

Evidencias de validez de la Prueba de Aptitud Académica de la Universidad de Costa Rica basadas en su estructura interna

Evidence of validity of the Scholastic Aptitude Test of the University of Costa Rica based on their internal structure

Luis Rojas Torres¹
Universidad de Costa Rica, Costa Rica

Resumen. El objetivo de este estudio es hallar evidencias de validez la Prueba de Aptitud Académica (PAA) de la Universidad de Costa Rica (UCR), basadas en su estructura interna. El constructo que mide esta prueba presenta dos hipótesis para su estructura: la unidimensionalidad y la bidimensionalidad, bajo ambas conjeturas se alcanzó el objetivo planteado. Para la unidimensionalidad, un análisis factorial exploratorio (AFE) mostró que el primer factor explicó un 23% de la variancia de todos los ítems de la PAA, mientras que el segundo factor solamente logró explicar un 2.8%; luego, un análisis factorial confirmatorio (AFC) evidenció que esta hipótesis se ajusta aceptablemente a los datos. En cuanto a la bidimensionalidad, la rotación Promax del AFE manifestó la asociación de los factores a las dimensiones planteadas para el constructo; además, el AFC asociado a esta hipótesis también mostró un buen ajuste. Estos resultados brindan al equipo desarrollador de la PAA un sustento teórico para realizar análisis psicométricos de la Prueba basados en cualquiera de las dos posiciones sobre la estructura del constructo evaluado.

Palabras clave. Prueba de Aptitud Académica, unidimensionalidad, razonamiento, contexto verbal, contexto matemático, correlación tetracórica, Análisis Factorial Confirmatorio.

Abstract. The aim of this study is to find validity evidence in the Scholastic Aptitude Test (PAA; Spanish acronym) of the University of Costa Rica (UCR) based on its internal structure. The construct measuring of this test presents two hypotheses for its structure: the one-dimensional and two-dimensional. For both guesses the stated goal was reached. For the one-dimensional hypothesis an exploratory factor analysis (EFA) showed that the first factor explained a 23% of the variance of the PAA's items while the second factor was only able to explain a 2.8%. Then, a confirmatory factor analysis (CFA) showed that this hypothesis is acceptably consistent with data. In terms of the two-dimensional hypothesis, the EFA Promax rotation displayed the association of factors to the suggested dimensions for the construct; in addition, the CFA associated with this hypothesis also showed a good fit. These results provide to the PAA developers team a theoretical support to make the Test's psychometric analysis based on either position proposed on the structure of the evaluated construct.

Keywords. Scholastic Aptitude Test, one-dimensional, reasoning, verbal context, math context, tetrachoric correlation, Confirmatory Factor Analysis.

¹ Luis Rojas Torres, Instituto de Investigaciones Psicológicas, Universidad de Costa Rica. Dirección postal: 1156-2060. E-mail: luisrojasxtorres@gmail.com



Introducción

Uno de los conceptos fundamentales en el campo de la medición es el de validez el cual hace referencia al conjunto de evidencias acumuladas (teóricas y empíricas) que fundamentan las interpretaciones de un test, utilizadas para un propósito determinado (AERA, APA & NCME, 1999).

Dentro de estas evidencias se encuentran las que están basadas en la estructura interna del test, estas se desprenden de las relaciones entre los ítems y los componentes del constructo que busca evaluar. Se espera que los reactivos asociados a una dimensión presenten correlaciones altas entre sí, y que estas superen a las obtenidas entre ítems de distintos componentes. Lo anterior indica que una evidencia de estructura interna del test, se obtiene al comprobar que los ítems presentan patrones de correlaciones que representan la estructura teórica del constructo a evaluar.

Debido a lo anterior, es importante recalcar que una de las evidencias asociadas a la validez del uso de una prueba es comprobar que las dimensiones teóricas del constructo medido se encuentren presentes en el test (Elosua, 2003). De hecho, una de las principales violaciones a la validez es la subrepresentación del constructo, esto es, que al menos una de las dimensiones del constructo no está representada en el test (Messick, 1989).

Las evidencias de validez basadas en la estructura interna del test, no sólo son útiles para la validación, sino que son necesarias para los análisis psicométricos de los tests. Modelos importantes en el campo de la medición como el modelo de Rasch o la Teoría Clásica de los Tests (TCT) requieren del cumplimiento del supuesto de la unidimensionalidad para sustentar las fórmulas matemáticas asociadas a sus planteamientos teóricos de la medición (Burga, 2005). En estos modelos la unidimensionalidad de un test se concibe como la medición de un único rasgo latente o constructo en todos los ítems de una prueba (Wright & Linacre, 1995). En constructos multidimensionales (como la actitud hacia la matemática, la cual presenta tres dimensiones

básicas: cognitiva, comportamental y afectiva (Bazan & Sotero, 2000)), los análisis generalmente se realizan de forma desagregada o por medio de análisis más novedosos como la TRI-multidimensional.

En resumen, las evidencias de validez de estructura interna del test, tienen repercusiones en la validez y en el análisis psicométrico. El desconocimiento de la estructura de un constructo, además de provocar problemas de tipo operativo, trae consigo inconvenientes asociados a la validez. Un caso típico es el de la variancia irrelevante al constructo, es decir, la variación de los puntajes del test que no está asociada a variaciones de los puntajes en el constructo (Montero, 2013).

Ahora, en el presente estudio se buscan evidencias de validez de estructura interna para el caso particular del examen de admisión a la Universidad de Costa Rica (UCR), denominado Prueba de Aptitud Académica (PAA). Este test evalúa el constructo llamado “habilidades generales de razonamiento en contextos verbales y matemáticos” (Programa Permanente de la Prueba de Aptitud Académica, 2012; Rojas, 2013). La medición de este constructo en la selección de estudiantes universitarios es respaldada por algunos estudios de validez predictiva (Rojas, 2013; Jiménez & Morales; 2010) –en los que se ha observado su utilidad para predecir el rendimiento académico de los estudiantes– y por las propuestas teóricas en las que se expone que el razonamiento es una de las características más importantes para el éxito en la educación superior (Powers & Dwyer, 2003).

Como puede observarse, el constructo medido por la PAA se deriva del razonamiento, el cual es “un proceso de pensamiento que produce una conclusión a partir de los preceptos, los pensamientos o afirmaciones” (Johnson-Laird, 1990, p. 109).

El hecho de que la Prueba mide un constructo asociado al razonamiento en dos tipos de contextos implica que el constructo pueda ser visto de cualquiera de las siguientes formas: a) unidimensional, lo cual se basa en que los tipos de contexto de los ítems no determinan dimensiones distintas, o b) que presenta

dos dimensiones asociadas a las dos partes de la prueba (razonamiento en contexto verbal y razonamiento en contexto matemático, las cuales se les llamará RCV y RCM respectivamente), que están altamente correlacionadas entre sí.

En PPPAA (en prensa) se menciona que en los análisis factoriales exploratorios se sugiere que el constructo medido por la Prueba es unidimensional, pero no se descarta la segunda conjetura, ya que en las soluciones oblicuas para dos factores se observan dos componentes asociados a RCM y a RCV, con una correlación superior a .5.

Ahora el objetivo de este estudio es generar evidencias de validez interna de la PAA, tanto para el enfoque unidimensional, como para el bidimensional. Además, se desea determinar cuál de las dos estructuras se ajusta mejor al diseño actual de la Prueba.

Método

Instrumento

El instrumento a utilizar en este estudio es el formulario 1 de la PAA aplicado en el 2012, este se compone de 85 ítems, de los cuales 50 corresponden a RCV y 35 a RCM. Todos los ítems de la PAA son de selección única y presentan un nivel de medición ordinal dicotómico (0 = incorrecto y 1 = correcto).

La prueba presenta tres categorías de ítems: los de RCM, los de completar oraciones (CO) y los de comprensión de lectura (CL); las dos últimas categorías se distribuyen equitativamente en RCV.

Los ítems de RCM buscan medir la habilidad de los examinados para aplicar los conceptos elementales de la matemática en la resolución de problemas numéricos. Por su parte, los reactivos de RCV pretenden evaluar la habilidad verbal; los ítems de CL son ítems en que se pretende que el examinado interprete y analice mensajes, establezca conclusiones válidas y resuma ideas a partir de textos cortos. En CO se presenta un texto corto con un(os) espacio(s) en blanco, que debe(n) ser rellenado(s) con la(s) opción(es) que le dé(n) sentido al texto (Brizuela et al, 2014).

Dentro de la PAA se incluyen nuevos ítems experimentales, los cuales no son considerados en la calificación ni en los análisis estadísticos generales. La cantidad de estos ítems en la fórmula utilizada fue de 15 reactivos; 5 en RCM, 5 en CO y 5 en CL, lo cual implica que la cantidad de ítems para analizar en RCM se redujo a 30, en CO a 20 y en CL a 20.

Esta prueba presentó un alfa de Cronbach de .876 para la parte de RCV y de .845 para la parte de RCM. Si se asumiera que el constructo medido por la prueba es unidimensional, se obtiene un alfa de Cronbach para toda la prueba de .92.

Para esta fórmula, se obtuvo que la proporción de aciertos (dificultad en TCT) en los ítems de RCV varió entre .22 y .71. Por otro lado, todos los ítems analizados mostraron correlaciones ítem-total superiores al umbral de .20 (valor mínimo para considerar que esta medida de discriminación es aceptable), al contrastarlos tanto con el total de aciertos en RCV como con el total en PAA (Muñiz, 1998).

Con respecto a los ítems de RCM de la fórmula en cuestión, todos obtuvieron correlaciones ítem-total RCM e ítem total-PAA mayores a .20 y dificultades que oscilaron entre .14 y .50.

Las correlaciones ítem-total PAA, tanto de los reactivos de RCV como de RCM, indican que si los ítems se analizaran como pertenecientes a un solo test igualmente se obtendrían estadísticas adecuadas para los reactivos utilizados en esta fórmula. Lo anterior, aunado al alfa de Cronbach global, sugiere que si la PAA se analizara de manera general los estándares psicométricos de la prueba se mantendrían.

Finalmente, es importante aclarar que a pesar de que la estructura de la PAA es nominalmente similar a la del SAT (Scholastic Aptitude Test) del ETS (Educational Testing Service) o a la presentada por la PAA construida por el College Board, las diferencias en los ítems de estas pruebas con respecto a la PAA de la UCR son muy marcadas, por tanto, las definiciones del constructo usadas no son aplicables directamente al constructo medido en la prueba de la UCR. El ejemplo más claro de estas diferencias son los contenidos

escolares utilizados en las pruebas: para la PAA de la UCR se utiliza a lo sumo conocimientos del séptimo año educativo costarricense (Programa Permanente de la Prueba de Aptitud Académica, 2012), mientras que en la prueba del College Board se requiere el dominio de contenidos como la factorización y el cálculo de volúmenes, que se enseñan hasta los últimos años de educación secundaria en Costa Rica (College Board, 2007; MEP, 2003)

Participantes

La población a la que se le aplicó el instrumento presentó un tamaño de 8190 individuos, distribuidos de la siguiente manera: 46.1% hombres y 53.9% mujeres; 78% matriculados (o egresados) en colegios ubicados en las provincias centrales (San José, Alajuela, Cartago y Heredia) y 22% de colegios en las provincias costeras (Guanacaste, Puntarenas y Limón), un 80.2% de colegios públicos y 19.8% de colegios privados.

Estrategia de análisis

Primeramente es importante mencionar que entre los métodos estadísticos más populares para analizar el número de dimensiones presentes en un test se encuentran el Análisis Factorial Exploratorio (AFE) y el Análisis Factorial Confirmatorio (AFC); Gómez, Chacón y Moreno (2000) mencionan que ambas son las técnicas por excelencia para la validación de constructo de una prueba, esto es, para evidenciar empíricamente que el test “corresponde fielmente al constructo que trata de medir” (García & Magaz, 2009, p. 23).

El AFE como su nombre lo indica, es una técnica exploratoria que permite elaborar hipótesis sobre las posibles dimensiones de un constructo. Esta técnica se basa en buscar conjuntos de variables (las cuales se supone que miden el constructo) que presenten ciertos patrones de asociación entre sí, los cuales se denominan factores; posteriormente se les asignan nombres con base en la teoría que fundamenta la construcción de los ítems. Esta técnica es útil cuando no se cuenta con una teoría consolidada sobre el constructo de interés, como el caso clásico del uso del AFE en el desarrollo de la Teoría de los Tres Estratos de la inteligencia Carroll (1993). Por otro lado, mediante el

AFC se busca comprobar el ajuste de los datos a un modelo teórico previamente establecido. Dicha técnica permite comprobar si los ítems se agrupan según las dimensiones que pretendidamente subyacen a su elaboración (Gómez, Chacón & Moreno, 2000).

Ahora, para la consecución del objetivo de este estudio se realizará un análisis factorial exploratorio a partir de la matriz de correlaciones tetracóricas de los ítems de la PAA, con el fin de determinar cómo se agrupan los ítems de la Prueba. Estos resultados se compararán con los obtenidos en un AFE, que supone que los datos son continuos; lo cual tiene como objetivo evidenciar las consecuencias de realizar un AFE sin considerar los supuestos.

Además, se realizarán tres análisis factoriales confirmatorios, dos de ellos basados en las estructuras del constructo planteadas en la introducción, y un tercero basado en otra posible estructura del constructo, la cual se define a partir del tipo de ítems de la Prueba (RCM, CO y CL). De esta manera, los AFC que se realizarán presentan las siguientes estructuras: a) todos los ítems de la PAA agrupados en único factor (unidimensionalidad), b) los ítems de RCV cargando en un factor A y los ítems de RCM cargando en un factor B y c) los ítems de CO cargando en un factor A, los de CL, en un factor B y los de RCM, en un factor C (ver figura 1).

En este punto cabe señalar que se considerará la naturaleza dicotómica de los reactivos de los test, ya que en ocasiones no se toma en cuenta la escala de los ítems y se utiliza la matriz de correlaciones de Pearson para analizar ítems dicotómicos, con lo cual se corre el riesgo de obtener resultados erróneos acerca de la estructura del constructo (Brown, 2006; McDonald, 1999).

Finalmente, se presentará una comparación de los tres modelos de AFC mediante sus índices de ajuste, para determinar cuál de estos se adapta mejor a los datos estudiados.

Resultados

En primer lugar se realizó un análisis factorial exploratorio, cuyos resultados se presentan en la tabla

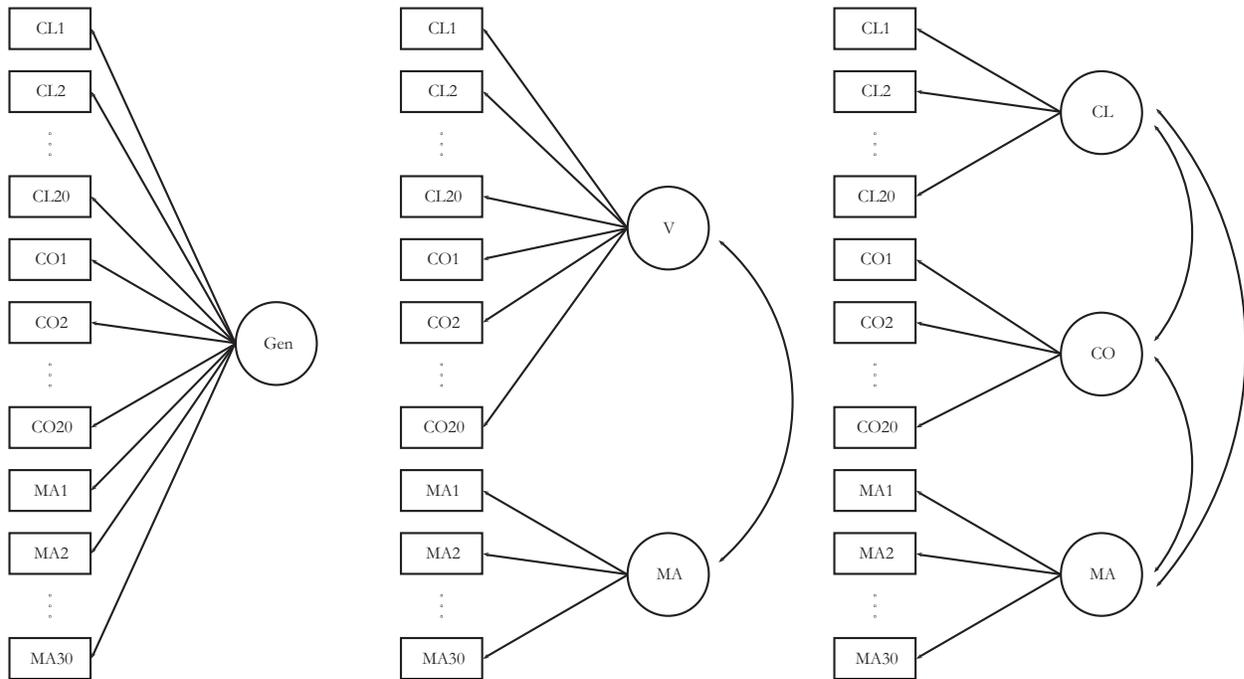


Figura 1. Modelos de AFC asociados a la estructura factorial de la PAA.

Nota. De izquierda a derecha: modelo 1 = unidimensional, modelo 2 = partes de la prueba, modelo 3 = tipos de reactivos. Gen. = Habilidad General, V. = Razonamiento en contexto verbal. MA = Razonamiento en contexto matemático, CL = Comprensión lectora, CO = Completar oraciones.

1. El algoritmo utilizado para la extracción de los factores se denomina Minimum Residual (MR) y el software utilizado para el análisis fue el paquete psych de R en su versión 2.15.3 (R Core Team, 2012).

Dada la hipótesis de que el constructo medido por la PAA es bidimensional, se decidió extraer dos factores. Además, se realizó un AFE con el algoritmo de Máxima Verosimilitud (MV), con el fin de ejemplificar las desventajas que presenta el uso de técnicas para datos continuos (que además deben cumplir una serie de supuestos) en datos dicotómicos.

Se puede observar que todos los ítems presentan cargas factoriales en el primer factor más altas que en el segundo, tanto por el método de Minimum Residual como por el de Máxima Verosimilitud. La ventaja del método Minimum Residual radica en que las cargas de los ítems en el factor general son más altas, por lo que la identificación factorial de los ítems resulta más

evidente. De hecho con el método MR todas las cargas factoriales indican que los ítems cargan en el factor general, dado que todas son mayores a .30 (Cea, 2002); mientras que con MV se encuentran nueve ítems que no cargan en ningún factor.

Por otro lado, en la solución inicial no rotada, la variancia explicada por el primer factor, con el método Minimum Residual es casi un 10% mayor que la obtenida en el factor preponderante de la solución por Máxima Verosimilitud. Luego, la diferencia de explicación de la variancia de los ítems, dada por el primer factor en comparación con la dada por el segundo factor, es de 20.2% para el método de Minimum Residual, mientras que con Máxima Verosimilitud es de apenas 12.4%.

Si se realiza una rotación Promax para tratar de identificar los factores, se obtiene que en un factor cargan la mayoría de ítems de RCV y en el otro cargan la mayoría de ítems de RCM. Según el método de

Tabla 1
Cargas factoriales de la solución inicial del AFE según método de extracción factorial

Ítem	MR1	MR2	MV1	MV2	Ítem	MR1	MR2	MV1	MV2	Ítem	MR1	MR2	MV1	MV2
CL1	.43		.32		CO5	.49		.38		MA9	.70	-.21	.56	
CL2	.53		.42		CO6	.60	.21	.47		MA10	.36		.29	
CL3	.40		.32		CO7	.62	.20	.45		MA11	.38	-.25	.30	.20
CL4	.51		.41		CO8	.40		.31		MA12	.57		.46	
CL5	.51		.41		CO9	.34		.27		MA13	.53		.42	
CL6	.47		.37		CO10	.36		.27		MA14	.39		.31	
CL7	.37		.30		CO11	.36		.29		MA15	.57	-.22	.45	
CL8	.44		.35		CO12	.50		.40		MA16	.67	-.25	.51	
CL9	.49		.39		CO13	.33		.26		MA17	.33	-.20	.25	
CL10	.41		.32		CO14	.57	.27	.46	-.21	MA18	.40		.31	
CL11	.45		.36		CO15	.59		.47		MA19	.64	-.25	.49	.22
CL12	.47	.23	.38		CO16	.57	.21	.46		MA20	.37	-.21	.26	
CL13	.37		.30		CO17	.47		.37		MA21	.34	-.22	.27	
CL14	.56		.44		CO18	.40		.32		MA22	.32		.25	
CL15	.55		.44		CO19	.50		.37		MA23	.42		.33	
CL16	.44		.36		CO20	.40		.32		MA24	.38	-.24	.30	
CL17	.53		.42		MA1	.45		.36		MA25	.50		.39	
CL18	.41		.32		MA2	.50		.40		MA26	.57	-.27	.43	.23
CL19	.60		.48		MA3	.52		.42		MA27	.39	-.24	.30	
CL20	.43		.34		MA4	.53		.42		MA28	.49	-.20	.36	
CO1	.59	.21	.46		MA5	.55		.44		MA29	.33		.25	
CO2	.58		.46		MA6	.48		.39		MA30	.34		.22	
CO3	.53		.41		MA7	.48		.38						
CO4	.31		.25		MA8	.55	-.24	.44	.20					

Nota. Los espacios en blanco indican que la carga es menor a .20 en valor absoluto. R1 y MR2 = Primer y segundo factor obtenidos con Minimum Residual. Variancia explicada: MR1 = .230, MR2= .028. MV1 y MV2 = Primer y segundo factor obtenidos con Máxima Verosimilitud. Variancia explicada: MV1 = .142, MV2 = .018.

Minimum Residual, la variancia explicada de RCV es de 13.1% y la de RCM es de 10.2%, mientras que la correlación entre ambos factores es de .73. Con Máxima Verosimilitud las varianzas explicadas de RCV y RCM son de apenas 8.4% y 6.3%, respectivamente.

El segundo conjunto de análisis elaborados consistió en la estimación de los modelos de AFC planteados en la metodología; al igual que los AFE, que se realizaron con el software R y se utilizó la librería *lavaan*. El

método de estimación utilizado fue la aproximación robusta de Mínimos Cuadrados Ponderados (DWLS). Este método es recomendado para datos dicotómicos u ordinales, ya que en muestras grandes proporciona errores típicos correctos (Recio, 2012). Para la estimación de los modelos se fijó las variancias de las variables latentes en 1.

Los resultados del modelo 1 se presentan en la tabla 2. En éste todos los ítems presentaron cargas

factoriales (coeficientes) significativos al 1%, por lo cual se puede decir que todos los ítems están significativamente relacionados con su constructo; además, todos los coeficientes fueron mayores a .30, el valor usado típicamente como referencia para indicar que una variable carga en un factor (Cea, 2002).

Por otro lado, los coeficientes varían entre .32 y .70, por lo cual el factor general explica entre el 10.24% y el 49% de la variancia de cada ítem, analizándolos de manera individual. Con respecto a la variancia explicada de los ítems en conjunto, el factor general explica un 23.1% de la variancia de todos los ítems. El coeficiente de fiabilidad del constructo fue de .952, lo cual indica que este fue medido de manera confiable, ya que supera el umbral de .70 (Cea, 2002).

En la tabla 3 se presentan los resultados del modelo 2. En éste se puede observar que, al igual que en el modelo 1, todas las variables resultaron significativas al 1% y las cargas factoriales, mayores a .30, lo cual indica que las variables cargan significativamente en los factores especificados.

Los dos constructos establecidos en el segundo modelo presentan altos índices de confiabilidad, superiores a .90 (fiabilidad de constructo de RCV = .93 y RCM = .91); además explican aproximadamente un 25% de la variancia global de los ítems que los definen (varianza extractada de RCV = .249 y RCM = .26). Por otro lado, los dos factores estudiados presentan una correlación bastante alta entre sí, la cual es superior a .80 ($\rho_{RCV, RCM} = .81$). Lo anterior indica que en este modelo se está realizando una distinción forzada entre

Tabla 2

Resultados del modelo de AFC 1: Solución completamente estandarizada

Ítem	Coef	Var									
CL1	.43	.81	CL19	.60	.64	CO17	.47	.78	MA15	.57	.67
CL2	.53	.72	CL20	.43	.82	CO18	.41	.84	MA16	.67	.55
CL3	.40	.84	CO1	.60	.64	CO19	.50	.75	MA17	.32	.90
CL4	.51	.74	CO2	.58	.67	CO20	.40	.84	MA18	.40	.84
CL5	.52	.74	CO3	.53	.72	MA1	.45	.79	MA19	.64	.60
CL6	.47	.78	CO4	.32	.90	MA2	.50	.75	MA20	.37	.87
CL7	.38	.86	CO5	.49	.77	MA3	.52	.73	MA21	.34	.89
CL8	.45	.80	CO6	.60	.64	MA4	.53	.72	MA22	.32	.90
CL9	.50	.75	CO7	.62	.62	MA5	.55	.69	MA23	.42	.82
CL10	.41	.83	CO8	.40	.84	MA6	.48	.77	MA24	.38	.85
CL11	.46	.79	CO9	.34	.88	MA7	.48	.77	MA25	.50	.75
CL12	.48	.77	CO10	.36	.87	MA8	.55	.7	MA26	.57	.67
CL13	.38	.86	CO11	.36	.87	MA9	.70	.51	MA27	.39	.85
CL14	.56	.69	CO12	.50	.75	MA10	.36	.87	MA28	.49	.76
CL15	.55	.70	CO13	.33	.89	MA11	.37	.86	MA29	.33	.89
CL16	.45	.80	CO14	.58	.67	MA12	.57	.67	MA30	.34	.88
CL17	.53	.72	CO15	.59	.65	MA13	.53	.72			
CL18	.41	.83	CO16	.57	.67	MA14	.39	.85			

Nota. Todos los coeficientes fueron significativos al 1%. Los coeficientes de los ítems representan la relación del factor general hacia estos. Fiabilidad del constructo = .95; variancia extractada = .23. La columna Var indica el porcentaje de variancia que no es explicada por la variable latente correspondiente.

Tabla 3
Resultados del modelo de AFC 2: Solución completamente estandarizada

Ítem	Coef	Var									
CL1	.45	.80	CL19	.62	.61	CO17	.49	.76	MA15	0.61	.63
CL2	.54	.70	CL20	.44	.80	CO18	.42	.82	MA16	0.71	.49
CL3	.42	.83	CO1	.62	.62	CO19	.52	.73	MA17	0.35	.88
CL4	.53	.72	CO2	.60	.65	CO20	.42	.83	MA18	0.43	.82
CL5	.53	.72	CO3	.55	.70	MA1	.48	.77	MA19	.68	.54
CL6	.49	.76	CO4	.33	.89	MA2	.53	.72	MA20	.39	.85
CL7	.39	.85	CO5	.50	.75	MA3	.56	.69	MA21	.37	.87
CL8	.46	.79	CO6	.62	.61	MA4	.56	.69	MA22	.35	.88
CL9	.52	.74	CO7	.64	.59	MA5	.59	.65	MA23	.45	.80
CL10	.43	.82	CO8	.42	.83	MA6	.51	.74	MA24	.41	.83
CL11	.47	.78	CO9	.36	.87	MA7	.51	.74	MA25	.53	.72
CL12	.50	.75	CO10	.37	.86	MA8	.59	.66	MA26	.61	.63
CL13	.39	.85	CO11	.38	.86	MA9	.75	.44	MA27	.42	.83
CL14	.58	.66	CO12	.52	.73	MA10	.39	.85	MA28	.52	.73
CL15	.57	.67	CO13	.34	.88	MA11	.40	.84	MA29	.35	.88
CL16	.46	.79	CO14	.60	.64	MA12	.61	.63	MA30	.37	.87
CL17	.55	.70	CO15	.61	.63	MA13	.56	.69			
CL18	.43	.82	CO16	.60	.65	MA14	.42	.83			

Nota. Todos los coeficientes fueron significativos al 1%. Los coeficientes de los ítems CL y CO representan la relación de RCV hacia estos. Los coeficientes de los ítems MA representan la relación de RCM hacia estos. Fiabilidad del constructo RCV = .93; varianza extractada RCV = .25. Fiabilidad del constructo RCM = .91; varianza extractada RCM = .27. Correlación RCV-RCM = .81.

los factores, cuando probablemente estos son uno solo (Moral & Segovia, 2013).

Los resultados del modelo 3 se presentan en la tabla 4. Al igual que en los otros dos modelos, todos los coeficientes son significativos al 1% y además son mayores a .30, lo cual evidencia que los ítems están asociados significativamente a los constructos establecidos. Por otro lado, las tres variables latentes presentan índices de confiabilidad aceptables (fiabilidad de constructo de CL = .87, CO = .87 y RCM = .91) y explican aproximadamente el 25% de la variancia global de los ítems que tienen asociados (varianza extractada de CL = .25, CO = .27 y RCM = .26). Con respecto a las correlaciones de CO y CL con RCM, se puede apreciar que estas son cercanas a .80, por lo

cual hay evidencia de que CO o CL conforman un solo factor junto con RCM ($\rho_{CO, RCM} = .87$ y $\rho_{CL, RCM} = .87$). Seguidamente, la correlación observada entre CO y CL ($\rho_{CO, CL} = .95$) evidencia que estas dos variables latentes realmente son una sola.

Por último, en la tabla 5 se presentan los índices de ajuste asociados a los modelos estimados, en la cual se observa que los índices de ajuste del modelo 2 (CFI = .98, TLI = .98, RMSEA = .01, valor p asociado a la $\chi^2 = .001$) y los del 3 (CFI = .98, TLI = .98, RMSEA = .01, valor p asociado a la $\chi^2 = .001$) son prácticamente iguales. También se puede apreciar que los índices de ajuste para el modelo 1 (CFI = .96, TLI = .96, RMSEA = .02, valor p asociado a la $\chi^2 = .001$) son cercanos a los estimados por los otros dos modelos.

Tabla 4

Resultados del modelo de AFC 3: Solución completamente estandarizada

Ítem	Coef	Var									
CL1	.45	.79	CL19	.63	.61	CO17	.50	.76	MA15	.61	.63
CL2	.55	.70	CL20	.45	.80	CO18	.43	.82	MA16	.71	.49
CL3	.42	.83	CO1	.63	.60	CO19	.53	.72	MA17	.35	.88
CL4	.53	.71	CO2	.61	.63	CO20	.43	.82	MA18	.43	.82
CL5	.54	.71	CO3	.56	.69	MA1	.48	.77	MA19	.68	.54
CL6	.49	.76	CO4	.34	.89	MA2	.53	.72	MA20	.39	.85
CL7	.39	.85	CO5	.51	.74	MA3	.56	.69	MA21	.37	.87
CL8	.46	.79	CO6	.64	.60	MA4	.56	.69	MA22	.35	.88
CL9	.52	.73	CO7	.65	.57	MA5	.59	.65	MA23	.45	.80
CL10	.43	.81	CO8	.43	.82	MA6	.51	.74	MA24	.41	.83
CL11	.48	.77	CO9	.36	.87	MA7	.51	.74	MA25	.53	.72
CL12	.50	.75	CO10	.38	.86	MA8	.59	.66	MA26	.61	.63
CL13	.40	.84	CO11	.39	.85	MA9	.75	.44	MA27	.42	.83
CL14	.59	.66	CO12	.53	.72	MA10	.39	.85	MA28	.52	.73
CL15	.58	.67	CO13	.35	.88	MA11	.40	.84	MA29	.35	.88
CL16	.47	.78	CO14	.61	.62	MA12	.61	.63	MA30	.37	.87
CL17	.55	.70	CO15	.62	.61	MA13	.56	.69			
CL18	.43	.82	CO16	.61	.63	MA14	.42	.83			

Nota. Todos los coeficientes fueron significativos al 1%. Los coeficientes de los ítems CL representan la relación de CL hacia estos, los de CO, la relación de CO hacia estos y los de MA, la relación de RCM hacia estos. Fiabilidad del constructo CL = .87; varianza extractada CL = .25. Fiabilidad del constructo CO = .87; varianza extractada CO = .27. Fiabilidad del constructo RCM = .91; varianza extractada RCM = .26. Correlaciones: RCM-CO = .78; RCM-CL = .81; CO-CL = .95.

Además, los tres modelos presentan índices de ajuste que indican que estas estimaciones son aceptables (CFI > .90; TLI > .90 y RMSEA < .05); únicamente los valores p asociados al estadístico chi cuadrado no apoyan el ajuste de los modelos, sin embargo, para muestras grandes este estadístico generalmente es significativo (Cea, 2002), por lo cual se puede pasar por alto.

Discusión

Al analizar los resultados recabados en este estudio, se pueden obtener varias evidencias de validez basada en la estructura interna del test desde el enfoque de la unidimensionalidad del constructo medido por

Tabla 5

Índices de ajuste para los modelos de AFC

Criterio	Modelo 1	Modelo 2	Modelo 3
CFI	.96	.98	.99
TLI	.96	.98	.98
RMSEA	.02	.01	.01
p asociada a χ^2	.001	.001	.001

Nota. CFI=Comparative fit index; TLI=Tucker Lewis index; RMSEA=Root mean square error of approximation.

la PAA. Primeramente, basándose en el AFE con el método MR, se puede observar que hay un factor general que explica más del 20% de la variancia de los ítems, contra un segundo factor que no logra explicar ni el 3%, lo cual según Reckase (1979), es evidencia de la unidimensionalidad del constructo.

Además, Hattie (1985) indica que un constructo es unidimensional si el cociente entre la diferencia del primer y el segundo factor y la diferencia entre el segundo y el tercero es mayor a 3; en la PAA, asumiendo el caso extremo de que el tercer factor explica un 0% de la variancia, se obtiene un cociente de 7.21, lo cual supera el doble del valor crítico propuesto por Hattie (1985).

Por otro lado, también se hallan evidencias de validez basadas en la estructura interna del test a partir de la hipótesis bidimensional, ya que la rotación del AFE muestra que los dos factores se asocian a las dos partes de la Prueba. Además, el AFC asociado a esta hipótesis presenta un buen ajuste.

Luego, a pesar de que el AFC con tres factores presenta un buen ajuste, se concluye que esta hipótesis no es confiable, dado que la correlación entre CO y CL es superior a .95, por lo cual no tiene sentido dividir el constructo RCV.

Estos resultados son similares a los observados en PPPAA(en prensa), ya que se hallan evidencias de validez basada en la estructura interna del test, desde los enfoques unidimensional y bidimensional.

Ahora, con base en los análisis factoriales confirmatorios, se concluye que en términos prácticos el modelo unidimensional y el bidimensional presentan un ajuste idéntico, con una leve ventaja del segundo, la cual es esperable porque están agrupando ítems que comparten características comunes entre sí, e hipotéticamente, esto contribuye a mejorar la explicación de los datos.

En el modelo 2 se forzó al software a estimar un AFC donde RCV y RCM son dos constructos separados. Este análisis indicó que la variabilidad explicada de cada ítem por RCV o por RCM, según sea el factor de pertenencia, resultó muy similar a la explicada por el factor general del modelo 1. De hecho, la variabilidad

global de los ítems explicada por el factor general es semejante a la variabilidad global de los ítems que es explicada por las variables latentes RCV y RCM.

Además, en el modelo 2 se presentó una correlación entre las dos variables latentes superior a .80, lo cual sugiere que esas dos variables podrían determinar un mismo factor general. Análogamente, en la rotación Promax realizada en el AFE, se obtuvo una correlación entre los dos factores (asociados a RCV y RCM) de .73.

Las evidencias mencionadas anteriormente muestran que ambas estructuras se ajustan aceptablemente a los datos, y que la elección de alguna de ellas como predominante depende del criterio seleccionado. Si se considera que dos constructos con las correlaciones obtenidas entre RCV y RCM determinan un único factor, se escogerá la hipótesis unidimensional; en caso contrario será la asociada a la bidimensionalidad.

Para corroborar estas tendencias se recomienda repetir estos análisis en las futuras aplicaciones; en este estudio no se replican los análisis en pruebas pasadas debido a que el equipo desarrollador de la Prueba, considera que esta ha sido modificada sustancialmente en los últimos años.

Con base a estos resultados, el equipo desarrollador de la PAA tiene la posibilidad de analizar los resultados de la Prueba desde dos posiciones distintas, considerando todos los ítems como una sola escala o utilizando los ítems de RCM y RCV, como dos escalas distintas. En el modelo unidimensional se observó que la confiabilidad del constructo es superior a la alcanzada por alguna variable latente en el modelo bidimensional, lo cual indica que los análisis desde el enfoque unidimensional son aceptables (Vargas & Hernández, 2010).

Hay que mencionar que el porcentaje de varianza explicada de los constructos es relativamente pequeña, lo cual genera un reto para los investigadores de la Prueba, el cual puede abordarse en primera instancia buscando las fuentes de varianza irrelevante al constructo.

Finalmente, se considera importante llamar la atención sobre la práctica poco recomendada de asumir variables ordinales u dicotómicas como continuas. En este estudio se evidenció que el AFE estimado con MV

no cumplía la condición de Reckase (1979) para concluir que el instrumento es unidimensional, contrariamente a lo obtenido con el AFE estimado con MR, que considera la naturaleza dicotómica de los ítems.

La aplicación de un AFE en ítems dicotómicos, con algún método de estimación que suponga continuidad de las variables, puede llevar a crear “factores de dificultad” en vez de factores asociados a una variable latente; es decir, factores que agrupen ítems por dificultad y no por una variable externa asociada a la variabilidad de estos (Nunally & Bernstein, 1995; Burga, 2005).

Según Recio (2012) y Breckler (1990) en la mayoría de trabajos con variables ordinales (datos provenientes de escalas Likert, parcelas, etc.) que utilizan AFC, se realizan estimaciones por Máxima Verosimilitud a pesar de que su uso no es el más apropiado; en Jöreskog & Sörbom (1993) y Rosseel (2012), indican la metodología correcta para trabajar estas variables en los software Lisrel y R, respectivamente.

Referencias

- AERA, APA & NCM. (1999). *Standards for Educational and Psychological Testing*. Estados Unidos: American Educational Research Association.
- Bazán, J. & Sotero, H. (2000). Una aplicación al estudio de actitudes hacia la matemática en la UNALM. *Anales Científicos UNALM*, 1, 60-72.
- Burga, A. (2005). *La unidimensionalidad de un instrumento de medición: perspectiva factorial. Informe técnico*. Ministerio de Educación de Perú.
- Breckler, S. J. (1990). Applications of covariance structure modeling in Psychology: Cause for concern? *Psychological Bulletin*, 107, 260-271.
- Brizuela, A., Cerdas, D., Fallas, S., Ordóñez, K., Pérez, N., Rojas, L. & Seas, G. (2014). *Folleto de Práctica de la Prueba de Aptitud Académica*. San José: EUCR.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- Carroll, J. (1993). *Human Cognitive Abilities*. Estados Unidos. Cambridge University Press.
- Cea, M. (2002). *Análisis multivariable: teoría y práctica en la investigación social*. Madrid: Síntesis.
- College Board. (2007). *Guía de estudio para presentar la nueva Prueba de Aptitud Académica*. Puerto Rico: College Board.
- Elosua, P. (2003). Sobre la validez de los tests. *Psicothema*, 15 (2) 315-321.
- Gómez, J., Chacón, S. y Moreno, R. (2000). Validez de constructo: el uso del análisis factorial exploratorio-confirmatorio para obtener evidencias de validez. *Psicothema*, 12 (2) 442-446.
- García, M. & Magaz, A. (2009). *¿Cómo valorar los test psicométricos?* España: Grupo Albor-Cohs.
- Hattie, J. (1985). Methodology review: Assessing unidimensionality of test and items. *Applied Psychological Measurement*, 9 (2), 139-164.
- Jiménez, K. & Morales, E. (2009-2010). Validez predictiva del Promedio de Admisión de la Universidad de Costa Rica y sus componentes. *Actualidades en Psicología*, 23, 21-55.
- Johnson-Laird, P. N. (1990). Deductive Reasoning. *Annual Review Psychological*, 50, 109-135.
- Jöreskog, K. & Sörbom, D. (1993). *Lisrel 8: Structural equation modeling with the SIMPLIS command language*. USA: Scientific Software International, Inc.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: LEA.
- MEP. (2003). *Programa de Estudios de Matemática: Educación Diversificada*. San José: MEP.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assesment. *Educational Researcher*, 18 (2), 5-11.
- Montero, E. (2013). Referentes conceptuales y metodológicos sobre la noción moderna de validez de instrumentos de medición. *Actualidades en Psicología*, 27 (114), 113-128.
- Moral, J. & Segovia, M. (2013). Propiedades Psicométricas de la escala de discriminación

- temida y percibida para mujeres con VIH. *Revista Iberoamericana de Psicología y Salud*, 4 (1) 37-62.
- Muñiz, J. (1998). *Teoría Clásica de los Tests*. Madrid: Pirámide.
- Nunnally, J. & Bernstein, I. (1995). *Teoría Psicométrica*. México: McGraw Hill.
- Programa Permanente Prueba de Aptitud Académica. (2012). *La Prueba de Aptitud Académica de la Universidad de Costa Rica*. Recuperado de <http://www.paa.iip.ucr.ac.cr/userfiles/La%20Prueba%20de%20Aptitud%20Academica%20de%20la%20UCR.pdf> el 26 de junio del 2013.
- Programa Permanente Prueba de Aptitud Académica [CD-ROM]. Prueba de Aptitud Académica. En Smith-Castro (Ed.), *Catálogo de Mediciones Psicológicas*. UCR, en prensa.
- Powers, D. & Dwyer, C. (2003). *Toward specifying a construct of reasoning*. Research Memorandum. ETS: Princeton.
- R Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Recio, P. (2012). *Equivalencia e invariancia de medida entre grupos: análisis factorial confirmatorio vrs teoría de respuesta al ítem*. Tesis para optar al grado de doctor de la Universidad Complutense de Madrid.
- Reckase, M. (1979). Unifactor latent trait models applied to multifactor test: results and implications. *Journal of Educational Statistics*, 4 (3) 207-230.
- Rojas, L. (2013). Validez predictiva de los componentes del promedio de admisión a la Universidad de Costa Rica utilizando el género y el tipo de sexo como variables control. *Revista Actualidades Investigativas en Educación*, 13 (1) 24 pp.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48 (2) 1-35.
- Vargas, C. & Hernández, L. (2010). Validez y confiabilidad del cuestionario: “Prácticas de cuidado que realizan consigo misma las mujeres en el posparto”. *Avances en enfermería*, 28 (1) 96-106.
- Wright, B. & Linacre, J. (1995). MESA Research memorandum 44. *Archives of physical, medicine and rehabilitation*, 70 (12) 857-860.

Recibido: 16 de agosto 2013

Aceptado: 6 de febrero 2014