



# INEVA en acción

<http://ineva.uprrp.edu>

Boletín informativo

Volumen 10, Número 1, Diciembre 2015

## LA CANTIDAD DE CATEGORÍAS DE RESPUESTA EN UNA ESCALA DE VALORACIÓN

María del R. Medina Díaz, Ph.D., J.D.

Tabla 1.

Algunos estudios publicados acerca de la cantidad de categorías en las escalas (en orden cronológico)

Autores/Año	Número categorías	Identificación o rotulación	Instrumento	Personas que respondieron	Resultados
Likert (1932)	Método sigma* y cinco categorías	Cinco categorías numeradas y rotuladas: 5 <i>Strongly approve</i> 4 <i>Approve</i> 3 <i>Undecided</i> 2 <i>Disapprove</i> 1 <i>Strongly disapprove</i> Siete categorías: Valores 7, 5, 4, 3 y 1 asignados a las categorías anteriores. Valores sigmas ( $\sigma$ ) o de desviaciones estándar en la curva normal que corresponden a cinco intervalos entre -3 a +3.	<i>International Scale</i> (15 ítems) <i>Negro Scale</i> (10 ítems)	95 estudiantes universitarios, divididos en tres grupos	Altos coeficientes de correlación entre los tres formatos de respuesta (mayores de 0.98) con ambas escalas. Formato con cinco categorías y método sigma producen coeficientes de confiabilidad similares. Por el principio de la parsimonia, favorece el uso del formato más simple (cinco categorías).
Jacobi y Matell (1971)	Entre dos y 19 puntos, divididos en dos	Categorías pares divididas en dos grupos ( <i>Agree</i> y <i>Disagree</i> ) y las impares en tres ( <i>Agree</i> ,	Versión modificada de <i>Allport-Vernon-Lindzey Scales of Values</i> (60	360 estudiantes subgraduados de la Universidad de Purdue	No hay una relación sistemática entre la cantidad de categorías y los coeficientes de confiabilidad

Autores/Año	Número categorías	Identificación o rotulación	Instrumento	Personas que respondieron	Resultados
	y tres grupos	<i>Uncertain y Disagree)</i>	afirmaciones) para indicar la importancia de las seis áreas de valores en la vida		(consistencia interna y estabilidad) y los de validez predictiva y concurrente. La conversión o división de las categorías en dos o tres grupos no redujo los coeficientes de confiabilidad y validez.
Garland (1991)	Escala tipo Likert con cuatro y cinco categorías	Cinco categorías rotuladas: <i>Very important; Important; Neither important nor unimportant; Unimportant; Not at all important</i> Cuatro categorías rotuladas: <i>Very important; Important; Unimportant; Not at all important.</i> Ambas incluyen <i>Don't know/refuse to answer.</i>	Pregunta acerca de la importancia de la rotulación (e.g., ingredientes, aditivos) de productos empacados. Entrevistadores mostraron una tarjeta con las categorías de respuesta.	225 compradores de alimentos respondieron a la pregunta con el formato de cinco categorías y 223 con el de cuatro, en el área urbana de Palmerston North en Nueva Zelanda	El formato con las cuatro categorías parece que inclina a las personas indiferentes a hacer una selección en las categorías positivas. La presencia o ausencia de una categoría intermedia podría distorsionar los resultados.
Chang (1994)	Escala tipo Likert con cuatro y seis categorías	Cuatro categorías numeradas y rotuladas: 1 <i>Disagree</i> ; 2 <i>Somewhat disagree</i> ; 3 <i>Somewhat agree</i> ; 4 <i>Agree</i> Seis categorías numeradas y rotuladas: 1 <i>Strongly disagree</i> ; 2 <i>Disagree</i> ; 3 <i>Somewhat disagree</i> ; 4 <i>Somewhat agree</i> ; 5 <i>Agree</i> ; 6 <i>Strongly agree</i>	Nueve ítems del <i>Quantitative Attitudes Questionnaire</i>	165 estudiantes de maestría en Educación tomando su primer curso en métodos cuantitativos	Escala de cuatro categorías produce coeficientes de correlación más altos, aplicando el procedimiento Multirasgo-Multimétodo. No necesariamente aumentar las categorías aumenta la confiabilidad de las puntuaciones. La evidencia de validez relacionada con el criterio no se afectó por la cantidad de categorías.
Hernández Baeza, Muñiz & García Castro (2000)	Escala Likert con cuatro, cinco y seis categorías	Escala Likert con cinco categorías de respuesta (no se indican), reagrupadas en cuatro, cinco y seis categorías	Cuestionario de evaluación del profesorado con 17 ítems, de los cuales se usaron 14 para el estudio	1,000 estudiantes de Psicología de la Universidad de Oviedo y el mismo número en la muestra de simulación	Mejor ajuste al modelo de respuesta graduada, propuesto por Muraki (1990), y precisión en las estimaciones de la habilidad $\theta$ se obtiene con 4, 5 y 6 categorías, cuando la escala es unidimensional. Se requieren

Autores/Año	Número categorías	Identificación o rotulación	Instrumento	Personas que respondieron	Resultados
Amoo & Friedman (2001)	Dos escalas de categorías numeradas +4 a -4 y de 9 a 1	<i>Extremely satisfied</i> +4 +3 +2 +1 0 -1 -2 -3 -4 <i>Not at all satisfied</i> <i>Extremely likely</i> 9 8 7 6 5 4 3 2 1 <i>Extremely unlikely</i> <i>Excellent</i> +4 +3 +2 +1 0 -1 -2 -3 -4 <i>Awful</i> <i>Extremely valuable</i> 9 8 7 6 5 4 3 2 1 <i>Not at all valuable</i>	Dos cuestionarios con ocho preguntas acerca de la satisfacción con una institución de educación superior, la probabilidad de cambiarse de institución, calidad del profesorado y el valor de la educación recibida.	139 estudiantes de bachillerato en New York	seis o más categorías si la escala es bidimensional. Estudiantes seleccionaron más los numerales positivos que los negativos Categorías de respuestas usando numerales negativos y positivos producen resultados distintos.
Pajares, Hartley & Valiente (2001)	Respuestas numéricas de 0 a 100 puntos vs. Escala tipo Likert con seis puntos (1 a 6)	Escalas rotuladas en los extremos: 0 ( <i>No confident at all</i> ) a 100 ( <i>Completely confident</i> ); 1 ( <i>No confident at all</i> ) a 6 ( <i>Completely confident</i> )	Dos escalas de auto-eficacia en la escritura con 10 ítems (una con 6 puntos y la otra con 100)	437 estudiantes de sexto a octavo grado en un curso de Artes del Lenguaje en una escuela intermedia en los Estados Unidos de América	Respuestas en el formato 0 a 100 producen mejores estadísticas en el análisis factorial y de la consistencia interna de las puntuaciones (aplicando el coeficiente Alfa de Cronbach). Además, produjo correlaciones más altas con el promedio académico y la evaluación de las maestras de la competencia en escritura.
Dawes (2008)	cinco, siete y diez categorías	Extremos numerados y rotulados, según la cantidad de categorías: 1 <i>Strongly disagree</i> y X (5, 7 ó 10) <i>Strongly agree</i>	Escala con ocho ítems acerca de la conciencia de los precios	300, 185 y 250 personas se entrevistaron por teléfono usado las escalas de diez, siete y cinco categorías, respectivamente. También se produjeron datos de tres grupos de 185 personas mediante simulaciones.	Escalas con cinco y siete categorías produjeron puntuaciones más altas. No hay diferencias significativas entre las medias aritméticas de las puntuaciones entre los grupos que usaron las escalas con cinco y siete categorías. No hay diferencias significativas entre las tres escalas en la desviación estándar, el sesgo y la curtosis de las distribuciones de las puntuaciones. Las personas usaron más categorías en las escalas cuando la escala las incluía.

Autores/Año	Número categorías	Identificación o rotulación	Instrumento	Personas que respondieron	Resultados
Lozano, García-Cueto & Muñiz (2008)	Escala tipo Likert con dos a nueva alternativas de respuestas			Simulación de las respuestas a 30 ítems con distintas cantidad de alternativas (de dos a nueve), siguiendo una distribución normal y cuatro tamaños de muestra (50, 100, 200 y 500 personas), así como correlaciones entre los ítems (0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 y 0.9). Doscintas cincuenta y seis condiciones en la simulación de los datos, representando un modelo unidimensional	La reducción en la cantidad de alternativas disminuyó su varianza y confiabilidad, Sin embargo, la varianza de las respuestas disminuyó cuando se aumenta de dos a tres alternativas. En general, se obtienen mejores coeficientes de confiabilidad (consistencia interna Alfa de Cronbach) y de validez factorial (porcentaje de la varianza explicada por el primer factor) en escalas con cuatro a siete categorías.
Adelson & McCoach (2010)	Escala tipo Likert con cuatro y cinco categorías	Escala Likert con cuatro categorías: SD ( <i>Strongly disagree</i> ), D ( <i>Disagree</i> ), A ( <i>Agree</i> ) y SA ( <i>Strongly disagree</i> ). La codificación fue -2, -0.667, 0.667 y 2, respectivamente. Escala Likert con cinco categorías: SD ( <i>Strongly disagree</i> ), D ( <i>Disagree</i> ), N ( <i>Neither agree or disagree</i> ), A ( <i>Agree</i> ) y SA ( <i>Strongly disagree</i> ). La codificación fue -2, -1, 0, 1, y 2, respectivamente.	<i>Math and Me Survey</i> , el cual incluye dos subescalas: <i>Enjoyment of Mathematics</i> (10 ítems) y <i>Mathematical Self-Perception</i> (8 ítems)	606 estudiantes de tercero a sexto grado (304 usaron la escala con cuatro categorías y 302 usaron la de cinco) de seis escuelas elementales e intermedias en los Estados Unidos de América.	Estudiantes respondieron de manera similar usando ambos formatos de respuesta. Las respuestas usando ambos formatos produjeron estructuras factoriales, varianzas y medias aritméticas de los ítems similares. Sin embargo, los resultados del análisis factorial confirmatorio indicaron que la escala de cinco puntos se ajustaba mejor a las respuestas. También el coeficiente de confiabilidad en la subescala de <i>Self-Perception</i> fue mayor con la escala de cinco categorías. Favorece el uso de cinco categorías de respuesta con estudiantes de tercero a sexto grado de escuela elemental.
Chomeya (2010)	Escala Likert con cinco y seis	Categorías no provistas	Evaluación de tres pruebas psicológicas acerca de la	180 estudiantes subgraduados de la Universidad de	Escala Likert con seis categorías produjo índices de

Autores/Año	Número categorías	Identificación o rotulación	Instrumento	Personas que respondieron	Resultados
	categorías		actitud hacia el consumo de alcohol, locus de control y motivación hacia el aprovechamiento (60 estudiantes contestaron cada prueba), junto con el nivel de facilidad y dificultad de tomar la decisión de contestar cada prueba.	Mahasarakham, Tailandia	discriminación de los ítems más altos en las dos de las tres pruebas (actitud hacia el consumo de alcohol y locus de control) y coeficientes de consistencia interna Alfa más altos en las pruebas de motivación hacia el aprovechamiento y actitud hacia el consumo de alcohol. La estructura factorial aplicando ambas escales fue bastante similar, en términos de la cantidad de componentes identificados y el porcentaje de varianza que pueden explicar. Encontraron mayor nivel de dificultad al contestar dos de las pruebas (actitud hacia el consumo de alcohol y locus de control) con la escala de seis categorías
Contractor & Fox (2011)	Estudio 1: cinco, seis, nueve y diez categorías Estudio 2: cinco, seis, nueve y diez categorías	Estudio 1: Cuatro escalas con cinco categorías (5 <i>Excellent</i> , 4 <i>Very good</i> , 3 <i>Good</i> , 2 <i>Fair</i> , 1 <i>Poor</i> ); seis (6 <i>Excellent</i> , 5 <i>Extremely good</i> ; 4 <i>Very good</i> ; 3 <i>Good</i> ; 2 <i>Fair</i> , 1 <i>Poor</i> ); nueve (9 <i>Like it Extremely</i> ; 8 <i>Like it Very Much</i> ; 7 <i>Like it Moderately</i> ; 6 <i>Like it Somewhat</i> ; 5 <i>Neither Like/Dislike</i> ; 4 <i>Dislike it Somewhat</i> ; 3 <i>Dislike it Moderately</i> ; 2 <i>Dislike it Very Much</i> ; 1 <i>Dislike it Extremely</i> ; diez (Solo los extremos se rotularon: 10 <i>Excellent</i> y 1	Llevaron a cabo dos estudios. En el Estudio 1 entrevistaron a consumidores para evaluar tres tipos de productos (instrumentos de escritura, jugo de china y pañuelo facial) de diferente calidad. El Estudio 2 realizaron una tarea donde indicaban su opinión en tres contextos (alquiler mensual, lluvia anual, puntuaciones de exámenes)	Estudio 1: Consumidores reclutados en seis regiones geográficas de los Estados Unidos de América Estudio 2: 450 estudiantes universitarios de Administración de Empresas	Escalas con cinco ó seis categorías son más sensitivas* que las de 9 y 10 categorías. La mayoría (63.7%) de los consumidores en el primer estudio prefirieron la escala con seis categorías. Los resultados del Estudio 2 coinciden con los del Estudio 1 en que la sensibilidad de las escalas no está asociada con la cantidad de categorías. También en el Estudio 2, las escalas con cinco o seis categorías demostraron mayor sensibilidad.

Autores/Año	Número categorías	Identificación o rotulación	Instrumento	Personas que respondieron	Resultados
		<i>Poor</i> Estudio 2: Cinco escalas con cinco, seis, siete, nueve y diez categorías con los extremos rotulados: <i>About the same</i> y <i>A whole lot less</i> .			
Freund, Tiejens & Strauss (2013)	Comparación entre cuatro y seis categorías de respuesta	Seis categorías de respuesta ( <i>false, mostly false, more false than true, more true than false, mostly true &amp; true</i> ) Cuatro categorías ( <i>false, mostly false, mostly true &amp; true</i> )	<i>Physical Self-Description Questionnaire</i> . Contiene cinco escalas ( <i>Coordination, Endurance, Flexibility, Sports competence &amp; Strenght</i> ) con seis ítems cada una	905 estudiantes alemanes (edad promedio 14.96 años) contestaron el instrumento con cuatro categorías de respuestas y 806 (edad promedio 14.91 años) con seis categorías	Coeficientes de confiabilidad de consistencia interna (Alfa de Cronbach) similares para las cinco escalas con cuatro y seis categorías (valores entre 0.81 y 0.91). También fueron comparables las medias aritméticas y desviaciones estándares de las puntuaciones en las subescalas. Mediante la aplicación <i>Mixture Distribution Rasch Model</i> hallaron que las cuatro categorías se ajustan mejor que las seis y que induce menos la confusión de los estudiantes.

\*Porcentaje de personas que seleccionan cada posición se convierten en valores sigmas ( $\sigma$ ) o de desviaciones estándar en la curva normal (-3 a +3) que corresponden a cinco intervalos.

+Sensitividad se refiere a detectar diferencias entre las calificaciones de los constructos u objetos que se están midiendo, aplicando procedimientos estadísticos (e.g., diferencias entre la media aritmética de las respuestas de dos grupos de personas mediante el Análisis de Varianza).

Si desea citar alguno de los artículos presentados en este boletín, recomendamos utilizar el formato que especifica el Manual de estilo de publicaciones de la *American Psychological Association* (2010). A continuación se presenta un ejemplo de cómo citar un artículo de un boletín electrónico.

Vázquez, J. P. (2007, marzo). Estudio de Evaluabilidad. *INEVA en acción*, 3(1). Recuperado el 1 de marzo de 2007, de <http://ineva.uprrp.edu/boletin/boletin8.pdf>

Las opiniones vertidas en esta publicación son de los (las) autores(as) y no representan las del Programa de INEVA. Las personas interesadas en escribir para esta publicación deben comunicarse con la Junta Editora a nuestra dirección electrónica ([ineva@uprrp.edu](mailto:ineva@uprrp.edu)). De la misma manera pueden enviarnos sus comentarios y sugerencias acerca de esta publicación.

#### **JUNTA EDITORA**

Carlos R. Carrasquillo Ríos  
Luz G. Cátala Rivera  
Angélica Cortés Ortiz  
Elba Ortiz Laureano  
Nellivette Prieto Vázquez  
Careli Sánchez Pabón  
María B. Serrano Abreu  
Sarai Torres Ruiz  
Rose M. Vincenty Colón  
Dra. Claudia X. Alvarez Romero  
Dr. Víctor E. Bonilla Rodríguez

*Conceptuación gráfica*  
*Víctor E. Bonilla Rodríguez, Ph. D.*