

El formato de la pregunta y los aspectos conativos en la evaluación del aprendizaje

Elsy J. Urdaneta Durán

elsy.urdaneta.d@gmail.com

Universidad de Los Andes, Núcleo Rafael Rangel. Trujillo, Venezuela

Recibido: 15/10/2014 **Aceptado:** 15/04/2014

Resumen

En este trabajo se investigó la posible interacción entre el formato del ítem y los aspectos conativos en los resultados obtenidos por los examinados en pruebas de conocimiento. Para el estudio se utilizaron tres preguntas formuladas en formato de selección, de respuesta corta y de respuesta abierta respectivamente y un cuestionario de preguntas no cognitivas orientado a recoger información sobre aspectos volitivos y conativos del examinado. A través de la elaboración de modelos de regresión logística se logró concluir que los aspectos conativos (motivación y volición) son factores que deben considerarse en la evaluación del aprendizaje mediante solución de problemas y que el comportamiento del examinado en relación a estas variables es diferencial en función del formato en que deba responder a la pregunta.

Palabras clave: formato del ítem, volición y motivación.

Item format and conatives aspects in assessment of learning

Abstract

In this work, we investigated the possible interaction between item format and conative aspects of the results obtained by the examinees in knowledge tests. For the study we used three questions in multiple choice format, short-answer and open response, respectively, and a series of questions designed to collect noncognitive information about volitional and conative aspects of the examinee. Through the development of logistic regression models were concluded that conative aspects (motivation and volition) are factors to be considered in the assessment of learning through problem solving and behavior examined in relation to these variables is differential depending on the format in which to respond to the question.

Key words: Item format, volition & motivation

La evaluación de los resultados obtenidos por los estudiantes representa una fase de suma importancia en el proceso de enseñanza-aprendizaje, de manera que ésta debe hacerse con los instrumentos más confiables y apropiados. Esta es la tarea a enfrentar para elaborar un buen examen y, considerando que los ítems son la unidad básica del mismo, la construcción de un buen cuestionario pasa por escribir buenos ítems o preguntas. Un buen ítem será aquel que nos permita obtener una puntuación que dé cuenta del constructo o variable que estamos interesados en medir. Para que esto sea posible, el conjunto de ítems que conforman el test deben abarcar todos los aspectos relevantes a ese constructo.

En este sentido, el formato de la pregunta que se utiliza en los exámenes es un elemento de interés relevante, pues no está claro que todos los formatos permitan evaluar cualquier constructo o, cuando menos, que lo hagan con la misma eficiencia; además, pueden

poner en juego procesos cognitivos diferentes al requerir del sujeto conductas distintas para responder y, de la misma manera, pueden no resultar igual de motivadores para todos los sujetos. Es decir, se debe tener muy en cuenta el formato de ítem a utilizar, pues como bien afirma Rodríguez (2002), esta elección lleva implícitos aspectos relacionados con el tema de la validez.

Esta es la razón por la cual la selección del formato de la pregunta ha motivado diversas investigaciones que intentan averiguar si miden lo mismo aquellas preguntas en las que el examinado debe seleccionar una respuesta entre varias (elección múltiple) o aquellas en las que el estudiante debe elaborar su respuesta (respuesta construida), ya sea ésta una simple palabra o frase, o una respuesta larga (e. g. Ackerman y Smith, 1988; Ayala, Shavelson, Yin y Shultz, 2002; Gadalla, 1999; Jodoin, 2003; Rodríguez; 2002, 2003; Urdaneta, 2013). La revisión de estos trabajos conduce a pensar que, en general, los ítems de elección múltiple no miden lo mismo que los ítems de respuesta construida.

En esta línea, los resultados obtenidos por Urdaneta (2012) ponen de manifiesto que las puntuaciones obtenidas con los distintos formatos estudiados tienen propiedades psicométricas diferente. En esta investigación se estableció la falta de equivalencia psicométrica entre los distintos formatos considerados, pero no se abordó el porqué de esta falta de equivalencia. Snow (1993) certeramente señala que la equivalencia del rasgo establecida a través de criterios psicométricos no tiene por qué ser la misma que la establecida a través de criterios psicológicos. En tal sentido, es necesario continuar investigando en esta dirección para tratar de determinar con claridad el potencial de cada formato para medir un determinado constructo y examinar no solo la equivalencia psicométrica sino también la equivalencia psicológica de las medidas obtenidas al administrar ítems de distinto formato.

Para poder interpretar adecuadamente las puntuaciones de una prueba en la que el estudiante debe afrontar la resolución de problemas, no solo es preciso recoger evidencias de validez procedentes de distintas fuentes sino que es también importante contar con una teoría psicológica acerca del constructo que se desea medir. Sostienen Borsboom y Mellenberg (2007) que el planteamiento de un modelo de medida requiere enunciar una teoría sustantiva que especifique la relación causal entre el atributo y las puntuaciones del test.

Una teoría acerca del constructo que se pretende medir brinda la posibilidad de identificar y comprender los mecanismos que operan en la mente humana cuando el

examinado ejecuta una determinada tarea. La identificación pormenorizada de los elementos o factores subyacentes a la solución de problemas provee un marco para la representación detallada del constructo que ayudará a identificar las fuentes de dificultad y el entramado de habilidades y destrezas concretas que debe poseer el sujeto para llevar a cabo la tarea, de modo tal que las puntuaciones obtenidas sean interpretables.

La propuesta de Kupermintz (2002) supone ampliar el concepto de aptitud reconociendo la naturaleza dinámica y compleja de las interacciones persona-situación e incluir procesos afectivos y conativos en la explicación de las diferencias individuales en aprendizaje y rendimiento. Esta propuesta es deudora de los planteamientos de Snow, quien ha recalcado la importancia de considerar lo conativo y afectivo como aspectos de la mente humana que afectan tanto el aprendizaje como la ejecución de las tareas. En el modelo de aptitud propuesto por Snow y cols. (Corno, Cronbach, Kupermintz, Lohman, Mandinach, Porteus y Talbert, 2002; Snow, 1992; Snow, Corno y Jackson, 1996) se revisa y actualiza el concepto de aptitud como un repertorio multivariado de potencialidades para aprender y ejecutar tareas, que incluyen no solamente las habilidades cognitivas sino también características afectivas y conativas.

Snow (1989) ha señalado, además, que el patrón de influencia de las variables afectivas y conativas en el rendimiento académico y la solución de problemas varía en función de características situacionales, tal como podría ser el formato de respuesta en un escenario de evaluación. En el mismo orden de ideas, Pintrich y Schunk (2002) afirman que el tipo de tareas es una importante influencia en la motivación y en la cognición, de modo que es lógico considerar que la tarea en sí misma representa un factor que puede influir a la hora de responder a un ítem.

En atención a lo anterior, se puede afirmar que la elaboración de las pruebas para evaluar rendimiento académico mediante la resolución de problemas debe estar soportada por la explicación de los procesos y estructuras del conocimiento que se desean evaluar, pero también debe considerar los aspectos de tipo volitivo y motivacional e incluso los aspectos contextuales relacionados con la forma de administración, formato y organización del test (Snow, 1993; Nickerson, 1994; Schneider y McGrew, 2013).

Así pues, se plantea como objetivo de esta investigación probar un modelo que nos informe en qué medida pueden considerarse las variables provenientes del ámbito conativo determinantes de las puntuaciones y como varían estas en función del formato de la pregunta.

Factores conativos

El aspecto conativo está relacionado con las conductas de toma, mantenimiento y ejecución de decisiones y cursos de acción para el logro de metas. La motivación y la volición representan el aspecto conativo del funcionamiento psicológico humano. Los procesos motivacionales median la toma de una decisión y la promueven y los volitivos median la puesta en marcha de esas decisiones.

Motivación

El dominio motivación está relacionado con la toma de decisiones y la selección de metas individuales. Puede considerarse el estado predecisional (Heckhausen, 1991). La motivación se hace presente en una gran cantidad de modelos sobre rendimiento, habilidad, desarrollo de la experiencia y solución de problemas (véase por ejemplo Kupermintz, 2002; Lau y Roeser, 2002; McGrew, 2013; Nickerson, 1994; Snow, 1989; Sternberg, 1998 y Sugrue, 1995). La motivación se refiere a deseos, apetencias, necesidades y metas, de modo que se pueden situar en esta categoría variables como orientación hacia el logro, auto-concepto, auto-eficacia, valor de la tarea, expectativas e intereses.

El trabajo de Pintrich y Schunk (2002) sobre la motivación como factor determinante del aprendizaje y de la ejecución resalta la importancia de algunas teorías desde las cuales se ha aproximado el estudio de este factor, en particular el modelo expectativa-valor de Eccles y Wigfield (1995) y la teoría cognitivo social de Bandura (1997).

El modelo expectativa-valor (Wigfield, 1994; Eccles y Wigfield, 1995, 2002) se centra en el rol de las expectativas de los estudiantes sobre su éxito académico y el valor percibido de la tarea como los dos más importantes predictores de conductas de logro. El constructo expectativa se refiere a la respuesta del examinado a la pregunta ¿soy capaz de hacer esta tarea?, que involucra juicios sobre su competencia y acerca de la dificultad percibida de la tarea. El constructo valor percibido de la tarea se refiere a la respuesta del estudiante a la pregunta ¿por qué debería hacer esta tarea?, la cual puede incluir interés en el tópico, creencias sobre su utilidad e importancia y beneficios que pueda reportarle la ejecución de la misma.

En la teoría cognitiva social de Bandura (1993, 1997) la auto-eficacia y las expectativas sobre los resultados son centrales como determinantes de la conducta. Define la auto-eficacia como el juicio del individuo sobre sus propias capacidades para organizar y ejecutar cursos de acción requeridos para el logro de metas fijadas, lo que permite asociarla con el conocimiento autopercebido señalado dentro de la categoría metacognitiva. La auto-eficacia no se refiere a las destrezas que el sujeto tiene sino a lo que cree que puede hacer con lo que tiene en una diversidad de circunstancias, operando como un factor clave en los procesos motivacionales.

Volición

El dominio volición está relacionado con los constructos involucrados en la consecución de metas y con las acciones seguidas para llevar a cabo, de la mejor manera, planes e intenciones. El estado postdecisional será lo que se considere como volición puesto que se da una vez que la acción comienza (Heckhausen, 1991; VandenBos, 2007). La volición se refiere a intenciones, esfuerzo, acciones y auto-regulación, de modo que se pueden situar aquí las variables control de acciones y aprendizaje auto-regulado (Kuhl y Beckmann, 1985; Kupermintz, 2002; Schunk y Zimmerman, 1994).

El aprendizaje auto-regulado se observa cuando los sujetos son capaces de aplicar deliberadamente estrategias que promueven un aprendizaje más profundo (Kuhl, 1984). Pero lo que más interesa en esta investigación es el hecho de si el sujeto se siente más eficiente para realizar la tarea más recursos utilizará para alcanzar la meta. Esta es la razón por la que este autor considera este aspecto –el conocimiento autopercebido– dentro de lo volitivo al promover acciones que contribuyen al logro de la meta.

Variables del ámbito motivacional y volitivo a considerar en el modelo.

Atendiendo a lo anterior se considerarán las variables relacionadas con la percepción del sujeto sobre la tarea (*valor percibido de la tarea y dificultad percibida de la tarea*) como los indicadores del ámbito motivacional. La variable *conocimiento autopercebido* representa en este estudio el indicador del ámbito volitivo.

El *valor percibido de la tarea* se expresa en función de:

- su interés o valor intrínseco, definido como el disfrute que experimenta un sujeto cuando realiza una tarea o como el interés subjetivo en su contenido;
- la importancia de la tarea, definida como la trascendencia que se le da al hecho de realizar correctamente la tarea; y

- su valor utilitario o extrínseco, que se define como la utilidad de la tarea para el individuo en términos de sus metas futuras.

La *dificultad percibida de la tarea* está relacionada con los juicios del examinado sobre el grado de dificultad y el nivel de esfuerzo que pueda demandar la tarea que debe ejecutar.

La variable *conocimiento autopercibido* o percepción sobre sí mismo revela los juicios que el examinado hace sobre su competencia en la ejecución de la tarea y que representan lo que él cree que es capaz de hacer, más no necesariamente sus competencias reales

Método

Para lograr el objetivo propuesto se realizó una investigación de campo con un diseño de medidas repetidas a fin de reducir la variabilidad atribuible a las diferencias individuales entre sujetos, garantizando de esta manera la precisión y potencia de la prueba.

Sujetos

Se aplicaron las pruebas a una muestra no probabilística formada por 556 participantes, estudiantes de 4° año de bachillerato de siete centros educativos ubicados en el Estado Trujillo en Venezuela. La selección de los estudiantes de 4° año obedeció al hecho de que la mayoría de los estudiantes de este curso tienen 15 años, que es la edad para la que están diseñadas las preguntas del estudio PISA utilizadas en la prueba de competencias.

La muestra tiene una distribución equitativa en relación al sexo de los examinados (50.2 % masculino y 49.8% femenino). También se logró que la mayoría de los examinados tuvieran una edad de 15 años (73.7 %) y el resto con edades muy próximas. En cuanto a la titularidad de los centros educativos se tienen porcentajes relativamente similares: centros privados (42.6 %) y centros públicos (57.4 %).

Instrumentos

Para hacer posible la recolección de la información sobre las variables que conformarían el modelo a probar se utilizaron dos instrumentos:

- Una prueba de competencias, para observar la ejecución de los estudiantes en cada tipo de formato de ítem y obtener la puntuación en la variable solución de problemas.
- Un cuestionario ad hoc para obtener información sobre las variables conativas.

Prueba de competencias

La prueba de competencias permitió obtener la respuesta del sujeto a cada uno de los ítems formulados en tres distintos tipos de formato. Se optó por incluir un solo ítem por cada

tipo de formato considerado en el estudio con el objeto de optimizar el tiempo de aplicación y de esta manera conseguir la máxima colaboración de los responsables de los centros escolares, al interrumpir lo menos posible el desarrollo de la dinámica escolar. Los formatos utilizados en esta prueba son: elección múltiple (EM), respuesta corta (RC) y respuesta abierta (RA).

Las preguntas seleccionadas fueron extraídas del grupo utilizado en la edición 2003 del estudio PISA (INECSE, 2005), dado que estas reunían en mayor medida las características antes señaladas, particularmente las del dominio o área de conocimiento denominada solución de problemas, al no estar asociadas directamente a ningún área curricular concreta y no demandar recursos especiales que pudieran encarecer los costos de aplicación, ni tampoco requerir tiempos excepcionales para realizar la tarea.

En particular, se seleccionaron la pregunta 8 del problema denominado *Campamento* y la pregunta 15 del problema denominado *Vacaciones*, ambas correspondientes al dominio de solución de problemas. Para representar el ítem de respuesta abierta se utilizó la pregunta 8 del problema denominado *Campamento*, el cual tenía varias respuestas correctas. En el problema *Vacaciones* se introdujeron unas mínimas modificaciones en la tabla de distancias, para forzar al estudiante a realizar algunos cálculos, utilizando la pregunta 15 como ítem de respuesta corta y a partir del mismo se elaboró el ítem de elección múltiple en el que se daba la respuesta correcta y tres distractores.

Para asignar puntuaciones a cada ítem se procedió tal como se hace en el estudio PISA. Para el problema *Campamento* se adjudicaban dos puntos para la respuesta correcta, que ha de satisfacer todas las condiciones planteadas en el enunciado. Se adjudicaba un punto para una respuesta parcialmente correcta en la cual se incumplen una o dos condiciones, si bien no cumplir la misma condición más de una vez se considera sólo como un incumplimiento. Se adjudicaba cero puntos para otras respuestas. Para el ítem de respuesta corta se considera *900 Km.* como la respuesta correcta y se le adjudicaba un punto, otras respuestas se consideraban incorrectas y se les adjudicaba cero puntos. Para el ítem de elección múltiple se considera la opción *b* como la respuesta correcta y se le adjudicaba un punto, el resto de opciones se consideran respuestas incorrectas y se les adjudica cero puntos.

Cuestionario

Para la elaboración de este cuestionario se partió del estudio de las teorías y modelos que sobre estas variables se han desarrollado. Se consideraron diferentes propuestas sobre

motivación, autoeficacia, metacognición (y demás constructos relacionados con las variables de interés, revisándose un amplio conjunto de instrumentos elaborados a partir de estas teorías y modelos.

Se comenzó redactando los ítems que servirían como indicadores de la percepción del sujeto sobre la tarea y sobre sí mismo puesto que para ambas variables la elaboración de los ítems estuvo basada en las categorías señaladas en la teoría de la autoeficacia de Bandura (1995, 1997, 2006) y en la teoría de expectativa-valor de Eccles y cols. (1983, 1994, 1995). Los ítems fueron redactados de modo que el examinado respondiera juzgando tanto el contenido de la tarea como el formato en el que ésta se presentaba.

Para medir el *valor percibido de la tarea* se elaboraron 8 ítems expresados en función del valor intrínseco, la importancia y el valor utilitario de la tarea. Para la medición de la variable *dificultad percibida de la tarea* se elaboraron 5 ítems y para estimar la variable *conocimiento autopercibido* o percepción sobre sí mismo se elaboraron 3 ítems. Estos ítems se agruparon en dos bloques, distribuidos en cada apartado en función de si el ítem se orienta a medir lo relativo al contenido de la tarea (1^{er} bloque) o lo relativo al formato en que ésta se presenta (2^o bloque); cada parte con 8 ítems con formato Likert con cuatro opciones de respuesta.

Plan de análisis

Para someter a comprobación empírica el modelo teórico formulado, se utilizó un modelo lineal generalizado con el logit como función de enlace. Dado que para los ítems de EM y RC solo existen dos categorías de respuesta (correcta o incorrecta), se ha trabajado con la regresión logística binaria. El ítem de EA cuenta con tres categorías de respuesta posibles (correcta, parcialmente correcta, incorrecta). Sin embargo, al violarse en el modelo finalmente obtenido el supuesto de líneas de regresión paralelas en las distintas categorías de respuesta, se optó por recodificar la respuesta (con 0 para respuestas incorrectas o parcialmente correctas y 1 para respuestas correctas) y se estimó un modelo de regresión logística binaria, siguiendo idéntico procedimiento que para los ítems de EM y RC.

Previo al modelado se hicieron algunos análisis que permitieron estimar la confiabilidad de las puntuaciones obtenidas y obtener algunas evidencias que garantizan la validez tanto interna como externa de los resultados obtenidos.

La variable respuesta está representada por las respuestas de los sujetos a los ítems de la prueba de competencia (tres preguntas cada una con un formato diferente) y las variables predictoras o independientes serían cada una de las correspondientes a la medida de los factores que se plantean en el modelo como posibles determinantes de la respuesta al ítem, a saber:

1. Dentro del ámbito motivacional las variables independientes son:
 - 1.1. Valor percibido de la tarea en relación al contenido (VTC): ítems 3, 4, 5 y 6.
 - 1.2. Valor percibido de la tarea en relación al formato (VTF): ítems 10, 11, 12 y 13.
 - 1.3. Dificultad percibida de la tarea en relación al contenido (DTC): ítems 7 y 8.
 - 1.4. Dificultad percibida de la tarea en relación al formato (DTF): ítems 14, 15 y 16.
2. Dentro del ámbito volitivo la variable independiente es el conocimiento auto-percibido o percepción sobre sí mismo (CAP): ítems 1, 2 y 9.

Para construir cada variable se optó por sumar las respuestas a los ítems correspondientes. Esto nos conduce a un modelo a probar con cinco variables independientes.

Para estimar el modelo se trabajó siguiendo el enfoque descrito por Ato y cols. (2005) que consta de tres fases: especificación, selección y evaluación.

En la fase de especificación se formuló el componente sistemático del modelo, definido por el conjunto de variables predictoras del modelo ya señaladas. En la fase de selección se procedió a la estimación del modelo trabajando paso a paso, utilizando siempre y de forma sistemática dos métodos distintos de estimación, a fin de verificar que se obtenían los mismos resultados: el método de Wald hacia delante y el método de razón de verosimilitudes hacia atrás.

Se valoraron los distintos modelos obtenidos examinando:

1. la contribución de cada variable al modelo mediante el estadístico de Wald,
2. el ajuste global del modelo mediante los estadísticos de Hosmer y Lemeshow (H-L) y el coeficiente pseudo R^2 de Nagelkerke,
3. la eficacia predictiva del modelo mediante el porcentaje de casos correctamente pronosticados (% CP).

Adicionalmente, se consideraron otros indicadores importantes que podrían revelar problemas de estimación, tales como los valores del error típico de estimación de los coeficientes de regresión.

Para la fase de evaluación se examinó en el modelo final el cumplimiento del supuesto de linealidad entre el logit y cada predictor mediante un análisis gráfico, el supuesto de ausencia de multicolinealidad examinando la matriz de correlaciones y, para el caso del formato EA, el supuesto de igualdad de las pendientes de las funciones de regresión para todas las categorías de la variable, llevando a cabo el test de las líneas paralelas (que ya se mencionó no se cumplió, razón por la cual el análisis se hizo transformando la variable de ordinal a dicotómica).

Seguidamente, se procedió a realizar una validación cruzada del modelo finalmente obtenido, dividiendo aleatoriamente la muestra en dos y utilizando una mitad como muestra de estimación y la otra como muestra de validación.

Resultados

Previo al modelado de los datos se procedió a realizar un análisis que permitiera dar evidencias acerca de la confiabilidad y la validez, tanto para la prueba de competencias, como para el cuestionario sobre los aspectos conativos.

Para la prueba de competencias el valor obtenido para el coeficiente alfa de Cronbach es de 0.519, que aunque resulta moderado no es alarmante, habida cuenta de que los ítems provienen de un estudio altamente calificado a nivel internacional. El análisis factorial exploratorio revela la presencia de un único factor, en el que cargan los tres ítems con pesos factoriales altos y bastante similares: 0.751 para EM, 0.792 para RC y .691 para EA. En resumen, las evidencias recogidas tanto acerca de la precisión como de la validez de las puntuaciones obtenidas con los ítems de esta prueba apoyan su utilización para dar respuesta a la pregunta de investigación planteada en este estudio.

En relación al cuestionario para obtener medidas sobre los aspectos conativos, se obtuvieron los siguientes valores de los coeficientes alfa: para el valor percibido de la tarea en relación al contenido un valor de 0.716, para la misma variable en relación al formato un valor de 0.660, para la dificultad percibida en relación al contenido un valor de 0.611 y relación al formato un valor 0.613. Las puntuaciones de la variable conocimiento autopercibido dieron valores de coeficiente de confiabilidad de 0.772. Cabe destacar que a pesar del pequeño número de ítems para cada dimensión los valores del coeficiente de confiabilidad obtenidos permiten afirmar que tenemos unas puntuaciones con una precisión de medida bastante aceptable.

En relación a la validez, los valores obtenidos en los estadísticos indicadores del ajuste para cada uno de los modelos planteados para los análisis factoriales confirmatorios (uno por cada formato) realizados permiten corroborar la estructura interna hipotetizada para cada variable medida. Los valores del error medio cuadrático de aproximación (RMSEA) están en el rango de 0 a 0.008 aceptado en la valoración de los modelos confirmatorios. De la misma forma el índice de la bondad del ajuste (GFI) tiene valores por encima de 0.999, indicando que las variables e indicadores en que se basa la construcción del instrumento se ajustan a los modelos teóricos especificados.

En resumen, los resultados obtenidos para el coeficiente alfa y para los análisis factoriales confirmatorios representan evidencias favorables acerca de la confiabilidad de las puntuaciones obtenidas con la prueba general y de la validez de las inferencias a que pueda dar lugar.

Modelos de regresión por formato

Todas las variables incluidas en los respectivos modelos tienen estadísticos de Wald significativos, lo que se interpreta como que cada una de ellas es un buen predictor en la ecuación modelada. Además para ninguno de los modelos estimados se presentaron errores típicos de los coeficientes estimados con valores inflados como para causar preocupación en cuanto a problemas en la estimación del modelo. También se verificaron los supuestos de linealidad y de ausencia de multicolinealidad, verificándose el cumplimiento de los mismos. Al dividir la muestra en dos partes para el proceso de validación cruzada se obtuvo aproximadamente la misma tasa de predicciones correctas. Los modelos finalmente seleccionados después de su evaluación se presentan a continuación.

Modelo EM

En el modelo de EM se observa que de las cinco variables introducidas en el modelo solamente cuatro de ellas muestran influencia en la respuesta, por lo cual el modelo será:

$$Y_{EM} = 0.582CAP + 0.282VTC - 0.320DTC - 0.400DTF - 2.703$$

El índice de Hosmer y Lemeshaw (6.196 con significación de 0.625) y el R^2 de Nagelkerke (0.268) dan valores que indican que el modelo trabaja adecuadamente bien. El porcentaje de casos correctamente pronosticados (71%) revela una capacidad predictiva aceptable del modelo finalmente seleccionado, habida cuenta que estas son sólo variables del

ámbito no cognitivo y que las respuestas obviamente tendrán mucho más que ver con los conocimientos.

Modelo RC

Para la pregunta con formato RC se observa que sólo dos variables de las estudiadas muestran influencia en la respuesta al problema planteado, obteniéndose el siguiente modelo de regresión:

$$Y_{RC} = 1.385CAP + 0.508 VTF - 10.510$$

El índice de Hosmer y Lemeshaw (11.946 con significación de 0.063) y el R^2 de Nagelkerke (0.527) dan valores más que aceptables indicando que el modelo reproduce los datos adecuadamente bien. El porcentaje de casos correctamente pronosticados (80.9%) revela una capacidad predictiva superior al del modelo obtenido para EM.

Modelo RA

El modelo de regresión finalmente obtenido para la variable puntuación en el problema evidencia que sólo tres variables son incluidas como buenos predictores de la variable respuesta y nos da el siguiente resultado:

$$Y_{RA} = 1.182CAP + 0.429VTF - 0.461DTF - 7.244$$

El índice H-L tiene un valor de 5.543 con una significación de 0.698 y el R^2 de Nagelkerke es 0.519, lo cual indica un buen ajuste del modelo. Adicionalmente tenemos un porcentaje de casos correctamente pronosticados 78.6 %, evidencia importante de que las variables que se seleccionaron para el modelo ejercen una influencia reconocible en la respuesta a la pregunta.

Discusión y conclusiones

Mediante el uso del análisis de regresión logística se han elaborado tres modelos, uno para cada formato. Los modelos nos indican que, efectivamente, los aspectos conativos (motivación y volición) son factores que deben considerarse en la evaluación del rendimiento mediante solución de problemas. Muestran además que el comportamiento del examinado en relación a estas variables es diferencial en función del formato en que deba responder a la pregunta, es decir, que existe interacción.

Como es lógico suponer, la respuesta a una pregunta o la posibilidad de resolver un problema estará asociada, en primer lugar al conocimiento y a las habilidades que sobre el problema maneje el examinado, pero se observa que esta competencia puede estar afectada por

esas variables no cognitivas y que esa relación es diferente en función del formato en que deba responder el sujeto.

Se evidencia que la respuesta al problema cuando está en formato de elección múltiple es influida por todos los factores de esta naturaleza que hemos considerado, a excepción del valor que se le da a la tarea en función del formato de la pregunta. La respuesta al ítem planteada en un formato de respuesta corta es más independiente de estos factores, sólo de la dificultad que se perciba en relación a ese formato y a la capacidad que el sujeto perciba que tiene para resolver el problema. En cuanto al formato de respuesta abierta se observa que el contenido de la pregunta influye menos que el formato en que está planteada.

Considerando cada una de las variables podemos observar que la relación con el conocimiento autopercebido, que es un indicador de la voluntad del sujeto a realizar la tarea (aspecto volitivo) es muy importante y que influye en la respuesta en cualquiera de los formatos utilizados. La relación es más fuerte cuando el formato utilizado es de respuesta corta y menor cuando el formato es de elección múltiple, lo cual podría considerarse como un indicador de que en la medida de que se sienta más competente se involucrará más en la tarea, mientras que un formato de elección no le exige mayor voluntad.

En cuanto al valor que le dan a la tarea considerando el formato en que está planteada observamos que la relación solo existe cuando las preguntas tienen formato de respuesta abierta o de respuesta corta, con una mayor relación con el formato de respuesta corta. En relación al valor que le dan a la tarea en función de su contenido sólo se observa relación cuando el formato es de elección. Para el resto de los formatos esta variable no entra en el modelo de regresión, es decir, no explica la variable respuesta, lo cual es perfectamente lógico dado que es prácticamente la misma tarea.

En relación a la dificultad, los hallazgos sugieren que cuando el formato es de respuesta corta no importa que tan difícil y esforzado encuentren realizar la tarea, pues no existe relación, mientras que cuando el formato es de elección múltiple o de respuesta abierta si se presenta la relación de dependencia. En este sentido, la consideran más difícil cuando el formato es de respuesta abierta.

En definitiva, se observa que la intensidad de la relación de las variables consideradas para la construcción de los modelos varía con el formato. Los examinados se consideran más competentes cuando enfrentan tareas de respuesta abierta, la valoración que le dan a la

pregunta basándose en el formato en que deban presentar la respuesta sólo influye cuando las respuestas no son de elección múltiple y la dificultad percibida de la tarea varía con su formato encontrándola más difícil para la respuesta abierta y más fácil para la respuesta en el formato de elección.

Las implicaciones de estos hallazgos son muy relevantes pues se pone de realce que lo relativo a la motivación y a la volición son aspectos que deben considerarse no solo en una fase del proceso educativo, tal como es la fase de impartir los contenidos, sino que es importante tener en cuenta que los elementos conativos deben mantenerse activados en el estudiante en todo momento y que la fase correspondiente a la evaluación también requiere de un estudiante motivado y con voluntad de afrontar la tarea que significa responder un examen. En este sentido, el estudio además revela que el formato en que se plantea la pregunta actúa como un factor regulador de la motivación y la volición.

De modo que, como lo plantean Corno et. al. (2002), Snow (1993), Nickerson (1994) y más recientemente Schneider y McGrew (2013), la elaboración de las pruebas para evaluar rendimiento académico debe considerar los aspectos de tipo volitivo y motivacional, lo cual va a estar relacionado en gran medida con los aspectos contextuales relacionados con la forma de administración, formato y organización de las evaluaciones. Mejorar el proceso de enseñanza aprendizaje pasa por mantener en todo momento la motivación y voluntad del estudiante y un momento clave en el proceso es la evaluación del logro de los objetivos curriculares, en razón de lo cual no se debe abandonar al estudiante en este momento crucial y si el formato de la pregunta ayuda a mantenernos en esta dirección, hay que beneficiarse de ello.

Avanzando en el estudio de la contribución a la medida del rendimiento estudiantil que puede hacer cada tipo de formato de pregunta, será más sencillo determinar cuál sería el más adecuado para medir cada dominio de interés, con un determinado objetivo y para un grupo de estudiantes y contexto de aplicación particulares. Esto debe contribuir a obtener evaluaciones que tengan una confiabilidad adecuada y que aporten el mayor número de evidencias que respalden su validez y de esta manera contribuir en el estudio de la evaluación de los aprendizajes con pruebas que nos reporten evidencia objetiva y con credibilidad tal que quede fuera de toda la duda que la evaluación ha sido justa e imparcial.

Referencias

Ackerman, T. y Smith, P. (1988). A comparison of the information provided by essay, multiple-choice and free response. *Applied Psychological Measurement*, 12(2), 117 -128.

- Ato, M., Losilla, J. M., Navarro, J. B., Palmer, A. y Rodrigo, M. F. (2005). *Análisis de datos. Modelo lineal generalizado*. Girona: Edicions a Petició, SL.
- Ayala, C., Shavelson, R., Yin, Y. y Shultz, S. (2002). Reasoning dimensions underlying science achievement: the case of performance assessment. *Educational Assessment*, 8, 101-121.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychology*, 28, 117 – 148.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.
- Borsboom, D. y Mellenbergh, G. J. (2007). Test validity in cognitive assessment. En J. Leighton y M. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge: Cambridge University Press.
- Corno, L., Cronbach, L., Kupermintz, H., Lohman, D., Mandinach, E., Porteus, A et al. (2002). *Remaking the concept of aptitude: extending the legacy of Richard E. Snow*. Mahwah, N J: Lawrence Erlbaum Associates, Publishers.
- Eccles, J. y Wigfield, A. (1995). In the mind of the actor: the structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology*, 21, 215-225.
- Eccles, J. y Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology*, 53, 109-132.
- Gadalla, T. (1999, Abril). *Multiple-choice versus constructed response tests in the assessment of mathematics computation skills*. Documento presentado en la reunión anual de la American Educational Research Association, Montreal, Québec.
- Jodoin, M. (2003). Measurement efficiency of innovative item format in computer-based testing. *Journal of Educational Measurement*, 40, 1 – 15.
- Heckhausen, H. (1991). *Motivation and action*. New York, NY: Springer-Verlag.
- Instituto Nacional de Evaluación y Calidad del Sistema Educativo. (2005). *Prueba de matemáticas y de solución de problemas*. Madrid: Autor.
- Kuhl, J. (1984). Volitional aspects of achievement motivation and learned helplessness: Toward a comprehensive theory of action control. En B. Maher (Ed.), *Progress in experimental personality research* (Vol. 12, pp. 99-170). Nueva York: Academic Press.
- Kuhl, J. y Beckmann, J. (Eds.) (1985). *Action control: From cognition to behavior*. Heidelberg, Germany: Springer-Verlag.
- Kupermintz, H. (2002). Affective and conative factors as aptitude resources in high school science achievement. *Educational Assessment*, 8, 123-137.
- Lau, S. y Roeser, R. (2002). Cognitive abilities and motivational processes in high school students' situational engagement and achievement science. *Educational Assessment*, 8, 139-162.
- McGrew, K. (2013). The Motivation and Academic Competence (MACM) Commitment Pathway to Learning Model: Crossing the Rubicon to Learning Action. *The MinHub*, 1, 2-11.
- Nickerson, R. S. (1994). The teaching of thinking and problem solving. En R. J. Sternberg (Ed.), *Thinking and problem solving* (pp. 409-449). San Diego, California: Academic Press, Inc.

- Pintrich, P. R. y Schunk, D. H. (2002). *Motivation in Education .Theory, Research, and Applications. 2nd Edition.* Upper Saddle River, N J: Merrill Prentice Hall.
- Rodriguez, M. (2002). Choosing an item format. En G. Tindal y T. Haladyna (Eds.), *Large-scale assessment programs for all students: validity, technical adequacy, and implementation.* (pp. 213-231). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Rodriguez, M. (2003). Construct equivalence of multiple-choice and constructed-response items: a random effects synthesis of correlations. *Journal of Educational Measurement, 40*, 163-184.
- Schunk, D. y Zimmerman, B. (Eds.) (1994). *Self-regulation of learning and performance: Issues and educational applications.* Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schneider, W. J. & McGrew, K. S. (2013). Cognitive performance models: Individual differences in the ability to process information. En B. Irby, G. Brown, R. Laro-Alecio y S. Jackson (Eds.), *Handbook of educational theories.* (pp. 767–782). Charlotte, NC: Information Age Publishing.
- Snow, R. E. (1989). Toward assessment of cognitive and conative structures in learning. *Educational Researcher, 18*(9), 8-14.
- Snow, R. E. (1992). Aptitude theory: Yesterday, today, and tomorrow. *Educational Psychologist, 27*, 5–32.
- Snow, R. (1993). Construct validity and constructed-response tests. En R. E. Bennett y W. Ward (Eds.), *Construction versus Choice in Cognitive Measurement: issues in constructed response, performance testing and portfolio assessment* (pp. 45-60). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Snow, R. E., Corno, L. y Jackson, D., III. (1996). Individual differences in affective and conative functions. En D. C. Berliner y R. C. Calfee (Eds.), *Handbook of Educational Psychology* (243–310). New York: Macmillan.
- Sternberg, R. J. (1998). Abilities are form of developing expertise. *Educational Researcher, 27*(3), 11-20.
- Sugrue, B. (1995). A theory-based framework for assessing domain-specific problem-solving ability. *Educational Measurement: Issues and Practice, 14*(3), 29-36.
- Urdaneta, E (2012). Equivalencia psicométrica de tres formatos de pregunta. *Ágora, Año15*(30),121 – 140.
- Urdaneta, E. (2013). ¿Miden lo mismo distintos formatos de pregunta? Una revisión sobre el tema. *Paradigma, 34*(1), 69 -82.
- VandenBos, G. R. (Ed.). (2007). *APA dictionary of psychology.* Washington, DC: American Psychological Association.
- Wigfiel, A. (1994). Expectancy-value theory of achievement motivation: a development perspective. *Educational Psychology Review, 6*, 49 -78.

Autora:

Elsy J. Urdaneta Durán elsy.urdaneta.d@gmail.com

Ingeniera (Universidad de los Andes, ULA; Venezuela). Doctora en Metodología de las Ciencias del Comportamiento (Universidad Autónoma de Madrid).

Profesora (ULA, Núcleo Rafael Rangel).

Línea de investigación: formatos de ítem, confiabilidad y validez.