



## Serie de Posts Número 6

**¿Cómo debiera realizarse la Investigación Educativa en torno a la relación: calidad del aprendizaje / de la formación vs. preparación profesional de los docentes?  
(4° parte)**



¡La encuesta ha sido todo un éxito! Estoy muy agradecido del esfuerzo de sus más de 150 respondientes; tanto compañeros cubanos (aún en medio de las marcadas afectaciones eléctricas de días recientes), como de los solidarios colegas de: España, República Dominicana, México, Honduras, Colombia, Venezuela, Ecuador y Brasil que, igual, la contestaron. Ello nos permitirá –de aquí en lo adelante– trabajar con datos reales, algo de gran valor en términos epistemológicos.

Dando continuidad a la parte tercera, y asistidos ahora de esta nueva base de datos, vamos a responder –sobre la base de los recursos más actuales de la Psicometría– la pregunta de: ‘¿cuán buenos son nuestros instrumentos?’. Esto, antes de contestar la pregunta científica principal de este Número 6, de: ‘¿en qué medida impacta la

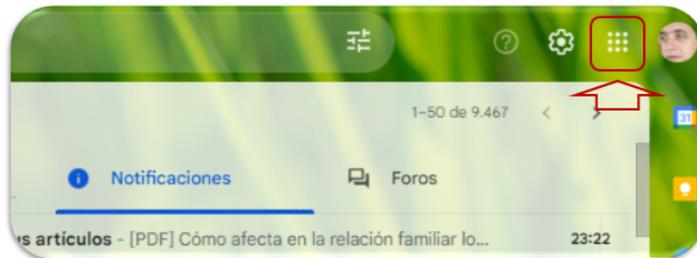
preparación profesional de los educadores, el aprendizaje y la formación de sus educandos?'; lo que haremos en la quinta y última parte de este importante Post.

Pero antes de responder la primera de las interrogantes anteriores, abriré un espacio para explicar –aunque sea muy brevemente– el procedimiento a seguir para hacer uso del recurso ‘Formulario’ de Google; algo que parece que llamó la atención a más de uno de mis lectores cubanos, por su factibilidad, elegancia y sencillez (en términos de programación). Sin más dilaciones, ¡comencemos!...

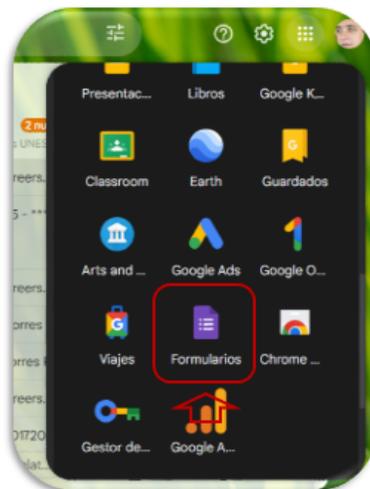
## ¿Cómo puedo desarrollar un cuestionario desde Google?

Pues comenzaré por revelar que –aunque ya conocía la opción– solo hasta unos días antes de escribir la segunda parte del Número de Post actual no lo había utilizado. Nada, que ‘cacharrear’ (como decimos por acá, en Cuba) y asistirse de ese ‘sabelotodo’ que es Internet siguen siendo recursos que nos llevan al éxito de casi cualquier empeño informático; solo hay que vencer ese temor a lo ignoto que suele paralizarnos. Si se dispone de una cuenta de correo en Gmail, el procedimiento es el siguiente.

1. Abrir la cuenta de correo en Google y ubicar en esa página Web el botón de las restantes aplicaciones correspondientes a Google, como se muestra en la imagen de abajo.

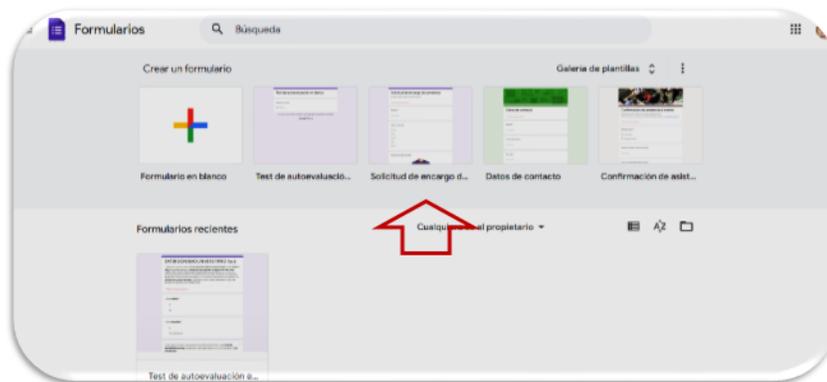


2. Buscar la aplicación “Formulario” en la lista desplegable (bien debajo).



3. Se abrirá una nueva página Web donde está la opción de utilizar una plantilla prototípica (es decir, según la intención que se tiene con el formulario), o crear una

totalmente nueva; si ya fue creada anteriormente una o varias plantillas, también puede acceder desde ahí a ellas y modificar la seleccionada.



4. Una vez elegida la plantilla, se abrirá otra página Web que dispone de las secciones y botones para diseñar un formulario personalizado; también se tienen en la parte superior varias pestañas para diferentes fases del proceso, aunque se abre –por defecto– en la página de la pestaña ‘Preguntas’. Aunque Google dispone de su propia ayuda, pueden encontrarse en Internet (por ejemplo, en YouTube) numerosos tutoriales de cómo proceder. Con el botón ‘Enviar’ se puede acceder a un link, que puede posteriormente ofrecer –junto con la invitación a que sea respondida– a través de los canales elegidos para difundirla (por ejemplo, a través de redes sociales).



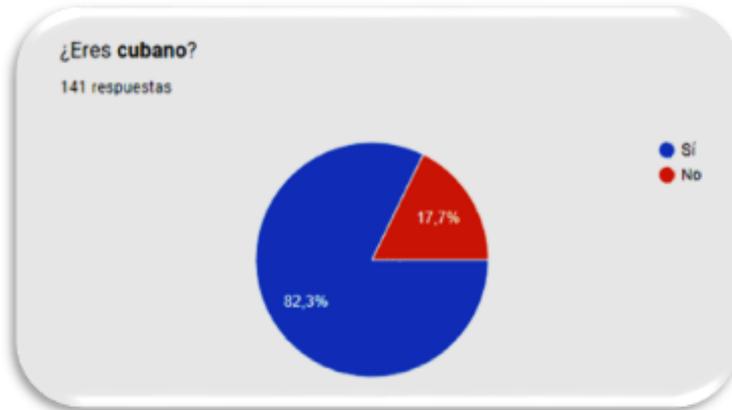
5. Después de vencida la etapa de captación de datos, se puede –desde la página de la pestaña ‘Respuestas’– impedir que se siga contestando (con un botón disponible), así como descargar la base de datos conformada (por ejemplo, con Microsoft Excel); también Google va generando gráficos estadísticos con los datos que se van acumulando (de pastel, si la variable subyacente en la pregunta es dicotómica (por ejemplo: ‘Sí’ o ‘No’); o un gráfico de barras, si la variable de referencia es ordinal (por ejemplo: ‘1’ [‘Muy en desacuerdo’], ..., ‘5’ [‘Muy de acuerdo’])).

152 respuestas [Ver en Hojas de cálculo](#)

No se aceptan más respuestas

Mensaje para los encuestados

Ya no se aceptan respuestas en este formulario



Ahora bien, a la hora de elaborar un instrumento de investigación es muy importante también el contenido de cada reactivo, y no solo su forma... ¿Qué aspectos se deben tener en cuenta, en ese sentido?... Bien, ya aquí no estamos hablando ni de Informática, ni –siquiera– de Psicometría, sino de Metodología de la Investigación Científica; y en ese paso del desarrollo investigativo –desafortunadamente– suelen existir notables falencias, al menos entre los investigadores educativos cubanos contemporáneos (Torres, 2016b).

Por eso –y aunque me siga separando un poco más del propósito de esta cuarta parte, del actual Número de la Serie de Posts– voy a realizar un breve recordatorio de cómo se debiera proceder. Yo lo he llamado ‘camino descendente’ (desde las variables/categorías, a los reactivos de los instrumentos de investigación), pero está esbozado en cualquiera de las obras clásicas de la investigación científica, como es el caso de (Kerlinger, 1985), o (Hernández-Sampieri, Fernández y Baptista, 2010).

Una vez establecido el marco teórico-referencial de la investigación –y teniendo en cuenta las categorías teórico-metodológicas del diseño que le precedió– se debe

trabajar con las variables y/o categorías principales de la investigación [hablo de variables para referirme a atributos esenciales del objeto de investigación que son percibidas bajo el enfoque cuantitativo; y de categorías para hacerlo con sus similares desde la perspectiva del enfoque cualitativo de investigación (Torres, 2016a)].

¿En qué consiste ese trabajo ‘post establecimiento de la plataforma teórica de partida’?... Pues, en el establecimiento de las definiciones teóricas y operacionales de aquellas. Las definiciones teóricas de las variables/categorías principales debieron quedar contextualizadas durante la delimitación del marco teórico-referencial, pero toca ahora explicar en qué consisten los términos con que las designamos.

Cada una de esas ‘definiciones teóricas’ están compuestas por atributos o rasgos esenciales del objeto o proceso, sujeto a estudio, y cada uno de ellos debiera constituir una dimensión (o primer nivel de división de la variable o categoría); dando lugar así al inicio de la correspondiente definición operacional (Torres, 2016b).

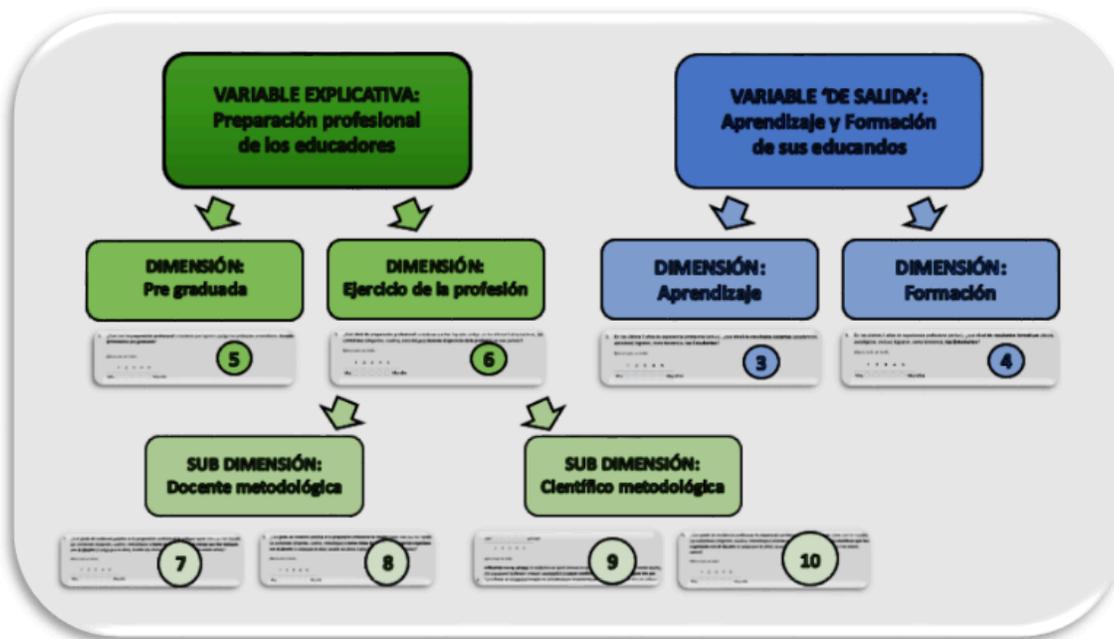
Uno de los errores más frecuentes identificados, entre los 1 377 resultados de investigación analizados para (Torres, 2016b), es salir en búsqueda de los indicadores de la variable/categoría principal de forma desligada de su definición teórica y, por tanto, del marco teórico-referencial. Vamos, ¡provocando toda una desarticulación del acto investigativo! (o ‘desmadre’, como dirían mis colegas mexicanos).

Y por ese camino –errado, muy errado– se puede terminar indagando (con los instrumentos de investigación elaborados) acerca de ‘cosas’ que no tienen que ver directamente con el objeto de investigación (y, por extensión, con el problema científico, con el objetivo de investigación y con la hipótesis general de trabajo).

En la primera parte de este Número destacamos un atributo esencial de una investigación científica es ‘buscar, en cada momento, la mayor objetividad posible’ (<https://drive.google.com/file/d/14reH5H00hXdcU-jlbqvWb6z2JdxsFiq7/view?usp=drivesdk>).

Cada decisión de tipo diferente que asumamos, durante el desarrollo de una investigación, le estará incorporando a ella más subjetividad (a la ya existente, por el hecho de ser los propios investigadores sujetos, con conocimientos, necesidades y conductas propios). Por ese camino, podemos estar distanciándonos del objeto de investigación; en vez de acercarnos a él, con el interés de ‘capturar’ su esencia, su ‘movimiento interno’. Y esto es válido durante la fase del diseño de los instrumentos de investigación.

En nuestro caso, aquí, en este estudio de contraste entre la calidad de la superación profesional de los educadores y la del aprendizaje y la formación de sus educandos, la lógica metodológica anterior se concretó de la siguiente forma: dos variables de investigación (una predictiva y otra ‘de salida’), y varias dimensiones en torno a cada una de ellas, como se muestra en la representación gráfica siguiente.



Representación de la definición operacional de las variables del estudio (Elaboración propia).

De la imagen anterior se desprende una decisión metodológica esencial: cada reactivo (o ítem) de un instrumento de investigación debe corresponderse (mejor, derivarse) de un único indicador resultante de la definición operacional de la variable o categoría; si esa relación estricta no resulta evidente, entonces no se ha llegado aún a un indicador, y se está aún a nivel de una dimensión o de una subdimensión (Torres, 2016b).

Lo dicho y representado más arriba constituye lo esencial del ‘camino descendente’; pero este nos habla tan solo de la validación de un instrumento de investigación (o sea, de la respuesta a la pregunta: ‘¿el instrumento medirá/valorará bien lo que se quiere medir/valorar, con la investigación?’).

Sin embargo, el principio metodológico asociado al trabajo con los instrumentos de investigación nos demanda –además– el análisis de su confiabilidad (el responder también la pregunta: ‘¿proporciona el instrumento diseñado resultados consistentes, ante su aplicación sucesiva en las unidades de análisis de la investigación?’).

Pero ya alertamos en la parte anterior que los que realizamos investigaciones en las ciencias sociales y humanísticas estamos en desventaja, en relación con nuestros pares de las ciencias naturales y tecnológicas, en lo que a la confiabilidad de los instrumentos de investigación se refiere. Ellos trabajan con objetos de investigación objetivos; a nosotros nos toca lidiar –en cambio– con el ‘mundo interior’ de las personas, o con la siempre imprecisa relación entre esas subjetividades –difíciles de ‘atrapar’– y nuestro objeto de investigación.

Al mismo tiempo, la ventaja anterior les ha permitido a los investigadores de las ciencias naturales y tecnológicas trabajar más intensamente en la calibración de sus instrumentos de investigación, al punto de disponer de un margen de error de medida probable; mucho más preciso que el que pudiéramos estimar los investigadores de las

ciencias sociales y humanísticas con los nuestros; este hace más convincente sus **criterios de confiabilidad**, en comparación con los nuestros.

Así, por ejemplo, pudiéramos hablar del **Microscopio de fuerza atómica (AFM)**, que utiliza una punta afilada para medir superficies a nivel atómico, con lo cual se mejora la resolución y la precisión de la medida, en comparación con microscopios electrónicos tradicionales; o de la **Espectroscopia de infrarrojos (IR)**, que mide la absorción de luz infrarroja por las moléculas, permitiendo mediciones más precisas y rápidas; o de la **Tomografía por coherencia óptica (OCT)**, que proporciona una imagen de alta resolución de los tejidos biológicos, haciendo que mejore la precisión de la imagen y la reducción de los errores de alineación; entre otros muchos ejemplos.

En cambio, nosotros –los investigadores sociales y humanistas– pretendemos obtener nuevos **hallazgos científicos**, en nuestros **campos** respectivos, con el diseño y empleo de indelebles **instrumentos** concebidos para **fuentes orales, visuales y documentales**, muy dependientes todas de decisiones y condiciones **subjetivas**. Es por ello que –como explicamos en la **parte anterior**– resulta muy importante este tema de **confiabilidad de los instrumentos de investigación científica**.

Regresemos sobre él –como previmos entonces– ahora con la ventaja de disponer **datos reales (tangibles)** sobre la encuesta diseñada.

## ¿Qué ‘dicen’ los datos captados sobre la confiabilidad de la encuesta?

Antes de responder la pregunta, debemos estar claro –primero– que la encuesta que hemos diseñado no es un **instrumento de investigación ‘canónico’**. Primero, porque hemos asumido la **medición de la ‘variable de salida’ (la calidad del aprendizaje y de la formación de los estudiantes)**, con los mismos informantes que los de la **‘variable explicativa’ (la calidad de la preparación profesional de los educadores)**; cuando lo razonable hubiera sido que para medir la primera se hubieran utilizado **pruebas objetivas**, aplicables directamente a los estudiantes.

Por otra parte, porque hemos incluido dos **variables concomitantes (‘nacionalidad’ y ‘género’)**, que –sí– resultarán de utilidad para la respuesta de la pregunta central del estudio (a atender en la quinta y última **parte** de este Número 6, próximamente), en tanto **variables** que nos ayudan a comprender mejor la **composición de la muestra** utilizada, pero que aquí –donde el foco es el **instrumento, su confiabilidad**– resultan algo desajustadas. No obstante, la utilizaremos para remarcar las diferencias en los **estudios psicométricos de reactivos dicotómicos** (como ellas dos) y de **reactivos** que puedan llamarse de **‘respuesta graduada’**, como ya esbozamos en la parte anterior.

Dicho esto, comencemos con los resultados de los **estudios psicométricos**, ya con los datos aportados directamente por ustedes, mis apreciados lectores. Ahora, como en cualquier otro **análisis estadístico** –y los **psicométricos** lo son– es aconsejable hacer primero **revisiones ‘descriptivas’** (esto es, de lo observado únicamente en la **muestra**),

antes de pasar a los generalizables ‘estudios inferenciales’ (o sea, de lo esperado en la población toda). Comencemos por una imagen (parcial) de la base de datos creada.

*Base de Datos. Primeros casos*

Nacionalidad	Genero	Logros_cognitivos	Logros_formativos	Formacion_posgrado	Formacion_permanente	Reuniones_metodologicas	Visitas_tecnicas	Estudios_cientificos	Conferencias_cientificas
0	0	3	3	1	1	2	3	3	3
1	0	5	5	5	5	3	3	2	2
1	1	3	3	5	3	3	3	2	2
1	0	3	3	3	3	3	3	3	3
1	0	4	3	4	4	3	3	3	3
								2	3

Vista de los seis primeros registros de la base de datos de las respuestas (Elaboración propia).

Y ahora veamos las principales estadísticas de resumen (tanto de tendencia central, como de dispersión) de las preguntas de la encuesta.

*Estadísticos de las respuestas por preguntas*

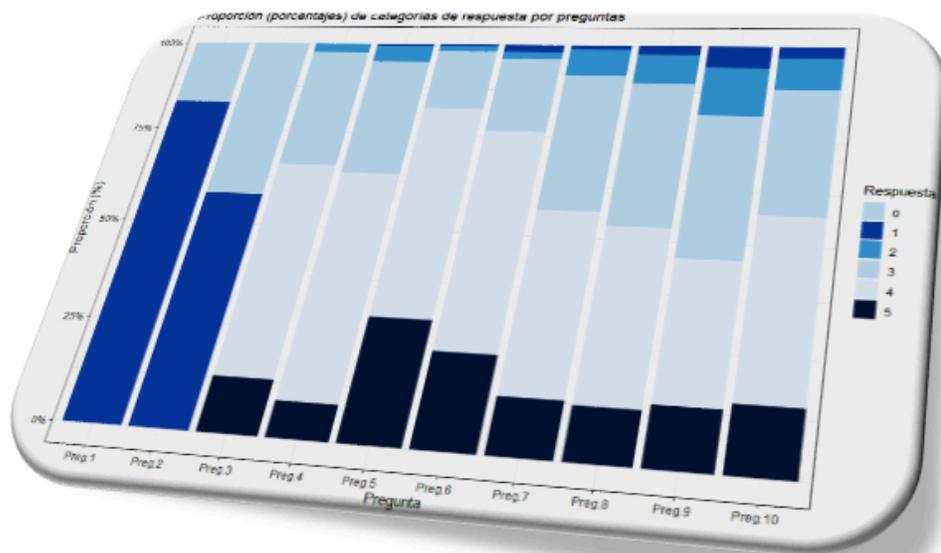
Pregunta	Media	Mediana	Moda	Desv_Est	Varianza	Rango
Preg.1	0.827	1	1	0.380	0.144	1
Preg.2	0.577	1	1	0.496	0.246	1
Preg.3	3.763	4	4	0.701	0.492	3
Preg.4	3.667	4	4	0.721	0.520	4
Preg.5	4.083	4	4	0.745	0.554	4
Preg.6	3.923	4	4	0.815	0.665	4
Preg.7	3.596	4	4	0.848	0.720	4
Preg.8	3.526	4	4	0.898	0.806	4
Preg.9	3.353	3	3	1.052	1.107	4
Preg.10	3.571	4	4	0.951	0.905	4

Representación tabular de estadísticas de resumen de las preguntas (Elaboración propia).

De donde podemos identificar que, entre las dos preguntas dicotómicas (es decir, de categorías 0 y 1), la de mejor comportamiento en la distribución de los datos es la ‘Pregunta 1’, pues tiene una media superior a la de la ‘Pregunta 2’ (0.827 vs. 0.577), y los datos se dispersan menos que en la ‘Pregunta 2’, en torno a la media (0.380 vs. 0.496).

Mientras que, entre las variables politómicas (o sea, con categorías 1, 2, 3, 4 y 5), la de mejor comportamiento de la distribución de sus datos es la ‘Pregunta 5’, con una media (o promedio) superior a la de las demás (4.083), a la vez que la segunda menor dispersión de los datos en torno a ella (0.745).

Estos comportamientos se pueden apreciar mejor gráficamente (con un gráfico de barras apiladas), representando en ellas las proporciones de las categorías que admiten los dos tipos de variables representadas en las preguntas de la encuesta.



Representación gráfica de las proporciones de categorías elegidas, por preguntas (Elaboración propia).

Hecho esto, podemos pasar al análisis psicométrico del instrumento y –asumiendo la Teoría de Respuesta al Ítem [TRI], además– el de sus reactivos, por separado. Ya sabemos que tenemos dos preguntas que no tienen una implicación importante para la pregunta principal del estudio (justo las que utilizan una escala dicotómica), pero vamos a aprovecharlas aquí para repasar los recursos de los análisis psicométricos para ese tipo de reactivo.

Es como suponer –aunque para nuestro ejemplo, ahora, no aplica– que el hecho de ser mujer (o ser cubano) tiene una implicación superior en la relación entre variables explicativas y ‘de salida’ (pudiera ser que para la variable ‘nacionalidad’, trascienda el hecho de que en Cuba se tiene diseñado un tipo muy peculiar de preparación profesional durante el ejercicio de la profesión, de modo que tendría sentido decir que el atributo ‘ser cubano’ (1) es –para un estudio interesado en el Trabajo Metodológico– superior a la cualidad ‘no ser cubano’ (0); pero –reitero– aquí eso no aplica; de modo que solo lo utilizaremos para repasar lo ya visto. Cuando los reactivos utilizan una escala de medición (o de valoración) dicotómica, su análisis descriptivo inicial es aconsejable utilizando las frecuencias relativas (o proporciones) de manifestación de sus dos categorías.

*Base de Datos. Por cientos en Preguntas dicotómicas*

Pregunta	Porcentaje_1	Porcentaje_0
Preg.1	82.7	17.3
Preg.2	57.7	42.3

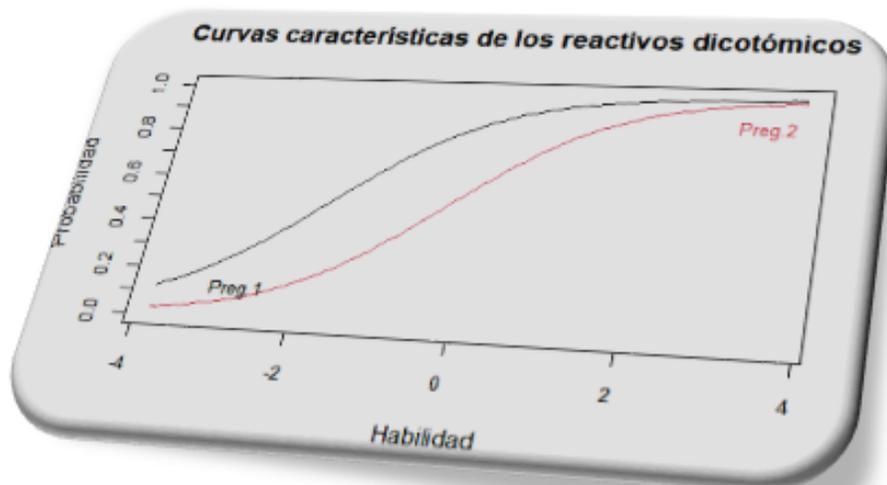
Representación tabular de las proporciones de las categorías en las preguntas dicotómicas (Elaboración propia).

Así como verificar que entre los datos acopiados no hay 'valores perdidos' o missing; en cuyo caso habría que estimarlos o excluir sus registros.



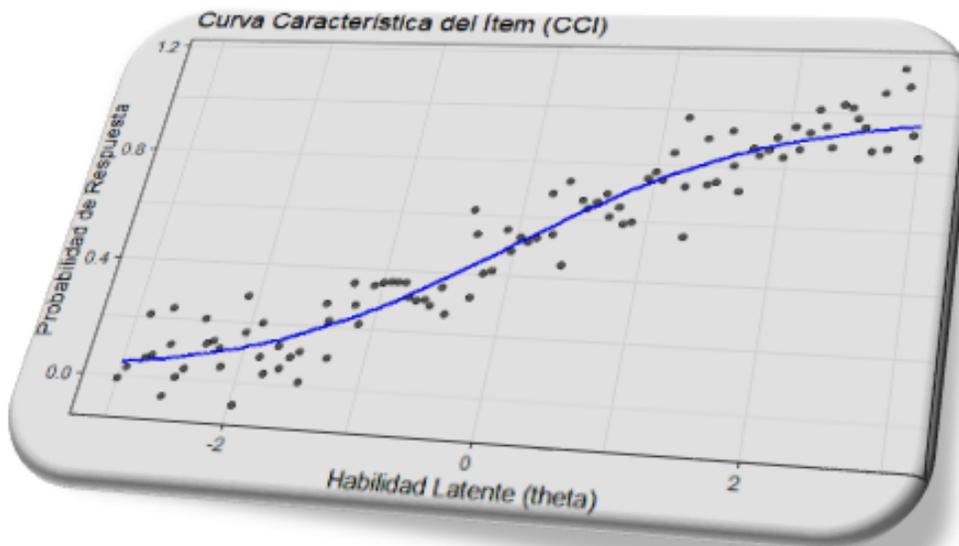
Representación gráfica de la revisión de la eventual existencia de missing dentro de la base de datos, para las preguntas dicotómicas, realizado con R-Project (Elaboración propia).

Ya vimos en la tercera parte –anterior– del actual Número, que lo óptimo es hacer el análisis psicométrico, en principio, con los reactivos del instrumento, sobre la base de la TRI, de modo que no dependamos de las características (siempre peculiares) de la muestra de respondientes. Y prestamos atención únicamente al parámetro de dificultad [b] de cada reactivo. En el gráfico siguiente podemos verificar que –muy probablemente– la 'Preg. 1' es más fácil que la 'Preg. 2'.



Representación de las curvas características de los ítems 1 y 2 (Elaboración propia).

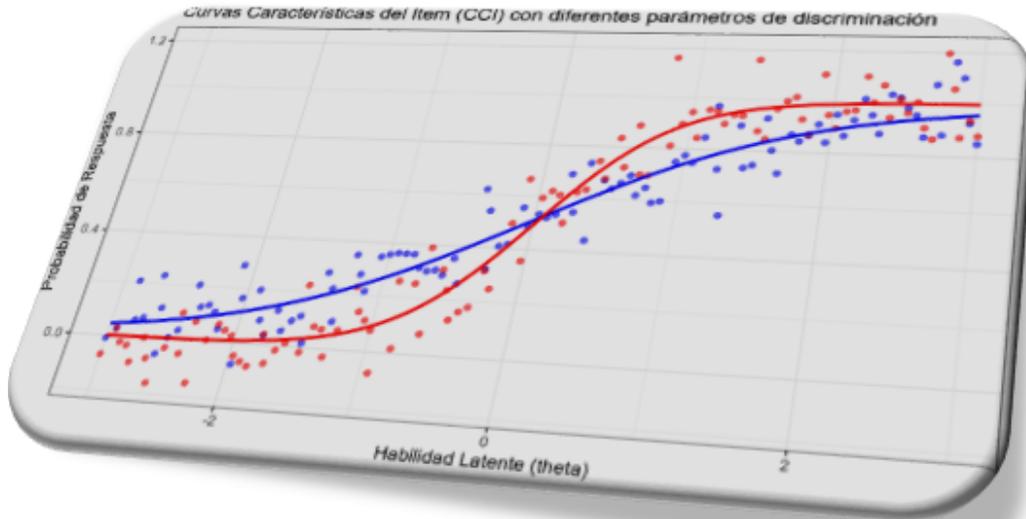
Puesto que había que repasar estos aspectos ahora, al disponer ya de una base de datos reales, aproveché para trabajar con R-Project (Torres, 2018) la idea básica de que esas curvas representan las líneas de mejor ajuste a la nube de los puntos del tipo ('habilidad latente'; 'probabilidad de respuesta exitosa'). Veámoslo gráficamente.



Representación de la curva característica del ítem como línea de mejor ajuste a la nube de puntos (Elaboración propia).

Es importante señalar que, en la teoría psicométrica asistida de la regresión binomial, es posible hablar de modelos de dos y hasta de tres parámetros. Los modelos estadísticos de dos parámetros estiman, además de la medida de la dificultad del ítem (o reactivo) [b], el parámetro de discriminación [a] que indica en qué medida el ítem logra separar a los individuos de alta probabilidad de éxito en sus respuestas, de los de baja probabilidad; de modo que ello queda determinado por el nivel de inclinación

que asume la curva característica en valores próximos a su punto de inflexión. Veámoslo en la siguiente representación gráfica construida con R-Project.



Representación de dos curvas características de ítems con igual parámetro de dificultad, pero diferentes parámetros de discriminación; donde la línea de color rojo discrimina mejor a los respondientes que la línea de color azul (Elaboración propia).

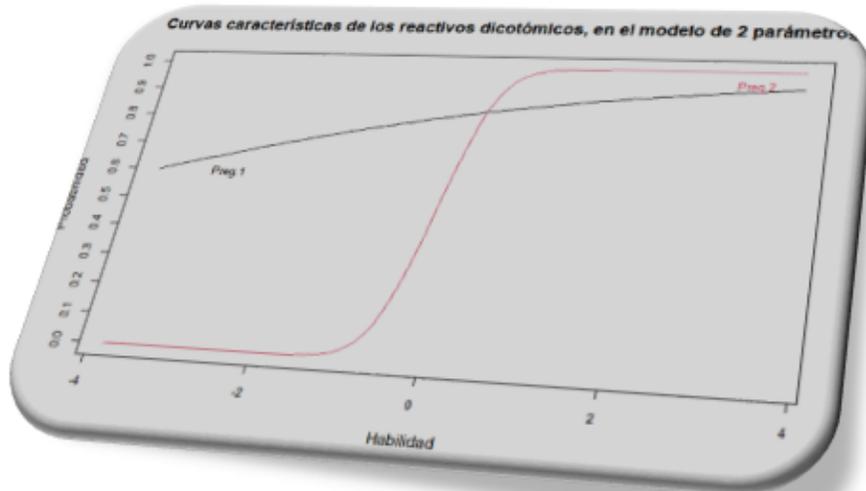
Si regresemos a nuestros datos reales, y al supuesto de que las preguntas 1 y 2 de nuestra encuesta conforman un instrumento de investigación a parte, se tendrían las siguientes estimaciones de parámetro de dificultad y de discriminación a la vez.

Estimaciones de los parámetros de dificultad, discriminación y adivinación, de las preguntas dicotómicas

Gussng	Dffclt	Dscrmn
0.0571	-0.871	8.39
0.4813	0.913	6.29

Representación tabular de las estimaciones simultáneas de los parámetros de dificultad y de discriminación de un 'instrumento' conformado por las preguntas 1 y 2 (Elaboración propia).

De modo que podemos apreciar que la 'Preg.2' es más difícil de responder y, a la vez, discrimina más que la otra pregunta del instrumento; o sea, que 'Preg.1'. Gráficamente sería como sigue.



Representación de las curvas de los ítems 1 y 2 con dos parámetros (Elaboración propia).

Ya habíamos dicho arriba que también pueden estudiarse las propiedades de los ítems dicotómicos a partir de modelos que estiman hasta tres parámetros. Se añade en ellos, a los parámetros de dificultad ('Dffclt') y de discriminación ('Dscrmn'), el parámetro de adivinación por azar ('Gussng').

Así, en nuestro ejemplo ficticio (aunque sustentado en datos reales) el primer reactivo del 'instrumento' es bastante fácil, pero con una considerable capacidad de discriminación y una muy baja probabilidad de adivinar su respuesta por azar; mientras que el segundo reactivo resulta ser algo difícil, así como con una considerable capacidad de discriminación, pero con un nivel medio de probabilidad de adivinación de su respuesta correcta.

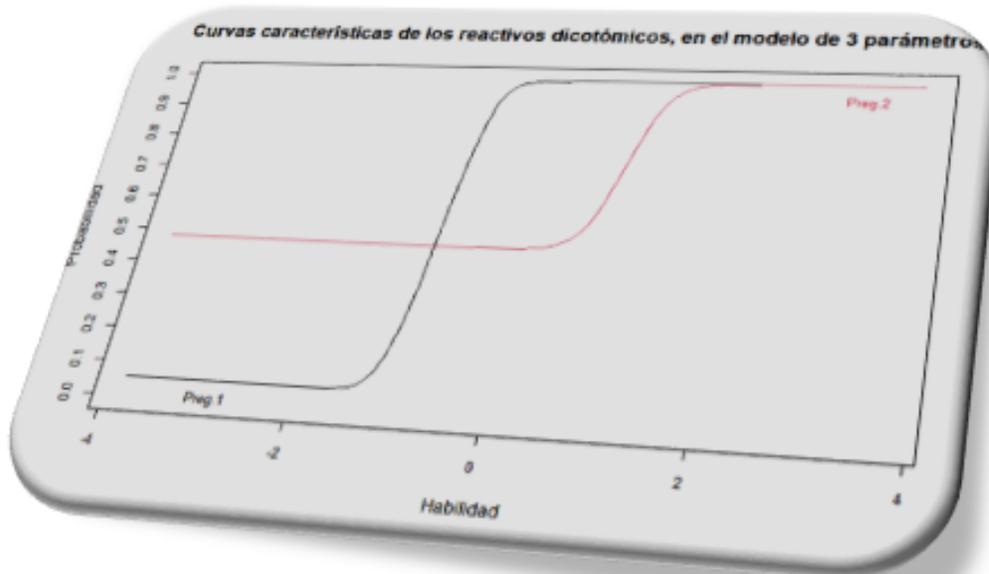
Estimaciones de los parámetros de dificultad, discriminación y adivinación, de las preguntas dicotómicas

Gussng	Dffclt	Dscrmn
0.0571	-0.871	8.39
0.4813	0.913	6.29

Representación tabular de los 3 parámetros estimados para los ítems dicotómicos 1 y 2 (Elaboración propia).

Y gráficamente tendríamos las siguientes curvas, como representación de los modelos de tres parámetros de los ítems dicotómicos de nuestro simulado 'instrumento'. Obsérvese como –a diferencia de las curvas características de los dos modelos anteriores– aquí las 'colas' de la izquierda no son asintóticas a la recta donde la probabilidad es siempre cero, sino que para respondientes con un

nivel de habilidad muy bajo (por ejemplo, -2 o -4) la probabilidad de responder de manera exitosa no está próxima a cero (como sería de esperar), pues el modelo detecta una cierta probabilidad de acertar por azar; en el caso del ítem 'Preg.2' (con el color rojo) esa probabilidad es 0.5; o sea, de un 50%.



Representación gráfica de los modelos de 3 parámetros para los ítems dicotómicos 1 y 2 (Elaboración propia).

En la siguiente tabla se pueden apreciar los criterios de aceptación de ítems dicotómicos en un instrumento de investigación, de acuerdo con los consensos más actualizados de la Psicometría; tanto desde la Teoría Clásica del Test (TCT), como desde la Teoría de Respuesta al Ítem (TRI).

En la tabla aparecen también los valores alcanzados por nuestros dos reactivos dicotómicos con las respuestas proporcionadas por nuestros lectores que pudieron, finalmente, participar. La mayoría de los indicadores de aceptación se cumplen, y solo unos pocos sugieren una revisión o exclusión de cierto ítem; pero se debe considerar— como ya habíamos aclarado— que las 'Preg.1' y 'Preg.2' no fueron elaboradas para conformar un instrumento de investigación propiamente, sino como referentes composicionales de la muestra, para ampliar la respuesta principal del estudio, en la quinta parte (y última) de este extenso Número 6 de la Serie de Post.

Criterios de aceptación de los reactivos dicotómicos, según los parámetros de la TCT y la TRI

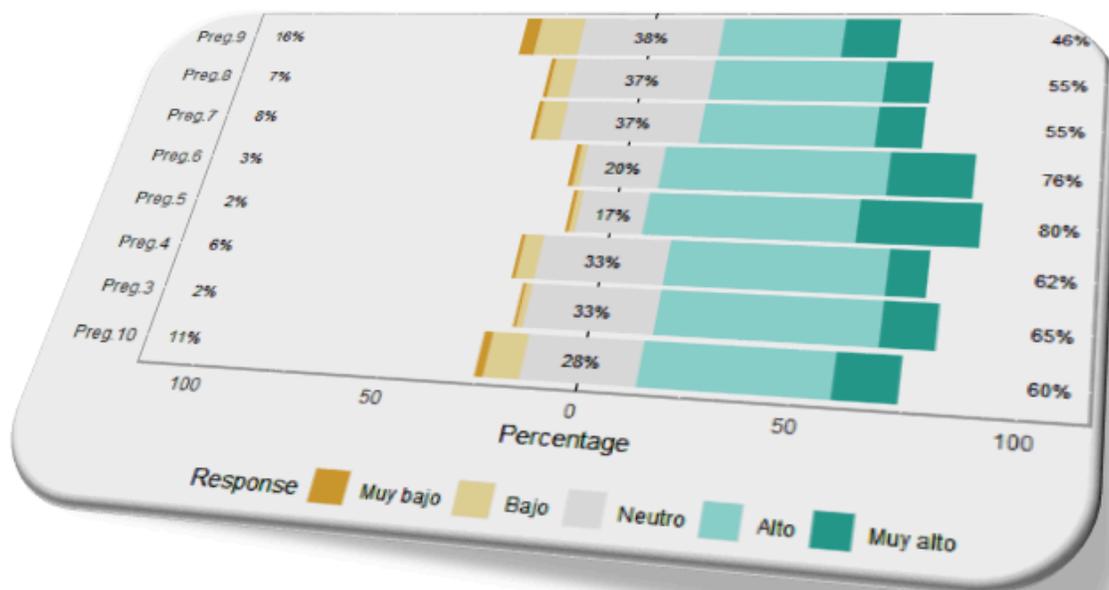
Aceptación	dificultad_TCT: a=0.6 o mayor	discriminación_TCT: item_test=0.7 o mayor	dificultad_TRI_Rasch: -3 < b1 < 3	dificultad_TRI_2PL: -3 < b2 < 3	discriminación_TRI_2PL: -3 < a2 < 3	Outfit_TRI: 0.75 < outfit < 1.3	Infit_TRI: 0.75 < infit < 1.3	Alpha-Conbrach_TCT: alpha=0.7 o mayor
Preg.1	82.7	0.75	-2.31	-5.09	0.31	0.97	1.10	0.09
Preg.2	57.7	0.75	-0.46	-0.20	4.76	0.82	0.95	0.09

Representación tabular de los criterios de inclusión de ítems en dicotómicos en instrumentos de investigación, de manera que puedan ser considerados confiables (Elaboración propia).

Como puede apreciarse, la tendencia es a combinar indicadores de la Teoría Clásica del Test (TCT) con similares de la Teoría de Respuesta al Ítem (TRI), como: **difficultad y discriminación**. Ahora bien, como estamos priorizando aquí el poder reportar argumentos convincentes de la ‘confiabilidad del instrumento’ (sin tener que repetir su aplicación varias veces) –y no tanto el ejecutar un **pilotaje** de este, con fines de conservar o sustituir **reactivos**– debemos prestar especial atención a los criterios de ‘**Outfit**’ (ajuste lejano) e ‘**Infit**’ (ajuste cercano) del modelo estadístico a los datos factuales; así como el coeficiente Alfa de Cronbach, inferido más allá de la muestra utilizada.

Pasemos ahora al análisis de lo que debe ser considerado nuestro ‘verdadero instrumento’: el conjunto de **reactivos politómicos** de la encuesta, compuesto por las ocho preguntas que van desde la número 3, a la número 10.

Del caso anterior, sabemos que lo aconsejable es primero hacer una **revisión de los datos disponibles desde la Estadística Descriptiva** (recuérdese, sin pretensiones de inferir más allá de la muestra). Veamos un resumen proporcionado por R-Project.

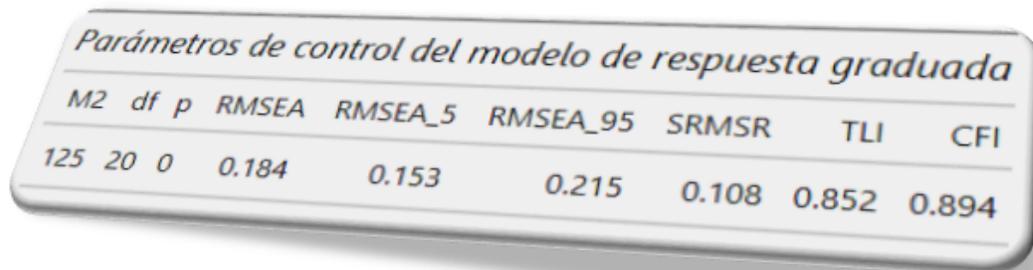


Representación gráfica de los porcentajes de repuestas en cada uno de los ítems politómicos (Elaboración propia).

También se debe verificar –recuerden– que no tenemos ‘valores perdidos’ o **missing** entre los datos de partida, algo que verifiqué con nuestro ‘software estrella’, quien generó un gráfico muy similar al homólogo para los reactivos dicotómicos, y que vimos más arriba.

A diferencia de los modelos de uno (o ‘modelo de Rasch’), dos o tres parámetros, que analizamos antes para **datos dicotómicos**; acá, en los reactivos de escalas

politómicas (como la típica ‘escala Likert’ que diseñamos para esta otra parte de nuestra encuesta) se debe utilizar el ‘modelo de respuesta graduada’ (GRM), ajustado igualmente a la Teoría de Respuesta al Ítem (TRI) y aportado por la Estadística Multivariada (Attorresi, Abal, Galibert, Lozzia, Aguerri, 2011). Las ‘librerías’ (o ‘paquetes’) especializados de R-Project se ocupan de estimar parámetros de la calidad del ajuste del modelo a los datos, entre los que destacan ‘M2’, ‘RMSEA’ (‘Root Mean Square Error of Approximation’) y el intervalo de confianza de este último; veamos los correspondientes a nuestros datos.

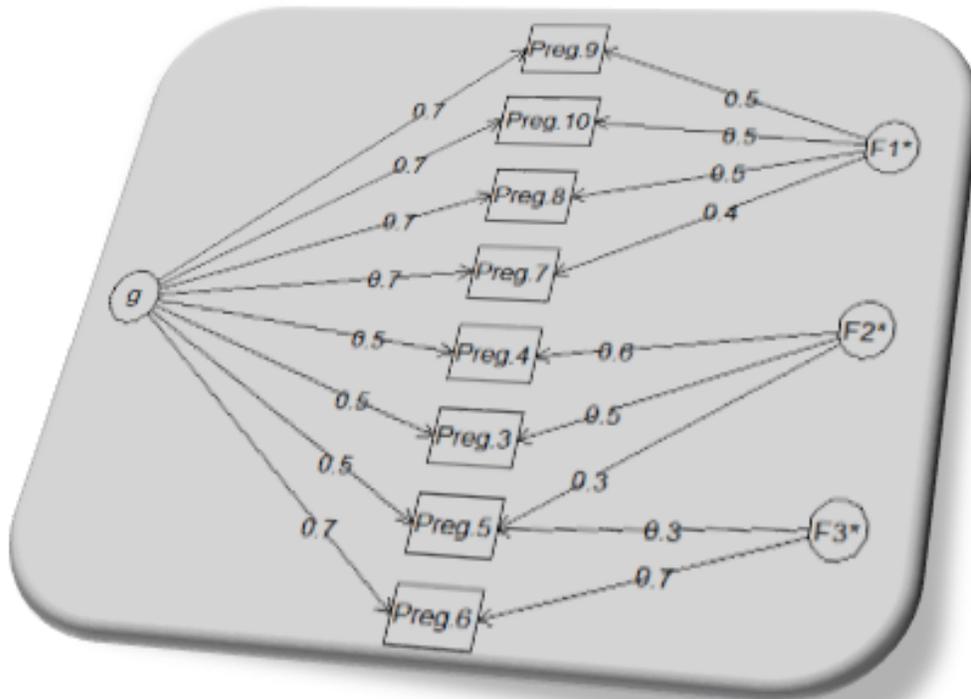


Parámetros de control del modelo de respuesta graduada								
M2	df	p	RMSEA	RMSEA_5	RMSEA_95	SRMSR	TLI	CFI
125	20	0	0.184	0.153	0.215	0.108	0.852	0.894

Representación tabular de los parámetros fundamentales de ajuste del modelo de respuesta graduada a nuestros datos politómicos, de reacción a la encuesta (Elaboración propia).

Ciertamente, en este caso ninguno de los tres parámetros resulta óptimo; posiblemente, porque estamos considerando dentro de nuestro ‘instrumento’ los reactivos ‘Preg.3’ y ‘Preg.4’ que –en rigor– debieron constituir sendas pruebas objetivas (Torres, 2016c), a aplicar –por demás– en robustas muestras de estudiantes; y no como dos simples preguntas de opinión a sus docentes, a través de la encuesta.

El gráfico siguiente –del Análisis Factorial Confirmatorio (AFC) asociado– nos muestra también la debilidad del ajuste a nuestros datos, puesto que en nuestro ‘factor g’ –quien debiera hablarnos a favor de la ‘unidimensionalidad de los reactivos’ del ‘instrumento’, en tanto supuesto (o exigencia previa) de la TRI– tenemos 3 cargas factoriales con valores inferiores a 0.7.



Representación gráfica de los resultados del AFC asociado al GRM (Elaboración propia).

Como quiera que lo que estamos queriendo en esta parte del Número es ilustrar procedimientos actualizados (y, por tanto, muy precisos), y posibilidades de hacer uso de ellos con un software libre y de código abierto, a nuestro alcance, seguiremos adelante sin acometer las rectificaciones aconsejables.

Ya sabemos que –aunque lo que nos interesar aquí es analizar la confiabilidad del instrumento diseñado– la TRI nos permite realizar análisis independientes para cada uno de los reactivos que ‘lo cargan’. Veamos ahora, entonces, el comportamiento del parámetro de discriminación [ $\alpha$ ] y de los ‘puntos de umbral’ o dificultad [ $b_1$ ,  $b_2$ ,  $b_3$  y  $b_4$ ] de cada uno de los reactivos del instrumento, que a diferencia de los del análisis anterior, están compuestos por escalas politómicas (específicamente, del tipo Likert). Los reactivos más robustos son los 4 últimos.

*Parámetros de localización del modelo de respuesta graduada*

Preguntas	a	b1	b2	b3	b4
Preg.3	1.07	-5.17	-3.84	-0.745	2.16
Preg.4	1.05	-5.20	-3.14	-0.683	2.68
Preg.5	1.16	-4.86	-3.89	-1.582	1.00
Preg.6	1.85	-2.90	-2.43	-0.949	1.06
Preg.7	2.96	-2.63	-1.60	-0.199	1.29
Preg.8	3.68	-2.15	-1.39	-0.113	1.28
Preg.9	3.33	-1.77	-1.01	0.099	1.22
Preg.10	3.35	-2.06	-1.28	-0.213	1.17

Representación tabular del parámetro de discriminación [a] y de los 'puntos de umbral' [b1, b2, b3 y b4] de cada uno de los reactivos del instrumento (Elaboración propia).

También los actuales (¡y potentes!) recursos resultantes de la conjunción de la Psicometría con la Estadística Inferencial Multivariante (pero puestos amigablemente a nuestro alcance con software como R-Project) nos proporcionan parámetros de control –no ya del modelo de ajuste a las respuestas del instrumento, sino también– de los reactivos politómicos, por separado.

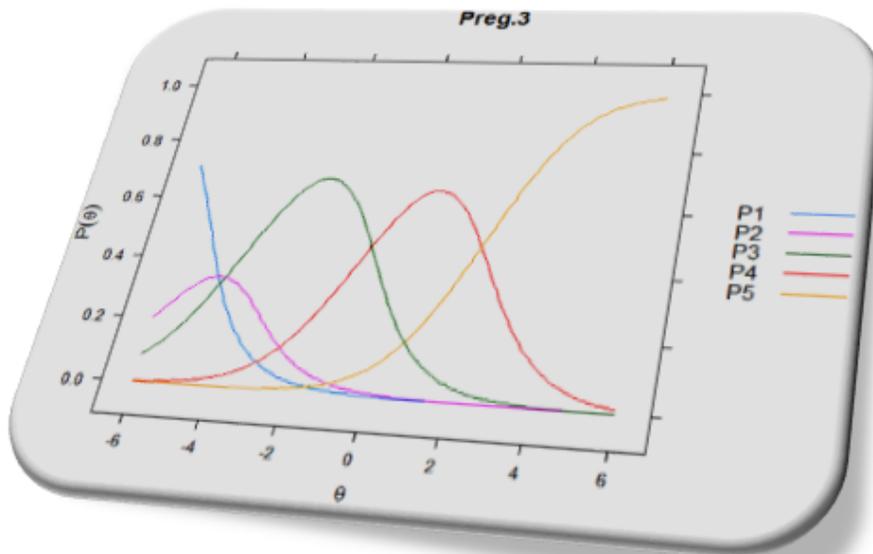
*Parámetros de control de los reactivos, según el modelo de respuesta graduada*

Item	S_X2	df.S_X2	RMSEA.S_X2	p.S_X2
Preg.3	31.63	12	0.1027	0.00158
Preg.4	20.06	12	0.0658	0.06602
Preg.5	27.24	15	0.0725	0.02687
Preg.6	9.76	13	0.0000	0.71367
Preg.7	14.62	9	0.0635	0.10191
Preg.8	5.66	8	0.0000	0.68493
Preg.9	23.42	12	0.0784	0.02438
Preg.10	9.55	9	0.0198	0.38836

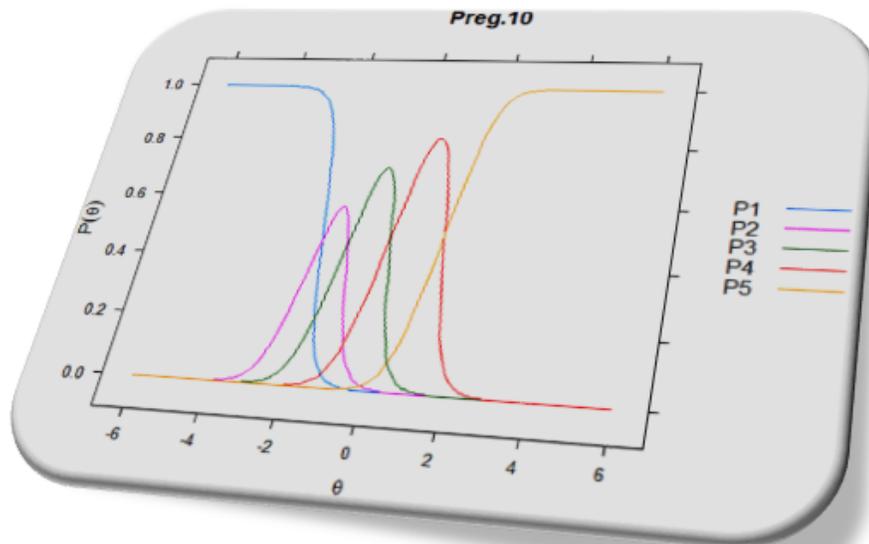
Representación tabular de parámetros de control del ajuste de los reactivos del instrumento a los datos (Elaboración propia).

De la tabla anterior puede inferirse que salvo la polémica 'Preg.3', los demás reactivos muestran un buen ajuste a los datos de las respuestas proporcionadas por los respondientes, con valores aceptables, buenos y hasta muy buenos del indicador de control 'RMSEA.S\_X2', según el caso.

Como en el caso de los modelos de uno, dos o tres parámetros de los reactivos dicotómicos –vistas arriba– también aquí podemos graficar, con R-Project, curvas del comportamiento de los reactivos, los que pasan a llamarse ahora 'Curvas de Probabilidad Acumulada' o 'Curvas Características de las Categorías' (CCC), con los modelos de respuesta graduada. Veamos solo los de algunos reactivos, por razones de extensión, de esta ya larga cuarta parte del Número 6.



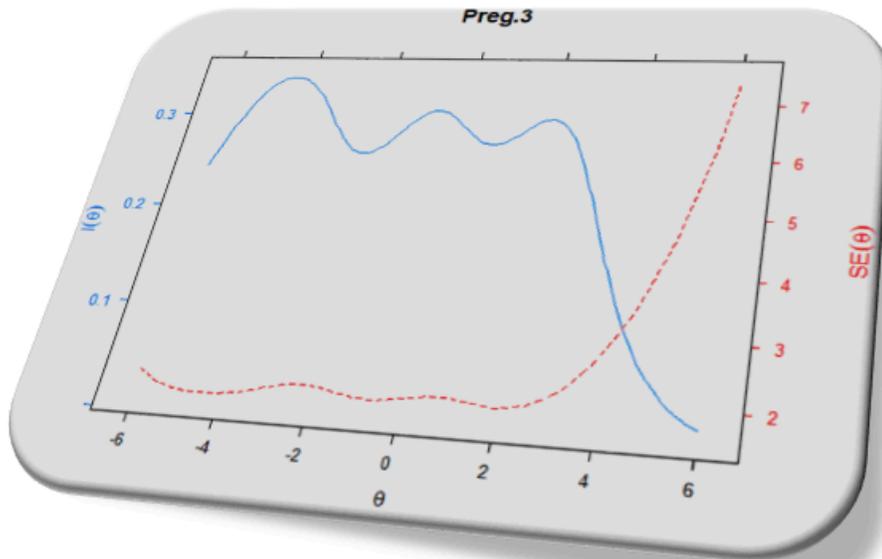
Representación gráfica de las 'Curvas Características de las Categorías' de la 'Preg.3' (Elaboración propia).



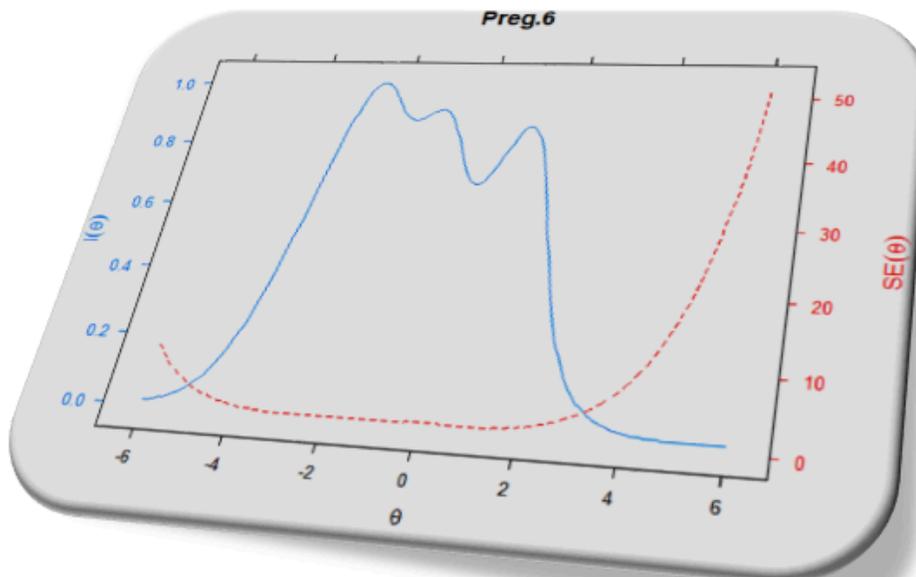
Representación gráfica de las 'Curvas Características de las Categorías' de la 'Preg.10'  
(Elaboración propia).

Véase la diferencia en el comportamiento de un reactivo tan robusto, como la 'Preg.10', en contraste con otro de diferente comportamiento, como el de la 'Preg.3'.

Además, se puede graficar fácilmente sus 'Curvas de Información del Reactivo'; algo que ilustraremos también contrastando ítems de comportamientos diferentes.



Representación gráfica de la 'Curva de Información del Reactivo' de la 'Preg.3'  
(Elaboración propia).



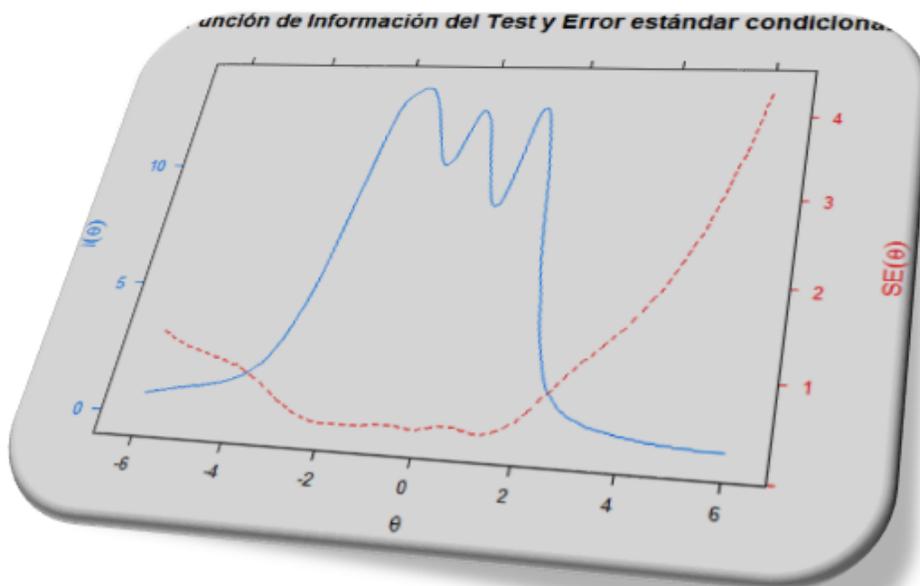
Representación gráfica de la 'Curva de Información del Reactivo' de la 'Preg.6'  
(Elaboración propia).

Obsérvese como el reactivo 'Preg.3' es relativamente informativo (y, por tanto, fácil) a lo largo de casi todo el rango de la escala de la habilidad latente ( $\Theta$  o Theta), a diferencia de lo que sucede con el reactivo 'Preg.6' (que lo es solo entre -2 y 2, o lo que es lo mismo, en un entorno cercano al comportamiento medio).

Este tipo de gráfico proporciona también información del 'error estándar' (de la habilidad  $\Theta$  ante cambios en la muestra de los respondientes; representado con la línea discontinua de color rojo); algo trascendente en términos de 'confiabilidad' (o de consistencia ante la administración repetida del reactivo). Ese error muestral es mayor en los niveles de habilidad latente donde el reactivo es menos informativo (o sea, donde hay mayor incertidumbre sobre el comportamiento de sus propiedades psicométricas).

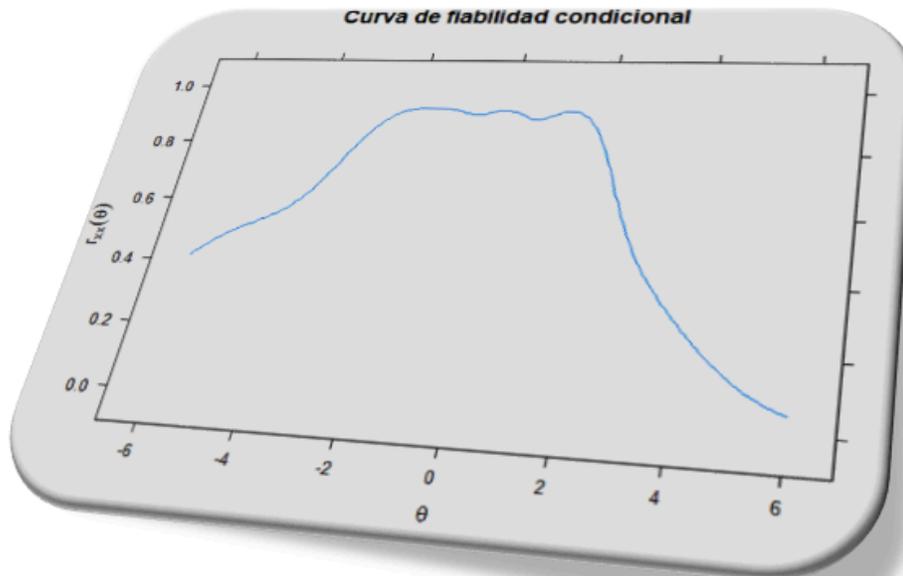
Recuérdese, la habilidad latente a la que se hace referencia está asociada aquí a la destreza o potencialidad del respondiente de tener una percepción más profunda (o sólida) acerca de lo que se indaga a través del reactivo (en general, del instrumento).

Ahora bien, en última instancia, lo que más nos interesa es el comportamiento global del 'instrumento' y R-Project nos puede proporcionar también un gráfico del nivel de información del test todo, para cada valor de la habilidad latente. Veámoslo.



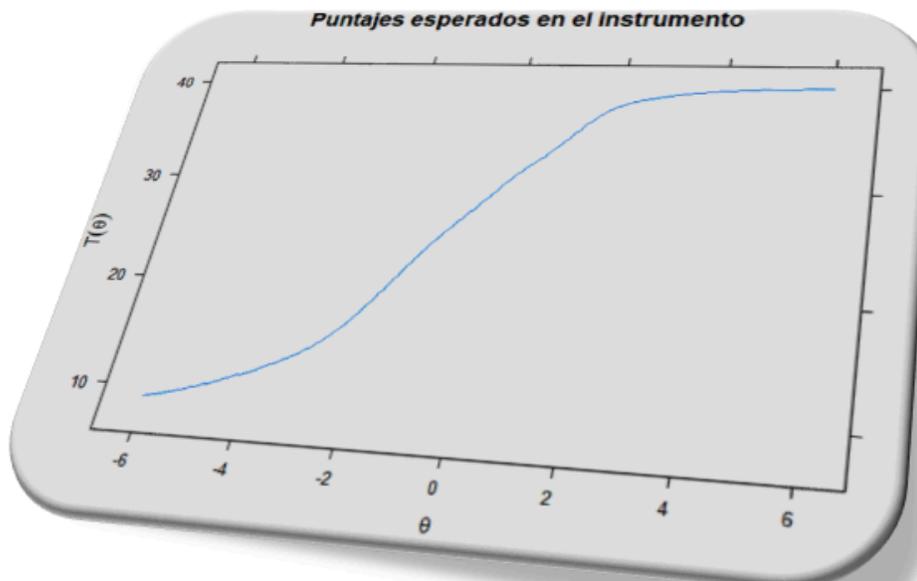
Representación gráfica de la 'Curva de Información del Test' (Elaboración propia).

Posiblemente, más importante –para los fines de esta parte del Número– resulta la 'curva de fiabilidad condicional' del test, que se muestra a continuación; ella nos indica que los resultados de la consistencia del instrumento son mucho mayores en respondientes con un desarrollo medio de la habilidad latente (o sea, entre -2 y 2).



Representación gráfica de la 'Curva de fiabilidad condicional' (Elaboración propia).

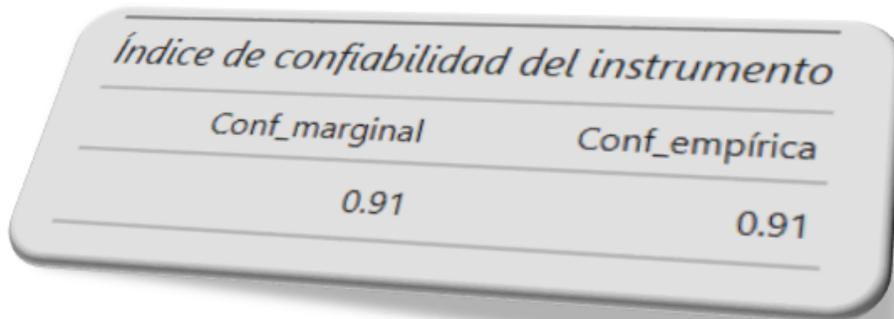
Al mismo tiempo, según la 'puntuación total' alcanzada por cada respondiente en el instrumento (un valor que oscila entre 0 y 40 puntos; esta última, si el encuestado responde siempre con la categoría máxima de 5 ['Muy alta']), se tendría -muy probablemente- una elevada 'confiabilidad condicional' del instrumento, que se alcanzará en 'puntuaciones totales' de entre 15 y 37 puntos, aproximadamente; esto, según la gráfica de 'puntajes esperados en el instrumento', siguiente.



Representación gráfica de los 'Puntajes esperados en el instrumento' (Elaboración propia).

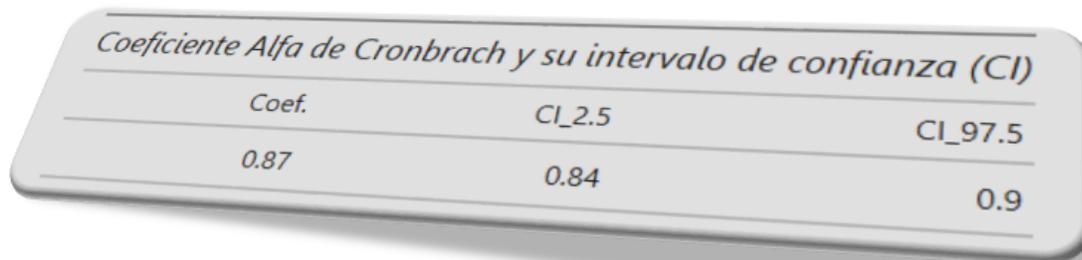
Finalmente, las 'librerías' especializadas de R-Project nos pueden proporcionar rápida y muy puntualmente los dos indicadores estadísticos que mejor nos hablan

de la confiabilidad de un instrumento de investigación: el 'índice de confiabilidad del instrumento' –propriadmente dicho– y el 'coeficiente de Alfa de Cronbach', ambos inferidos con un 95% o más de confianza.



Índice de confiabilidad del instrumento	
Conf_marginal	Conf_empírica
0.91	0.91

Representación tabular de los índices de confiabilidad marginal y empírica del instrumento (Elaboración propia).



Coeficiente Alfa de Cronbrach y su intervalo de confianza (CI)		
Coef.	CI_2.5	CI_97.5
0.87	0.84	0.9

Representación tabular del índice Alfa de Cronbach y su intervalo de confianza con solo 2.5% de error (Elaboración propia).

Como puede apreciarse, ambos con valores muy altos (pues corresponden a escalas de 0 a 1); ello, a pesar de las deficiencias en torno al ajuste del modelo asumido ('de Respuesta Graduada') a las respuestas recibidas, y que no debieran desatenderse.

Bien, ha sido muy extensa esta cuarta parte del Número 6, pero creo que lo ameritaba. Al resumirlo, deberíamos destacar las siguientes ideas básicas:

- Las ciencias naturales y tecnológicas han dispuesto –históricamente– de mejores condiciones para calibrar y aplicar el 'principio de la validez y la confiabilidad' de sus instrumentos de investigación; pero desatenderlos por parte de nosotros, los investigadores de las ciencias sociales y humanísticas –a causa de esa desventaja– sería exponerse aún más a la influencia nociva de la subjetividad en el acto científico; a caer en la engañosa área de influencia de las pseudociencias.

- Asumir el riesgo de la desatención al tema es aún más injustificado en la actualidad, cuando la **Psicometría**, de la mano de la **Estadística Inferencial Multivariada** (en particular, con los **Modelos de regresión binomial** y el **Análisis Factorial Confirmatorio**), ha hecho tanto por dotarnos de **recursos potentes** para enfrentar ese reto.
  
- Como si fuera poca esa generosa disponibilidad, los **desarrolladores informáticos contemporáneos** han simplificado notablemente esos engorrosos procesos estadísticos (e incluso su estudio teórico subyacente); los cubanos tenemos a **R-Project**, un **software libre y de código abierto** que lidia muy bien con esos **modelos sofisticados** y con las **salidas tabulares y gráficas** requeridas para la redacción sustentada de nuestros **informes de investigación**.
  
- El **‘mundo interior’** y las **relaciones sociales** que conforman nuestras **unidades de análisis** –en las investigaciones que practicamos– son marcadamente **multifactoriales, dinámicos y hasta perturbados**, como para pretender seguirlos estudiando **‘científicamente’** con instrumentos imbuidos de empírea y de procesamientos estadísticos apoyados únicamente en palotes y **por cientos**, e incluso en limitadas **pruebas de hipótesis**, ansiosas de encontrarse con alguna evidencia de **significación estadística** (que no por ello resultará, necesariamente, una **significación factual**); dejemos atrás la **‘edad antigua’** y el **medievo** de las **investigaciones sociales y humanísticas**, y sepamos ponernos a la **‘altura de nuestros tiempo’**.

Bueno, todo queda listo para nuestro encuentro con la quinta y última **parte** (esta vez, sí) de este **Número 6** de la **Serie de Post**, dedicado nuevamente al campo de la **investigación científica** en nuestras universidades e institutos. En aquella daremos respuesta a la **pregunta principal**, de cuál es la **magnitud del impacto** de la **preparación profesional** de los educadores sobre la **calidad del aprendizaje** y la **formación** de sus educandos, sobre la base del nivel de desarrollo de nuestras **habilidades latentes** en torno a esos dos temas vitales de la vida universitaria.

Al cerrar este **Número** tendremos la posibilidad de rotar nuevamente por el **área de la docencia**; para la que les propongo analizar el imprescindible **tratamiento de las funciones didácticas** del **‘trabajo para la nueva materia’**; con énfasis en la **creación de una motivación real** entre nuestros educandos al momento de apropiarse de los **nuevos contenidos**; siempre teniendo presente esa alerta lapidaria de **Roger Schank**, de que: **‘el aprendizaje ocurre cuando alguien quiere aprender, no cuando alguien quiere enseñar’**.

¡Los espero, entonces, en la **parte 5°** y final de este **Número 6!**

/continuará



**Paul Antonio Torres Fernández**

**Licenciado en Educación (1980). Doctor en Ciencias Pedagógicas (1993). Profesor Titular (2001). Investigador Titular (2013). Doctor en Ciencias (2016). Profesor Invitado de la Universidad de Matanzas (2023). [orcid.org/0000-0002-7862-2737](https://orcid.org/0000-0002-7862-2737)**

**4 de noviembre de 2024**

## Referencias Bibliográficas

1. Hernández-Sampieri, R., Fernández, C. y Baptista, M. P. (2010). *Metodología de la investigación (Quinta edición)*. McGraw-Hill.  
[https://drive.google.com/file/d/1OzAyRwb\\_hGWHFOuhs6iWpFv8bstIXLfs/view](https://drive.google.com/file/d/1OzAyRwb_hGWHFOuhs6iWpFv8bstIXLfs/view)
2. Kerlinger, F. N. (1985). *Investigación del comportamiento (2º edición)*. Nueva Editorial Interamericana. <https://idoc.pub/documents/kerlinger-1985-investigaciondelcomportamientopdf-d47e1r238yn2>
3. Torres, P. (2016a). Acerca de los enfoques cuantitativo y cualitativo en la investigación educativa cubana actual. *Atenas, Vol. 2, No.34, pp. 1-15*.  
<https://www.redalyc.org/jatsRepo/4780/478054643001/478054643001.pdf>
5. \_\_\_\_\_ (2016b). *Retos de la investigación educativa actual. Aportes a su tratamiento*. [Tesis doctoral de segundo grado]. Universidad en Ciencias Pedagógicas “Enrique José Varona”.  
<https://drive.google.com/file/d/17w13EvbTYYMR266KHI2UsFw9VDyCXwaI/view?usp=sharing>
6. \_\_\_\_\_ (2016c). Acerca de las pruebas objetivas y la enseñanza desarrolladora. *Temas de Educación / Vol.22, No.1: 115-129*.  
<https://revistas.userena.cl/index.php/teeducacion/article/view/740>
7. \_\_\_\_\_ (2018). Lo que todo investigador educativo cubano debiera conocer: el entorno informático R. *Atenas, Vol. 4, Núm. 44*.  
<https://www.redalyc.org/articulo.oa?id=478055154001>
8. Attorresi, H. F.; Abal, F. J. P.; Galibert, M. S.; Lozzia, G. S.; Aguerri, M. E. (2011). Aplicación del modelo de respuesta graduada a una escala de voluntad de trabajo. *Interdisciplinaria, Vol. 28, Núm. 2, pp. 231-244*. <http://www.redalyc.org/articulo.oa?id=18022339005>