

## El instrumento de observación ante el panel de expertos

### The observation instrument before the expert panel

Idalia Gabriela de Castro | Antonio Humberto Closas

#### RESUMEN

Cadena de bloques de seguridad criptográfica (*blockchain*) aporta a la contabilidad, la confiabilidad del registro distribuido y se presume que podría generar cambios significativos en la práctica contable ante la extendida opinión que supondría el paso a la triple entrada. El relevamiento bibliográfico evidenció las bondades atribuidas a dicha tecnología, su carácter experimental, las expectativas generadas en la disciplina contable y la carencia de instrumentos científicos. El objetivo de este artículo es realizar, a partir del juicio de expertos y técnicas estadísticas, el proceso de validación de contenido del instrumento de observación. La investigación sigue una estructura clásica de análisis teórico-empírico. Las decisiones metodológicas incluyen un diseño *no experimental, descriptivo, transversal, cualicuantitativo y prospectivo* que recolecta datos mediante *encuesta*. El *juicio de expertos* es una práctica generalizada de evaluación externa que pondera la representatividad de los ítems del instrumento. El *grado de acuerdo* se determinó mediante el coeficiente *kappa de Fleiss*. El *proceso de validación del documento entrevista* logró consenso entre los evaluadores respecto a la *validez de su contenido* para producir información sobre cadena de bloques, criptomonedas e información contable; aunque, las evidencias deberían admitirse con cautela dada la selección no probabilística de la muestra. Los ceñidos valores de las pruebas se atribuyen al carácter emergente de las temáticas abordadas, sin embargo, superan el umbral de acuerdo requerido al panel de expertos habiéndose verificado los criterios de claridad, coherencia, relevancia y suficiencia, utilizados para evaluar el documento entrevista.

*Palabras clave: validez de contenido, kappa de Fleiss, grado de acuerdo, estudio psicométrico*

**Idalia Gabriela de Castro**

[idecastro@eco.unne.edu.ar](mailto:idecastro@eco.unne.edu.ar)

[orcid.org/0000-0002-1359-080X](https://orcid.org/0000-0002-1359-080X)

Facultad de Ciencias Económicas  
Universidad Nacional del Nordeste  
ARGENTINA

**Antonio Humberto Closas**

[hclosas@hotmail.com](mailto:hclosas@hotmail.com)

[orcid.org/0000-0001-7144-068X](https://orcid.org/0000-0001-7144-068X)

Facultad de Ciencias Económicas  
Universidad Nacional del Nordeste  
ARGENTINA

#### COMO CITAR ESTE ARTÍCULO

de Castro, I. G.; Closas, A.H. (2023).  
El instrumento de observación ante el  
panel de expertos.  
*Revista de la Facultad de Ciencias  
Económicas*, 30(1), 61-76  
[http://dx.doi.org/10.30972/  
rfce.3016691](http://dx.doi.org/10.30972/rfce.3016691)



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Revista de la Facultad de Ciencias Económicas  
ISSN 1668-6357 (formato impreso) ISSN  
1668-6365 (formato digital) por Facultad de  
Ciencias Económicas Universidad Nacional  
del Nordeste (UNNE) Argentina se distribuye  
bajo una Licencia Creative Commons  
Atribución – No Comercial – Sin Obra  
Derivada 4.0 Internacional.

## ABSTRACT

Cryptographic security blockchain brings to accounting, the reliability of distributed recording and it is presumed that it could generate significant changes in accounting practice in the face of the widespread opinion that the move to triple entry would entail. The bibliographic survey showed the benefits attributed to this technology, its experimental nature, the expectations generated in the accounting discipline and the lack of scientific instruments. The aim of this article is to carry out, on the basis of expert judgement and statistical techniques, the process of validating the content of the observation instrument. The research follows a classical structure of theoretical-empirical analysis. The methodological choices include a *non-experimental, descriptive, cross-sectional, qualitative-quantitative* and *prospective* design that collects data by means of a survey. *Expert judgement* is a widespread external evaluation practice that weights the representativeness of the instrument's items. The *degree of agreement* was determined by *Fleiss' kappa* coefficient. The *validation process* of the *interview document* achieved consensus among the evaluators regarding the *validity of its content* for producing information on blockchain, cryptocurrencies and accounting information; however, the evidence should be admitted with caution given the non-probabilistic selection of the sample. The narrow values of the evidence are attributed to the emerging nature of the topics addressed, however, they exceed the threshold of agreement required of the expert panel, having verified the criteria of clarity, coherence, relevance and sufficiency, used to evaluate the interview document.

*Keywords: content validity, Fleiss kappa, degree of agreement, psychometric study*

---

## 1. INTRODUCCIÓN

En la Facultad de Ciencias Económicas de la Universidad Nacional del Nordeste, Argentina se desarrolla el proyecto de investigación: La contabilidad como ventaja competitiva (PI 19M004) y una línea de trabajo se refiere a *la relación entre la contabilidad y cadena de bloques*, la cual es incipiente, aunque se presume prometedora.

La tecnología de *cadena de bloques de seguridad criptográfica (blockchain)* fue comunicada en 1991, sin embargo, su difusión, promoción y desarrollo ha estado asociada a las criptomonedas surgidas desde 2009 y se ha extendido a ecosistemas de finanzas descentralizadas conocidos como De-Fi<sup>1</sup>. Dada las bondades que se le atribuyen, las valoraciones positivas que se expresan de ella y las expectativas de distintos sectores su uso se ha expandido con rapidez a proyectos<sup>2</sup> del sector público y privado.

Cadena de bloques aporta a la contabilidad, la confiabilidad del registro distribuido (transparencia-trazabilidad, certeza e inmutabilidad). Es igualmente cierto que, sin consolidarse, las incertidumbres y riesgos son numerosos, aunque se cree que podría generar cambios significativos en la práctica contable, la cual ha registrado un lento avance tecnológico ([Deloitte, 2016](#); [Argañaraz et al., 2019](#)).

Por ello, resulta prudente estudiar las características de la tecnología y monitorear sus avances para confirmar las implicancias que, se presume, tendrá en la disciplina contable y en el ejercicio profesional en ciencias económicas; ante la extendida opinión que supondría el paso desde la doble a la triple entrada<sup>3</sup> (de Castro y D'Andrea, 2021).

Así las cosas, adquiere relevancia aparear investigación y experiencia para revelar certezas que superen los aportes hipotéticos. Entonces, se han buscado herramientas utilizadas previamente que pudieran replicarse en trabajo de campo; sin embargo, *blockchain* y criptomonedas son temas emergentes para la disciplina contable, y la búsqueda no ha arrojado resultados. Asimismo, el relevamiento ha detectado literatura que explica la validación de instrumentos de diferentes áreas de conocimiento, aunque no en contabilidad. En ese contexto, se ha considerado adecuada la entrevista semiestructurada como técnica de recolección de datos empíricos y; a fin de mantener el diálogo encauzado en los temas consultados y reducir la dispersión de respuestas a sistematizar, se ha elaborado un documento *ad hoc* que guíe el encuentro previamente acordado.

Estas acciones, pertinentes y relevantes, ayudan a incrementar el conocimiento sobre la contribución de la disciplina contable en los procesos decisorios y su aporte al desempeño del ente como fin último de la investigación.

El objetivo de este artículo es realizar, a partir del juicio de expertos y técnicas estadísticas, el proceso de validación de contenido del instrumento de observación.

La *validación* es un proceso continuo que incluye diferentes procedimientos para comprobar si el instrumento captura realmente lo que pretende medir (Nunnally, 1967). Los estudios de *validez* y *fiabilidad* se centran en demostrar la capacidad de vincular conceptos teóricos a indicadores empíricos (aptitud psicométrica) y la precisión de los resultados (cuantía en que las medidas están libres de errores aleatorios) lo que concede y, a la vez es necesaria para, una interpretación teórica coherente de las observaciones recogidas.

El estudio de la *validez* (o *autenticidad: certeza y verdad en el ámbito científico*) de los instrumentos, en general, considera tres aspectos: contenido, criterio y constructo.

La *validez de contenido*, refiere a la relevancia y suficiencia (o representatividad) de los ítems. Surgida de la corriente racionalista se evalúa mediante *juicio de expertos*<sup>4</sup> como también de *métodos estadísticos* que miden la idoneidad de las respuestas.

El *juicio de expertos* se define como una opinión informada de personas con trayectoria en el tema y que pueden dar información, evidencia, juicios y valoraciones respecto del mismo. Su *objetivo* es validar el contenido de una prueba diseñada (o adaptada) por un grupo de investigadores (Escobar-Pérez y Cuervo-Martínez, 2008). Esto es, comprobar si el contenido es representativo de la realidad que pretende observar; como también la forma en que los ítems fueron agrupados (estructura).

Escobar-Pérez y Cuervo-Martínez (2008) señalan su mayor relevancia cuando el instrumento ha sido diseñado para: a) cierta oportunidad específica (*ad hoc*), b) una población diferente y tuvo que ser adaptado<sup>5</sup> y/o traducido<sup>6</sup>; o c) cuando las observaciones experimentales son limitadas (Utkin, 2006). En tanto, las restricciones se dan en relación a la validez de constructo (mismo nombre y distintas dimensiones) y el propósito de la evaluación (e.g.: diagnóstico, desempeño).

La *validez de criterio*, entendida como posibilidad de equiparar las puntuaciones obtenidas con otros resultados (validez concurrente) o la capacidad del instrumento para anticipar medidas (validez predictiva). Derivada de la tradición empirista, busca evidencia de relación con variables externas mediante técnicas estadísticas de análisis multivariante tales como: el *análisis factorial exploratorio* (examina la relación existente entre un conjunto de variables observables y factores subyacentes inobservables sin depender de la especificación previa del modelo teórico) y; la aproximación exploratoria mediante *coeficientes de correlación* (e.g.: Rho, Pearson, Tau-b) del constructo; pruebas que podrían realizarse cuando el instrumento ha sido puesto al servicio de la investigación.

La *validez de constructo* (o *estructural*), referida a la coherencia del instrumento con la teoría de la que parte y su aptitud para medir la construcción teórica (concepto). Pondera su *organización interna* (dimensiones y funcionamiento diferencial de ítems) mediante técnicas de análisis multivariante ([Martínez Gómez et al., 2012](#)) tales como: *Análisis factorial confirmatorio e índice de homogeneidad* (o correlación dimensión-total corregida). La contrastación de hipótesis estructurales determina su validez y un pseudo procedimiento para confirmar el constructo desde la teoría hacia los hechos ([Pérez Gil et al., 2000](#)).

La confiabilidad (fiabilidad o *precisión*) indica la estabilidad y predictibilidad de una medida. Se determina con técnicas que evalúan la consistencia interna de la prueba; por ejemplo, *coeficiente Alfa de Cronbach*, según Martínez Gómez et al. (2000). El indicador señala la cuantía en que las observaciones únicas de una prueba, a partir de la cual se generaliza, están libres de errores casuales o aleatorios (varianza de error). Puede aplicarse a la muestra y en cada dimensión (Alfa de Cronbach sin la dimensión). Este último revela la correlación lineal entre el ítem y la puntuación total (sin el ítem evaluado) señalando la magnitud y dirección de la relación. Alfa de Cronbach se basa en tres premisas: a) observaciones independientes y errores no correlacionados; b) escalas aditivas de cada categoría linealmente relacionada con la puntuación total; y c) cada par de elementos debe tener distribución normal bivariada.

## 2. METODOLOGÍA

La investigación se desarrolla siguiendo una estructura clásica, de análisis teórico-empírico. Las características metodológicas se indican seguidamente:

### 2.1. Diseño

La investigación, en función del objetivo, es de naturaleza *no experimental*, de estilo *descriptivo*, de corte *transversal* y de línea *cuali-cuantitativa*. Recolecta datos mediante la técnica de *encuesta* administrada en forma personal a través del correo electrónico y, conforme a su propósito, es de tipo *prospectivo*.

### 2.2. Participantes

Son partícipes de este estudio tres investigadores del PI 19MOO4 involucrados en la generación de procedimientos y documentos como también cuatro expertos que han evaluado el instrumento

denominado: entrevista semiestructurada sobre *blockchain*, criptomonedas e información contable. Del total de sujetos que intervienen, seis poseen formación de posgrado a nivel de maestrías o doctorados y todos con carrera académica en docencia universitaria.

El panel de expertos se ha conformado por especialistas que acreditan trayectoria científica en estadística multivariante aplicada a estudios sociales, responsabilidad institucional en planeamiento y administración financiera y, ejercicio profesional en administración de negocios o en ingeniería en sistemas de información como antecedentes destacados de quienes han aceptado voluntaria y amablemente compartir su experiencia en la temática.

Se ha definido el criterio trayectoria científica, académica y/o profesional para elegir los expertos configurando una selección intencional y por conveniencia al incluir a referentes locales. Así, el grupo aceptante quedó integrado por cuatro profesionales que proporcionaron datos para las pruebas estadísticas que se aplicaron y cuyos resultados respaldan la validez del instrumento de medida.

### 2.3. Instrumentos

El material al que se ha denominado “Entrevista semiestructurada sobre *blockchain*, criptomonedas e información contable” (en adelante, *documento entrevista*), consta de la introducción donde se explican aspectos del proceso<sup>7</sup>, cuatro secciones con 20 preguntas que capturan 36 datos que se utilizan exclusivamente con propósitos científicos en un tiempo estimado de respuesta de 30 minutos (véase [Tabla 1](#)).

**Tabla 1.** Estructura de preguntas y datos de la entrevista<sup>8</sup>

Sección	Contenido	Preguntas	Respuestas
Presentación del entrevistado	Identificación del entrevistado y, si corresponde, la organización a la que representa. Su experiencia en los temas consultados.	2	6
Cadena de bloques de seguridad criptográfica	Uso y/o implementación de cadena de bloques. Efectos sobre la competitividad, ventajas y desventajas.	6	10
Criptomonedas	Utilización, acceso, regulación y motivación que impulsa las transacciones con criptomonedas.	6	8
Información contable	Datos sobre la incorporación de recursos digitales a la economía formal, información para tomar decisiones, evaluación de riesgo y rentabilidad, como también procesamiento, medición y exposición contable.	6	12
	Total	20	36

Fuente: Elaboración propia.

Al final se dispone de un espacio de expresión libre y el compromiso de los investigadores de enviar a los participantes un informe con las conclusiones obtenidas.

El documento “Instructivo de validación del instrumento de observación”, tiene una presentación general y una “Planilla de valoración del instrumento: Entrevista semiestructurada sobre *blockchain*, criptomonedas e información contable” (en adelante, *planilla de valoración*) donde los especialistas han registrado su intervención.

## 2.4. Variables

El documento entrevista define las *variables teóricas-temáticas*: cadena de bloques de seguridad criptográfica, criptomonedas e información contable.

La planilla de valoración refleja *variables operativas* relativas al diseño del instrumento de observación: claridad, coherencia, relevancia y suficiencia (véase Tabla 2).

**Tabla 2.** Criterios para valorar las preguntas

Criterios	Calificación	Indicador
<b>Claridad</b> La pregunta se comprende fácilmente, es decir, su sintáctica y semántica son adecuadas.	1. No cumple el criterio	La pregunta no es clara.
	2. Bajo nivel	La pregunta requiere bastantes modificaciones o una modificación muy grande en el uso de las palabras de acuerdo con su significado o por la ordenación de las mismas.
	3. Moderado nivel	La pregunta requiere una modificación específica de algunos de los términos.
	4. Alto nivel	La pregunta posee claridad, semántica y sintaxis adecuada.
<b>Coherencia</b> La pregunta tiene relación lógica con la dimensión que está observando.	1. No cumple el criterio	La pregunta no tiene relación lógica con la dimensión.
	2. Bajo nivel	La pregunta tiene una relación tangencial con la dimensión.
	3. Moderado nivel	La pregunta tiene una relación adecuada con la dimensión que está midiendo.
	4. Alto nivel	La pregunta se encuentra completamente relacionada con la dimensión que está midiendo.
<b>Relevancia</b> La pregunta es esencial o importante, es decir debe ser incluido.	1. No cumple el criterio	La pregunta puede ser eliminada sin que se vea afectada la medición de la dimensión.
	2. Bajo nivel	La pregunta tiene alguna relevancia, pero otra pregunta puede estar incluyendo lo que mide esta.
	3. Moderado nivel	La pregunta es relativamente importante.
	4. Alto nivel	La pregunta es relevante y debe ser incluida.
<b>Suficiencia</b> Las preguntas de una misma dimensión bastan para obtener la observación de esta.	1. No cumple el criterio	Las preguntas no son suficientes para medir la dimensión.
	2. Bajo nivel	Las preguntas miden algún aspecto de la dimensión pero no la dimensión en su totalidad.
	3. Moderado nivel	Se deben incrementar algunas preguntas para poder evaluar la dimensión completamente.
	4. Alto nivel	Las preguntas son suficientes para medir la dimensión.

Fuente: Escobar-Pérez y Cuervo-Martínez (2008).

## 2.5. Procedimiento

La construcción del marco teórico ha permitido comprender distintos aspectos de la tecnología cadena de bloques y su vinculación con las criptomonedas, como también formular conjeturas sobre los avances que, en la disciplina contable, a ella se vinculan lo que ha orientado el diseño del instrumento de observación de la investigación.

La *validación* se ha aplicado sobre el documento entrevista y su evaluación se hizo utilizando la planilla de valoración cuyos criterios y calificaciones se han reseñado en 2.4 agregándose las observaciones cualitativas, para los 20 ítems del instrumento sometido a arbitraje. Finalmente, se ha evaluado globalmente la adecuación de preguntas considerando el perfil de quienes serán entrevistados, la integridad del constructo en cuanto a la amplitud y pertinencia de su contenido.

A mediados de mayo de 2022 se ha enviado a los expertos los instrumentos indicados en la sección 2.3, por correo electrónico personal nominado. Se les indicó que la planilla de valoración tendría que ser remitida a la brevedad y considerando como fecha límite un periodo máximo de quince días posteriores a su recepción.

Las observaciones recogidas se tabularon en archivo del programa Microsoft Office Excel 2019 donde se las ordenó vinculándolas con cada experto del panel generando así una matriz electrónica de datos que ha facilitado su posterior incorporación al programa de análisis estadístico.

## 2.6. Análisis de datos

En el estudio de *validez de contenido*, participaron expertos que evaluaron cada ítem, según se anticipara, con base en criterios de claridad, coherencia, relevancia y suficiencia, determinándose el grado de acuerdo mediante el coeficiente *kappa de Fleiss*.

El procesamiento de datos numéricos fue realizado con el programa informático IBM SPSS Statistic 26 (George y Mallery, 2020), ampliamente utilizado por investigadores en áreas de psicología, educación, economía, entre otros campos disciplinares.

## 3. RESULTADOS

En este apartado se comparten los hallazgos del *proceso de validación* del documento entrevista.

Las evidencias iniciales sobre la utilidad, aptitud y eficacia han sido provistas por la técnica *juicio de expertos*, que muchas veces proporciona los *únicos indicadores subjetivos sobre la validez de contenido*, determinando si la prueba respeta los propósitos para los que ha sido diseñada. Esta faceta cualitativa resulta valiosa a fin de minimizar los márgenes de error durante su utilización en cada espacio cultural.

Los especialistas, en general, han asignado (ver criterios de valoración en Tabla 2) calificaciones de 3 = *Moderado* y 4 = *Alto*. Ninguno ha utilizado 1 = *No cumple* y, en contadas ocasiones, han atribuido valor 2 = *Bajo nivel*; dando evidencia preliminar respecto de la calidad del instrumento y la afinidad de los jueces.

Sin embargo, esta medida resulta insuficiente ya que no incluye: a) el posible sesgo del investigador en la medición del constructo que puede limitarse mediante métodos objetivos, por

ejemplo, las combinaciones binarias de Thurstone ([Pedrosa et al., 2013](#)); b) el consenso esperado por el azar cuya consideración requiere determinar el coeficiente kappa entre otros indicadores posibles.

En efecto, para producir evidencia respecto del grado de acuerdo logrado entre los cuatro jueces que evaluaron el documento entrevista se ha empleado el índice kappa de Fleiss ([Fleiss et al., 2003](#)). La interpretación se realiza vinculando su valor a una escala cualitativa de seis niveles generada por [Landis y Koch \(1977\)](#), presentada en [Tabla 3](#), e indica la medida del consenso que supera al que se esperaría por casualidad, sin aportar certeza sobre la calidad (los evaluadores pueden estar *concordantemente equivocados*).

**Tabla 3.** Estadístico kappa de Fleiss

Valor <i>kappa</i>	Interpretación	
< 0	Concordancia por azar	El coeficiente kappa varía entre -1 y +1. Cuanto más cercano a +1, mayor el grado de concordancia. Su proximidad a -1, indica acuerdo inferior al esperado por el azar.
.01 – .20	Concordancia ligera	
.21 – .40	Concordancia justa	
.41 – .60	Concordancia moderada	
.61 – .80	Concordancia sustancial	
.81 – 1.00	Concordancia perfecta	

Fuente: Landis y Koch (1977).

Según se sabe, el instructivo de validación posee cuatro criterios: claridad, coherencia, relevancia, suficiencia y 20 ítems en total. A su vez, en cada uno de ellos se pueden distinguir los ejes temáticos o dimensiones: Presentación del entrevistado (2 ítems), *Blockchain* (6 ítems), Criptomonedas (6 ítems), Información Contable (6 ítems).

Se brindan a continuación los resultados obtenidos para cada criterio, así como para las categorías más relevantes que formaron parte de las respuestas proporcionadas por los expertos que participaron en la validación del instrumento de observación.

Por cierto, las categorías en las que interesa medir el nivel de coincidencia entre los observadores son los dos de mayor calificación; esto es, moderado y alto.

En la [Tabla 4](#) se observan los resultados relativos al acuerdo global obtenido a partir de las respuestas dadas por los cuatro jueces a los 20 ítems del documento entrevista. Se puede interpretar que, para los criterios claridad, coherencia y relevancia, el valor de acuerdo global medido a través del coeficiente kappa, como su intervalo de confianza para el 95%, de acuerdo con [Landis y Koch \(1977\)](#), se encuentran en el rango denominado *concordancia justa* [.21 – .40]. Además, los valores para los criterios coherencia (kappa = .334; IC 95% [.329, .339]) y relevancia (kappa = .350; IC 95% [.344, .355]), superan la media del rango concordancia justa.



**Tabla 4.** Acuerdo global entre evaluadores

Criterio	Coeficiente <i>kappa</i>	Concordancia	Estadístico <i>z</i>	Valor <i>p</i>	Intervalo de confianza al 95%	
					Lím. inferior	Lím. Superior
Claridad	.232	Justa	3.219	.001	.227	.236
Coherencia	.334	Justa	4.219	.000	.329	.339
Relevancia	.350	Justa	4.231	.000	.344	.355
Suficiencia	.410	Moderada	1.996	.046	.395	.420

Fuente: Elaboración propia.

En el criterio suficiencia, si bien participan los cuatro evaluadores, no se tiene respuesta sobre los 20 ítems con los que se ha trabajado en los criterios anteriores, sino sobre las cuatro dimensiones que conforman el documento entrevista. Cabe señalar que tampoco se modifica la cuantificación de las categorías que se utilizan para evaluar, en esta instancia, las dimensiones del cuestionario. Así pues, se detallan en [Tabla 4](#), los datos del acuerdo global logrado para la muestra conformada por las variables (jueces) y observaciones (dimensiones). El valor kappa = .410, se encuentra al principio del rango *concordancia moderada*, siendo su IC 95% [.395, .420].

El resultado obtenido para el índice kappa de Fleiss, en cada una de las pruebas, se transforma a valores normalizados del estadístico *z*; siendo este último a través de su correspondiente valor *p* el que participa en el contraste de hipótesis, las cuales se formulan a continuación.

$H_0$ : Los datos observados indican concordancia por azar entre los expertos

$H_1$ : Los datos observados no indican concordancia por azar entre los expertos

Así, se observa que el índice kappa de Fleiss es estadísticamente significativo para los tres primeros criterios, no solo al nivel  $\alpha = 5\%$ , sino cumpliendo también con el criterio de Benjamin et al. (2017), ya que valor *p* ha resultado inferior a .005. En cambio, para el criterio suficiencia el índice kappa es estadísticamente significativo solo al nivel de confianza del 95%, puesto que en esta ocasión valor *p* < .05. Ciertamente, en todos los casos analizados se rechaza la hipótesis nula de que la concordancia entre los expertos sea debida a un hecho casual o aleatorio.

En [Tabla 5](#), se encuentran los valores que se obtuvieron del análisis acuerdo entre evaluadores para las dos categorías de interés: moderado nivel (3) y alto nivel (4).

**Tabla 5.** Acuerdo entre evaluadores en categorías individuales

Criterio	Punt. Categ.	Prob. Cond.	<i>Kappa</i>	Concordancia	Estadístico <i>z</i>	Valor <i>p</i>	Intervalo de confianza al 95%	
							Lím. inferior	Lím. superior
Claridad	3	.2000	.375	Justa	4.108	.000	.369	.381
	4	.7125	.247	Justa	2.710	.007	.242	.253
Coherencia	3	.1875	.426	Moderada	4.663	.000	.420	.431
	4	.7750	.331	Justa	3.625	.000	.325	.337
Relevancia	3	.2000	.427	Moderada	4.678	.000	.421	.433
	4	.7750	.331	Justa	3.625	.000	.325	.337
Suficiencia	3	.4375	.410	Moderada	1.996	.046	.395	.420
	4	.5625	.410	Moderada	1.996	.046	.395	.420

Fuente: Elaboración propia.

*Nota:* Punt. Categ. (Puntuación categoría), Prob. Cond. (Probabilidad condicional).

### Interpretación de los datos de la Tabla 5

Claridad: las probabilidades condicionales indican que los cuatro expertos están bastante más de acuerdo, a la hora de valorar los 20 ítems, con la calificación alto nivel (57/80, es decir 71.25%), que con la categoría moderado nivel (16/80, esto es 20.00%). La probabilidad experimental acumulada entre estas dos categorías es realmente elevada (91.25%, o sea que 73 de 80 respuestas calificaron la comprensión de las preguntas con la categoría alto o moderado nivel).

Además, como se aprecia en la [Tabla 5](#), cada índice kappa, junto con el correspondiente intervalo de confianza, son indicadores que se sitúan todos en el rango concordancia justa. También, el valor  $p$  relativo al estadístico de contraste  $z$ , en cada categoría, permite rechazar la hipótesis nula.

En atención a los resultados obtenidos en el análisis estadístico de: a) acuerdo global entre evaluadores, y b) acuerdos individuales para las categorías alto (4) y moderado (3), es posible afirmar que existen evidencias suficientes para sostener la concordancia justa entre los jueces participantes sobre el criterio claridad.

Coherencia: los resultados logrados para las calificaciones individuales (3 y 4) fueron similares a los obtenidos en el estudio del criterio claridad. En efecto, en esta ocasión las probabilidades condicionales evidencian que nuevamente la categoría alto nivel es la seleccionada por los evaluadores con mayor frecuencia relativa (62/80, 77.50%); en cambio, mucho menos acuerdo hubo entre los jueces sobre la elección de la modalidad moderado nivel (15/80, 18.75%). La probabilidad empírica acumulada para estas dos categorías, volvió a ser muy importante, incluso superó a la registrada en el criterio anterior (96.25%; es decir, 77 de 80 respuestas coinciden en que la relación de la pregunta con la dimensión que se evalúa tenía un nivel alto o moderado).

En la [Tabla 5](#), se distingue que el índice kappa relativo a la categoría moderado nivel (3), junto con el correspondiente intervalo de confianza, se encuentran entre .41 y .60, identificado como rango de concordancia moderada ([Landis y Koch, 1977](#)). Por su parte, el coeficiente kappa para la modalidad alto nivel (4) y su respectivo intervalo de confianza, están entre .21 y .40; es decir, en el rango conocido como concordancia justa.

A su vez, el estadístico de prueba  $z$  para ambas categorías (3 y 4) resultó significativo al nivel  $\alpha = .005$ , puesto que valor  $p < \alpha$  ([Benjamin et al., 2017](#)), lo que permite rechazar la hipótesis nula.

En virtud de lo que antecede, se da por probada la concordancia entre los observadores en la selección de las categorías (3 y 4) sobre el criterio coherencia.

Relevancia: los resultados para las calificaciones individuales 3 y 4 fueron similares a los hallados en el estudio de la dimensión coherencia. Efectivamente, las probabilidades experimentales muestran que la categoría alto nivel es la opción preferida por los observadores (62/80, 77.50%); mientras que para la modalidad moderado nivel el acuerdo entre los evaluadores ha sido bastante más bajo (16/80, 20.00%). La probabilidad condicional acumulada para estas dos categorías volvió a ser muy importante, incluso superó a la registrada para el criterio anterior (97.50%; es decir, 78 de 80 respuestas coinciden en que la importancia de la pregunta objeto de análisis poseía un alto o moderado nivel).

En la [Tabla 5](#), se distingue que el índice kappa relativo a la categoría moderado nivel (3), junto con el correspondiente intervalo de confianza, se encuentran entre .41 y .60, identificado

como rango de concordancia moderada. Por su parte, el coeficiente kappa para la modalidad alto nivel (4) y su respectivo intervalo de confianza, están entre .21 y .40; es decir, en el rango según sabemos llamado concordancia justa.

A su vez, el estadístico de prueba  $z$  resultó significativo para las dos modalidades, puesto que el valor  $p$  es inferior a  $\alpha = .005$  en ambos casos, motivo por el que es posible rechazar la hipótesis nula.

En razón de los resultados precedentes, se puede afirmar que entre los jueces hubo acuerdo en la selección de las categorías 3 y 4 relativas al criterio relevancia.

Suficiencia: En la [Tabla 5](#), se hallan los datos obtenidos para las categorías de interés: moderado (3) y alto (4). Según se puede ver el coeficiente kappa de Fleiss y su intervalo de confianza, el estadístico de contraste  $z$  y su valor  $p$ , para ambas modalidades resultaron iguales a los hallados en el análisis del acuerdo global ([Tabla 4](#)). Sin embargo, se observa una diferencia poco relevante en las probabilidades condicionales, ya que para la categoría moderado nivel es 43.75% (7 de 16 respuestas), mientras que para la modalidad alto nivel la probabilidad empírica resultó 56.25% (9 de 16 respuestas). La frecuencia relativa acumulada para estas dos categorías fue 16/16; es decir, el 100.00% de las respuestas coinciden en que la cantidad de preguntas utilizadas para evaluar la dimensión objeto de análisis poseen un nivel moderado o alto de suficiencia.

En atención a lo que precede, la concordancia moderada entre los observadores sobre el criterio suficiencia ha quedado debidamente verificada.

Desde luego, a partir de los resultados logrados en las pruebas de hipótesis relativas a las puntuaciones de las categorías de interés (niveles moderado y alto) para cada uno de los cuatro criterios estudiados, se puede sostener en cada caso el rechazo de que el acuerdo entre los jueces es debido a circunstancias fortuitas o azarosas.

#### 4. DISCUSIÓN

Desde el punto de vista científico, la evaluación externa del instrumento a cargo de especialistas es de gran relevancia, concede validez a los datos obtenidos a través del mismo y mejora la eficiencia del trabajo de recolección.

El *juicio de expertos* es una práctica generalizada que evalúa la representatividad de los ítems del instrumento para obtener evidencias sobre su *validez de contenido*. Las numerosas circunstancias que inciden en el proceso de emisión de la opinión –en esencia subjetiva– resaltan la importancia de estimar su confiabilidad mediante el indicador: *grado de acuerdo* interevaluadores que, en este estudio, ha sido medido a través del test *kappa de Fleiss*; considerado adecuado por tratarse de cuatro evaluadores y de variables categóricas, de escala ordinal.

Las decisiones metodológicas asumidas para determinar la muestra de expertos que integraron el panel a través de métodos no probabilísticos de selección implicaron un esfuerzo deliberado para obtener un grupo representativo. Las mismas se han basado en la optimización de los recursos disponibles y, por tal motivo, las conclusiones tendrían un alcance limitado.

Sin embargo, las evidencias deberían admitirse con cierta racionalidad, puesto que la interpretación de los indicadores de contraste de hipótesis (estadístico  $z$  y su valor  $p$ ) han descansado en su vinculación con el índice kappa, utilizado en el cálculo del acuerdo entre los cuatro evaluadores

(global y en categorías individuales de calificación 3 y 4). El estadístico kappa, desarrollado por [Fleiss et al. \(2003\)](#), es una extensión o generalización del índice *pi de Scott* ([Scott, 1955](#)) y resulta en la actualidad, de acuerdo con [Gwet \(2014\)](#), una adecuada opción entre las pruebas del mismo tipo.

En el acuerdo global entre evaluadores ([Tabla 4](#)), se obtuvo fuerza de concordancia justa e inicialmente moderada, siendo más alta la relativa al criterio de suficiencia ( $\text{kappa} = .410$ ). El coeficiente kappa resultó estadísticamente significativo para los cuatro criterios analizados (valor  $p < \alpha$ ), al menos para el nivel  $\alpha = .05$ .

En el acuerdo entre evaluadores en categorías individuales, se obtuvo fuerza de concordancia justa y moderada en cuatro ocasiones para cada una de las categorías en las ocho pruebas cuyos resultados se muestran en la [Tabla 5](#). La calificación 3 = Moderado nivel ha obtenido los más altos valores en los criterios de coherencia ( $\text{kappa} = .426$ ) y relevancia ( $\text{kappa} = .427$ ). El coeficiente kappa resultó estadísticamente significativo en las pruebas de hipótesis de las ocho categorías de puntuaciones 3 y 4, como se puede observar en la tabla (valor  $p < \alpha$ ), al menos para el nivel  $\alpha = .05$ .

En los análisis realizados, el test kappa ha provisto evidencias sobre la calidad del consenso y el valor  $p < .05$  del estadístico z indicó que el entendimiento interevaluadores es significativamente mejor del acuerdo esperado por azar. Sin embargo, estos estadísticos no informan si dicha conformidad es una decisión correcta.

En atención a los resultados obtenidos en el análisis estadístico de: a) acuerdo global entre evaluadores, y b) acuerdos individuales para las categorías alto y moderado, por cierto, es posible afirmar que existen evidencias suficientes para sostener la concordancia entre los jueces sobre los criterios de claridad, coherencia, relevancia y suficiencia, utilizados para evaluar el documento entrevista.

Los ceñidos valores logrados en la mayoría de las pruebas se atribuyen al carácter emergente de las temáticas abordadas en la disciplina contable y dado que todas las mediciones superan el umbral de acuerdo requerido al panel de expertos el instrumento podrá ser utilizado para sustentar una investigación empírica futura sobre cadenas de bloques de seguridad criptográfica y criptomonedas e información contable. En particular para conocer sobre el grado de avance en las provincias de Chaco y Corrientes, como también un diagnóstico sobre la funcionalidad del sistema contable como generador de información para la toma de decisiones.

## 5. CONCLUSIONES

El relevamiento bibliográfico puso en evidencia las bondades que se atribuyen a la tecnología de cadena de bloques, el carácter experimental y las expectativas que ha generado en el ámbito de la disciplina contable como también la carencia de instrumentos de observación, aspectos que han justificado el diseño metodológico adoptado.

El instrumento *ad hoc*, elaborado, discutido y adecuado internamente por investigadores del proyecto, inevitablemente contiene cierto grado de sesgo; la intervención de expertos le concede: objetividad, validez y confiabilidad. La pertinencia del contenido de los ítems propuestos, como

la conformación del instrumento, han resultado adecuadas para recoger información sobre los temas objeto de interés.

En la evaluación externa han participado referentes del área de estadística, finanzas, ingeniería en sistemas y administración. Sus recomendaciones y sugerencias han enriquecido el documento objeto de arbitraje, aportando tanto a la disciplina contable como al fortalecimiento de la comunidad científica local.

Se ha llevado a cabo el proceso de validación del documento entrevista el cual ha logrado consenso entre los evaluadores respecto a la validez de su contenido para generar información sobre la temática abordada.

Los valores de los test realizados superan el umbral de acuerdo requerido ([Landis y Koch, 1977](#)) al panel de expertos, habiéndose verificado los criterios de claridad, coherencia, relevancia y suficiencia, utilizados para evaluar el documento entrevista.

Las distintas pruebas de *validez* que se conocen: contenido, criterio y constructo, la primera realizada, las demás en vista de realizarse sobre el documento entrevista, proporcionarán evidencias que consolidarán su utilidad, aptitud y eficacia. A su vez, la *confiabilidad*, es requisito necesario pero no suficiente, que ciertamente podrá ser evaluada con posterioridad a la aplicación del instrumento. Así, la *aptitud psicométrica* de una herramienta de medición depende de la presencia de ambas características conformando estos indicadores una interpretación combinada que es objetivamente valorada en el área de conocimiento.

Más allá de las limitaciones que toda elaboración científica lleva consigo, se anhela que este estudio constituya un aporte sobre la generación de instrumentos científicos de observación como también al avance futuro de la relación entre la contabilidad, cadena de bloques y criptomonedas.

## REFERENCIAS BIBLIOGRÁFICAS

Argañaraz, A., Mazzuchelli, A., Albanese, D. y López, M. A. (2019, 12 de diciembre). *Blockchain: un nuevo desafío para la contabilidad y auditoría (ponencia)*. Trabajo presentado en XV Simposio Regional de Investigación Contable y XXV Encuentro Nacional de Investigadores Universitarios del Área Contable. <https://backend.congresos.unlp.edu.ar/index.php/encuentro-simposio-fce/article/view/777/207>

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6-10. <https://doi.org/10.1038/s41562-017-0189-z>

De Castro, I. G. y D'Andrea, N. A. (2021). *Cadena de bloques. Su vinculación con la contabilidad (ponencia)*. Trabajo presentado en 17º Simposio de Investigación Contable, Buenos Aires. <http://sedici.unlp.edu.ar/handle/10915/132212>

Deloitte Deutschland. (2016). *Blockchain Technology. A game-changer in accounting?*  
[https://www2.deloitte.com/content/dam/Deloitte/de/Documents/Innovation/Blockchain\\_A%20game-changer%20in%20accounting.pdf](https://www2.deloitte.com/content/dam/Deloitte/de/Documents/Innovation/Blockchain_A%20game-changer%20in%20accounting.pdf)

Escobar-Pérez, J. y Cuervo-Martínez, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición*, 6, 27-36.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions*(3rd. ed.). Wiley.

George, D., & Mallery, P. (2020). *IBM SPSS Statistics 26 Step by Step* (16th ed.). Routledge.

Gwet, K. L. (2014). *Handbook of Inter-Rater Reliability*(4th ed.). Advanced Analytics.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.

Martínez Gómez, J. A., Guerrero Rodríguez, S. I. y Rey Anacona, C. A. (2012). Evaluación de la validez de constructo y la contabilidad del inventario de masculinidad y femineidad en adolescentes y adultos jóvenes colombianos. *Avances en Psicología Latinoamericana*, 30(1), 170-181.

Nunnally, J. (1967). *Psychometric Theory*. McGraw-Hill.

Pedreño, E. P. (2019). *La tecnología blockchain y su aplicación a la contabilidad (y II)*. CONT4BL3. Núm. LXX, 15-23. Asociación Profesional de Expertos Contables y Tributarios de España (AECE). [https://www.aece.es/descargararchivo\\_docnoticias\\_2007](https://www.aece.es/descargararchivo_docnoticias_2007)

Pedrosa, I., Suarez Álvarez, J. y García Cueto, E. (2013). Evidencias sobre la Validez de Contenido: Avances Teóricos y Métodos para su Estimación. *Acción Psicológica*, 10(2), 3-18.

Pérez Gil, J. A., Chacón Moscoso, S., Moreno Rodríguez, R. (2000). Validez de constructo: el uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez. *Psicothema*, 12(2), 442-446.

Scott, W. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly*, 19(3), 321-325.

Utkin, L. V. (2006). A method for processing the unreliable expert judgments about parameters of probability distributions. *European Journal of Operational Research*, 175(1), 385-398.

## CURRICULUM VITAE

### Idalia Gabriela de Castro

Magíster en Metodología de la Investigación Científica. Especialista en Contabilidad Superior y Auditoría. Profesora adjunta, con dedicación exclusiva, en la cátedra Fundamentos de Contabilidad de la Facultad de Ciencias Económicas (FCE) de la Universidad Nacional del Nordeste (UNNE). Director e integrante de proyectos de investigación acreditados por la UNNE, que se llevan a cabo en el ámbito de la FCE. Investigadora del Programa de Incentivos del Ministerio de Educación de la Nación Argentina, categoría III. Coautor de publicaciones en revistas científicas nacionales y extranjeras.

📄 <https://orcid.org/0000-0002-1359-080X>  
[idecastro@eco.unne.edu.ar](mailto:idecastro@eco.unne.edu.ar)

### Antonio Humberto Closas

Doctor en Estadística e Investigación Operativa por la Universidad Pública de Navarra (Pamplona, España). Profesor Titular de Inferencia Estadística y Director de Proyectos de Investigación en la Facultad de Ciencias Económicas de la Universidad Nacional del Nordeste. Investigador, categoría II, en el Programa de Incentivos del Ministerio de Educación de la Nación Argentina. Coautor de libros y de diversas publicaciones en revistas de impacto científico nacionales y extranjeras. Integrante de distintos comités de arbitraje y consejos editoriales de revistas científicas. Actúa como director o miembro de comisiones de seguimiento y evaluación de tesis de maestrías y doctorados.

📄 <https://orcid.org/0000-0001-7144-968X>  
[hclosas@hotmail.com](mailto:hclosas@hotmail.com)

## NOTAS

<sup>1</sup> De-Fi: abreviatura en inglés de *decentralized finance*.

<sup>2</sup> *International Business Machines Corporation* (IBM) informa desarrollos para el sector de salud, suministro de alimentos, logística, gestión de proveedores, garantías, financiamiento y credenciales de aprendizaje e identidad digital. <https://www.ibm.com/ar-es/blockchain/solutions>

<sup>3</sup> En la doble entrada los registros de las empresas carecen de conexión y, una o ambas pueden falsear datos. En *blockchain* las partes ratifican cada evento con firma digital única e intransferible y se sella criptográficamente por una *tercer entrada* de la red, convirtiéndose en prueba valiosa (Pedreño, 2019).

<sup>4</sup> Pedrosa et al. (2013) han elaborado un recorrido histórico por los métodos basados en el juicio de expertos: 1) Análisis factorial (Tucker, 1961); 2) Índice de validez de contenido (Lawshe, 1975); 3) Índice de congruencia ítem-objetivo (Rovinelli y Hambleton, 1977); 4) Índice de congruencia (Hambleton, 1980, 1984); 5) V de Aiken (Aiken, 1980); 6) Escalamiento multidimensional y análisis de clusters (Sireci y Geisienger, 1992); 7) Capacidades mínimas (Levine et al., 1997); 8) Rango interpercentil ajustado a la simetría (Fitch et al., 2001); 9) Coeficiente de validez de contenido (Hernández-Nieto, 2002); 10) Índice de validez factorial (Rubio et al., 2003); 11) Índice promedio de la desviación media (Claeys et al., 2012).

<sup>5</sup> Equivalencia cultural: barreras del idioma, significados diferentes de un constructo particular o una variedad de interpretaciones de un comportamiento observado basado en normas culturales.

<sup>6</sup> Equivalencia semántica: expresiones idiomáticas, eliminar ítems irrelevantes y modificar ítems.

<sup>7</sup> Son principios de carácter irrenunciable y aplicación universal, que rigen en cualquier actividad de investigación en ciencia, tecnología e innovación tales como: el respeto por la privacidad e intimidad de los sujetos de investigación y la confidencialidad de los datos obtenidos; la participación libre y voluntaria de las personas...y; la obligatoriedad de utilizar procesos de consentimiento informado en forma previa al reclutamiento de sujetos, entre otros criterios que se establecen en el art. 3° de la Ley 25467.

<sup>8</sup> Para acceder al documento entrevista completo: [idecastro@eco.unne.edu.ar](mailto:idecastro@eco.unne.edu.ar)