



Para citar este artículo, le recomendamos el siguiente formato:

Cruz, M. y Martínez, M. (2012). Perfeccionamiento de un instrumento para la selección de expertos en las investigaciones educativas. *Revista Electrónica de Investigación Educativa*, 14(2), 167-179. Consultado en <http://redie.uabc.mx/vol14no2/contenido-cruzmtnz2012.html>

Revista Electrónica de Investigación Educativa

Vol. 14, Núm. 2, 2012

Perfeccionamiento de un instrumento para la selección de expertos en las investigaciones educativas

Refining a Tool for the Selection of Experts in Educational Research

Miguel Cruz Ramírez (1)
mcruzr@facinf.uho.edu.cu

Mayelín Caridad Martínez Cepena (2)
cepena@ucp.ho.rimed.cu

(1) Universidad de Holguín “Óscar Lucero Moya”

Carr. vía Guardalavaca s/n
Rpto. Piedra Blanca
Holguín, Cuba

(2) Universidad de Ciencias Pedagógicas “José de la Luz y Caballero”

Av. de los Libertadores km 3½
Rpto. Pedro Díaz Coello
Holguín, Cuba

(Recibido: 17 de febrero de 2012; Aceptado para su publicación: 16 de mayo de 2012)

Resumen

Se reporta una investigación sobre el perfeccionamiento de un instrumento empírico para la selección de expertos en las investigaciones educativas, de acuerdo a su fiabilidad y consistencia interna. Para ello se aplica el método Delphi a tres rondas y los resultados se someten al análisis factorial. Se determinan variables latentes que explican la naturaleza

de las fuentes de argumentación necesarias para lograr un adecuado nivel de competencia de los expertos.

Palabras clave: Método de expertos, test estandarizados, métodos de investigación, método Delphi.

Abstract

In this paper we report a research study geared toward refining an empirical instrument for the selection of experts for educational research, according to its reliability and internal consistency. To this end we used a three-round Delphi technique and subjected the results to a factor analysis. Latent variables were determined that explain the nature of the sources of argumentation necessary for ensuring an adequate level of competence on the part of the experts.

Key words: Expertise evaluation method, standardized test, methods research, Delphi technique.

I. Introducción

Al igual que la entrevista, la encuesta y la prueba pedagógica, el criterio de expertos ocupa un lugar importante entre los métodos de investigación empírica en las Ciencias de la Educación. Por lo regular, este último se basa en la consulta a personas que tienen profundos conocimientos sobre el objeto de estudio. La síntesis, el consenso y la estabilidad del juicio colectivo pueden ofrecer una visión verosímil del futuro, combinando la imaginación y el talento individual (Konow y Pérez, 1990). Existe asentimiento en el hecho de que el juicio colectivo es superior a la suma trivial de resultados individuales, ya que la información disponible está siempre más contrastada que aquella de que dispone el participante mejor preparado (Linstone y Turoff, 1975). Cada experto podrá aportar la idea que tiene sobre el tema debatido, desde su área de conocimiento, su experiencia e inteligencia.

En la literatura científica no existe acuerdo pleno respecto a la noción de experto. Hoffman, Shadbolt, Burton y Klein (1995) arguyen que subsisten casi tantas definiciones de experticia como investigadores que la estudian. Sin embargo, algunos estudios conceptuales han identificado aspectos comunes y, en general, los autores enfocan este concepto desde la óptica de la competencia profesional, de la pericia y capacidad para prever, evaluar, ofrecer valoraciones conclusivas y hacer recomendaciones viables (Landeta, 1999). Crespo aporta una definición que unifica aspectos diversos y se ajusta al contexto educativo. Según este autor:

Se entiende por experto a un individuo, grupo de personas u organizaciones capaces de ofrecer con un máximo de competencia, valoraciones conclusivas sobre un determinado problema, hacer pronósticos reales y objetivos sobre el efecto, aplicabilidad, viabilidad y relevancia que pueda tener en la práctica la solución que se propone, y brindar recomendaciones de qué hacer para perfeccionarla (2007, p. 13).

Esta posición reconoce tanto el experto individual como colectivo y enfatiza la posibilidad de aportar valoraciones sobre un aspecto determinado, como las causas de un problema, la calidad de una fundamentación científica, la novedad de un aporte teórico-práctico, etcétera. El experto puede incluso pronosticar lo que sucedería de implementarse la solución que el investigador da al problema y que ha sido sometida a su consideración.

Respecto a la búsqueda de una clasificación de experticia o bien, en última instancia, de una tipología que facilite su estudio, tampoco puede afirmarse que exista unanimidad en los criterios. Weinstein sostiene que existen dos clases de experticia: una basada en el conocimiento (experticia epistémica) y otra basada en la actividad (experticia ejecutiva). La primera es la capacidad para proveer justificaciones sólidas para un rango de proposiciones en cierto dominio, mientras la segunda es la capacidad para ejecutar bien una habilidad de acuerdo con las reglas y virtudes de una práctica. “Tanto los expertos epistémicos como los ejecutivos legítimamente pueden estar en desacuerdo el uno con el otro, y las dos percepciones son conceptualmente y lógicamente bien definidas” (Weinstein, 1993, p. 57).

En pocos años la clasificación anterior ha encontrado objeciones más flexibles que consideran también posturas intermedias. Por ejemplo, para Ténrière-Buchot (2001) pueden distinguirse tres tipos de expertos: tácticos, conciliadores y comunicadores. Los primeros se seleccionan de acuerdo al grado de especialización y experiencia en el tema y refieren desde estándares de calidad, relevancia, impacto. En los segundos se apuesta por la presencia de equilibrio, imparcialidad y sentido común. Los terceros suelen estar implicados en el objeto de investigación y su percepción aporta criterios sobre viabilidad, contextualización, pertinencia, entre otros aspectos. De esta manera, el colectivo aportará una información sumamente útil, con un balance adecuado entre objetividad y subjetividad.

Si bien son polémicos el concepto de experto y su clasificación, todavía es más complejo el proceso sistemático de selección, orientación y empleo de los expertos en la investigación científica (Grant y Davis, 1997; Scapolo y Miles, 2006). Dificultades referidas a la evaluación del nivel de competencia, así como a la determinación del número de expertos, suelen pasar desapercibidos en algunas investigaciones. Es justo señalar que estos problemas ya vienen siendo debatidos desde hace varios años (cf. Dalkey y Helmer, 1963; Landeta, 1999; Hasson y Keeney, 2011).

Respecto a los criterios de selección de expertos, la literatura muestra algunos tan difusos como la capacidad prospectiva y otros tan pragmáticos como el coste y la proximidad del experto. En general, pueden tomarse en consideración aspectos tales como la ética profesional, la profundidad del conocimiento, la amplitud de enfoques, el nivel de motivación y disposición a participar, la independencia de juicios, etcétera. Al análisis de estos aspectos suelen asociarse problemáticas de búsqueda de indicadores medibles, especialmente en la elaboración de

instrumentos empíricos. El presente trabajo se centra, precisamente, en el problema referido a la selección de los expertos. Un instrumento de selección se somete a juicio crítico y se presentan los resultados de su perfeccionamiento en el campo de la investigación educativa.

II. El instrumento y sus orígenes

En Cuba se conoce desde hace varios años una metodología implementada por el Comité Estatal para la Ciencia y la Técnica de la extinta URSS para la elaboración de pronósticos científico-técnicos, cuyo objetivo consiste en determinar el nivel de competencia de un candidato a experto (Evlanov y Kutusov, 1978). En esta metodología la competencia se determina por un coeficiente $k = \frac{1}{2} (k_c + k_a)$, donde k_c representa una medida del nivel de conocimientos sobre el tema investigado y k_a una medida de las fuentes de argumentación.

El cálculo de k_c requiere de la autoevaluación del candidato en una escala de 0 a 10, donde el valor seleccionado se divide por 10 para lograr cierta normalización. Respecto al cálculo de k_a , también es necesario que el encuestado se autoevalúe, pero atendiendo a seis posibles fuentes de argumentación en una escala tipo Likert. Para ello debe completar marcando con equis en cada fila de la Tabla I, donde los números constituyen los pesos asignados a cada fuente y aparecen ocultos en el instrumento.

Tabla I. Escala tipo Likert para el cálculo de k_a en el instrumento original

Fuentes de argumentación	Grado de influencia de las fuentes en sus criterios		
	Alto	Medio	Bajo
Análisis teóricos realizados por usted	.30	.20	.10
Su experiencia obtenida	.50	.40	.20
Trabajos de autores nacionales	.05	.05	.05
Trabajos de autores extranjeros	.05	.05	.05
Su propio conocimiento del estado del problema en el extranjero	.05	.05	.05
Su intuición	.05	.05	.05

Una vez efectuadas las autoevaluaciones es posible calcular el valor individual de k_a , a partir de la suma de los pesos correspondientes. Si el encuestado marca todas las posibles fuentes de argumentación en el nivel alto, entonces la suma de $k_a = .30 + .50 + 4 \times .05 = 1$; en cambio si las marca todas en un nivel bajo, entonces $k_a = .10 + .20 + 4 \times .05 = .50$. Como $0 \leq k_c \leq 1$ y $.50 \leq k_a \leq 1$, puede deducirse que la semisuma comprende el intervalo $.25 \leq k \leq 1$, donde el punto de corte aparece fijado en el valor .80 (nivel medio de competencia).

Desde hace una década esta metodología ha encontrado aplicación en disímiles investigaciones educativas cubanas, pero en opinión de varios especialistas no se logran los mejores resultados. En un sentido no peyorativo puede afirmarse que el instrumento no discrimina bien. Nótese que la igualdad de todos los números correspondientes a los diferentes grados de influencia ya suscita serias dudas.

Basta notar que se obtiene igual valor de ka , independientemente de que el encuestado marque cualquier categoría de las últimas cuatro filas.

En varias ocasiones, la mencionada metodología fue objeto de mejoramiento en alguna de sus partes. Por ejemplo, en una investigación en la que se sometió a criterio de expertos un aporte científico de carácter didáctico, las fuentes de argumentación se modificaron en función de acercarlas más a los requerimientos del objeto investigado (Aguilasocho, 2004). Asimismo, en investigaciones del Instituto Central de Ciencias Pedagógicas se perfeccionó el procesamiento estadístico de la escala Likert antes descrita (Campistrous y Rizo, 1998).

En el marco de la investigación que aquí se reporta, se concibió el objetivo de perfeccionar el citado instrumento desde el punto de vista empírico. Específicamente, se puso mayor atención a la escala Likert, en cuanto a la pertinencia de las fuentes de argumentación y a la objetividad de los pesos establecidos. Un argumento más a favor de la investigación consistió en que dicho instrumento tiene su origen en campos ajenos al ámbito educativo, de manera que resulta razonable explorar su consistencia tras un proceso natural de transferencia de tecnología.

El instrumento no sólo presentaba las dificultades antes descritas, sino que también carecía de fundamento metodológico. Durante el transcurso de la investigación no fue posible llegar a la génesis de la metodología, pues no se hallaron evidencias de su concepción original en la institución gestora que aún existe. A pesar de todo, muchos investigadores han hecho uso del instrumento, lo cual da crédito sobre su utilidad práctica. Estos hechos motivaron un estudio dirigido a subsanar las dificultades detectadas y a contar con evidencias empíricas contrastables que aseguren un proceso continuo de perfeccionamiento.

III. Metodología

Se aplicó un estudio Delphi a tres rondas. En la primera ronda se explicó el objetivo de la investigación y se pidió criterio sobre las posibles fuentes de argumentación que perfeccionarían el cálculo de ka . De forma preliminar se ofrecieron 15 posibles fuentes de argumentación, procedentes de diversas obras relacionadas con la determinación de expertos (Evlanov y Kutusov, 1978; Grant y Davis, 1997; Landeta, 1999; Germain, 2006). Se solicitó a los encuestados la propuesta de otras fuentes permisibles, así como la reescritura o perfeccionamiento de las ya existentes (búsqueda de validez). Con la ronda intermedia se optimizó el número de ítems y se reconcibió la escala de medición desde argumentos objetivos (búsqueda de fiabilidad). La ronda final constituyó un primer acercamiento a la determinación de un nuevo punto de corte.

3.1 Participantes y contexto

Para el estudio se solicitó la participación de investigadores de Ciencias Pedagógicas en ocho universidades cubanas. De 71 candidatos se seleccionó un

total de 56, los cuales conformaron el panel. En la selección se establecieron dos premisas fundamentales: el nivel de familiarización con el instrumento descrito y la voluntad por participar (experticia epistémica y ejecutiva, conforme a la clasificación de Weinstein, 1993).

Se estableció como vía fundamental para la comunicación el correo electrónico. Durante el transcurso del estudio algunos abandonaron el proceso y, por tal motivo, se determinó procesar la información de 52 panelistas (27 magísteres y 25 doctores). En su mayoría, los integrantes seleccionados estaban vinculados a proyectos de investigación, lo cual favorecía una discusión colectiva con incidencia directa o indirecta en la objetividad de sus juicios.

IV. Resultados

Después de la aplicación de la primera encuesta se pudo constituir un conjunto de 22 fuentes de argumentación, las cuales se evaluaron atendiendo a su grado de importancia. Para ello, en la segunda encuesta se empleó una escala de evaluación ordinal con diez categorías, donde a menores valores corresponde mayor importancia y viceversa. El cómputo de los resultados mostró una fiabilidad aceptable, con coeficiente de Cronbach $\alpha = .95$. El cálculo del coeficiente de Kendall reveló una ligera concordancia $W = .17$, lo cual es comprensible al observar que el número de ítems es casi la mitad del número de panelistas. No obstante, el resultado es significativo ($\chi^2 = 185.26$, $gl = 21$; $p < .01$). En la Figura 1 se representan los rangos medios de cada fuente de argumentación, a través de un diagrama de barras.

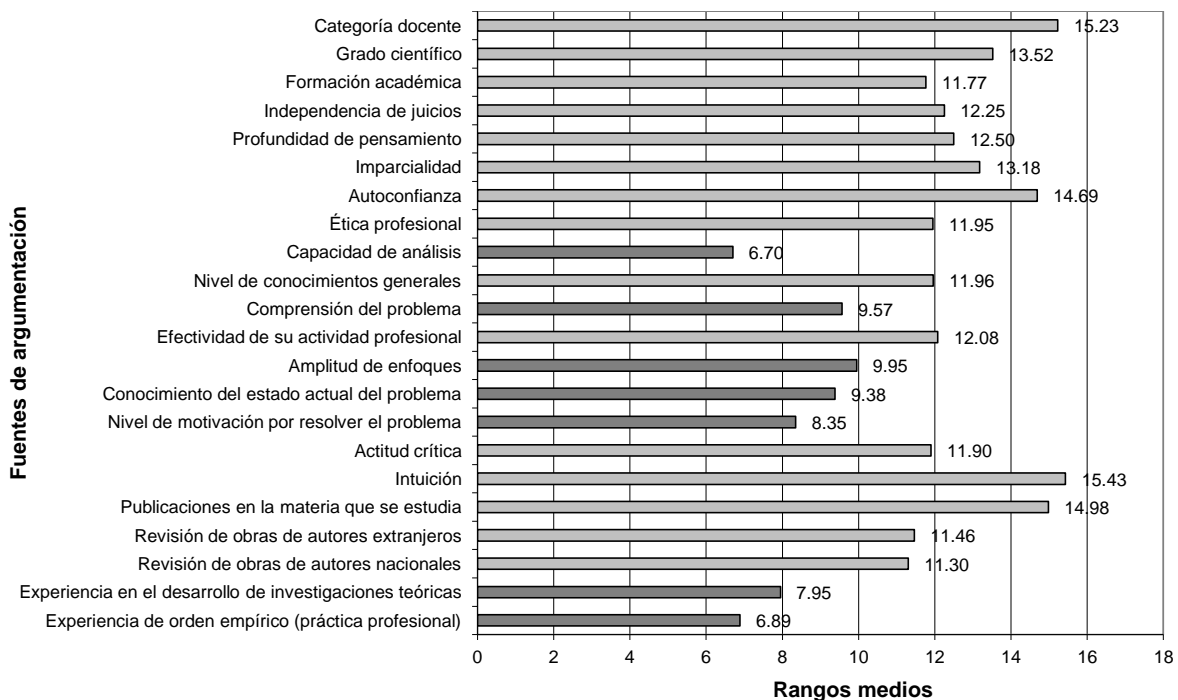


Figura 1. Rangos medios de las fuentes de argumentación

Por las características de la escala (1-10) el rango medio es inversamente proporcional al grado de importancia, asignado colectivamente a cada posible fuente de argumentación. Como puede observarse, a pesar de que “la intuición” constituye uno de los indicadores del instrumento original, tras el estudio quedó relegada al último puesto de entre el total de fuentes posibles. Las fuentes de rango medio inferior a 10 han sido destacadas con barras más oscuras. Este conjunto fue aislado del resto y reanalizado por el test de Kendall, el cual nuevamente indicó baja concordancia respecto al orden asignado ($W = .09$; $\chi^2 = 29.02$; $gl = 6$, $p < .01$). El bajo valor de W aquí es realmente favorable, pues constituye un indicio de pertinencia para cada fuente por separado.

Para la determinación de los pesos por categorías, se decidió partir de los nuevos rangos medios que resultaron de aislar las siete fuentes antes mencionadas. Seguidamente, se aplicó una simetría a cada uno de los rangos resultantes, respecto al promedio, de manera que a mayor jerarquía correspondería mayor peso y viceversa. Los pesos obtenidos se ilustran en la Tabla II.

Tabla II. Cálculo del peso de cada fuente de argumentación

No.	Fuentes de argumentación aisladas	Rangos medios	Pesos*
1	Capacidad de análisis	3.10	4.90
2	Comprensión del problema	4.61	3.39
3	Amplitud de enfoques	4.65	3.35
4	Conocimiento del estado actual del problema	4.47	3.53
5	Nivel de motivación por resolver el problema	3.97	4.03
6	Experiencia en el desarrollo de investigaciones teóricas	3.80	4.20
7	Experiencia de orden empírico (práctica profesional)	3.40	4.60

* Valor simétrico respecto al promedio de los rangos

Para medir el grado de influencia de las fuentes de argumentación se tomaron seis categorías desde nulo hasta muy alto, lo cual es permisible en escalas tipo Likert (Elejabarrieta e Íñiguez, 1984, p. 32). También se supuso que estos valores son puntos divisorios de un continuo subjetivo con escala de intervalo, donde es igual la distancia entre dos categorías limítrofes (intervalos aparentemente iguales). Dividiendo por cinco cada peso fue razonable asignar a la escala los valores expresados en la Tabla III desde cero hasta el propio peso como máxima categoría.

Tabla III. Escala Likert para el grado de influencia de las fuentes

Fuentes	Muy alto	Alto	Medio	Bajo	Muy bajo	Nulo
1	4.90	3.92	2.94	1.96	0.98	0.00
2	3.39	2.72	2.04	1.36	0.68	0.00
3	3.35	2.68	2.01	1.34	0.67	0.00
4	3.53	2.82	2.12	1.41	0.71	0.00
5	4.03	3.22	2.42	1.61	0.81	0.00
6	4.20	3.36	2.52	1.68	0.84	0.00

7	4.60	3.68	2.76	1.84	0.92	0.00
Suma	28.00	22.40	16.80	11.20	5.60	0.00

Como la suma de los valores de la columna “Muy alto” es 28, se dividió cada uno de los pesos por este número, de manera que el valor máximo de las sumas posibles quedara fijado en 1 (normalización). En la Tabla IV se ilustra la escala definitiva para la determinación del coeficiente k_a en el instrumento.

Tabla IV. Una modificación al instrumento descrito

Fuentes de argumentación	Grado de influencia de las fuentes en sus criterios					
	MA	A	M	B	MB	N
Capacidad de análisis	.18	.14	.11	.07	.04	.00
Comprensión del problema	.12	.10	.07	.05	.02	.00
Amplitud de enfoques	.12	.10	.07	.05	.02	.00
Conocimiento del estado actual del problema	.13	.10	.08	.05	.03	.00
Nivel de motivación por resolver el problema	.14	.12	.09	.06	.03	.00
Experiencia en el desarrollo de investigaciones teóricas	.15	.12	.09	.06	.03	.00
Experiencia de orden empírico (práctica profesional)	.16	.13	.10	.07	.03	.00

Leyenda: MA: muy alto A: alto M: medio B: bajo MB: muy bajo N: nulo

Cuando los encuestados evaluaron los ítems, por lo regular lo hicieron circunscribiéndose a cada uno de forma aislada, sin embargo es permisible que entre las fuentes exista cierta correlación objetiva. En la matriz de correlaciones se pudo observar que sólo cuatro parejas de ítems (de 231) tenían coeficientes de correlación de Spearman superiores a $\rho = .70$, con valores significativos. Consecuentemente, se indagó sobre la posible existencia de relaciones entre los ítems y ciertas variables latentes. Esta problemática se examinó desde la óptica del análisis factorial.

Tomando en consideración la elevada proporción de ítems respecto al número de encuestados, se aplicó la prueba de Kaiser-Meyer-Olkin (KMO) de adecuación de la muestra, obteniéndose el valor de .75 lo cual da una medida de la proporción de varianza producida por factores subyacentes (García, Gil y Rodríguez, 2000). De igual manera, la prueba de esfericidad de Bartlett que contrasta la hipótesis de que la matriz de correlación es una matriz identidad produjo resultados significativos ($\chi^2 = 1204.99$, $gl = 231$; $p < .01$).

Tras el análisis factorial se reveló cierta estructura latente. En efecto, la matriz de coeficientes rotados por el método Varimax sugiere la presencia de cuatro componentes. En la Tabla V se ilustran los coeficientes de correlación de cada ítem respecto a los componentes, donde los valores superiores a .50 aparecen en negritas y las filas correspondientes a los siete ítems seleccionados han sido sombreadas.

Tabla V. Matriz de componentes rotados

	Componentes			
	1	2	3	4
Categoría docente	.41	.10	.83	.08
Grado científico	.20	.15	.86	-.06
Formación académica	.47	.17	.76	-.06
Independencia de juicios	.73	.36	.34	.04
Profundidad de pensamiento	.80	-.08	.23	.20
Imparcialidad	.78	.39	.25	.20
Autoconfianza	.52	.38	.28	.15
Ética profesional	.69	.44	.30	-.04
Capacidad de análisis	.27	.11	-.25	.63
Nivel de conocimientos generales	.65	.25	.34	.00
Comprensión del problema	.70	.08	.08	.39
Efectividad de su actividad profesional	.58	-.04	.13	.45
Amplitud de enfoques	.84	.15	.30	.11
Conocimiento del estado actual del problema	.06	.48	.10	.70
Nivel de motivación por resolver el problema	.10	.04	.06	.91
Actitud crítica	.78	.54	.08	.00
Intuición	.76	.33	.11	.18
Publicaciones en la materia que se estudia	.38	.38	.37	.31
Revisión de obras de autores extranjeros	.13	.83	.35	.24
Revisión de obras de autores nacionales	.12	.82	.33	.22
Experiencia en el desarrollo de investigaciones teóricas	.21	.82	.08	.08
Experiencia de orden empírico (práctica profesional)	.35	.73	-.18	-.03

Nota:

Método de extracción: Análisis de componentes principales.

Método de rotación: Varimax con normalización de Kaiser.

Convergencia en 9 iteraciones.

Como puede observarse, el primer componente provee explicación para casi la mitad de los ítems. Aquí, en su mayoría, son fuentes de argumentación asociadas a aspectos de orden intrapersonal, como cognitivos, afectivos y motivacionales. El segundo componente explica aquellos ítems de naturaleza básicamente fáctica, basados en la influencia de los hechos. El tercer componente apunta hacia aspectos que dan crédito externo del grado de competencia profesional, como es el caso de las titulaciones. Finalmente, el cuarto componente se centra en el problema, ora por su conocimiento, ora por su comprensión.

Llama la atención que el conjunto formado por el primer, segundo y cuarto componente no discrepa, en lo esencial, de las siete fuentes definitivas del instrumento modificado. En cambio, el tercer componente constituye un conjunto rechazado por el panel, probablemente por disociar lo externo de lo esencial en la argumentación científica. De aquí puede concluirse que los ítems de la Tabla IV constituyen no sólo los elementos más significativos del conjunto original, sino que sirven de representantes de una especie de tres factores esenciales latentes. Por tanto, la modificación del instrumento puede considerarse efectiva.

En la tercera ronda se suministró un sumario de los resultados al panel de expertos y se aplicó una encuesta final. En este caso, se solicitó una autoevaluación por medio del instrumento modificado, a fin de juzgar la competencia individual respecto al uso de la metodología rusa. Cada experto calculó su propio coeficiente de competencia y debió de calificarlo de aceptable o inaceptable, conforme a su autovaloración personal. El 82.69% de los integrantes del panel calificó de aceptable su respectivo coeficiente. Luego, el conjunto de valores aceptables (43 de 52) fue promediado para obtener un punto de corte preliminar: $\mu = .75$ ($\sigma = .12$), ligeramente inferior al de la metodología original (.80). De acuerdo a la prueba t de una muestra, las diferencias respecto al punto de corte previo son significativas ($t = -2.57$, $gl = 42$, $p < .05$).

Para complementar los resultados, también se pidió a los panelistas una apreciación general sobre el nuevo instrumento. Ante ello las opiniones fueron favorables en su mayoría, siendo la sencillez y la objetividad los aspectos más frecuentes. Sin embargo, algunos panelistas expresaron sus reservas respecto a la forma de computar el coeficiente k , a partir de k_c y k_a , por considerar el promedio una medida demasiado simplificada y carente de un fundamento claro.

V. Discusión

Como paso primario para el aseguramiento de la validez de un instrumento, todo conjunto de indicadores debe reflejar el dominio específico que se mide. Las fuentes de argumentación seleccionadas caracterizan el concepto estudiado, lo cual es premisa para asegurar la validez de contenido. Esta afirmación se apoya en tres argumentos:

El conjunto de ítems original fue tomado de investigaciones afines a la experticia, y contó con la apreciación razonada del panel respecto al ámbito educativo.

El subconjunto seleccionado cuenta con elevada preferencia en una escala evaluativa.

Los ítems definitivos son representativos de ciertas variables latentes, las cuales explican la argumentación desde una concepción más general de la experticia.

Los componentes que emergieron tras el análisis factorial sugieren la existencia de cierta propensión por rechazar factores externos y enfatizan tres dimensiones: una personológica, una fáctica y otra problémica. En la primera dimensión confluyen aspectos objetivos y subjetivos, pero en la escala definitiva solo figura la motivación como elemento subjetivo.

Germain (2006) subraya la necesidad de lograr un balance adecuado entre objetividad y subjetividad durante la selección de los ítems. Esta autora introduce una dimensión "conductual" en el estudio de la experticia profesional, en el marco del empleo. Para ello incluye en su escala varios indicadores predominantemente subjetivos, tales como la autoconfianza, el carisma, la extroversión y (nuevamente) la intuición.

En la presente investigación, tanto la autoconfianza como la intuición experimentaron rangos medios superiores a 14, lo cual los posiciona en el último cuartil. Obrando con parsimonia, se decidió excluir estos indicadores de la escala definitiva, pero no es posible desechar la existencia de variables ajenas causantes de una especie de sesgo antisubjetivo en el panel.

Además de juzgar la validez de un instrumento por la naturaleza del contenido, también es necesario hacerlo desde algún criterio externo y desde evidencias referidas al grado en que la medición se relaciona consistentemente con otras mediciones (Ary, Cheser y Razavieh, 1990, pp. 256-268). La efectividad de la autoaplicación del test, por parte del panel, en alguna medida sirve de argumento para sostener cierto grado de validez de criterio. En cambio, la búsqueda de correlaciones con variables afines a la experticia (validez de constructo) no ha sido explorada en esta investigación. Se trata de un problema complejo, por implicar en primera instancia la determinación y justificación teórica de tales variables, para luego someterlas a estudios de orden empírico.

El proceso de consecución de las modificaciones se apoyó en un estudio Delphi a tres rondas, lo cual asegura un proceso óptimo en la comunicación con el panel (Landeta, 1999). La aplicación reiterada de las rondas no se basó en la búsqueda de un consenso puntual, sino en una estructuración lógica para la concepción y evaluación del conjunto de ítems. Para la etapa de evaluación solo se concibió una ronda, pues el coeficiente alfa de Cronbach sugirió elevada confiabilidad. De aquí que la inclusión de una nueva ronda sería innecesaria, más aún si otros estudios han experimentado efectos negativos en el panel cuando el número de rondas tiende a crecer (Linstone y Turoff, 1975).

Si bien la fiabilidad ha sido controlada en el sentido del número y diversidad de panelistas, del total de rondas y la claridad de los instrumentos, y del procesamiento de la información, todo ello se refiere al proceso de perfeccionamiento y no al instrumento mismo. Para investigaciones futuras es necesario seguir profundizando en la fiabilidad y validez del instrumento modificado. Las evidencias de la última ronda vaticinan resultados favorables, pero todavía debe prestarse mayor atención a ciertas cuestiones sutiles, de las cuales no se tiene toda claridad en esta investigación:

El promedio tácito de k_c y k_a podría sustituirse por una media ponderada. El alcance de este estudio solo se limitó al perfeccionamiento de k_a , aceptando de forma provisional la aparente igualdad en el peso objetivo de ambos coeficientes.

Las bases conceptuales para la integración del nivel de conocimientos y las fuentes de argumentación no están suficientemente establecidas. Incluso surgen inquietudes epistémicas respecto a la combinación de una escala de diferencial semántico para k_c y otra de tipo Likert para k_a , lo cual todavía requiere de mayor argumentación teórica.

La naturaleza controvertida y diversificada de la experticia motiva el análisis de sensibilidad de la escala, ante cualquier adopción de una tipología de expertos.

Persiste la necesidad de profundizar en los fundamentos metodológicos del instrumento, su aplicación y el procesamiento de la información empírica que brinde.

Agradecimientos

La presente investigación ha sido posible gracias al proyecto A2/039476/11 de la Universidad Politécnica de Valencia, el cual se desarrolla de conjunto con la Universidad de Holguín "Oscar Lucero Moya". Los autores desean agradecer a todos los panelistas que colaboraron con el estudio Delphi. Especialmente se agradecen los comentarios críticos de los árbitros anónimos, los cuales contribuyeron a enriquecer el presente trabajo.

Referencias

Aguilasocho, D. (2004). *Propuesta metodológica para la enseñanza de la programación visual en el bachillerato mexicano*. Tesis doctoral no publicada. Universidad de Ciencias Pedagógicas "Félix Varela" Santa Clara, Cuba.

Ary, D., Cheser, L. y Razavieh, A. (1990). *Introduction to research in education*. Nueva York: The Dryden Press: Holt, Rinehart and Winston.

Campistrous, L. y Rizo, C. (1998). *Indicadores e investigación educativa*. Instituto Central de Ciencias Pedagógicas. La Habana: Ministerio de Educación.

Crespo, T. (2007). *Respuestas a 16 preguntas sobre el empleo de expertos en la investigación pedagógica*. Lima, Perú: San Marcos.

Dalkey, N. y Helmer, O. (1963). An experimental application of the Delphi Method to the use of experts. *Management Science*, 9, 458. Consultado en <http://www.links.jstor.org/journals/00251909.html>

Elejabarrieta, F. J. e Íñiguez, L. (1984). *Construcción de escalas de actitud tipo Thurst y Likert*. Consultado en <http://antalya.uab.es/liniguez/Materiales/escalas.pdf>

Evlanov, L. G. y Kutusov, V. A. (1978). *Ehkspertnie otsenki v upravlenii* [Valoraciones de expertos en la dirección]. Moscú: Ehkonomika.

García, E., Gil, J. y Rodríguez, G. (2000). *Análisis factorial*. Cuadernos de Estadística (Núm. 7). Madrid: La Muralla.

Germain, M. L. (2006). *Development and preliminary validation of a psychometric measure of expertise: The Generalized Expertise Measure (GEM)*. Tesis doctoral no publicada. Barry University, Florida.

- Grant, J. S. y Davis, L. L. (1997). Selection and use of content experts for instrument development. *Research in Nursing & Health*, 20(3), 269-274.
- Hasson, F. y Keeney, S. (2011). Enhancing rigour in the Delphi technique research. *Technological Forecasting & Social Change*, 78, 1695-1704.
- Hoffman, R. R., Shadbolt, N., Burton, A. M. y Klein, G. A. (1995). Eliciting knowledge from experts: A methodological analysis. *Organizational Behavior and Human Decision Processes*, 62, 129-158.
- Konow, I. y Pérez, G. (1990). Método Delphi. En *Métodos y técnicas de investigación prospectiva para la toma de decisiones*. Universidad de Chile. Consultado en <http://geocities.com/Pentagon/Quarters/7578/pros01-03.html>
- Landeta, J. (1999). *El Método Delphi: Una Técnica de Previsión para la Incertidumbre*. Barcelona: Ariel.
- Linstone, H. A. y Turoff, M. (1975) (Eds.). *The Delphi Method. Techniques and Applications*. Reading, Massachusetts: Addison-Wesley.
- Scapolo, F. y Miles, I. (2006). Eliciting experts' knowledge: A comparison of two methods. *Technological Forecasting & Social Change*, 73, 679-704.
- Ténière-Buchot, P. F. (2001). Décision, expertise, arbitraire et transparence: éléments d'un développement durable. *Le courrier de l'environnement de l'INRA*, 44. Consultado en <http://www.inra.fr/dpenv44.htm>
- Weinstein, B. D. (1993). What is an expert? *Theoretical Medicine*, 14, 57-73.