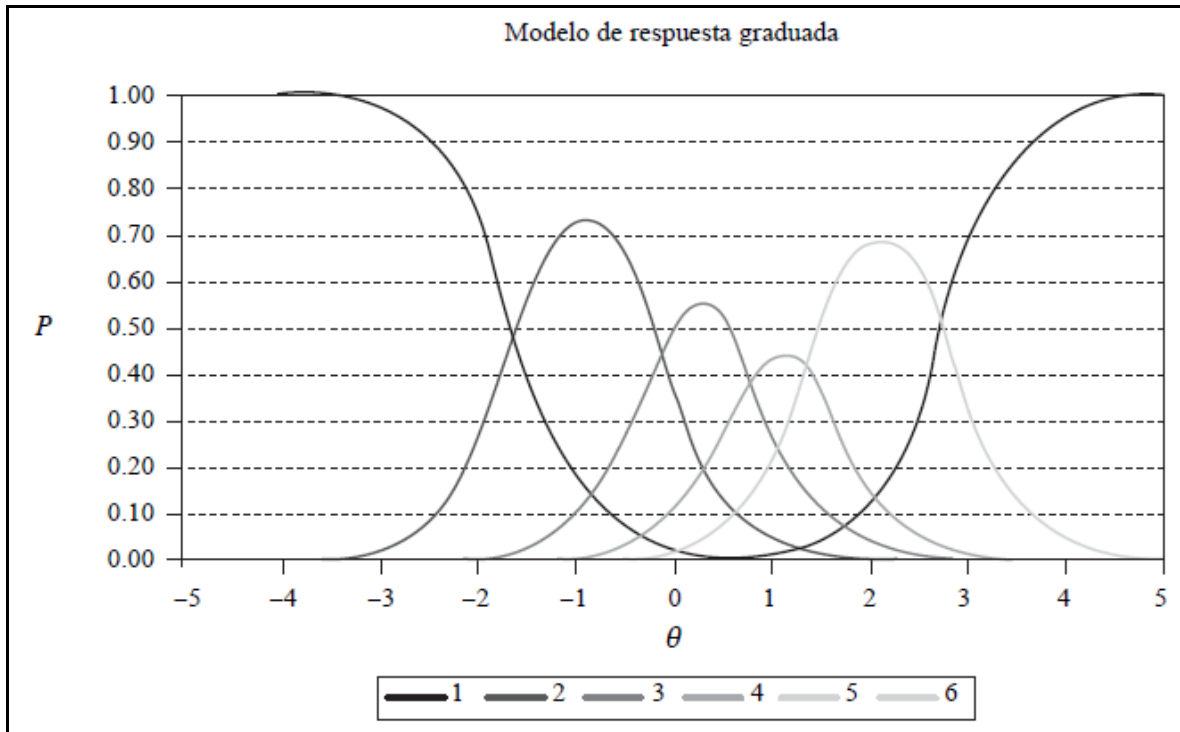


Figura 4. Curva característica en el modelo de respuesta graduada

Dicho modelo es adecuado para ítems de escalas de actitud donde la elección de una categoría más alta representa una actitud más favorable o un nivel de rasgo mayor. Una propiedad relevante en este modelo es que se pueden unir categorías de los ítems cuando sea necesario, como el caso de que no se cuente con la cantidad de individuos suficiente para realizar las estimaciones de los parámetros de una categoría, en otras palabras, sea una categoría poco seleccionada por los individuos y por consiguiente no se puedan realizar las estimaciones correspondientes. Un ejemplo de curvas características para un ítem con 6 categorías en escala ordinal es como la que se muestra en la figura 4. (Abad, Ponsoda, & Revuelta, 2006)



III. MARCO CONTEXTUAL

III.1 Herramientas para la confiabilidad y validez

Los términos validez y confiabilidad hacen énfasis al contenido exigente de un instrumento de medición, ya que al tratar temas muy serios se tiene que tener una estructura muy específica que mida lo que se pretende medir. Un concepto de fiabilidad desde la Teoría Clásica de los Test (TCT) de Spearman es que dicho coeficiente representa la correlación existente entre las puntuaciones obtenidas por individuos en dos formas paralelas del test, donde si las puntuaciones de un individuo son iguales en ambas formas paralelas, por consiguiente existía una fiabilidad perfecta.

Según la TCT la fiabilidad está asociada a los errores de medida aleatorios de puntuaciones provenientes de los sujetos en la aplicación de un test, donde si el error es pequeño la fiabilidad es alta o viceversa, en otras palabras la fiabilidad es la capacidad de realizar medidas libres de errores. En el campo de la psicometría se tiene que la fiabilidad toma en cuenta la precisión y consistencia de la medida, es por ello que si ésta es precisa y consistente por consiguiente se puede confiar en los resultados obtenidos en la aplicación de un test. (Martínez-Arias, Hernández-Lloreda, & Hernández-Lloreda, 2014)

Respecto a la validez, sus primeras definiciones se remontan en los años 30, en el desarrollo del conductismo y con la influencia del positivismo lógico en el campo de la psicología, en el momento en que lo referente a aspectos no observables se consideraban como subjetivos y no científicos, es por ello que su definición se establecía como la “*capacidad de un test para poder predecir conductas criterio observables*”. Respecto a lo anterior se vio la necesidad recurrir a procedimientos estadísticos para la validación, tales como técnicas de correlación y de diferencia entre grupos, por consiguiente varios autores empezaron a definir validez en términos de correlación.



Existen tres tipos de validez que se crearon dentro del campo de la psicología, los cuales son la validez de contenido, criterio y constructo. El término de validez de constructo fue establecido en el léxico psicométrico en 1954 en las "Recomendaciones técnicas para las pruebas psicológicas y las técnicas de diagnóstico" (APA, 1954), la cual constituyeron la primera edición de los estándares de examinación.

Posteriormente, Cronbach y Meehl en 1955 expusieron detalladamente lo que se refiere a validación de constructo. La validez de constructo tuvo y sigue teniendo mucha importancia, ya que se ha centralizado en la función que tienen los test, basándose en la teoría psicológica. Dicha validez se refiere al grado en que se mide un constructo o rasgo teórico. Algunos ejemplos de constructos son la aptitud académica, la comprensión mecánica, la fluidez verbal, etc.

Éste tipo de validez se llegó a reconocer como fundamental y aunque algunas de las técnicas que se emplean dentro de la investigación de la validez de constructo se conocían desde tiempo atrás, el campo de aplicación se ha ampliado para conocer una mayor diversidad de procedimientos. Uno de los primeros métodos para validar escalas de medición de los ítems se presentan en los trabajos de Hoyt en 1941 y de Guttman en 1945, surgió así el método Alpha de Cronbach en 1951, que mide la fiabilidad de los ítems en un constructo. (Anastassi & Urbina, 1998)

A principios del siglo XX con los trabajos de Spearman y Thurstone en los años cuarenta, el análisis factorial fue establecido como la herramienta estadística utilizada en psicología para tratar de identificar factores que influyen en la inteligencia, por consiguiente Thurstone propuso utilizar ésta herramienta para explicar las correlaciones entre diferentes ítems de un test de inteligencia, permitiéndole observar y diferenciar las capacidades espaciales, verbales y numéricas como factores de inteligencia. (Meneses et al., 2013)



En la década de los 60's surge de manera vertiginosa en áreas de caracteres político, social, psicológico y educativo, el estudio de posibles sesgos en los ítems de un test en determinados grupos de una población específica. No todas las diferencias entre grupos son consecuencia del sesgo sino que pueden ser legítimas. Las diferencias entre grupos es una cuestión que pocos investigadores han analizado, la mayoría de estos estudios se basan en el funcionamiento diferencial de los ítems.

En psicometría se ha trabajado en el desarrollo de técnicas de detección de funcionamiento diferencial entre grupos. Los procedimientos se basan en general en la Teoría de Respuesta al Ítem y tablas de contingencia, también últimamente se ha utilizado en análisis Factorial Confirmatorio, donde como variables grupo representativas son el género, cultura, etc. (Gómez-Benito & Navas-Ara, 1998)

Asimismo, existen métodos que sirven para realizar una mejor y detallada validez de un cuestionario, es decir, se adentran más a detalle a los ítems para sacar mejor provecho de la evidencia con la que se cuenta, pero que comúnmente no se aplican ya que su difusión no es muy grande, pero el beneficio es mayor en la obtención mejores resultados. La poca verificación del contenido de un instrumento es un gran problema, ya que si no se hace de la manera adecuada la información recabada no será muy buena y por consiguiente no se podría hacer inferencia a cierta población.

En la actualidad, existen varios estudios donde no solo utilizan diferentes herramientas estadísticas para conocer la confiabilidad y la validez de los cuestionarios, sino que recalcan la importancia de realizar varios procedimientos conjuntamente que ayuden a una mejor validación de constructo, claros ejemplos de aplicación de diferentes métodos estadísticos así como la unión de técnicas son las siguientes:

- Sanchez-Barba (2008) realizó una versión abreviada de un instrumento llamado QUALEFFO, de veintitrés ítems en lugar de treinta y cinco, con menos categorías de



respuesta en once de ellos, y alta validez Factorial y fiabilidad, lo cual propuso un importante ahorro en el tiempo, tanto del médico de atención primaria, como del paciente y un descenso considerable en los gastos asociados, esto a través de la Teoría de Respuesta al Ítem.

- Vicente-Galindo (2011) realizó un estudio donde el objetivo principal fue comparar diferentes procedimientos utilizados para simplificar cuestionarios de calidad de vida relacionada con la salud y analizar si proporcionan resultados equivalentes. Los tres procedimientos, que hasta ahora se han usado por separado, deberían utilizarse conjuntamente para poder disponer de cuestionarios con alta validez factorial, alta validez de contenido y alto poder discriminante.
- Batista-Foguet y otros (2014) destacan la necesidad de precisar los términos esenciales de la medición (fiabilidad y validez) clarificando las deficiencias en que incurre la metodología tradicional en su uso. Señala las limitaciones de la evaluación de la fiabilidad mediante el Alpha de Cronbach o la laxitud en la valoración cuantitativa de la validez mediante el Análisis Factorial Exploratorio.
- Escurra-Mayaut y Salas-Blas (2014) realizaron un estudio con el propósito de diseñar, construir y validar el cuestionario de Adicción a Redes Sociales (ARS) mediante la aplicación del modelo de la Teoría de Respuesta al Ítem (TRI) para ítems politómicos de respuesta graduada.
- Huerta-Pacheco en 2015 propuso una evaluación estadística basada en cuestionarios diseñados para distinguir características esenciales de un juego de razonamiento espacial que den pauta a la identificación de patrones de comportamiento de la interacción de los usuarios con un juego serio de razonamiento espacial. Utilizando conjuntamente correlaciones policóricas, alpha de Cronbach, impacto de los ítems y Teoría de Respuesta al Ítem.



- Vicente-Galindo y otros (2015) realizaron un estudio comparativo de tres métodos estadísticos como el Análisis Factorial, Análisis del Impacto y TRI, utilizados en la simplificación de cuestionarios genéricos o específicos de Calidad de Vida Relacionada con la Salud. Donde se pone de manifiesto una clara discordancia entre los resultados de los tres métodos de análisis y sugiere la conveniencia de utilizarlos, no individualmente, sino conjuntamente.

Las investigaciones anteriormente mencionadas, muestran como la Teoría de Respuesta al Ítem es una herramienta utilizada en el ámbito de la fiabilidad y validez de instrumentos no solo en el contexto de la psicometría sino que se ha extendido a otros ámbitos de estudio. La Teoría de Respuesta al Ítem se ha comparado con el análisis factorial encontrando que se puede llegar al mismo resultado utilizando una u otra herramienta. Además, se ha mencionado la necesidad de utilizar conjuntamente varias herramientas estadísticas para verificar la fiabilidad y validez de los instrumentos de medición, ya que las diversas herramientas proporcionan características diferentes de un cuestionario.

III.2 Importancia de las encuestas en estudios de opinión

La opinión pública se ha vuelto objeto de estudio el cual ha adquirido un concepto controvertido en el que su significado es difícil de expresar, debido al contexto histórico social. Desde hace tiempo la representación de la opinión pública está muy cristalizada y apegada a los instrumentos de medición, ya que es difícil imaginar la información sobre opinión sin que mantenga un formato de datos recabados de un sondeo, esto quiere decir que la opinión pública incluye el método de estudio. El proceso de investigación de la opinión pública se entiende como pasos en donde los datos son recabados a través de los sondeos de opinión. Respecto a lo anterior existe una fuerte relación entre la opinión pública y las encuestas. (Mañas, 2012)



El uso de las encuestas como formas adecuadas dentro de los estudios de opinión es consecuente de un proceso social que tiene referencia histórica. En dicho contexto histórico se encuentra una alineación entre la estadística y la medición de sondeos de actitudes y opiniones, esta asociación tiene lugar en las primeras décadas del siglo XX, en donde con el desarrollo del mercado interno en los Estados Unidos y con el consumo de masas, se da la aparición de institutos de investigación en estadística con el fin de aplicar la metodología estadística a los estudios de mercado para poder aplicarlo en el análisis de la demanda.

Otra área que adquirió el mismo interés fue la psicología social donde la estadística estaba ligada a la investigación de la mente humana. Debido al auge de la estadística en los estudios de marketing y su utilización en la medición de actitudes de la población, surgió el interés de los partidos y gobiernos para poder conocer la inclinación de las personas en los votos electorales para poder anticipar los resultados. Los sondeos electorales terminan ocupando un lugar muy importante en las encuestas de opinión, ya que los resultados recabados serían los más presentados por los medios de comunicación.

Mediante el uso de muestreos estadísticos representativos de la población, para conocer los resultados de las encuestas tuvo un efecto positivo ante el público para demostrar que las encuestas son un instrumento apropiado y eficaz para los estudios electorales. Respecto a lo anterior, es muy evidente que la utilización de metodología estadística en los estudios de la demanda del mercado en la sociedad del consumo de masas y en la investigación, incluyendo la psicología social de las actitudes de la población, dan uso de la encuesta estadística como un método institucional y académicamente adecuado que fue socialmente aceptado en la opinión pública desde la Segunda Guerra Mundial hasta la actualidad. (Mañas, 2012)

La utilización de la estadística en muestras que son representativas de una población tienen fundamentos matemáticos de confiabilidad, lo cual se proyectará como método sólido y científico confiable. El rigor científico y estricto de las matemáticas favoreció en gran medida a



la percepción que se tenía de la estadística como una herramienta útil y adecuada para la investigación social. Con lo anterior se puede decir que utilizar la estadística en la investigación de cuestiones subjetivas como las actitudes y opiniones, se puede llegar a conocer la realidad de la opinión pública como objeto de estudio.



IV. METODOLOGÍA

En este apartado se mencionan algunos pasos utilizados para conocer la confiabilidad y la validez de un instrumento. Las técnicas presentadas son métodos estadísticos empleados por algunos autores que hacen hincapié en utilizar diferentes técnicas en conjunto para conocer la confiabilidad y validez de los instrumentos, y así encontrar los ítems que aportan mayor información para realizar una nueva versión del instrumento de medición reducido pero sin perder información. Por otro lado, también se describe el cuestionario que se utilizará como ejemplo para la aplicación de las herramientas estadísticas, basado en un caso real.

IV.1 Descripción del instrumento

El instrumento utilizado proviene de una encuesta de opinión generado en el proyecto PROMEP del Dr. Claudio Rafael Castro López de la Universidad Veracruzana, el cual está denominado “Calidad de vida de los habitantes de la ciudad de Xalapa, Veracruz (Definición de sus dimensiones, creación, confiabilidad, prueba y aplicación del constructo)” realizado en el periodo de julio del 2007 a junio del 2008, dicha encuesta se aplicó como piloteo para evaluar la confiabilidad, calibración y validez del instrumento. La selección de los encuestados se realizó a través de un muestreo estadístico de 315 individuos de ambos sexos.

El instrumento consta de 37 preguntas y varias sub-preguntas con diferentes escalas de medición. Inicialmente contiene datos demográficos y tres constructos, donde el primero corresponde a Salud, el segundo a Bienestar Personal y por último Entorno y Medio Ambiente. Éste último constructo es el que se utilizó como ejemplo, ya que la pregunta 37 corresponde a un sub-constructo de 16 ítems que están en escala ordinal, característica necesaria para realizar las técnicas mencionadas en este proyecto (Ver Anexo 1). Los ítems del constructo se pueden dividir en 4 contextos:



- 1) Los ítems del 1 al 4 hacen referencia al grado de satisfacción con diferentes sistemas como el educativo, judicial, seguridad social y de asistencia médica en los centros de salud.
- 2) En los ítems del 4 al 8 se habla del grado de satisfacción con la seguridad, limpieza, vialidad peatonal e infraestructura vial.
- 3) Los ítems 9 al 12 se presenta el grado de satisfacción con diferentes infraestructuras públicas.
- 4) Por último, el contexto del ítem 13 al 16 es sobre el transporte público y la gestión del gobierno municipal, estatal y federal.

IV.2 Descripción de los datos

El constructo seleccionado contiene 16 ítems (Anexo 1) donde todas son variables cualitativas medidas en escala ordinal, teniendo como posibles respuestas once categorías que están jerarquizadas de cero a diez donde cero representa totalmente insatisfecho y diez totalmente satisfecho. Para facilitar su tratamiento estadístico en la paquetería correspondiente, se decidió unir las categorías de los extremos, es decir, la categoría cero con la uno y la nueve con la diez, teniendo una nueva jerarquización en la escala que ahora va de 1 a 9, esto no alteró el comportamiento de la información. Dicha descripción de variables (Cuevas-Acosta & Hernández-González, 2013) se muestra en la tabla siguiente:

Tabla 4. Descripción de variables

Grado de satisfacción:	Tipo de variable	Escala de medición	Codificación
1. Con la asistencia médica de los centros de salud, hospitales, centros especializados.	Cualitativa	Ordinal	De 1 a 9
2. Con el sistema educativo.	Cualitativa	Ordinal	De 1 a 9
3. Con el Sistema Judicial (Justicia)	Cualitativa	Ordinal	De 1 a 9
4. Con el Sistema de Seguridad Social (pensiones y protección social).	Cualitativa	Ordinal	De 1 a 9
5. Con la Seguridad Ciudadana.	Cualitativa	Ordinal	De 1 a 9



6. Con la limpieza y decoración de los espacios públicos.	Cualitativa	Ordinal	De 1 a 9
7. Con los cambios que ha habido en la infraestructura vial	Cualitativa	Ordinal	De 1 a 9
8. Con los cambios que ha habido en la vialidad peatonal	Cualitativa	Ordinal	De 1 a 9
9. Con las Infraestructuras Públicas, para el entretenimiento y esparcimiento.	Cualitativa	Ordinal	De 1 a 9
10. Con la Infraestructura Pública, orientada para los ancianos (as).	Cualitativa	Ordinal	De 1 a 9
11. Con las Infraestructuras Públicas, dedicadas a los discapacitados(as) o minusválidos(as).	Cualitativa	Ordinal	De 1 a 9
12. Con los puntos de recreación existentes en su comunidad (restaurantes, teatros, cines)	Cualitativa	Ordinal	De 1 a 9
13. Con el Transporte Público.	Cualitativa	Ordinal	De 1 a 9
14. Con la Gestión (trabajo) del Gobierno Municipal.	Cualitativa	Ordinal	De 1 a 9
15. Con la Gestión (trabajo) del Gobierno Estatal.	Cualitativa	Ordinal	De 1 a 9
16. Con la Gestión (trabajo) del Gobierno Federal.	Cualitativa	Ordinal	De 1 a 9

IV.3 Herramientas estadísticas

Las técnicas empleadas son las estadísticas descriptivas e inferenciales. Gran parte de la descripción de los métodos estadísticos se mencionaron anteriormente, otros se describen brevemente en este apartado. Se muestran los pasos de cómo fueron utilizadas cada una de ellas para cumplir los objetivos de este proyecto. Para la ejecución de cada técnica estadística se utilizaron los paquetes SPSS v16.0 (SPSS Inc. Released, 2007), MULTILOG 7.03 (Thissen, 1991) y hojas de cálculo de Excel (Microsoft Excel, 2010).



IV.3.1 Estadística univariante

De manera inicial y descriptiva se realizaron tablas de frecuencia para conocer las categorías con mayor puntuación por los individuos. También se empleó un histograma para representar de manera resumida el comportamiento de todo el constructo.

IV.3.2 Estadística bivariante

Subcuentemente, se realizó el ji-cuadrado que es un contraste de hipótesis estadístico para conocer si dos variables categóricas están asociadas o no (Wackerly, Mendenhall III & Scheaffer, 2010), ya que permite conocer si existe suficiente evidencia para decir que hay asociación o no entre los ítems, dado que son variables cualitativas, esto como requisito para técnicas que posteriormente se utilizaran.

Luego se emplearon correlaciones policóricas, ya que son un tipo especial de correlación que permite obtener un coeficiente de relación entre dos variables en escala ordinal (Abad, Ponsoda, & Revuelta, 2006), también se realizaron la asociación de Spearman (Spearman, 1904) que es un método no paramétrico que utiliza rangos ordenados para mostrar con coeficiente de correlación alternativo al r de Pearson (Siegel & John, 1998), debido a que es una escala ordinal se pueden establecer pesos a cada categoría y así establecer un orden entre ellos.

IV.3.3 Estadística multivariante

Como coeficiente de fiabilidad se utilizó el Alpha de Cronbach para conocer la consistencia interna del cuestionario, a su vez se calculó el mismo coeficiente en el caso de que un ítem fuera eliminado para identificar cuales no están aportando al instrumento.

Se usó la técnica de Impacto de los Ítems con el fin de conocer la frecuencia, importancia y posteriormente el impacto que posee cada uno de los ítems donde el valor a considerar nulo



será la puntuación cinco de la escala, la cual no presenta tendencia de opinión ya sea positiva o negativa.

Por último, se hará uso del modelo de Respuesta Graduada de Samejima (1969, 1972, 1997) para ítems con escala ordinal de la Teoría de Respuesta al Ítem, para conocer la fiabilidad de la escala empleada y la validez del mismo. Para poder emplear esta técnica se comprobaron los supuestos de independencia local y unidimensionalidad, realizando un gráfico de sedimentación representando los porcentajes de variabilidad existentes en cada dimensión. Se utilizó el criterio de Carmines y Zeller (1979) para comprobar el supuesto de unidimensionalidad, dicho criterio consiste en que el primer factor debe contener al menos el 40% de la variabilidad total y que los componentes siguientes tienen una tendencia en declive.

Inicialmente, se empleará la técnica con los datos originales para conocer los parámetros resultantes, donde en base al parámetro de discriminación y el valor promedio de la función de información general se seleccionaran los ítems que aportan mayor información y se descartaran los que no. De manera visual a través de las curvas características se observará el número de categorías ideal que describe el comportamiento de cada ítem para posteriormente emplear de nuevo la Teoría de Respuesta al Ítem a los ítems con las categorías reducidas.

Al final de este, se propondrá una nueva versión del constructo que contenga cada una de las características encontradas en las herramientas empleadas, para que se tome en cuenta en la reestructuración del instrumento.



V. RESULTADOS

En este apartado se muestran los resultados de cada una de las herramientas estadísticas aplicadas en este proyecto para conocer la confiabilidad y validez de un conjunto de ítems pertenecientes al constructo de entorno y medio ambiente de un instrumento de calidad de vida. Como primeros resultados se muestran las estadísticas univariantes que dan a conocer el comportamiento individual de los ítems que conforman el constructo, para dar paso a responder los objetivos planteados en este proyecto.

En la Tabla 5 se observan las frecuencias y porcentajes de los ítems del 1 al 4 correspondientes al nivel de satisfacción con la asistencia médica y diferentes sistemas como el educativo, judicial y de seguridad social. Se muestran niveles de satisfacción altos para tres de los 4 ítems, ya que la tendencia de respuesta en la escala corresponde a calificaciones entre 8 y 9 (Muy satisfecho). Respecto a la satisfacción con el Sistema Judicial, la mayor proporción de individuos puntuó en la categoría 5, que hace referencia a una satisfacción regular.

Tabla 5. Porcentajes y frecuencias de los ítems del 1 al 4.

Grado de Satisfacción		1	2	3	4	5	6	7	8	9
1. Con la asistencia médica de los centros de salud, hospitales, centros especializados.	<i>Frec</i>	23.00	3.00	6.00	8.00	46.00	37.00	45.00	85.00	62.00
	<i>%</i>	7.30	0.95	1.90	2.54	14.60	11.75	14.29	26.98	19.68
2. Con el sistema educativo.	<i>Frec</i>	10.00	3.00	6.00	5.00	25.00	36.00	56.00	78.00	96.00
	<i>%</i>	3.17	0.95	1.90	1.59	7.94	11.43	17.78	24.76	30.48
3. Con el Sistema Judicial (Justicia)	<i>Frec</i>	29.00	12.00	9.00	12.00	65.00	43.00	60.00	55.00	30.00
	<i>%</i>	9.21	3.81	2.86	3.81	20.63	13.65	19.05	17.46	9.52
4. Con el Sistema de Seguridad Social (pensiones y protección social).	<i>Frec</i>	28.00	8.00	5.00	11.00	56.00	51.00	53.00	67.00	36.00
	<i>%</i>	8.89	2.54	1.59	3.49	17.78	16.19	16.83	21.27	11.43

Se presenta en la Tabla 6, las estadísticas descriptivas de los ítems del 5 al 8 que hacen referencia al grado de satisfacción con la seguridad, limpieza, infraestructura vial y vialidad peatonal de la ciudad de Xalapa, observando que la mayor proporción de personas tienen niveles de satisfacción altos, debido a que los porcentajes más altos se encuentran en las calificaciones 7, 8 y 9.



Tabla 6. Porcentajes y frecuencias del ítem 5 al 8

Grado de Satisfacción		1	2	3	4	5	6	7	8	9
5. Con la Seguridad Ciudadana.	<i>Frec</i>	25.00	11.00	12.00	13.00	50.00	61.00	64.00	50.00	29.00
	<i>%</i>	7.94	3.49	3.81	4.13	15.87	19.37	20.32	15.87	9.21
6. Con la limpieza y decoración de los espacios públicos.	<i>Frec</i>	22.00	6.00	5.00	7.00	48.00	42.00	68.00	73.00	44.00
	<i>%</i>	6.98	1.90	1.59	2.22	15.24	13.33	21.59	23.17	13.97
7. Con los cambios que ha habido en la infraestructura vial	<i>Frec</i>	15.00	4.00	4.00	5.00	39.00	40.00	56.00	66.00	86.00
	<i>%</i>	4.76	1.27	1.27	1.59	12.38	12.70	17.78	20.95	27.30
8. Con los cambios que ha habido en la vialidad peatonal	<i>Frec</i>	12.00	5.00	4.00	9.00	45.00	42.00	59.00	75.00	64.00
	<i>%</i>	3.81	1.59	1.27	2.86	14.29	13.33	18.73	23.81	20.32

La Tabla 7 contiene porcentajes y frecuencias de los ítems del 9 al 12 correspondientes al nivel de satisfacción con los puntos recreación e infraestructuras públicas orientadas al entretenimiento, los ancianos y los discapacitados. Las personas calificaron con puntajes de 7, 8 y 9 a los ítems presentados, mostrando niveles entre satisfechos y muy satisfechos.

Tabla 7. Porcentajes y frecuencias del ítem 9 al 12

Grado de Satisfacción		1	2	3	4	5	6	7	8	9
9. Con las Infraestructuras Públicas, para el entretenimiento y esparcimiento.	<i>Frec</i>	16.00	6.00	5.00	11.00	50.00	54.00	47.00	73.00	53.00
	<i>%</i>	5.08	1.90	1.59	3.49	15.87	17.14	14.92	23.17	16.83
10. Con la Infraestructura Pública, orientada para los ancianos (as).	<i>Frec</i>	26.00	9.00	9.00	18.00	44.00	48.00	63.00	56.00	42.00
	<i>%</i>	8.25	2.86	2.86	5.71	13.97	15.24	20.00	17.78	13.33
11. Con las Infraestructuras Públicas, dedicadas a los discapacitados(as) o minusválidos(as).	<i>Frec</i>	35.00	9.00	6.00	17.00	43.00	50.00	59.00	52.00	44.00
	<i>%</i>	11.11	2.86	1.90	5.40	13.65	15.87	18.73	16.51	13.97
12. Con los puntos de recreación existentes en su comunidad (restaurantes, teatros, cines)	<i>Frec</i>	34.00	7.00	10.00	15.00	26.00	41.00	41.00	60.00	81.00
	<i>%</i>	10.79	2.22	3.17	4.76	8.25	13.02	13.02	19.05	25.71

Por último, las estadísticas descriptivas de los ítems del 13 al 16 referentes al grado de satisfacción con el transporte público y gestión del gobierno municipal, estatal y federal, muestran un comportamiento similar a los ítems anteriores, esto es, la mayor proporción de personas que puntuaron en la escala, calificaron con grados altos de satisfacción (7 y 9) cada uno de los ítems. (Ver Tabla 8)



Tabla 8. Porcentajes y frecuencias del ítem 13 al 16

Grado de Satisfacción		1	2	3	4	5	6	7	8	9
13. Con el Transporte Público.	<i>Frec</i>	22.00	7.00	8.00	18.00	38.00	32.00	52.00	54.00	84.00
	%	6.98	2.22	2.54	5.71	12.06	10.16	16.51	17.14	26.67
14. Con la Gestión (trabajo) del Gobierno Municipal.	<i>Frec</i>	22.00	5.00	14.00	10.00	38.00	51.00	66.00	64.00	45.00
	%	6.98	1.59	4.44	3.17	12.06	16.19	20.95	20.32	14.29
15. Con la Gestión (trabajo) del Gobierno Estatal.	<i>Frec</i>	17.00	6.00	14.00	10.00	35.00	31.00	63.00	62.00	77.00
	%	5.40	1.90	4.44	3.17	11.11	9.84	20.00	19.68	24.44
16. Con la Gestión (trabajo) del Gobierno Federal.	<i>Frec</i>	28.00	6.00	14.00	10.00	36.00	37.00	72.00	57.00	55.00
	%	8.89	1.90	4.44	3.17	11.43	11.75	22.86	18.10	17.46

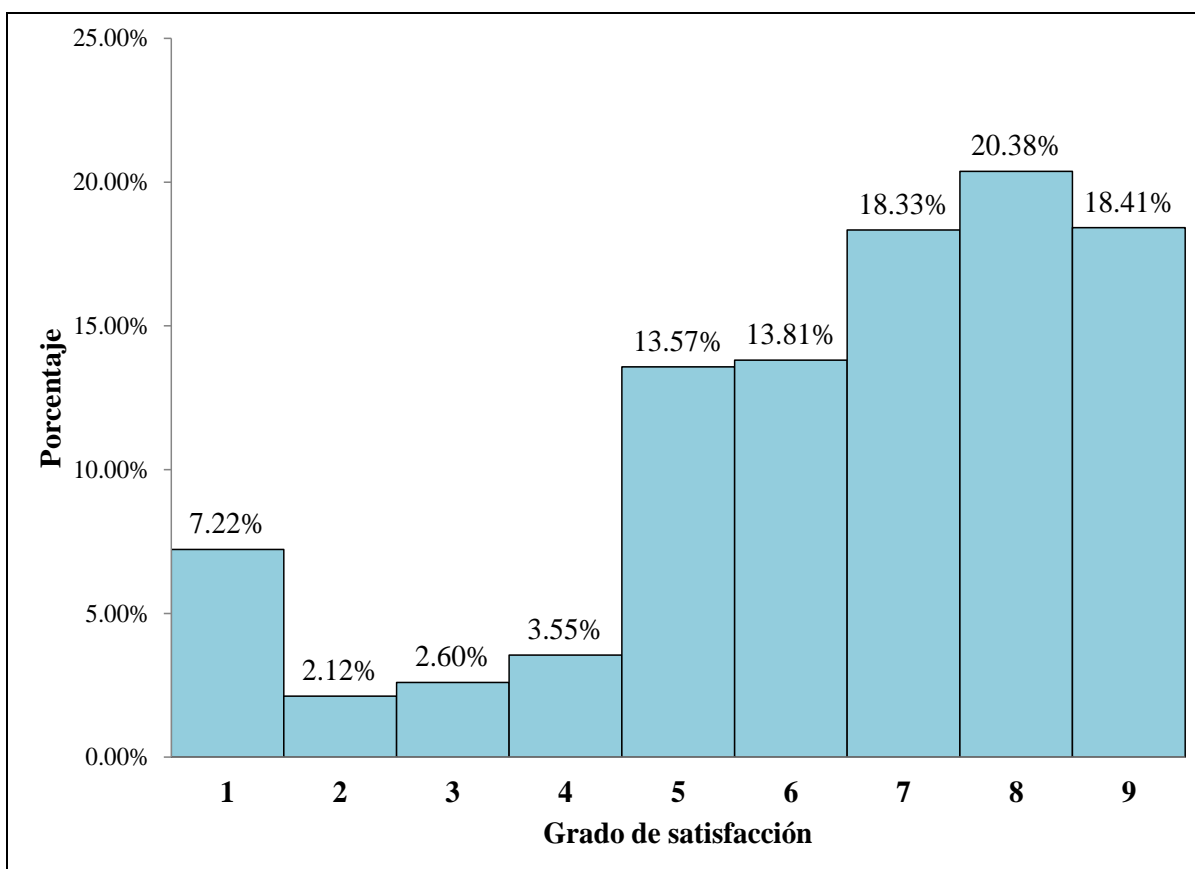


Figura 5. Distribución porcentual promedio de puntuaciones en la escala ordinal

En la Figura 5 se muestra la distribución porcentual promedio de todos los ítems en conjunto y se aprecia un comportamiento positivo, esto quiere decir que a valores bajos de la escala menores puntuaciones de los individuos y a valores altos de la escala mayores puntuaciones de las personas. Los encuestados puntuaron más alto la calificación 1 referente a nada satisfecho



que el grado de satisfacción 2, 3 y 4. La calificación 8 fue la más seleccionada por los ciudadanos. En general se observa que las personas se mostraron satisfechas y muy satisfechas con cada una de las preguntas (ítems) presentadas.

Luego de conocer las estadísticas univariantes, se presentan a continuación las herramientas bivariantes para poder apreciar cómo se relacionan las variables entre ellas mismas. Inicialmente se observa en la Tabla 8 las asociaciones de Spearman entre pares de ítems del constructo, identificando con color rojo varias correlaciones positivas con valores de .5 en adelante, esto da indicación de poder aplicar la Teoría de Respuesta al Ítem, ya que se requiere que exista relación entre varios pares de ítems.

Tabla 8. Asociaciones de Spearman entre los ítems

Ítem	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1.00															
2	0.52	1.00														
3	0.47	0.48	1.00													
4	0.47	0.53	0.56	1.00												
5	0.40	0.42	0.57	0.67	1.00											
6	0.33	0.38	0.46	0.54	0.62	1.00										
7	0.34	0.41	0.45	0.50	0.51	0.61	1.00									
8	0.36	0.44	0.43	0.50	0.53	0.56	0.77	1.00								
9	0.40	0.37	0.39	0.51	0.46	0.52	0.61	0.63	1.00							
10	0.48	0.37	0.41	0.58	0.50	0.53	0.51	0.53	0.67	1.00						
11	0.43	0.34	0.33	0.56	0.48	0.42	0.46	0.50	0.62	0.85	1.00					
12	0.23	0.33	0.35	0.46	0.43	0.38	0.46	0.42	0.55	0.52	0.53	1.00				
13	0.37	0.42	0.47	0.42	0.39	0.42	0.43	0.46	0.40	0.38	0.33	0.26	1.00			
14	0.36	0.34	0.40	0.50	0.47	0.44	0.46	0.50	0.46	0.54	0.55	0.43	0.40	1.00		
15	0.27	0.32	0.37	0.44	0.42	0.36	0.40	0.45	0.43	0.44	0.42	0.42	0.36	0.77	1.00	
16	0.34	0.38	0.42	0.51	0.47	0.39	0.44	0.47	0.37	0.43	0.41	0.35	0.36	0.69	0.79	1.00

También se muestran las correlaciones policóricas (Tabla 9), para tener más evidencia de la presencia de relaciones entre los ítems del constructo, ya que este tipo de correlaciones se ajusta bien al tipo de escala empleada en cada pregunta (Ordinal), mostrando con color rojo



aqueellos pares de ítems con una correlación igual o mayor a .5. Con esto se sustenta la presencia de correlaciones positivas entre los ítems.

Tabla 9. Correlaciones policóricas entre los ítems

Ítem	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1.00															
2	0.57	1.00														
3	0.52	0.56	1.00													
4	0.51	0.60	0.65	1.00												
5	0.43	0.47	0.63	0.71	1.00											
6	0.37	0.42	0.51	0.58	0.64	1.00										
7	0.38	0.48	0.47	0.53	0.51	0.66	1.00									
8	0.39	0.49	0.48	0.53	0.54	0.58	0.83	1.00								
9	0.46	0.42	0.45	0.56	0.48	0.58	0.67	0.69	1.00							
10	0.52	0.42	0.47	0.63	0.53	0.55	0.56	0.57	0.71	1.00						
11	0.46	0.39	0.37	0.59	0.50	0.45	0.50	0.54	0.65	0.86	1.00					
12	0.31	0.39	0.41	0.49	0.45	0.44	0.50	0.46	0.58	0.56	0.54	1.00				
13	0.39	0.50	0.51	0.46	0.40	0.45	0.48	0.53	0.47	0.43	0.37	0.34	1.00			
14	0.42	0.38	0.46	0.55	0.52	0.45	0.47	0.53	0.53	0.59	0.61	0.49	0.46	1.00		
15	0.30	0.33	0.39	0.44	0.43	0.36	0.41	0.48	0.47	0.47	0.46	0.45	0.39	0.77	1.00	
16	0.38	0.41	0.46	0.52	0.50	0.40	0.46	0.48	0.43	0.47	0.44	0.39	0.38	0.70	0.82	1.00

Por otro lado, se realizó el estadístico ji-cuadrado para todos los pares de ítems, de los cuales se contó con suficiente evidencia para decir que todos los contrastes de hipótesis estadísticos resultaron ser significativos a una confiabilidad del 95%, por lo que se omitió presentarlo.

Consiguientemente se presentan los resultados multivariantes, empezando con mostrar el Alpha de Cronbach general del constructo (Tabla 10) que es de .93 el cual indica una alta confiabilidad, mientras que en la Tabla 11 se muestra el valor correspondiente de Alpha en el caso de que un ítem sea eliminado teniendo que los valores son alrededores de .92 los cuales son inferiores al general por lo que no se identifican qué ítems podrían ser eliminados del constructo en una aplicación posterior para mejorar su fiabilidad.



Tabla 10. Alpha de Cronbach general

	Valor estadístico
Alpha de Cronbach	.931

Tabla 11. Alpha de Cronbach si se elimina alguno de los ítems

Ítems de grado de satisfacción	Alpha de Cronbach si el ítem es eliminado
1. Con la asistencia médica de los centros de salud, hospitales, centros especializados.	0.929
2. Con el sistema educativo.	0.928
3. Con el Sistema Judicial (Justicia)	0.927
4. Con el Sistema de Seguridad Social (pensiones y protección social).	0.924
5. Con la Seguridad Ciudadana.	0.925
6. Con la limpieza y decoración de los espacios públicos.	0.926
7. Con los cambios que ha habido en la infraestructura vial	0.925
8. Con los cambios que ha habido en la vialidad peatonal	0.925
9. Con las Infraestructuras Públicas, para el entretenimiento y esparcimiento.	0.925
10. Con la Infraestructura Pública, orientada para los ancianos (as).	0.924
11. Con las Infraestructuras Públicas, dedicadas a los discapacitados(as) o minusválidos(as).	0.925
12. Con los puntos de recreación existentes en su comunidad (restaurantes, teatros, cines)	0.929
13. Con el Transporte Público.	0.929
14. Con la Gestión (trabajo) del Gobierno Municipal.	0.925
15. Con la Gestión (trabajo) del Gobierno Estatal.	0.927
16. Con la Gestión (trabajo) del Gobierno Federal.	0.927

Después de conocer la confiabilidad se presenta el análisis del Impacto de los Ítems (Tabla 12), el cual consiste en el cálculo de la frecuencia e importancia para después conocer el impacto tomando como valor nulo la categoría 5 ya dicha categoría no muestra un grado de satisfacción o insatisfacción. El valor promedio del impacto corresponde a 573.29, donde aquellos que son mayores a este valor son los que mayor impacto tienen en el constructo desde vista de los sesgos que existen en ellos. Dichos ítems son el 1, 2, 7, 8, 12, 13 y 15.



Tabla 12. Impacto de los ítems

Ítem	Frecuencia	Importancia	Impacto
1	85.397	6.892	588.571
2	92.063	7.414	682.540
3	79.365	6.064	481.270
4	82.222	6.332	520.635
5	84.127	6.075	511.111
6	84.762	6.685	566.667
7	87.619	7.207	631.429
8	85.714	7.078	606.667
9	84.127	6.796	571.746
10	86.032	6.266	539.048
11	86.349	6.118	528.254
12	91.746	6.505	596.825
13	87.937	6.773	595.556
14	87.937	6.495	571.111
15	88.889	6.882	611.746
16	88.571	6.448	571.111

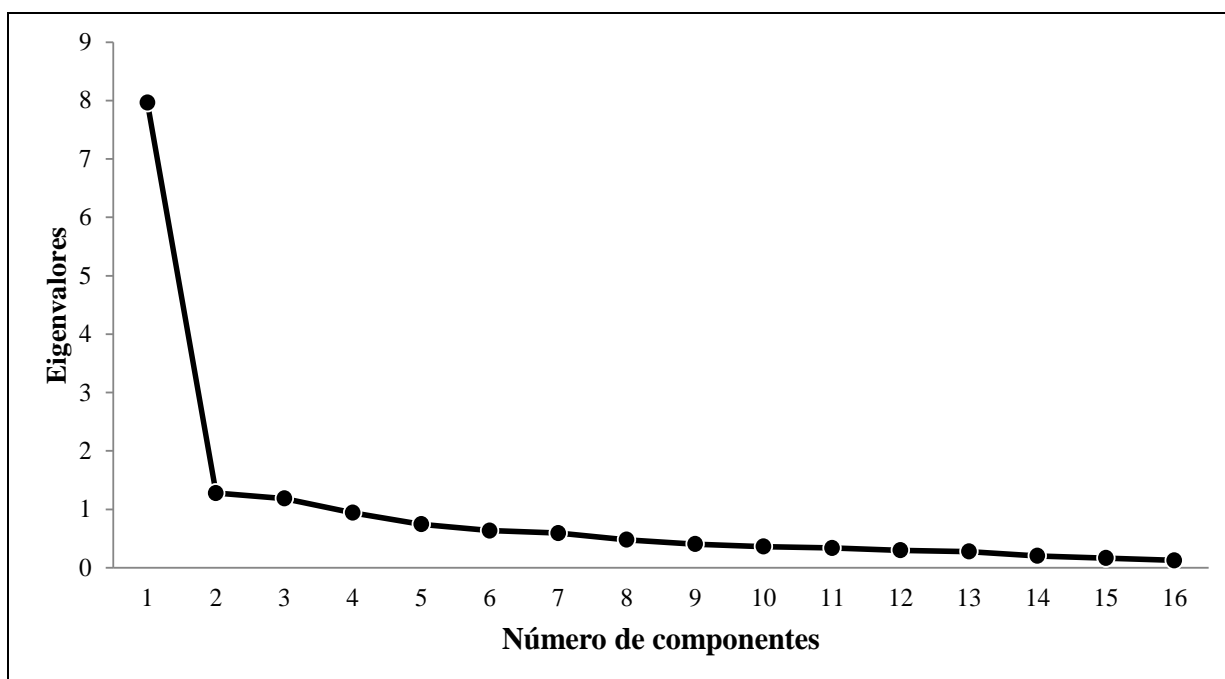


Figura 6. Valores propios asociados a cada componente representados en el gráfico de sedimentación



Uno de los requisitos esenciales para poder emplear la Teoría de Respuesta al ítem es el supuesto de unidimensionalidad, para comprobarlo en la Figura 6 se muestra el gráfico de sedimentación con los eigenvalores correspondientes a cada componente, en el que se observa que la mayor varianza explicada se encuentran en los primeros componentes y que con 3 componentes es donde se da la inclinación de la tendencia.

El supuesto requiere que todos los ítems estén contenidos en un solo factor, pero dado que el constructo utilizado tiene diferentes mediciones como lo son los grados de satisfacción con sistemas, infraestructuras públicas, entre otros, se utilizó el criterio mencionado por Carmines y Zeller (1979) que dice que para que existe unidimensionalidad el primer factor debe contener al menos el 40% de la varianza. La Tabla 13 muestra numéricamente los eigenvalores, el porcentaje de varianza y el porcentaje de varianza acumulada de cada componente, mostrando que el primer componente tiene un porcentaje de varianza de 49.77% por lo que el supuesto de unidimensionalidad se cumple. Además, se identifican 3 eigenvalores con valores arriba de 1 donde en conjunto contienen una varianza acumulada del 65.17%. A su vez con estos resultados se pueden dar alusión a la evidencia de validez de constructo.

Tabla 13. Eigenvalores asociados a cada uno de los componentes

Componente	Eigenvalor	% de varianza	% de varianza acumulada
1	7.963	49.769	49.769
2	1.278	7.989	57.758
3	1.186	7.414	65.171
4	0.942	5.888	71.059
5	0.744	4.649	75.709
6	0.637	3.979	79.688
7	0.594	3.716	83.404
8	0.479	2.996	86.400
9	0.404	2.526	88.926
10	0.363	2.267	91.193
11	0.339	2.119	93.312
12	0.298	1.861	95.174
13	0.277	1.731	96.904



14	0.202	1.265	98.169
15	0.166	1.038	99.207
16	0.127	0.793	100.000

La Tabla 14 contiene los parámetros de discriminación de cada ítem en el que se identifican con color naranja los ítems que mejor discriminan donde el grado de satisfacción con la infraestructura pública dedicada a las personas de la tercera edad es la que mayor puntaje tuvo, mientras que los ítems que discriminan poco están señalados con amarillo mostrando que el que tuvo menor puntaje es la satisfacción con la asistencia médica de los centros de salud.

Tabla 14. Parámetro de discriminación de la Teoría de Respuesta al Ítem

Ítems de grado de satisfacción	Discriminación
1. Con la asistencia médica de los centros de salud, hospitales, centros especializados.	1.27
2. Con el sistema educativo.	1.41
3. Con el Sistema Judicial (Justicia)	1.64
4. Con el Sistema de Seguridad Social (pensiones y protección social).	2.30
5. Con la Seguridad Ciudadana.	1.96
6. Con la limpieza y decoración de los espacios públicos.	1.85
7. Con los cambios que ha habido en la infraestructura vial	2.15
8. Con los cambios que ha habido en la vialidad peatonal	2.31
9. Con las Infraestructuras Públicas, para el entretenimiento y esparcimiento.	2.37
10. Con la Infraestructura Pública, orientada para los ancianos (as).	2.41
11. Con las Infraestructuras Públicas, dedicadas a los discapacitados(as) o minusválidos(as).	2.10
12. Con los puntos de recreación existentes en su comunidad (restaurantes, teatros, cines)	1.58
13. Con el Transporte Público.	1.38
14. Con la Gestión (trabajo) del Gobierno Municipal.	2.12
15. Con la Gestión (trabajo) del Gobierno Estatal.	1.54
16. Con la Gestión (trabajo) del Gobierno Federal.	1.71

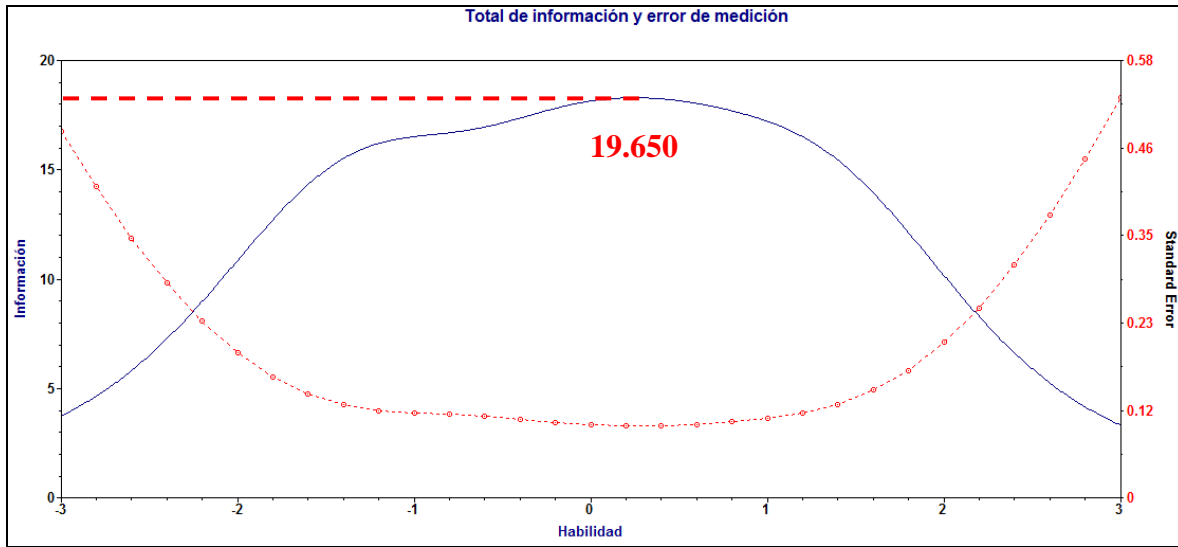


Figura 7. Total de información del constructo

Como se puede observar en la Figura 7 la máxima información proporcionada por el constructo es de 19.650, que al dividirlo entre 16 que es el número de ítems total se obtiene un valor de 1.228, lo cual nos sirve como indicador para conocer si cada ítem proporciona o no información, en el que si el ítem está por debajo de la media de información esperada (.865) no proporciona información relevante y viceversa.

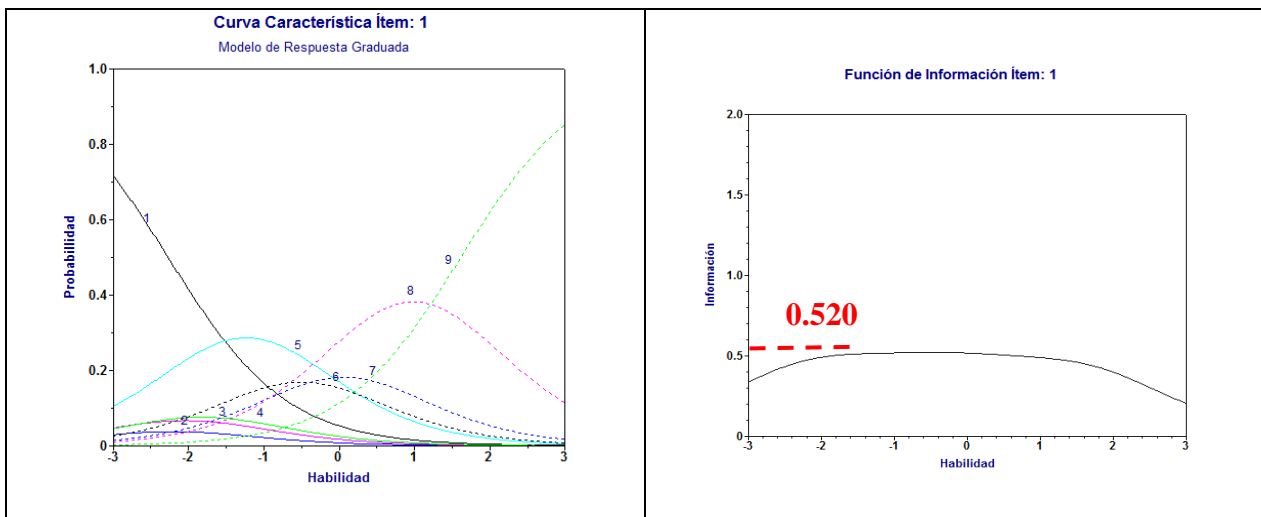


Figura 8. Curva característica y función de información del grado de satisfacción con la asistencia médica de los centros de salud, hospitales, centros especializados

En las curvas característica de la Figura 8 se puede observar el comportamiento de cada una de las categorías del grado de satisfacción con la asistencia médica de los centros de salud, teniendo que la categoría 1, 5, 8 y 9 están bien definidas ya que sobresalen de las demás, mientras que las categoría 2, 3, 4, 6 y 7 están contenidas dentro de otras curvas por lo que éste ítem se comporta como si tuviera menos categorías. Por otro lado el máximo valor en la función de información es de .520 menor a la media de información esperada (1.228) por lo que éste ítem no proporciona mucha información.

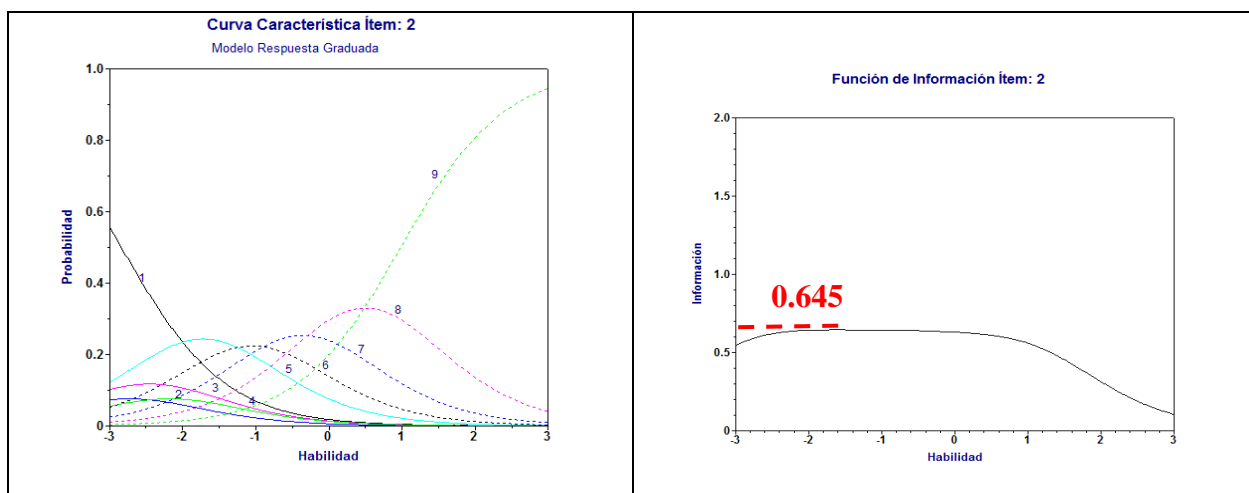


Figura 9. Curva característica y función de información del grado de satisfacción con el sistema educativo

Respecto al grado de satisfacción con el sistema educativo, la Figura 9 contiene las curvas características de este ítem, en el que se puede apreciar que las categorías 2, 3 y 4 están contenidas dentro de la categoría 1, mientras que las demás están bien definidas. Además si máximo información es de .645 menor que la media de información esperada por lo que el ítem no proporciona mucha información.

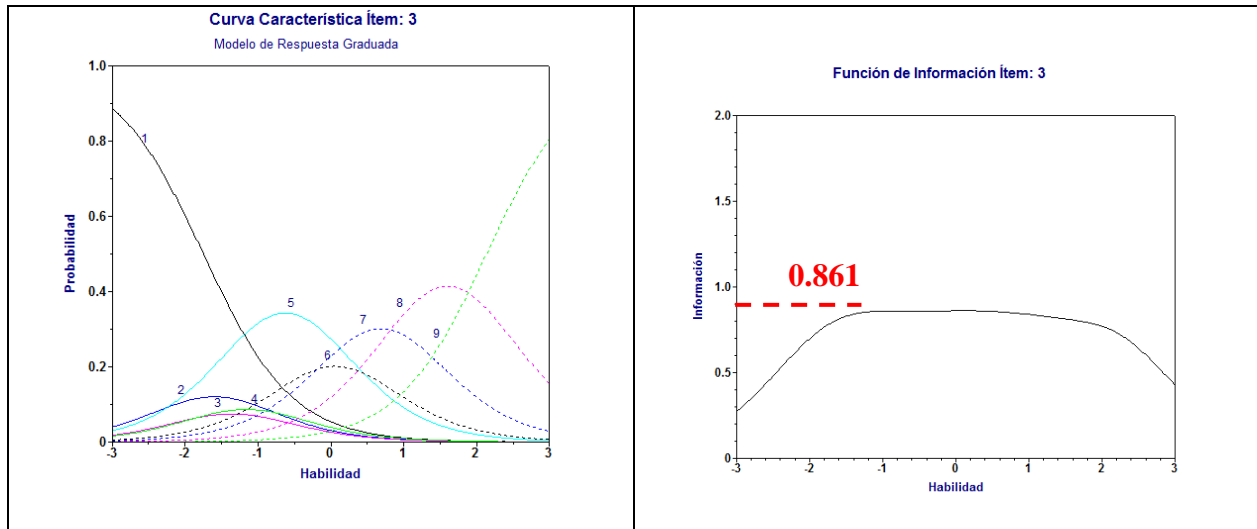


Figura 10. Curva característica y función de información del grado de satisfacción con el Sistema Judicial

En las curvas características de la Figura 10 se observa que las categorías 2, 3 y 4 están contenidas en la curva de la categoría 1, la 6 dentro de la categoría 5 y las demás tienen un comportamiento diferente. Su función de información indica que su valor máximo es de .861 por lo que no proporciona mucha información ya que es menor a 1.228.

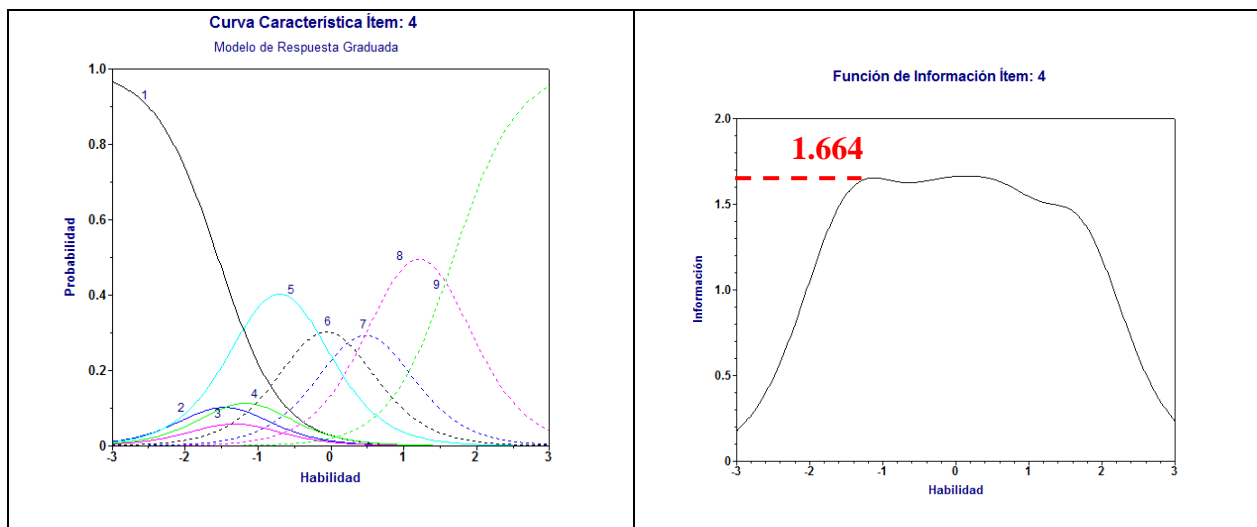


Figura 11. Curva característica y función de información del grado de satisfacción con el Sistema de Seguridad Social

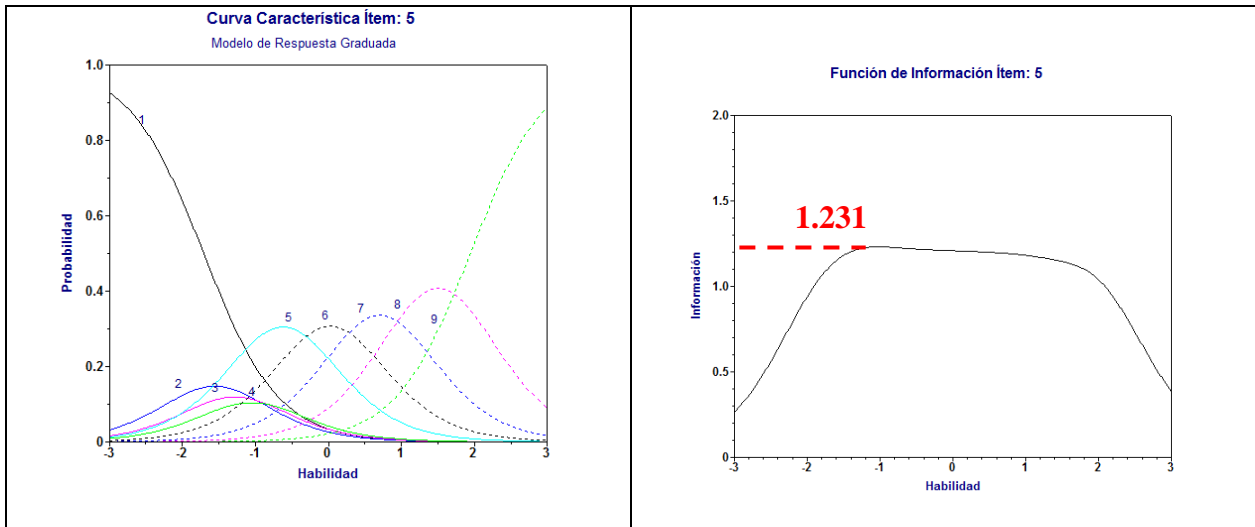


Figura 12. Curva característica y función de información del grado de satisfacción con la Seguridad Ciudadana

Las Figuras 11 y 12 muestran que las curvas de las categorías 2, 3 y 4 están debajo de la curva característica de la categoría 1 y las demás están bien definidas, comportándose como si solo tuviera 6 categorías de respuesta. Ambas funciones de información muestran un valor máximo arriba de la media de información esperada (1.228) por lo que los dos ítems proporcionan mucha información.

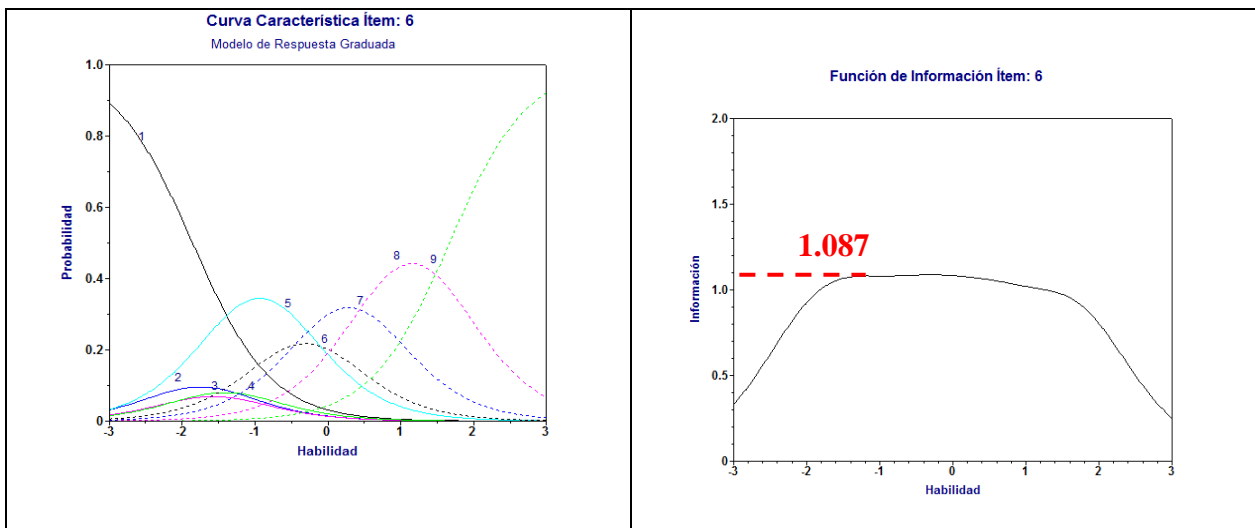


Figura 13. Curva característica y función de información del grado de satisfacción con la limpieza y decoración de los espacios públicos

Conforme al grado de satisfacción con la limpieza pública, se puede observar en la Figura 13 que las categorías 1, 2, 3 y 4 se comportan como una misma categoría ya que la curva característica de la categoría 1 contiene a las otras. La categoría 5 contiene a la categoría 6 por lo que esas dos categorías también se comportan como una sola, las demás tienen un comportamiento diferente, teniendo que éste ítem se comporta como si solo tuviera 5 ítems. Respecto a su función de información, se observa que el valor máximo que alcanza es de 1.087 siendo inferior a 1.228 por lo que no aporta mucha información.

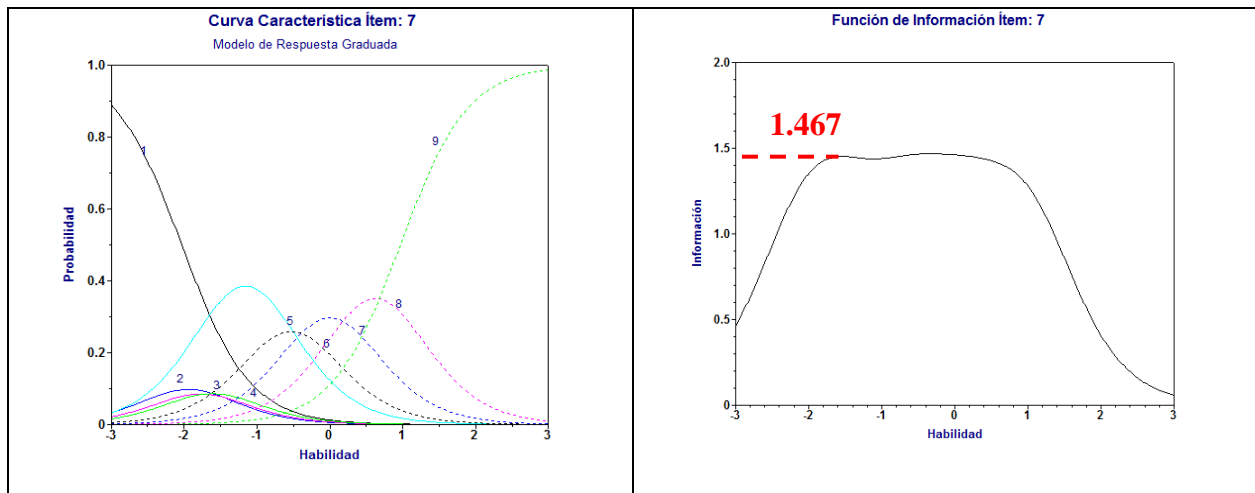


Figura 14. Curva característica y función de información del grado de satisfacción con los cambios que ha habido en la infraestructura vial

Este otro ítem (Figura 14) tiene un comportamiento similar al anterior, las categorías 2, 3 y 4 están por debajo de la curva característica de la categoría 1 y la 6 está dentro de la 5 por lo que el grado de satisfacción con los cambios que ha habido en la infraestructura vial se comporta como si solo hubiese 5 categorías de respuesta. Este ítem aporta mucha información dentro del constructo ya que visualizando su función de información se observa un valor máximo de 1.467, mayor que la media de información esperada.

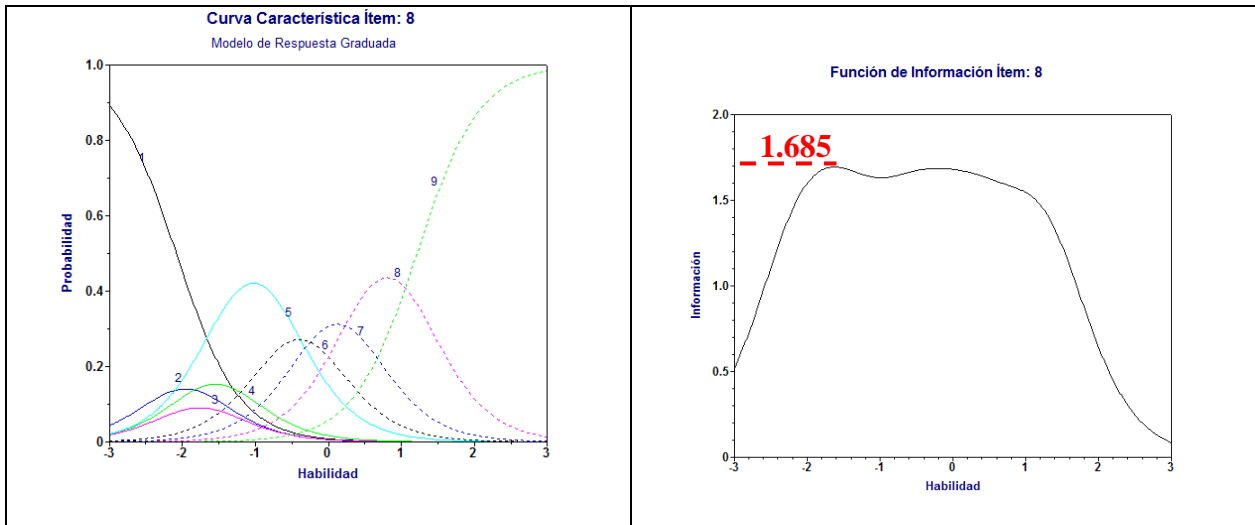


Figura 15. Curva característica y función de información del grado de satisfacción con los cambios que ha habido en la vialidad peatonal

Conforme al grado de satisfacción con los cambios que ha habido en la vialidad peatonal, en la Figura 15 se pueden apreciar las curvas características de cada categoría de respuesta en el que varias categorías están por debajo de otras y solo 5 son las sobresalientes por lo que el ítem se comporta como si hubiera menos categorías de respuesta. Además, su valor máximo de información es de 1.685 menor que 1.228 por lo que aporta información relevante.

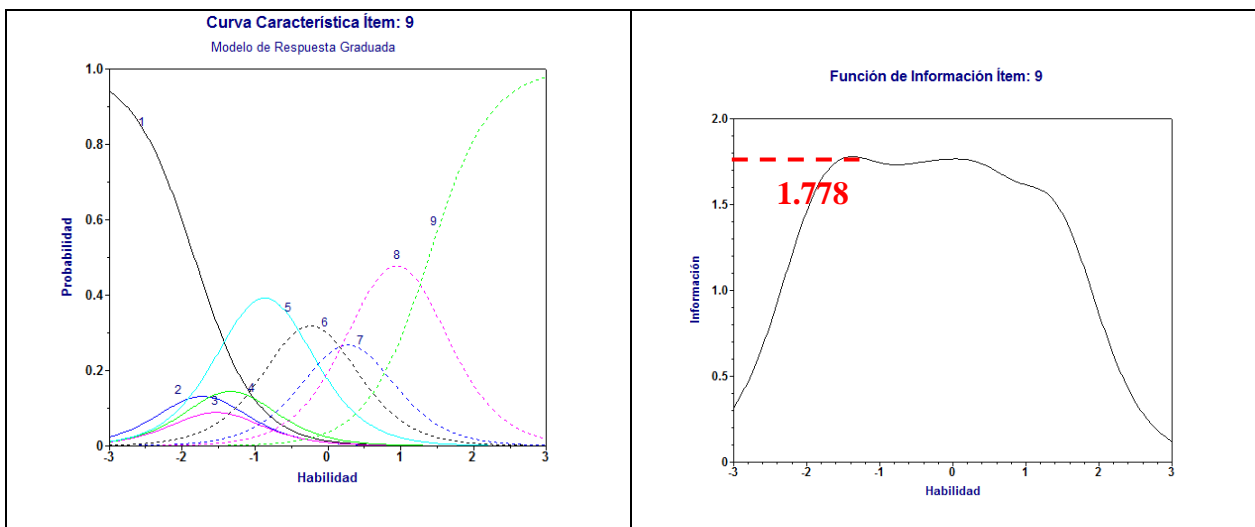


Figura 16. Curva característica y función de información del grado de satisfacción con las Infraestructuras Públicas, para el entretenimiento y esparcimiento

En las curvas características de la Figura 16 se observa que hay 5 curvas características que están bien definidas, mientras que las otras están contenidas en otras curvas. También muestra que su valor máximo de información es superior a la media esperada por lo que este ítem aporta mucha información.

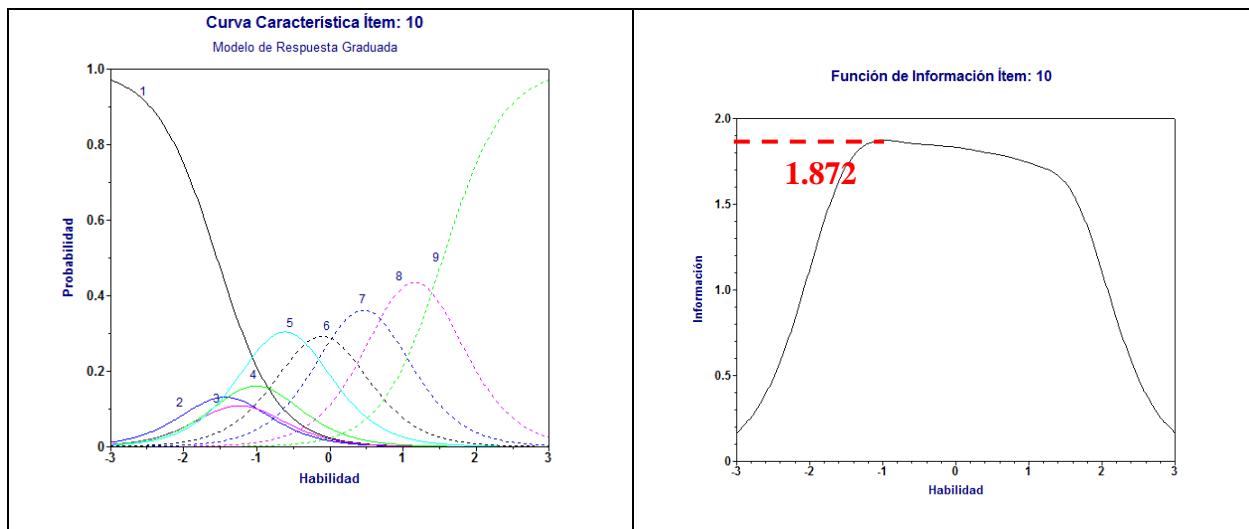


Figura 17. Curva característica y función de información del grado de satisfacción con la Infraestructura Pública, orientada para los ancianos (as)

En esta otro ítem (Figura 17) se puede observar que las categorías 2, 3 y 4 están debajo de la curva característica de la categoría 1, mientras que las demás están bien representadas. También se puede observar que éste ítem proporciona información importante en el constructo debido a que su valor máximo de información es mayor a 1.228 de la media de información esperada.

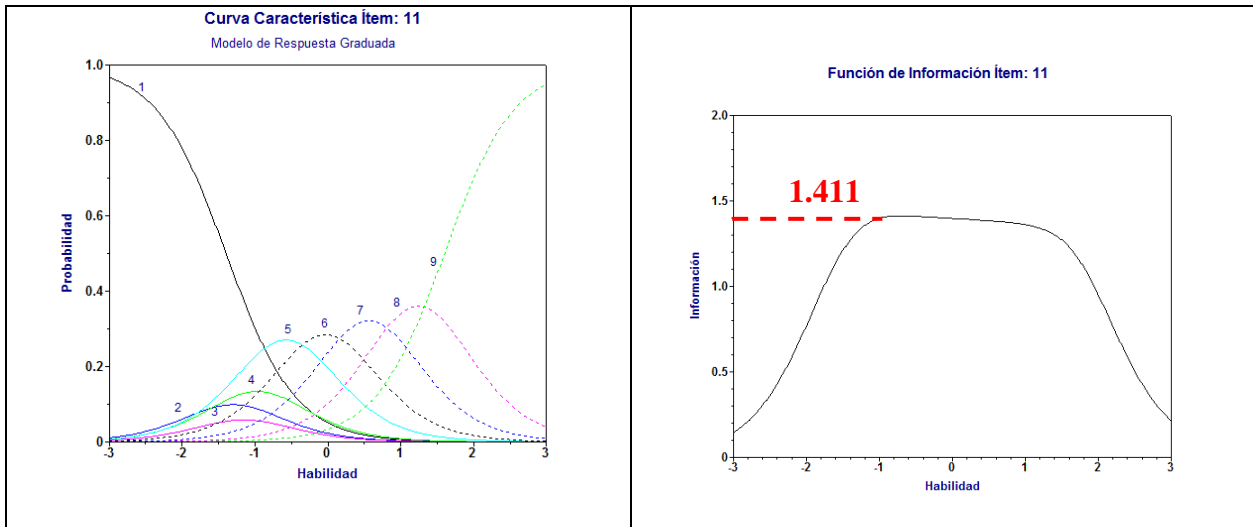


Figura 18. Curva característica y función de información del grado de satisfacción con las Infraestructuras Públicas, dedicadas a los discapacitados(as) o minusválidos(as)

Igual que la Figura 18 anterior en la Figura se observa que el ítem referente al grado de satisfacción con las infraestructuras públicas dedicadas a los discapacitados, las puntuaciones que los usuarios dieron parecen estar definidos en solo 6 categorías. Este ítem aporta mucha información debido a que su valor máximo es mayor a 1.228.

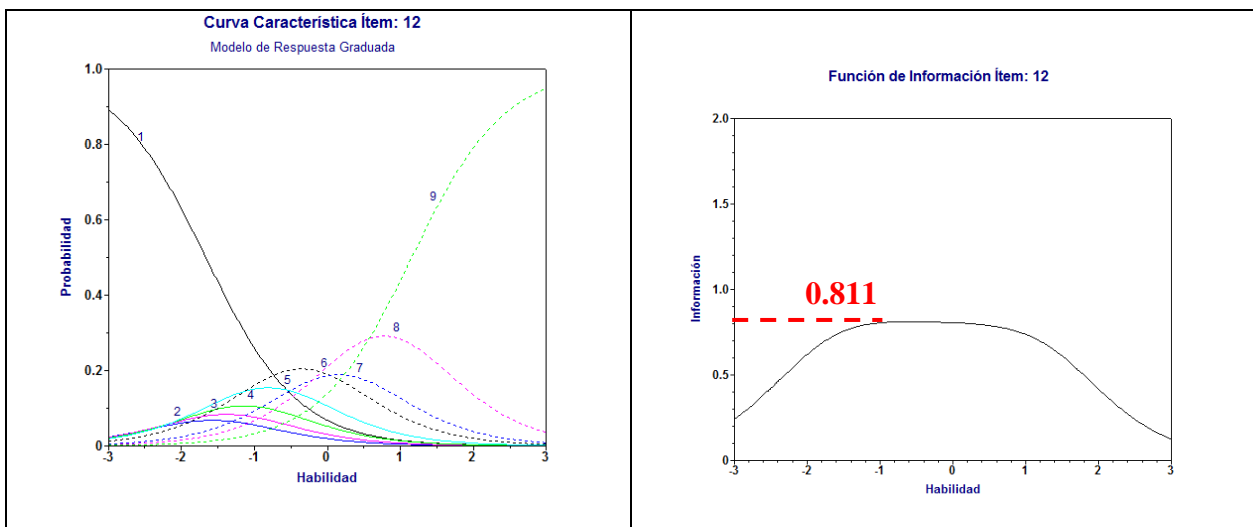


Figura 19. Curva característica y función de información del grado de satisfacción con los puntos de recreación existentes en su comunidad (restaurantes, teatros, cines)

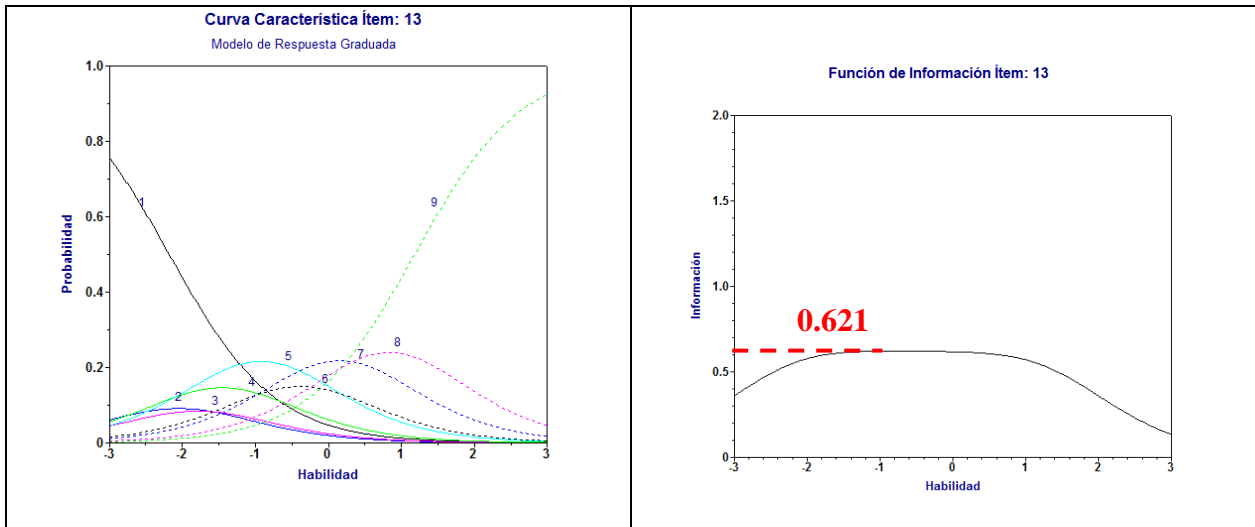


Figura 20. Curva característica y función de información del grado de satisfacción con el Transporte Público

Las curvas características de las Figuras 19 y 20 muestran un comportamiento muy diferente a todos los demás ya que el comportamiento de las categorías representadas en sus respectivas curvas características muestran que solo 4 curvas son las que se encuentran bien definidas mientras que las otras 5 están contenidas en otras. Además, ambos ítems poseen un valor máximo menor al esperado por lo que no aporta mucha información en el constructo.

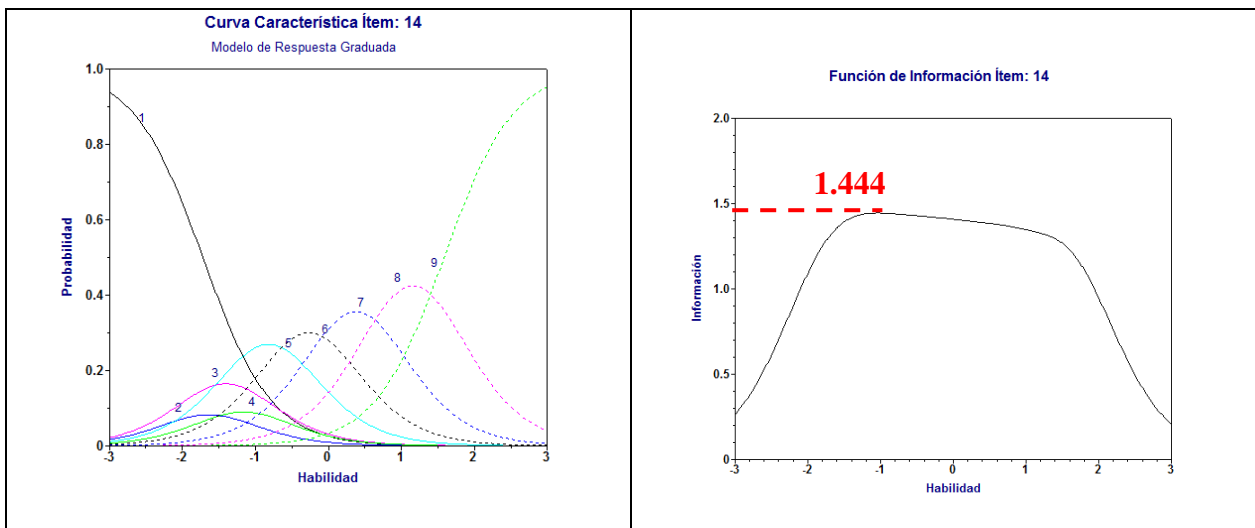


Figura 21. Curva característica y función de información del grado de satisfacción con la Gestión (trabajo) del Gobierno Municipal

Mediante las curvas características de las categorías del grado de satisfacción con la gestión del gobierno municipal, se puede observar en la Figura 21 que las curvas 2, 3 y 4 están contenidas en la curva de la categoría 1 y las demás muestran un comportamiento definido. Se muestra también que su valor máximo de información (1.444) es mayor al esperado por lo que el ítem es relevante.

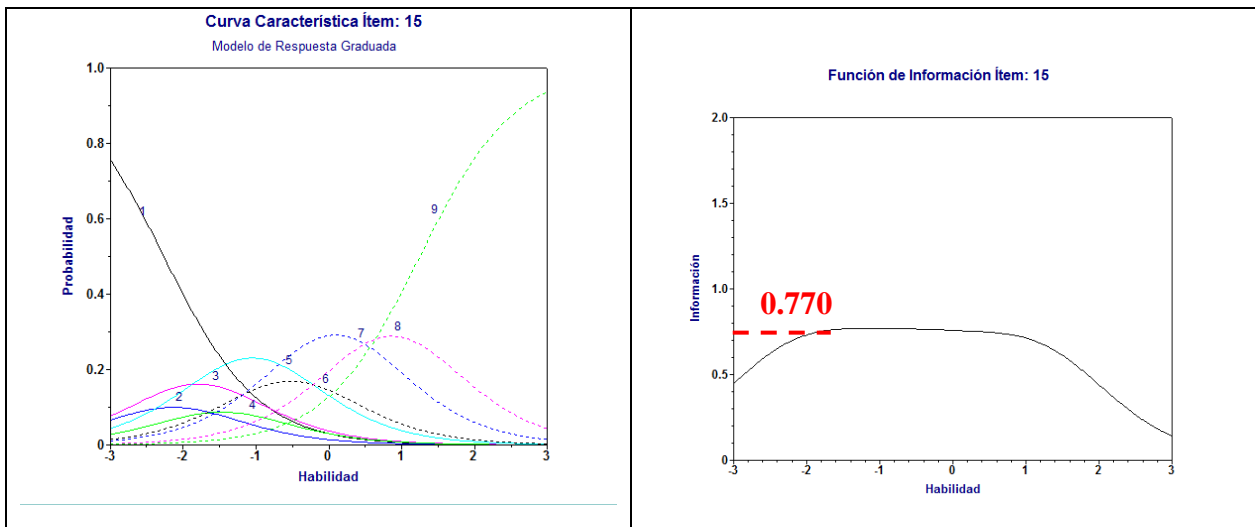


Figura 22. Curva característica y función de información del grado de satisfacción con la Gestión (trabajo) del Gobierno Estatal

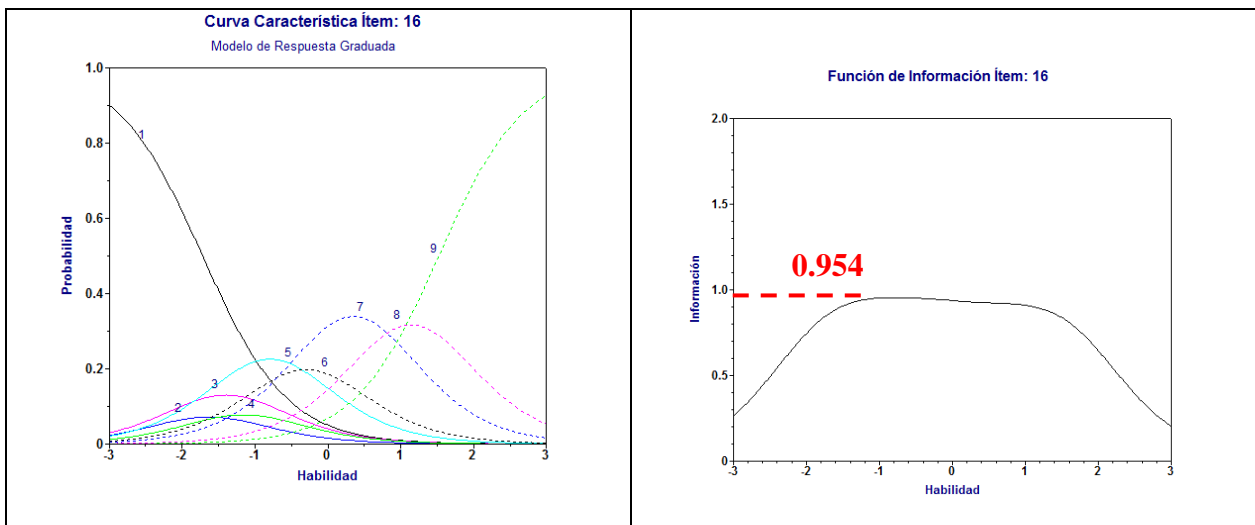


Figura 23. Curva característica y función de información del grado de satisfacción con la Gestión (trabajo) del Gobierno Federal



El grado de satisfacción con las gestión del gobierno estatal y federan representados en sus curvas características (Figuras 22 y 23) muestran un comportamiento similar, las categorías 2, 3 y 4 de cada ítem están debajo de la curva de la categoría 1, la 6 sobre la 5, por lo que ambos ítems funcionarían mejor con menor número de categoría de respuesta. A su vez se observa que el ítem aporta poca información ya que su valor máximo en la función de información es menor a la esperada.

Con las figuras anteriores de cada uno de los ítems se pudo apreciar que 7 de los 16 ítems son los que mayor información dan dentro del constructo ya sus valores máximos de información eran superiores al esperado. Y que se comportaban como si tuvieran menos categorías de respuesta, donde las curvas características bien definidas en algunos ítems eran 6 y en otros 5. Tomando en cuenta las curvas características bien definidas se recodificaron los ítems juntando las categorías que estaban contenidas en una misma curva, dicha re categorización se muestra en la Tabla 15 donde 1 es nada satisfecho y 6 junto con 5 son muy satisfecho respectivamente en ambas propuestas.

Tabla 15. Propuesta de número de categorías

Categorías de respuesta	Grado de Satisfacción								
	Nada satisfecho							Muy satisfecho	
General (9 categorías)	1	2	3	4	5	6	7	8	9
6 categorías	1	1	1	1	2	3	4	5	6
5 categorías	1	1	1	1	2	2	3	4	5



En las curvas características resultantes con 6 categorías (Anexo 2) se logran definir las curvas de cada categoría mejor que las de 9 categorías pero una curva parece estar dentro de otra, por lo que utilizando 1 categoría menos (5 categorías) se lograron definir bien las curvas características. A su vez la Teoría de Respuesta al Ítem proporciona un indicador de fiabilidad de la escala empleada en el constructo observándose en la Tabla 16 que la fiabilidad permanece estable tanto en la escala general como en las propuestas. Finalmente con las características encontradas en las herramientas estadísticas se realizó una nueva versión del constructo (Ver anexo 3).

Tabla 16. Fiabilidad marginal respecto al número de categorías propuesto

Categorías de respuesta	Fiabilidad Marginal
General (9 categorías)	0.945
6 categorías	0.941
5 categorías	0.937



CONCLUSIONES

En este trabajo se logró aplicar satisfactoriamente las herramientas estadísticas que usualmente se utilizan en el campo de la psicometría mostrando la eficiencia de las herramientas en cualquier campo de estudio como lo es en el caso de mediciones sociales como los estudios de opinión. El caso de ejemplo sirvió para emplear la metodología expuesta en este trabajo donde se encontraron muy buenos resultados.

Desde el punto de vista univariante se encontraron resultados favorables que permitieron conocer el comportamiento de cada uno de los ítems del constructo, donde de manera inicial se observó que existía una tendencia ascendente de los datos, lo que representa que a niveles bajos de satisfacción con cada uno de los ítems había pocas puntuaciones de los encuestados y que a mayores niveles de satisfacción en la escala, mayores eran las puntuaciones en la escala, esto quiere decir que gran parte de las personas que conformaron la muestra se inclinaron por dar calificaciones altas de satisfacción.

En la parte bivariante se encontraron relaciones positivas entre varios de los ítems donde el utilizar herramientas estadísticas como las asociaciones de Spearman y correlaciones policóricas fueron la mejor opción para este tipo de datos ya que se ajustan adecuadamente a variables categóricas en escala ordinal. Además se encontraron asociaciones entre los mismos a través de la ji-cuadrada lo cual permitió evidenciar las relaciones entre los ítems que a su vez dieron paso a encontrar buenos resultados.

Para conocer la fiabilidad y validez de los instrumentos se utilizaron herramientas multivariantes que permitieran responder esos objetivos. Empleando el alpha de Cronbach se demostró una alta confiabilidad ya que el valor de alpha resultó ser de .93. En el análisis del impacto se pretendió que de manera descriptiva se pudieran observar aquellos ítems que mayor información aportan dentro del constructo encontrando que el grado de satisfacción con el sistema educativo, con los cambios que ha habido en la infraestructura vial, los cambios que ha



habido en la vialidad peatonal y con la Gestión (trabajo) del Gobierno Estatal fueron las que mayor impacto tienen desde el punto de vista de los sesgos existentes en los ítems.

Para dar alusión a la validez del constructo se empleó la Teoría de Respuesta al Ítem (TRI) ya que como requisito esencial es que los ítems midan un mismo constructo, para ello se realizó el gráfico de sedimentación donde la mayor carga de variabilidad está representada en un solo factor por lo que a su vez el supuesto de unidimensionalidad se cumple. Para aplicar la Teoría de Respuesta al ítem los supuestos resultaron ser favorables para poderla emplear.

La TRI en comparación del análisis factorial para validez de cuestionarios, analiza los constructos desde una perspectiva unidimensional mientras que el análisis factorial lo hace desde una visión multidimensional, por lo que es posible analizar el constructo desde la TRI donde las cargas de importancia de cada uno de los ítems en el análisis factorial se encontraban en los parámetros arrojados por la TRI. En este trabajo se hizo hincapié en la importancia de la aplicación de la TRI ya que esta técnica no es muy usual aplicarla en mediciones sociales, además, permitió encontrar resultados relevantes.

Como primeros resultados de la TRI se obtuvo como parámetro de interés el de discriminación ya que permitió identificar aquellos ítems que son o no relevantes en el constructo, donde para su mejor interpretación se visualizaron las funciones de información de cada uno de los ítems teniendo que 7 de los 16 ítems están por arriba de la media de información esperada por lo que si se quiere reducir el constructo se pueden eliminar aquellos que aporten menor información. Además, el grado de satisfacción con los cambios que ha habido en la infraestructura vial y los cambios que ha habido en la vialidad peatonal aparecen importantes tanto en la TRI como de manera descriptiva en el Análisis del Impacto por lo que éstos 2 ítems son muy importantes en el constructo.

Respecto a las curvas características de cada uno de los ítems se encontró que la escala empleada es muy grande para conocer el grado de satisfacción de las diferentes características que



se incluían en el constructo y que con menos categorías se podría conocer ese objetivo, ya que en el comportamiento de las curvas características unas curvas estaban por debajo de otras lo que representaban que esas categorías en realidad parecieran ser categorías unidas y no diferentes, uno de los factores por los que se observaba este comportamiento es que existían categorías que fueron poco seleccionadas por las personas haciendo que sus probabilidades de puntuarlas eran muy pequeñas en comparación con otras por lo que no aportaban mucha información.

Para proponer el número de categorías ideal para las mediciones se tomaron en cuenta las curvas características bien definidas principalmente en los ítems que mayor información aportan dentro del constructo, encontrándose que entre 5 y 6 eran las que mejor comportamiento demostraban. Al realizar nuevamente las curvas características, con 6 categorías de respuesta se lograron definir mejor el comportamiento de cada una de ellas aunque había algunas en las que una curva pareciera estar contenida en otra, es por ello que en las curvas características con 5 categorías de respuesta se observaba un mejor comportamiento.

Conforme a lo anterior se propone utilizar 5 opciones de respuesta en la escala de satisfacción para una nueva aplicación del instrumento ya que las curvas características se lograron definir mejor. La Teoría de Respuesta al Ítem también proporciona un grado de confiabilidad del constructo teniendo que tanto en la escala original como en las propuestas con 6 y 5 categorías el valor permaneció estable de .94 que es muy semejante al arrojado por el alpha de Cronbach por lo que se demuestra que el utilizar menos categorías no se alteran los resultados además de que facilita su interpretación.

La metodología empleada sirvió no solo para sustentar la presencia de confiabilidad y validez del instrumento sino que también se lograron encontrar los ítems de mayor y menor impacto, así como el número de categorías ideal para tener buenas mediciones en los grados de satisfacción, con ello se propuso un nuevo constructo que se puede utilizar en una nueva aplicación.



Se resaltó la importancia de la TRI en el estudio social tal como en los estudios de opinión en específico el estudio de Calidad de Vida, para demostrar que no solo en el contexto de educación y salud, que es donde mayormente se aplica esta herramienta, se puede encontrar buenos resultados que permitan al o los investigadores construir y diseñar instrumentos en otros campos de estudio y obtener resultados relevantes.

Además se hizo hincapié en utilizar herramientas estadísticas conjuntamente esto para tener más evidencia y veracidad de los resultados, otra herramienta que se puede emplear es el análisis Factorial para conocer el comportamiento multidimensional de los ítems que conforman el constructo.

Cabe señalar que el problema que existe dentro de la confiabilidad y validez de instrumentos aun es mayor y requiere utilizar más herramientas ya que un caso particular es conocer cómo repercute la diferencia de grupos (género, edad, etc.) en el comportamiento de los ítems, para ello se requiere utilizar herramientas estadísticas de funcionamiento diferencial, en este trabajo se recalcan las herramientas necesarias para tener un sustento sólido de validez y confiabilidad pero no se limita a utilizar más herramientas.



REFERENCIAS

- Abad, F. J., Ponsoda, V., & Revuelta, J. (2006). *Modelos politómicos de respuesta al ítem*. Madrid: La Muralla.
- Allen, F., & Locker, D. (2002). A Modified Short Version of the Oral Health Impact Profile for Assessing Health-Related Quality of Life in Edentulous Adults. *The International Journal of Prosthodontics*, 15(5), 446-450.
- Anastasi A., & Urbina S. (1998). *Tests Psicológicos*. México: Prentice Hall.
- Batista-Foguet, J. M., Coenders, G., & Alonso, J. (2014). Análisis factorial confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud. *Med Clin (Barc)*, 21-27.
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. U.S.A.: SAGE Publications.
- Celina-Oviedo, H., & Campo-Arias, A. (2005). Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría*, 572-580.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cuevas-Acosta, J. H., & Hernández-González, S. (2013). Propuesta didáctica para caracterizar variables y datos. *Revista investigación operacional*, 266-273.
- Escurrea-Mayaute, M., & Salas-Blas, E. (2014). Construcción y validación del cuestionario de adicción a redes sociales (ARS). *LIBERABIT*, 73-91.



Gómez-Benito, J., & Navas-Ara, M. J. (1998). Impacto y funcionamiento diferencial de los ítems respecto al género en una prueba de aptitud numérica. *Psicothema*, 685-696.

Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2010). *Metodología de la investigación*. México D.F.: McGraw-Gill.

Huerta-Pacheco, N. S. (2014) *Propuesta de evaluación estadística para cuestionarios aplicados a un juego de razonamiento espacial que sirvan para distinguir características esenciales del mismo*. Maestría en Estadística Aplicada, Universidad Veracruzana, México.

Juniper, E. F., Guyatt, G. H., Streiner, D. L., & King, D. R. (1997). Clinical impact versus factor analysis for quality of life questionnaire construction. *Journal of Clinical Epidemiology*, 233-238.

Kerlinger, F. N., & Lee, H. B. (2002). *Investigación del comportamiento*. México: McGraw-Hill.

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, 140, 1-55.

Mañas, B. (2012): Mediciones objetivas de opiniones subjetivas: la construcción de un modelo legitimado para el estudio de la opinión pública, en Arribas, J.M. et al., *Historia de la probabilidad y de la Estadística VI*. Madrid. UNED-AHEPE, 289-304

Martínez-Arias, M. R., Hernández-Lloreda, M. V., & Hernández-Lloreda, M. J. (2014). *Psicometría*. Madrid: Alianza Editorial.



-
- Martínez-Arias, R. (1995). *Psicometría: Teoría de los tests psicológicos y educativos*. Síntesis Psicología. España.
- Meneses, J., Barrios, M., Bonillo, A., Cosculluela, A., Lozano, L., Turbany, J., y otros. (2013). *Psicometría*. Barcelona: Editorial UOC.
- Morales-Vallejo, P., Urosa-Sanz, B., & Blanco-Blanco, A. (2003). *Construcción de escalas de actitudes tipo Likert*. Madrid: La Muralla.
- Pérez-Tejada, H. E. (2008). *Estadística para las ciencias sociales, del comportamiento y de la salud*. México: Cengage Learning Editores.
- Real Academia Española, Asociación de Academias de la Lengua Española. (2014). *Diccionario de la lengua española*. (23 ed.). Madrid: Edición del Tricentenario.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, 17.
- Samejima, F. (1972). A general model for free response data. *Psychometrika, Monograph Supplement*, 18.
- Samejima, F. (1997). Graded response model. En Van der Linden, W.J. & Hambleton R.K. (Eds.), *Handbook of modern item response theory* (pp. 85-100). New York: Springer.
- Sánchez-Barba, M. (2014) *Aportaciones al análisis de datos de Calidad de vida relacionada con la salud, desde una perspectiva multivariante*. Máster Universitario en Análisis Avanzado de Datos Multivariantes, Departamento de Estadística, Universidad de Salamanca, España.



-
- Sánchez-Rivero, M. (2004). *Introducción a la Teoría de Respuesta al Ítem, una herramienta para el análisis de variables latentes: Aplicación a la medición de la Calidad de Vida de la Infancia*. Ponencia presentada en la XVII Reunión Anual. León, España.
- Siegel, S., & John, N. C. (1998). *Estadística no paramétrica aplicada a las ciencias de la conducta*. España: Trillas.
- Spearman, C. (1904). The proof and measurement of association between two things. *Am. J. Psychol.* 15: 72-101.
- SPSS Inc. Released 2007. SPSS for Windows, Version 16.0. Chicago, SPSS Inc.
- Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 677-680.
- Thissen, D. (1991). *MULTILOG: Multiple category item analysis and test scoring using item response theory*. Chicago, U.S.A.: Ed. Scientific Software International, Inc.
- Vicente-Galindo, E. D. (2011). *Análisis del Impacto frente a Teoría de Respuesta al Ítem (Master Tesis)*. Máster Universitario en Análisis Avanzado de Datos Multivariantes, Departamento de Estadística, Universidad de Salamanca, España.
- Vicente-Galindo, E., Vicente-Galindo, P., Galindo-Villardón, M. P., & Hernández-González, S. (2015). *Análisis Factorial, Análisis del Impacto y Teoría de Respuesta al Ítem: aplicación al estudio de calidad de vida relacionada con la salud de pacientes osteoporóticos*. España: RIDECA.
- Wackerly, D. D., Mendenhall III, W., & Scheaffer, R. L. (2010). *Estadística Matemática con aplicaciones* (Septima Edición ed.). México: Cengage Learning.



ANEXOS

Anexo 1. Versión original del constructo

37. ¿Qué grado de satisfacción tengo?	Totalmente Insatisfecho							Totalmente Satisfecho						
	SATISFACCION													
	0	1	2	3	4	5	6	7	8	9	10			
37.1 Con la asistencia médica de los centros de salud, hospitales, centros especializados.														
37.2 Con el sistema educativo.														
37.3 Con el Sistema Judicial (Justicia)														
37.4 Con el Sistema de Seguridad Social (pensiones y protección social).														
37.5 Con la Seguridad Ciudadana.														
37.6 Con la limpieza y decoración de los espacios públicos.														
37.7 Con los cambios que ha habido en la infraestructura vial														
37.8 Con los cambios que ha habido en la vialidad peatonal														
39.9 Con las Infraestructuras Públicas, para el entretenimiento y esparcimiento.														
37.10 Con la Infraestructura Pública, orientada para los ancianos (as).														
37.11 Con las Infraestructuras Públicas, dedicadas a los discapacitados(as) o minusválidos(as).														
37.12 Con los puntos de recreación existentes en su comunidad (restaurantes, teatros, cines)														
37.13 Con el Transporte Público.														
37.14 Con la Gestión (trabajo) del Gobierno Municipal.														
37.15 Con la Gestión (trabajo) del Gobierno Estatal.														
37.16 Con la Gestión (trabajo) del Gobierno Federal.														

Anexo 2. Curvas características con la cantidad de categorías propuestas

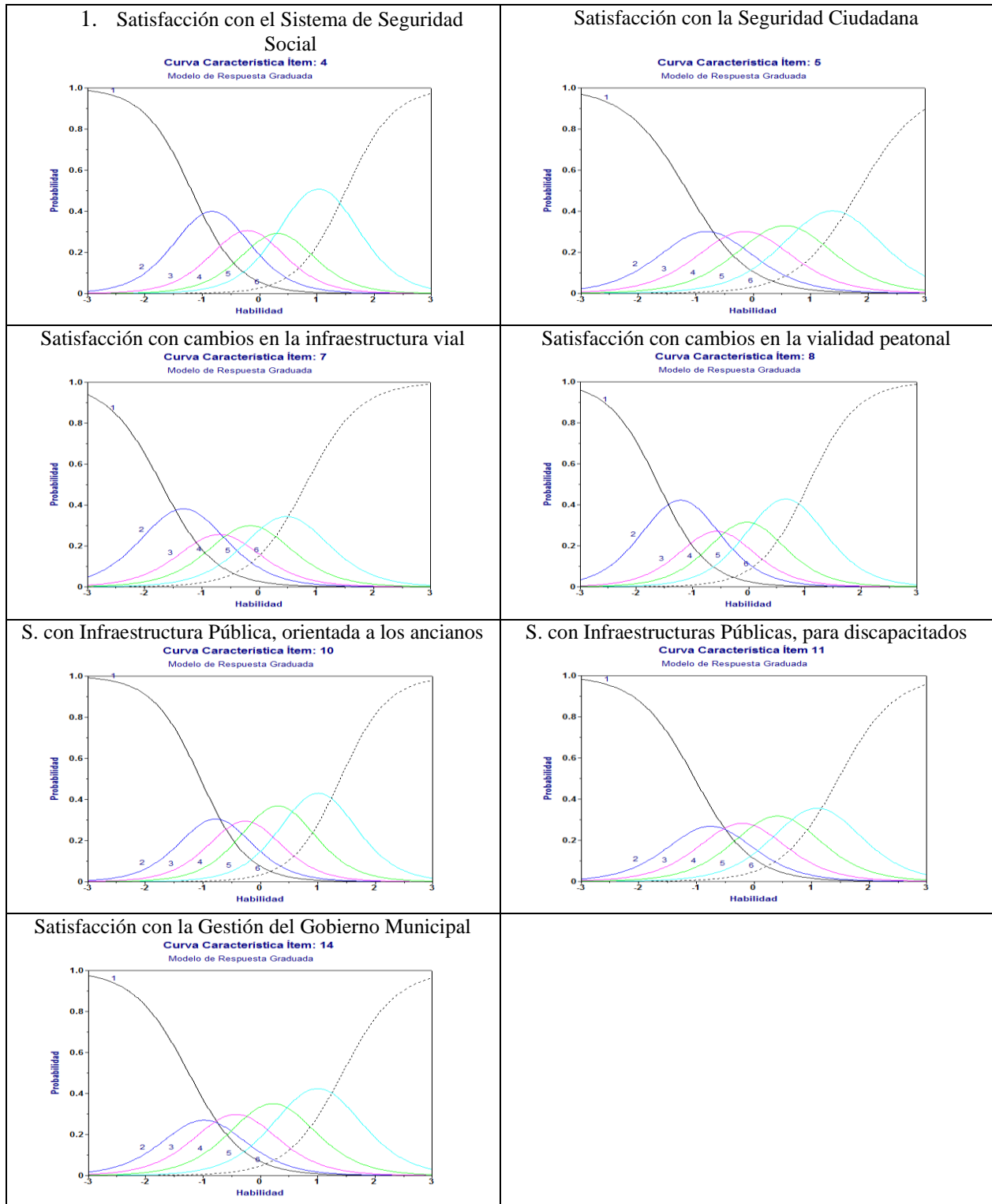


Figura 24. Curvas características con seis categorías de respuesta

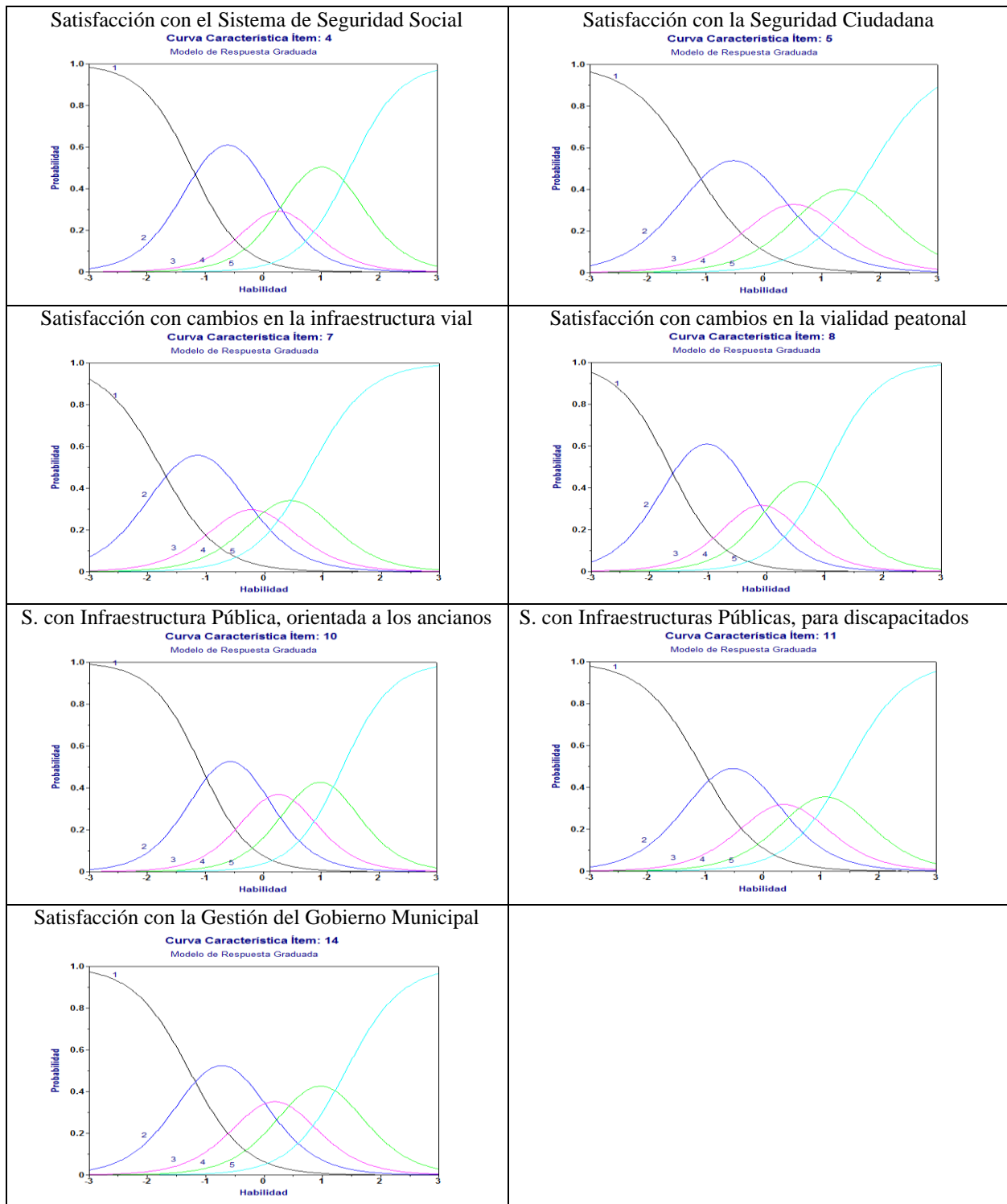


Figura 25. Curvas características con cinco categorías de respuesta



Anexo 3. Versión propuesta del constructo

37. ¿Qué grado de satisfacción tengo?	SATISFACCIÓN				
	Totalmente Insatisfecho			Totalmente Satisfecho	
	↓				↓
	1	2	3	4	5
37.1 Con el Sistema de Seguridad Social (pensiones y protección social).					
37.2 Con la Seguridad Ciudadana.					
37.3 Con los cambios que ha habido en la infraestructura vial					
37.4 Con los cambios que ha habido en la vialidad peatonal					
37.5 Con la Infraestructura Pública, orientada para los ancianos (as).					
37.6 Con las Infraestructuras Públicas, dedicadas a los discapacitados(as) o minusválidos(as).					
37.7 Con la Gestión (trabajo) del Gobierno Municipal.					