

UNIVERSIDAD NACIONAL DE EDUCACIÓN

Enrique Guzmán y Valle

Alma Máter del Magisterio Nacional

ESCUELA DE POSGRADO



Tesis

**La Validez y Confiabilidad de los Instrumentos de Recolección Predicen
la Efectividad de los Programas de Enseñanza-Aprendizaje Reportados
en las Tesis Doctorales Sustentadas en la Escuela de Posgrado de la
Universidad Nacional de Educación 2005-2015**

Presentada por

Carlos Manuel VILCHEZ ROMAN

Asesor

Guillermo Pastor MORALES ROMERO

Para optar al Grado Académico de

Doctor en Ciencias de la Educación

Lima - Perú

2021

**La Validez y Confiabilidad de los Instrumentos de Recolección Predicen
la Efectividad de los Programas de Enseñanza-Aprendizaje Reportados
en las Tesis Doctorales Sustentadas en la Escuela de Posgrado de la
Universidad Nacional de Educación 2005-2015**

A Romina, quien me acompañó durante casi 20 años y me enseñó el significado del amor y la entrega. Siempre estarás conmigo mi Romi.

A Alma Lucía, cuyo cariño y amor de hija me impulsan a seguir adelante cada día. Yo también te quiero mucho mi Almita.

A mi Gurú, quien siempre me dio la oportunidad de ser una mejor persona y me llevó de la oscuridad a la luz.

Reconocimientos

A la Universidad Nacional de Educación *Enrique Guzmán y Valle*, Alma Máter del Magisterio Nacional, por brindarnos este espacio de formación académica para seguir aportando a las ciencias de la educación.

Al personal de la Biblioteca Especializada de la Escuela de Postgrado UNE, quienes siempre me atendieron con gentileza e hicieron posible que obtuviese los datos necesarios para culminar este trabajo de investigación.

Al doctor Guillermo Pastor Morales Romero, el asesor de mi tesis doctoral, quien con sus valiosos aportes y reflexiones me permitió concluir este viaje académico que inicié hace diez años.

Tabla de Contenidos

Título	ii
Dedicatoria	iii
Reconocimiento	iv
Tabla de Contenidos	v
Lista de Tablas	viii
Lista de Figuras	ix
Resumen	x
Abstract	xi
Introducción	xii
Capítulo I. Planteamiento del Problema	
1.1 Determinación del Problema	1
1.2 Formulación del Problema: General y Específicos	7
1.3 Objetivos: General y Específicos	8
1.4 Importancia y Alcances de la Investigación	9
1.5 Limitaciones de la Investigación	10
Capítulo II. Marco Teórico	
2.1 Antecedentes de Investigación	12
2.1.1 Antecedentes Nacionales	12
2.1.2 Antecedentes Internacionales	15
2.2 Bases Teóricas	19
2.2.1 Tesis Doctorales en Educación	19
2.2.2 Diseños de Investigación en Educación	25
2.2.3 Validez y Confiabilidad de los Instrumentos de Recolección de Datos	30
2.3 Definición de Términos Básicos	37

Capítulo III. Hipótesis y Variables

3.1	Hipótesis: General y Específicas	40
3.1.1	Hipótesis General	40
3.1.2	Hipótesis Específicas	40
3.2	Variables	41
3.2.1	Definición Conceptual	41
3.2.2	Operacionalización de Variables	43

Capítulo IV. Metodología

4.1	Enfoque de Investigación	45
4.2	Tipo de Investigación	45
4.3	Diseño de Investigación	46
4.4	Método	46
4.5	Población y Muestra	47
4.6	Técnicas de Recolección de Datos	48
4.7	Tratamiento Estadístico	49

Capítulo V. Resultados

5.1	Validez y Confiabilidad del Instrumento	51
5.2	Presentación y Análisis de los Resultados	52
5.2.1	Análisis de Validez Empleados en las Tesis Doctorales	52
5.2.2	Análisis de Confiabilidad Empleados en las Tesis Doctorales	52
5.2.3	Tamaño del Efecto de los Programas de Enseñanza-Aprendizaje	56
5.2.4	Contraste de las Hipótesis de Investigación	62
5.3	Discusión	72
5.4	Limitaciones	77
	Conclusiones	79

Recomendaciones	81
Referencias	83
Apéndices	
Apéndice A. Matriz de Consistencia de la Tesis	91
Apéndice B. Instrumento de Recolección de Datos	93
Apéndice C. Ficha Técnica del Instrumento Original de Recolección	95
Apéndice D. Validación del Instrumento de Recolección de Datos	97
Apéndice E. Datos Registrados de las Tesis Doctorales	102
Apéndice F. Tamaños del Efecto e Intervalos de Confianza Originales	108
Apéndice G. Tamaños del Efecto Luego de la Transformación Logarítmica	112

Lista de Tablas

Tabla 1.	Planes de estudio del doctorado en educación de las universidades de Lima	5
Tabla 2.	Semejanzas y diferencias entre los principales enfoques de investigación	29
Tabla 3.	Vínculo entre diseños, preguntas, resultados y marcos de trabajo	30
Tabla 4.	Principales métodos para obtener evidencia de validez	32
Tabla 5.	Estudios que examinan las propiedades métricas de los instrumentos	35
Tabla 6.	Resultados de la validación según la opinión de los expertos	51
Tabla 7.	Número de ítems usados para medir cada variable	54
Tabla 8.	Datos descriptivos de las variables empleadas para calcular el tamaño del efecto	56
Tabla 9.	Los diez tamaños del efecto con los valores más altos	57
Tabla 10.	Datos descriptivos de las variables luego de la transformación logarítmica	58
Tabla 11.	Los veinte tamaños del efecto transformados con los valores más altos	59
Tabla 12.	Matriz de correlación de las variables incluidas en la hipótesis general	64
Tabla 13.	Coefficientes del modelo de regresión lineal para probar la hipótesis general	65
Tabla 14.	Coefficiente del modelo de regresión lineal para probar la 1° hipótesis específica	67
Tabla 15.	Coefficiente del modelo de regresión lineal para probar la 2° hipótesis específica	68
Tabla 16.	Matriz de correlación de las variables incluidas en la tercera hipótesis específica	70
Tabla 17.	Coefficientes del modelo de regresión para probar la 3° hipótesis específica	71
Tabla 18.	Propuesta de temas a incluirse en los cursos de Tesis	81

Lista de Figuras

Figura 1.	Histograma de la magnitud de los coeficientes de confiabilidad	53
Figura 2.	Histograma de frecuencias de la cantidad de ítems incluidos en cada escala	54
Figura 3.	Gráfico de cajas del número de ítems para medir cada variable	55
Figura 4.	Histograma de frecuencias del número de participantes para medir cada efecto	56
Figura 5.	Gráfico de cajas del número de participantes para medir cada efecto	56
Figura 6.	Gráfico de floresta de los tamaños del efecto reportados el año 2005	59
Figura 7.	Gráfico de floresta de los tamaños del efecto reportados el año 2006	60
Figura 8.	Gráfico de floresta de los tamaños del efecto reportados el año 2007	60
Figura 9.	Gráfico de floresta de los tamaños del efecto reportados el año 2008	60
Figura 10.	Gráfico de floresta de los tamaños del efecto reportados el año 2009	61
Figura 11.	Gráfico de floresta de los tamaños del efecto reportados el año 2010	61
Figura 12.	Gráfico de floresta de los tamaños del efecto reportados el año 2013	62
Figura 13.	Gráfico de floresta de los tamaños del efecto reportados el año 2014	63
Figura 14.	Matriz de correlación de las variables incluidas en la hipótesis general	64
Figura 15.	Gráfico de dispersión entre el tamaño del efecto y la validez del instrumento	66
Figura 16.	Gráfico de dispersión entre el tamaño del efecto y coeficiente de confiabilidad	68
Figura 17.	Matriz de correlación de las variables incluidas en la 3° hipótesis específica	70

Resumen

Para hacer aportes que contribuyan a expandir el conocimiento, la investigación en ciencias de la educación debe emplear un método de investigación riguroso. Ello supone usar procedimientos robustos e instrumentos de recolección de datos válidos y confiables. Sin embargo, en el caso de las tesis doctorales en educación este supuesto no siempre se cumple por la presión que tienen las instituciones académicas por elevar su producción científica. El propósito de este estudio es determinar qué relación existe entre el tamaño del efecto y la validez y confiabilidad de los instrumentos de recolección, así como con el tamaño de muestra empleada, en aquellas tesis doctorales enfocadas en los programas de enseñanza-aprendizaje orientados a mejorar el logro de aprendizaje de los estudiantes de educación superior. Para ello se examinaron 54 tesis con diseño (cuasi) experimental que se sustentaron entre los años 2005 y 2015 en la Escuela de Posgrado de la UNE. Dichas tesis reportaron un total de 153 efectos. La ficha de recojo de datos permitió registrar información referida a la validez y confiabilidad de los instrumentos empleados, así como datos sobre los grupos experimental y control (e.g., media, desviación estándar y número de participantes por cada grupo). El tamaño del efecto promedio y para cada intervención se calculó usando la d de Cohen y la g de Hedges, incluyendo su p -valor e intervalo de confianza al 95%. Para la meta-regresión se usaron como covariadas las variables tipo de validez, magnitud del coeficiente de confiabilidad y tamaño de cada grupo. Los resultados mostraron magnitudes de tamaño del efecto inusuales, que no guardan correspondencia con lo reportado en la literatura académica. Para explicar este hallazgo se planteó la hipótesis de inflación del tamaño del efecto, producto de cómo se diseñaron los instrumentos de recolección de datos. Al final se presentan algunas recomendaciones para diseñar instrumentos de recolección de datos válidos y confiables.

Palabras clave: Validez, Confiabilidad, Tamaño del efecto, Tesis doctorales.

Abstract

To contribute to the expansion of knowledge, educational research must use a rigorous research method. This goal requires using robust procedures, and valid and reliable data collection tools. However, regarding doctoral dissertations in education, this premise is not always fulfilled because academic institutions face a huge pressure for increase their scientific production. The purpose of this investigation is to determine what is the relationship between the effect size and the validity and reliability of data collection tools, as well as the used sample size, in those doctoral dissertations focused on learning teaching programs targeted to improve learning outcomes of higher education students. For that reason, 54 dissertations with a (quasi) experimental design and that were defended at the Postgraduate School of the Universidad Nacional de Educación, between the years 2005 and 2015 were analyzed. Those dissertations reported 153 interventions. The data collection form enabled the registration of information referred to the validity and reliability of the data collection tools, as well as data on experimental and control groups (e.g., mean, standard deviation, and the number of participants for each group). The average and specific effect size were computed using the Cohen's d and Hedge's g , including its p-value and confidence interval at 95%. For the meta-regression, the type of validity, the magnitude of the reliability indicator and size of each group were used as covariates. Results showed unusual magnitudes for the effect sizes, in sharp contrast with the effects reported in the academic literature. To explain this finding it was proposed the hypothesis of effect size inflation, a consequence of how were designed these data collection tools. Finally, are detailed some recommendations to design valid and reliable data collection tools.

Keywords: Validity, Reliability, Effect size, Doctoral dissertations

Introducción

La tesis está compuesta por cinco capítulos y seis apéndices. En el primer capítulo se describe el planteamiento del problema, incluyendo la formulación de los objetivos y la pregunta de investigación. En el segundo capítulo se presentan los antecedentes nacionales e internacionales de investigación y se describen los fundamentos teóricos del estudio referidos a las tesis doctorales en educación, los diseños de investigación y las propiedades métricas de los instrumentos de recolección de datos. En el tercer capítulo se detallan las hipótesis, general y específicas, junto con las definiciones conceptuales y operacionales. Por otro lado, en el cuarto capítulo se examina el enfoque analítico empleado, detallando el tipo y método de investigación, junto con las técnicas de recolección y el tratamiento estadístico de los datos. En el quinto capítulo se discuten los resultados del estudio, mostrando los hallazgos referidos a la validez y confiabilidad de los instrumentos empleados, así como los resultados propios del estudio: análisis de validez y confiabilidad en las tesis doctorales, los tamaños de efecto de las intervenciones reportadas en las tesis doctorales y el contraste de las hipótesis de investigación. Luego de ello se discute la importancia de los hallazgos reportados y se señalan las limitaciones del estudio. Posteriormente, se presentan las conclusiones y recomendaciones. Finalmente, en los apéndices se detallan la matriz de consistencia, los instrumentos de recolección empleados, la evidencia de la validación por jueces expertos, así como la base de datos usada para los análisis y los tamaños del efecto, tanto los originales como los obtenidos con la transformación logarítmica.

Capítulo I. Planteamiento del Problema

1.1. Determinación del Problema

El estatus científico de las ciencias de la educación es un tema que genera polémica porque, mientras por un lado están quienes sostienen que su estatus epistemológico está fuera de toda discusión y no hay nada que debatir, por otro lado están quienes consideran que la educación es una técnica que carece de un objeto y método de investigación propio, por lo que se limita a transmitir conocimientos, así como desarrollar habilidades y destrezas, usando y perfeccionando los medios disponibles en cada época.

Desde el ámbito académico la respuesta a este cuestionamiento se ha dado a través de la creación de los doctorados en educación, orientados a la creación de nuevos enfoques y modelos teóricos que permitan ampliar las fronteras del conocimiento en educación. Sin embargo, los programas doctorales no siempre se han adherido a los estándares de rigurosidad académica, en cuanto a la formulación de nuevos modelos conceptuales.

Precisamente, un elemento clave en la validación de los marcos teóricos de las tesis doctorales es contar con instrumentos de recolección de datos válidos y confiables, que permitan determinar la efectividad de los programas de enseñanza-aprendizaje en los diferentes niveles educativos, ya que no basta con diseñarlos, sino que es igual de importante determinar en qué medida dichos programas realmente son efectivos.

A nivel internacional, la mayoría de los estudios sobre tesis doctorales en educación se han enfocado en los temas abordados en ellas o en las características de los diseños de investigación, para lo cual se ha empleado análisis de contenido en base a categorías previamente definidas por los investigadores. (Davies et al., 2010; Durak et al., 2018; Ferreira-Vila et al., 2013; Richards et al., 2016). Por ejemplo, el profesor Gurhan Durak y su equipo de investigadores de la Universidad Balikesir de Turquía analizaron las áreas de investigación, disciplinas, diseños de investigación, instrumentos de recolección de datos,

análisis estadísticos y modelos de diseños instruccionales en 137 tesis doctorales. Por su parte, Judith Richards, Cass Dykeman y Sara Bender, procedentes de la Oregon State University y Central Washington University, examinaron 200 tesis doctorales sobre tutoría académica, sustentadas entre 1927 y 2014. Los resultados de sus hallazgos mostraron que no hubo diferencias significativas en contenido, método de investigación y diseño del estudio en las siete décadas analizadas: 1950-1959, 1960-1969, 1970-1979, 1980-1989, 1990-1999, 2000-2009 y 2010-2014. Cabe agregar que la revisión de la literatura académica mostró que solo dos estudios examinaron los resultados originales (e.g., medidas de tendencia central y de dispersión) presentados en las tesis doctorales sustentadas en universidades de Estados Unidos, enfocándose en la significancia estadística o el tamaño del efecto de los hallazgos reportados en dichas tesis (Pigott et al., 2013; Polanin et al., 2016).

Con relación a los estudios realizados en el Perú, un trabajo reciente da cuenta que la mayor cantidad de estudios doctorales por nivel educativo se concentra en la educación superior: de 554 tesis doctorales peruanas analizadas, los investigadores reportaron que 303 de ellas (55%) se realizaron con estudiantes universitarios de pre y posgrado (Díaz & Sime, 2016). A pesar del alto volumen de estudios doctorales producidos en Perú, llama la atención que muy pocos de ellos terminen publicados como investigaciones integradas en la corriente principal de la ciencia (por ejemplo, trabajos publicados en revistas indizadas en Scopus), ya que en el estudio de Díaz & Sime (2016) se señala que de las 99 tesis doctorales en educación sustentadas el año 2009, ninguna de ellas se publicó como artículo de investigación en alguna de las diez revistas académicas de educación vigentes el año 2016, aunque en Google Scholar aparecen dos tesis citadas de aquellas 99, una publicada como artículo en una revista indizada en Dialnet. Ninguna de las 99 tesis se publicó en revistas indizadas en SciELO o Redalyc.

En lo que se refiere a las tesis doctorales sustentadas en la Universidad Nacional de Educación ‘Enrique Guzmán y Valle’ (UNE, de aquí en adelante) durante 2000-2015, una búsqueda en la base de datos Scopus realizada el 27 de febrero de 2020 muestra que de las 22 investigaciones con afiliación institucional UNE (es decir, la declaración del autor de su vínculo con alguna institución durante la realización del estudio), solo cinco tuvieron como autor corresponsal (el que lidera la investigación) a un académico de la UNE, mientras que solo una tesis doctoral (Mazzi, 2003) terminó publicándose como artículo de investigación en la revista *Utopía y Praxis Latinoamericana* (Mazzi, 2017). Dado que en las publicaciones en coautoría pueden participar varios académicos como coautores, la manera de determinar quien lideró la investigación es identificar al autor corresponsal, ya que es él quien mantiene la comunicación con el editor de la revista y lleva adelante el proceso de revisión por pares.

Por otro lado, si tenemos en cuenta que el año 2021 una de las áreas con mayor déficit de doctores graduados será ciencias sociales (que incluye educación), habiéndose estimado en 4,107 el número de investigadores faltantes (Granda, 2013), queda clara la importancia de analizar la rigurosidad de las tesis doctorales, en especial de los instrumentos empleados para la recolección de datos. Por ello, examinar el método de recolección y análisis de datos empleado en la tesis doctoral (el producto académico que condensa y representa la formación recibida por el doctorado en educación) permite determinar en qué medida están articulados los planes de estudio con los conocimientos impartidos por los docentes sobre los diseños de investigación, ya que en ocasiones se observan severas limitaciones en la validación de los instrumentos o el análisis de datos (Cruz et al., 2010).

Hacia el año 2015 se apreciaba una falta de continuidad en los doctorados en educación ofrecidos en las universidades de Lima. Por ejemplo, de las siete universidades que promocionaban el doctorado en educación a través de sus sitios web (Pontificia

Universidad Católica del Perú [PUCP], Universidad Alas Peruanas [UAP], Universidad Femenina del Sagrado Corazón [UNIFE], Universidad de San Martín de Porres [USMP], Universidad Nacional de Educación [UNE], Universidad Nacional Mayor de San Marcos [UNMSM] y Universidad Nacional Federico Villarreal [UNFV]), sólo cuatro de ellas lo ofrecían de forma regular: PUCP, UNE, UNMSM y USMP.

Esta falta de continuidad contrastaba con la alta demanda de las maestrías en educación y revelaba la falta de docentes con nivel doctoral que pudieran asumir el dictado de los cursos y, sobre todo, el insuficiente interés de los investigadores que contaban con una maestría en educación en perfeccionar y mejorar sus habilidades como investigadores.

La exigencia por mejorar el nivel académico de los profesionales especializados en investigación educativa resultaba doblemente importante dado que para ingresar a los doctorados en educación ofrecidos por las universidades limeñas bastaba con que el postulante hubiera egresado de alguna maestría (porque no era requisito contar con el grado académico de magíster). Como de cada 100 egresados de las maestrías, solo 3-5 obtenía el grado académico luego de sustentar la tesis, en la práctica los ingresantes a los doctorados en educación no solían tener experiencia realizando investigaciones. Aun cuando ésta no era una limitación seria, no podía convertirse en excusa para una formación de baja calidad ya que el nivel de rigurosidad académica esperada en un doctorado debe ser el mismo, sea que hablemos de una universidad de países industrializados o países en vías de desarrollo.

Al revisar los planes de estudios de los doctorados en educación ofrecidos por algunas universidades de Lima el año 2015 parecía que eran estudios de posgrado intensivos en investigación, ya que entre el 33% y 40% de los cursos de dichos programas doctorales tenían que ver con los métodos de investigación en educación (ver tabla 1). Dado que las sumillas y sílabos de los cursos no estaban disponibles en los sitios web de los programas

doctorales, solo se pudo inferir el contenido de los cursos por el nombre de los mismos y porque los doctorados suelen seguir un enfoque similar: en el primer curso de Tesis (o Seminario de Tesis) se abordan las bases teóricas, en el segundo curso se desarrollan los métodos de recolección y análisis de datos, mientras que en el tercero se desarrolla la investigación y en el cuarto curso se consolidan los hallazgos y se redacta el informe final de la tesis. Sin embargo, en la práctica solía ocurrir lo contrario ya que al comprobar que muchos ingresantes a los doctorados no tenían experiencia previa en investigación, los profesores a cargo de los cursos tenían que replantear el contenido para empezar desde las bases de la investigación.

Tabla 1

Planes de estudio del doctorado en educación de las universidades de Lima

Plan de estudios del Doctorado en Ciencias de la Educación: UNE	
Ciclo I	Ciclo II
Historia crítica de la educación peruana	Gestión del sistema educativo
Seminario de Estadística aplicada a la educación	Seminario de corrientes filosóficas en Educación
Tesis 1	Tesis 2
Ciclo III	Ciclo IV
Currículo y didáctica universitaria	Proyecto educativo nacional
Tesis 3	Tesis 4
Plan de estudios del Doctorado en Educación: UNFV	
Ciclo I	Ciclo II
Epistemología	Currículo
Investigación educacional	Seminario taller de tesis 1 (proyecto)
Ciclo III	Ciclo IV
Estrategias didácticas modernas	Gestión educativa
Seminario taller de tesis 2 (desarrollo)	Seminario taller de tesis 3 (análisis)
Plan de estudios del Doctorado en Educación: USMP	
Ciclo I	Ciclo II
Filosofía de la Educación	Currículo
Economía de la Educación	Sociología de la Educación
Seminario taller de tesis 1	Seminario taller de tesis 2
Ciclo III	Ciclo IV
Administración de la Educación	Política educativa
Didáctica	Evaluación educativa
Seminario taller de tesis 3	Seminario taller de tesis 4
Plan de estudios del Doctorado en Educación: UAP	
Ciclo I	Ciclo II
Seminario de Historia y Filosofía universitaria	Seminario de diseño curricular

Estadística aplicada a la investigación educativa Tesis 1 Ciclo III Seminario de didáctica universitaria Seminario de problemática de la educación Tesis 3	Seminario de tecnologías en educación superior Tesis 2 Ciclo IV Seminario de gestión universitaria Seminario de evaluación y acreditación Tesis 4
Plan de estudios del Doctorado en Educación: UNIFE Ciclo I Seminario de enfoques teóricos del aprendizaje Epistemología de la Educación Seminario de tesis 1 Ciclo III Seminario de planeamiento de la Educación Problemática de la docencia en la educación Seminario de tesis 3	Ciclo II Seminario de Administración de la Educación Seminario de problemática educativa Seminario de tesis 2 Ciclo IV Evaluación de programas y proyectos educativos Seminario de tesis 4

Dicho replanteamiento del contenido de los cursos permite que los futuros doctorandos tengan un conocimiento básico del proceso de investigación, plasmando el producto de su aprendizaje en la tesis doctoral. Al revisar varias de esas tesis se observó que muchas cumplían los requerimientos formales de un estudio doctoral, pero no se había profundizado en la esencia misma de la actividad investigadora, ya que muchos de los trabajos citados eran libros de texto o páginas web, los análisis de datos correspondían a un nivel descriptivo o bivariado, o había confusión al analizar la validez y confiabilidad de los instrumentos de recolección de datos. Que una tesis doctoral muestre esas características suele ser consecuencia de un diseño de investigación que requiere mayor claridad.

A las limitaciones en el diseño de la investigación de las tesis, se agrega el hecho que no contar con un método de análisis claramente definido suele prolongar la duración del estudio, generando un aumento en los costos totales de la investigación. En ocasiones, dichos incrementos llegan a ser tan altos que llevan al investigador a dejar de recoger información u omitir algunas tareas incluidas en el diseño original del estudio. De allí que resulte importante también entender cómo se relacionan el diseño de los instrumentos de recolección de datos con el impacto de las intervenciones educativas (Johnson & Christensen, 2007).

Por ello, es importante conocer en qué medida la efectividad de los programas de enseñanza-aprendizaje (medida a través de un indicador conocido como tamaño del efecto) analizadas en las tesis doctorales sustentadas en la UNE durante el periodo 2005-2015 se relaciona con la validez y confiabilidad de los instrumentos de recolección de datos. Ese análisis permitirá ver si la efectividad reportada en las tesis doctorales es producto de los programas antes señalados, o si obedece a factores no considerados previamente.

1.2. Formulación del Problema: General y Específicos

El problema general y los problemas específicos de esta investigación son formulados en los siguientes términos:

Problema General.

¿Qué relación existe entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la Universidad Nacional de Educación (UNE) 2005-2015?

Problema Específico 1.

¿Qué relación existe entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015?

Problema Específico 2.

¿Qué relación existe entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015?

Problema Específico 3.

¿El tamaño de la muestra modera la relación entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015?

1.3. Objetivos: General y Específicos

El objetivo general y los objetivos específicos de esta investigación son formulados en los siguientes términos:

Objetivo General.

Establecer la relación que existe entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la Universidad Nacional de Educación (UNE) 2005-2015.

Objetivo Específico 1.

Establecer la relación que existe entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.

Objetivo Específico 2.

Establecer la relación que existe entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.

Objetivo Específico 3.

Determinar si el tamaño de la muestra modera la relación entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.

1.4. Importancia y Alcances de la Investigación

El estudio se enmarca dentro de una perspectiva cuantitativa ya que la revisión teórica y los métodos de análisis discutidos en este estudio corresponden a dicha perspectiva investigadora. Dentro de la tradición cuantitativa, se ha optado por el enfoque basado en evidencia, el cual se sustenta en resultados de investigaciones empíricas; es decir, emplea datos verificados y verificables, de acuerdo con los criterios propios de la perspectiva antes mencionada. Dado que en esta investigación se emplearán datos disponibles públicamente y se trabajará con una metodología explicitada paso a paso, las restricciones de presupuesto no serán un impedimento, aunque sí constituirán una limitación, como se explica más adelante.

En ese sentido, el estudio muestra que en momentos en que el financiamiento para las investigaciones se reduce, una alternativa viable es trabajar con información de libre acceso, mejor aún si está basada en instrumentos con altos niveles de validez y confiabilidad.

El estudio es importante porque cubre un vacío en cuanto a la evaluación de la producción académica de los doctorados de educación impartidos por las unidades de posgrado de las universidades peruanas, si bien es cierto nuestro centro de interés son las tesis doctorales aprobadas en la unidad de posgrado de la UNE. Dicha evaluación es también necesaria ya que a partir del segundo semestre del año 2016 todas las universidades peruanas deben implementar repositorios de tesis digitales, tal como lo señala el reglamento del Registro Nacional de Trabajos de Investigación (RENATI), emitida el 26 de julio de 2016 por la Superintendencia Nacional de Educación Superior Universitaria (SUNEDU). Por ejemplo, a diciembre de 2016, la UNE aún no publicaba sus tesis doctorales a través de su repositorio institucional, a diferencia de otras universidades públicas y privadas que ya lo hacían.

Para impulsar una cultura académica respaldada en la evaluación y una práctica profesional basada en evidencia, es necesario contar no solo con normas y procedimientos que fomenten una producción académica de alta calidad, sino que también debe existir una comunidad de investigadores decididos a mantener altos estándares de rigurosidad en su producción intelectual, tal como se observa en doctorados intensivos en investigación.

En ese sentido, el presente trabajo incluirá una sección de recomendaciones con propuestas tales como la incorporación de nuevos enfoques analíticos en el plan de estudios del doctorado en educación, a fin de asegurar que los egresados estén en capacidad de realizar aportes al conocimiento gracias a su manejo de la literatura académica y dominio de métodos avanzados para el análisis de datos.

1.5. Limitaciones de la Investigación

Tres limitaciones afectaron la recolección y análisis de datos de esta tesis doctoral.

Financieras.

Actualmente existen programas estadísticos que facilitan el cálculo de los indicadores de tamaño del efecto, como el *Comprehensive Meta-Analysis*, producto comercial elaborado por la empresa BioStats, el cual tiene una interfaz de usuario fácil de usar y documentación detallada, así como video tutoriales y un amplio programa de capacitación para usuarios. Sin embargo, la licencia anual de este programa cuesta casi US\$ 500 para usuarios académicos y US\$ 900 para usuarios comerciales. Dicho monto resulta bastante caro si tomamos en cuenta los bajos ingresos de los investigadores en educación y los gastos en que se incurre por temas de actualización profesional. Por ese motivo, para este proyecto se trabajó con el programa R, versión 3.6.0, el cual es de libre acceso y cuenta con módulos especiales para calcular el tamaño del efecto, en función del tipo de investigación empleado.

Temporales.

La revisión exhaustiva del contenido de cada tesis doctoral demandaría bastante tiempo, ya que varias de ellas tienen más de 200 o 300 páginas, en las cuales se detallan los elementos teóricos y los métodos empleados para la recolección y análisis de los datos, incluyendo la descripción detallada de los hallazgos del estudio. Como la lectura completa de cada tesis tomaría más tiempo del que el tesista tiene disponible para llevar a cabo este estudio, se empleó una ficha registro de datos (ver Apéndice B) que permitió recoger la información relevante sobre la validez y confiabilidad de los instrumentos de recolección, así como obtener los datos necesarios para calcular el tamaño del efecto.

Logísticas.

Dado que esta tesis corresponde al tipo de investigación conocido como reseña de la literatura, sub tipo revisión sistemática (Cooper, 1988, 2010), lo ideal hubiera sido tener acceso a las tesis doctorales de forma permanente a través de un repositorio de tesis digitales. Sin embargo, dado que todas las tesis doctorales del periodo 2005-2015 aún no están disponibles en el repositorio institucional de la UNE, el único momento para acopiar todos los datos requeridos fue durante el horario de atención de la biblioteca de posgrado, en la medida en que las tesis solicitadas estuviesen disponibles para su consulta en sala.

Capítulo II. Marco Teórico

2.1. Antecedentes de Investigación

2.1.1. Antecedentes Nacionales

Con relación a los estudios previos, un grupo de investigadores de la UNE (Cruz et al., 2010) analizó 30 tesis de doctorado, sustentadas durante los años 2004-2009, para determinar la relación entre los indicadores bibliométricos con el enfoque metodológico de las tesis doctorales y los descriptores temáticos. Para ello diseñaron una ficha técnica de análisis bibliométrico y una lista de cotejo para el enfoque metodológico de las tesis. La ficha incluyó nueve criterios: citas en la tesis, uso de norma internacional de citación, uso de tesis, uso de revistas, procedencia de revistas, uso de libros, idioma de las referencias bibliográficas y áreas temáticas. Por su parte, la lista de cotejo tenía 150 indicadores agrupados en 40 categorías y 15 dimensiones: título, paradigma, marco teórico, planteamiento del problema, sistema de objetivos-hipótesis-variables, tipo, instrumentos de recolección de datos, población y muestra, diseño, estadística, resultado y discusión, conclusiones, referencias bibliográficas, anexos, redacción y presentación. De acuerdo con los resultados de dicho estudio, los indicadores bibliométricos y el enfoque metodológico de las tesis doctorales en educación no se relacionan con los descriptores temáticos de las mismas (Cruz et al., 2010).

En un segundo estudio (Díaz & Sime, 2016), los investigadores revisaron 599 tesis doctorales en educación sustentadas, entre los años 2009 y 2013, en 13 programas de posgrado de universidades peruanas públicas y privadas. Para ello, analizaron 29 indicadores agrupados en cinco variables: aspectos generales, temática, metodología, referencias bibliográficas y visibilidad e impacto de la tesis. El estudio se enfocó en el diseño de investigación y la dimensión bibliométrica, ya que se midieron aspectos como la actualidad de las citas, cantidad de referencias citadas, citas a artículos académicos, citas a

tesis doctorales peruanas o trabajos de los profesores de posgrado, así como el número de tesis digitalizadas, citas recibidas por las tesis en artículos académicos o el número de tesis presentadas en eventos académicos nacionales. Ambos investigadores encontraron que en las tesis doctorales peruanas predomina el interés por los métodos de enseñanza y aprendizaje, así como los diseños curriculares en educación. En cuanto a los diseños de investigación, hay una marcada preferencia por los diseños correlacionales o cuasi experimentales. Por el lado de las citas otorgadas, prevalece lo que los autores de dicho estudio llaman “la cultura del libro”; es decir, el citar mayormente libros de texto o divulgación, a pesar que en muchas disciplinas –incluyendo educación– el conocimiento actualizado se difunde a través de las revistas académicas con revisión por pares. El limitado uso de estas fuentes de información explica, entre otras razones, la poca o escasa cantidad de citas recibidas en bases de datos como Scopus o *Web of Science* (WoS), lo que se traduce en un casi inexistente impacto académico de las tesis doctorales peruanas en educación (Díaz & Sime, 2016).

En otra investigación se analizaron 116 tesis doctorales sustentadas en el posgrado de la UNE entre los años 2000-2010 (Bustinza, 2014). La hipótesis de dicho estudio fue que la calidad científica de las tesis se relacionaba con la dinámica productiva, la pertinencia local del tema estudiado y el rigor científico esperado, a pesar de lo complejo que resulta medir variables como calidad científica o rigor esperado. Para llevar a cabo el estudio se diseñaron tres instrumentos para medir tres dimensiones de cada tesis: bibliométrica, conceptual y metodológica. En la ficha de registro de datos, además de la información del asesor, se registró el número de referencias, las citas por idiomas, el número de revistas citadas, las revistas citadas por idioma, el número de libros citados, los libros citados por idioma, así como la antigüedad y variabilidad de las citas.

Por el lado de la dimensión conceptual, el instrumento verificó si los temas de las tesis coincidieron con la clasificación de ciencias de la educación propuesta por la UNESCO. La dimensión metodológica verificó la presencia de los siguientes indicadores: diseño del estudio, hipótesis, variables, definición de términos, instrumentos de recojo de datos, validez y confiabilidad, unidades de análisis, técnicas de muestreo, tamaño de la muestra, inferencia estadística, junto con los hallazgos, conclusiones y recomendaciones. Los resultados son similares a las dos investigaciones anteriormente descritas: en las tesis doctorales en educación se observa una preferencia por citar libros en español y una omisión casi total de citas a artículos publicados en revistas académicas indizadas en Scopus o WoS, los temas abordados se concentran en el proceso de enseñanza y el aprendizaje en la educación superior; en cuanto a la dimensión metodológica, predominan los estudios cuantitativos que emplean técnicas de análisis estadístico y se constata una presencia menor de enfoques cualitativos, tales como el estudio de caso o la investigación-acción (Bustanza, 2014).

En el estudio más reciente sobre tesis doctorales peruanas en educación se analizaron 109 tesis de tres universidades públicas de la región central del Perú (Universidad Nacional del Centro del Perú, Universidad Nacional Daniel Alcides Carrión y Universidad Nacional Hermilio Valdizán), sustentadas entre los años 2010-2015 (Cencia, 2017). Las tesis se evaluaron mediante una ficha de contenido organizada en seis dimensiones: el problema objeto de investigación, la metodología de investigación utilizada, la discusión de los resultados, los aspectos formales del informe y la evaluación del impacto teórico y práctico. También se empleó una rúbrica para evaluar la discusión de los resultados presentados en cada tesis. De acuerdo con los hallazgos del estudio, las tesis lograron un nivel adecuado en los aspectos formales y la delimitación del objeto del estudio. Sin

embargo, en las demás dimensiones de evaluación la mayoría de las tesis se ubicaron en el nivel deficiente, en particular en lo referido al impacto teórico y práctico.

Si bien la mitad de los estudios descritos emplearon indicadores cuantitativos, un rasgo que caracteriza estas evaluaciones de tesis doctorales es el predominio de criterios cualitativos difíciles de replicar y escalar, ya que en la asignación de los códigos de clasificación pesa mucho el juicio particular de cada investigador. Si a este hecho se agrega el que los estudios antes descritos no reportaron el índice de confiabilidad intercodificador, resulta un verdadero desafío agregar estos resultados, de tal manera que pueda mejorar nuestra comprensión sobre la rigurosidad de las tesis doctorales peruanas en educación. Por ello, resulta pertinente explorar otros enfoques de evaluación que sean replicables y que no dependan de manera tan marcada del criterio particular de cada investigador. En ese sentido, el uso de esquemas de codificación validados en la literatura académica y el empleo de los tamaños del efecto ofrecen una aproximación complementaria al clásico abordaje cualitativo.

2.1.2. Antecedentes Internacionales

Con relación a los estudios previos a nivel internacional, en España se encontraron dos investigaciones que analizaron los métodos de investigación empleados en tesis doctorales en educación (Ferreira-Villa et al., 2013; Torralbo et al., 2004). Investigadores de las universidades de Córdoba y Granada analizaron 135 tesis doctorales sobre enseñanza de las matemáticas, sustentadas entre los años 1976 y 1998. Para tal fin emplearon una ficha de recojo de datos con 41 indicadores, organizados en cinco dimensiones: ubicación de la investigación, definición del problema, diseño y trabajo de campo, tratamiento y análisis de los datos, discusión de los resultados. Con relación al tipo de validez realizada, 44.4% empleó la validez de contenido, 17.8% validez de criterio y 17% trabajó con validez de constructo. En cuanto al tipo de confiabilidad, 25.2% analizó la consistencia interna, 4.4%

la estabilidad, 2.2% la equivalencia y el 1.5% la confiabilidad intercodificador. En lo que se refiere al tipo de diseño general, 36.3% utilizó el diseño pro-experimental, 18.5% el descriptivo, 15.6% estudio de caso/diseño etnográfico y 6.7% analítico-correlacional (Torralbo et al., 2004).

En la segunda investigación se examinaron 90 tesis doctorales sustentadas entre los años 2001-2012 y registradas en la base de datos Teseo. A diferencia del estudio anterior, se empleó un esquema de codificación con tres dimensiones (productividad del área, características generales y metodológicas y temas prioritarios estudiados) en dos áreas (asesoramiento y orientación educativas y orientación vocacional y profesional). El enfoque metodológico no fue similar en las dos áreas, por ejemplo, en asesoramiento y orientación los métodos cuantitativos representaron el 28% de las tesis, mientras los cualitativos el 23%, por su parte los métodos mixtos alcanzaron 9%. Por otro lado, en orientación vocacional y profesional, los métodos cuantitativos representaron 33%, los cualitativos 20% y los métodos mixtos subieron a 11% (Ferreira-Vila et al., 2013).

En Turquía se llevó a cabo un estudio en el que se examinaron 137 tesis doctorales en tecnología educativa para determinar cuáles fueron las disciplinas, líneas de investigación, herramientas de recolección de datos, modelos de diseños instruccionales y foco de interés, entre otras variables (Durak et al., 2018). Para ello se empleó un análisis de contenido con una ficha de recojo *ad-hoc*. Los autores de dicho estudio reportaron una κ de Cohen = 0.890, que representa una muy buena confiabilidad intercodificador. Con relación a los métodos empleados, por el lado cualitativo predominaron el estudio de caso y la investigación-acción, mientras que por el lado cuantitativo destacaron el diseño experimental, correlacional y los trabajos de campo basados en encuestas. En cuanto a las técnicas de análisis de datos, desde la aproximación cualitativa se emplearon principalmente el análisis de contenido (46.4%) y el análisis temático (53.6%). Desde el

enfoque cuantitativo, en la mayoría de las tesis doctorales se trabajó con técnicas paramétricas (87%), con un predominio de: análisis de varianza, también conocido como ANOVA ($n = 68$), t de student ($n = 49$), análisis de correlación ($n = 45$), análisis factorial ($n = 34$), modelamiento con ecuaciones estructurales ($n = 28$) y análisis de regresión ($n = 10$). Por el lado de las técnicas no paramétricas (13%), las dos más usadas fueron la U de Mann-Whitney ($n = 22$) y la K de Kruskal-Wallis ($n = 11$).

Los demás estudios que analizan tesis doctorales en educación se realizaron en universidades de Estados Unidos. En el primero se examinaron 308 tesis de doctorado y maestría en educación a distancia, remitidas a *ProQuest Dissertation Theses Database* (PQDT) en 1998, 2002 y 2007. Se diseñó una ficha de recojo de datos y se empleó el análisis de contenido. Con relación a los temas de investigación, predominaron tópicos referidos a los estudiantes, profesores y la evaluación de los métodos. En cuanto a las técnicas de análisis de datos, la mayoría de tesis empleó estadística descriptiva y ANOVA (Davies et al., 2010).

En el segundo estudio en universidades estadounidenses se empleó el análisis de contenido para examinar 203 tesis doctorales en consejería educativa, sustentadas entre los años 1947-2014 en la Oregon State University (Richards et al., 2016). Se empleó una ficha de registro de datos que alcanzó una alta confiabilidad intercodificador, con un alfa de Krippendorff = 0.94. La mayoría de las tesis se enfocó en consejería escolar ($n = 44$), multiculturalismo y consejería multicultural ($n = 28$), consejería universitaria ($n = 19$) y desarrollo humano ($n = 10$). Por el lado de los diseños de investigación destacaron los estudios transversales y correlacionales ($n = 65$), los estudios de caso ($n = 25$) y los experimentos aleatorizados o diseños cuasi experimentales ($n = 17$, respectivamente).

Los últimos dos estudios estadounidenses tuvieron un abordaje distinto, ya que se enfocaron en los resultados presentados en las tesis doctorales (e.g., el puntaje promedio,

la desviación estándar y el número de participantes). En ese sentido, son las dos investigaciones que más se aproximan al enfoque analítico de la presente tesis doctoral. En el primero de dichos estudios, los autores examinaron 621 tesis doctorales sustentadas el año 2005 en 96 universidades estadounidenses muy intensivas en investigación. De ese grupo, 79 fueron publicadas posteriormente en revistas académicas. Al comparar qué porcentaje de los resultados reportados en las tesis también aparecieron en la versión artículo, en primer lugar, se encontró que en las tesis doctorales el 46% de los hallazgos alcanzó significancia estadística. Ello explica que, en promedio, la versión publicada de la tesis doctoral incluyó solo la mitad de los resultados comunicados originalmente. En términos de probabilidades, los hallazgos no significativos tuvieron 30% menos probabilidades de aparecer en la versión publicada, mientras que los resultados significativos tuvieron 22% menos probabilidades de ser omitidos de la publicación al compararse con los no significativos (Pigott et al., 2013).

En un segundo estudio, investigadores de la Vanderbilt University examinaron los tamaños de efecto de 81 meta-análisis que incluían 6,392 estudios primarios: 27 meta-análisis publicados en *Review of Educational Research* y 54 en *Psychological Bulletin* (Polanin et al., 2016). La meta-revisión incluyó consideró tipos de publicación, entre los que destacaron los artículos de revistas (68.5%), reportes y *datasets* no publicados (31.5%) y tesis doctorales (18%). En primer lugar, se encontró que los tamaños de efecto reportados en artículos de revistas académicas eran mucho más grandes que los reportados en informes técnicos y tesis doctorales. En segundo lugar, al analizar los factores asociados con la publicación de los tamaños de efecto (e.g., tipo de tamaño de efecto [diferencia estandarizada de media o coeficiente de correlación], solo tesis doctorales como fuente de literatura gris, reportes no publicados, estudios financiados y fecha de publicación), la meta-regresión mostró que el haber incluido tesis doctorales como única

fuentes de literatura gris redujo el número de tamaños de efecto reportados en el artículo de investigación ($\beta = -0.77$, $p < 0.05$ para el modelo univariado y $\beta = -0.83$, $p < 0.05$ para el modelo completo).

En resumen, la mayoría de los estudios que examinaron las tesis doctorales emplearon el análisis de contenido como estrategia analítica ($n = 9$), de los cuales solo dos de ellos reportaron el índice de confiabilidad intercodificador: el alfa de Krippendorff y el kappa de Cohen. Dada la gran cantidad de meta-análisis publicados en revistas de investigación educativa y del conocido sesgo de dichas revistas en publicar principalmente resultados con significancia estadística, resultó inesperado que solo dos estudios se enfocaran en los tamaños de efecto reportados originalmente en las tesis doctorales. Por esa razón, en esta investigación examinamos si los tamaños de efecto reportados en las tesis doctorales sustentadas varían según la validez y confiabilidad de los instrumentos o el tamaño de la muestra.

2.2. Bases Teóricas

2.2.1. Tesis Doctorales en Educación

Así como ocurre en otras especialidades, el posgrado en educación puede orientarse hacia la profesionalización o la generación de conocimientos e innovaciones educativas. En el primer caso, el énfasis suele estar en la adquisición y dominio de nuevos enfoques pedagógicos, mientras que en el segundo el interés está en lograr que el estudiante de posgrado adquiera las competencias propias de un investigador capaz de producir nuevos conocimientos, modelos teóricos y enfoques innovadores. A nivel doctoral, esto se traduce en la obtención del grado Ed.D. (*Doctor of Education*) de carácter profesionalizante, o del Ph.D. (*Doctor of Philosophy* in Education), que supone un dominio de diseños de investigación unido a la capacidad de generar conocimientos e innovación en la práctica profesional.

En el campo de las publicaciones académicas, los artículos de investigación y las tesis doctorales son los productos más adecuados para determinar el nivel de competencia científica alcanzado por el investigador que llevó a cabo el estudio. No obstante, existen importantes diferencias entre las tesis de pregrado y posgrado, mientras que en el pregrado basta con que el investigador muestre capacidad para comprender los conceptos básicos del proceso de investigación, en el posgrado el tesista deberá además dejar constancia de un manejo experto de la literatura académica y de los métodos de recolección y análisis de datos. En cuanto al nivel de exigencia, las tesis de maestría deben mostrar que el investigador domina las técnicas de recolección y análisis de datos, mientras que las tesis doctorales –además de los requerimientos previos– deben mostrar que el tesista es capaz de realizar aportes al conocimiento o que puede innovar los enfoques y técnicas de análisis empleados en este nivel de investigación (Anderson, 2002; Capraro & Thompson, 2008; Eisenhart & DeHaan, 2005; Schulman et al., 2006).

La elaboración de la tesis doctoral es el resultado de un largo proceso de adquisición y dominio de las competencias necesarias para llevar a cabo investigaciones de alto nivel. Dado que en la investigación educativa existen diferentes enfoques epistemológicos, la realización de estudios rigurosos, conducentes a la generación de nuevos conocimientos e innovaciones educativas puede tomar diferentes rumbos, según la orientación teórica y metodológica de cada investigador. Más allá del problema analizado o del diseño de investigación empleado, todas las tesis doctorales buscan llenar un vacío, explorar un campo que no había sido abordado con mayor profundidad o evaluar la efectividad de un nuevo enfoque didáctico o programa curricular; en esa medida, tratan de generar nuevos conocimientos.

Una vez identificados los propósitos de la tesis doctoral, viene la discusión sobre cuáles son las competencias necesarias para elaborar tesis doctorales. Para que los egresados de

los doctorados en educación estén en capacidad de elaborar una tesis, al final de su formación deben haber desarrollado la habilidad de traducir un problema de investigación en su dimensión empírica a través de la formulación de hipótesis, identificación de variables y construcción de las definiciones operativas. Ello le permitirá emplear las estrategias de recolección y análisis de datos más apropiadas para probar sus hipótesis de investigación.

Los especialistas que se han ocupado del tema enfatizan las diferentes competencias que deben desarrollar los estudiantes de doctorados en educación (Anderson, 2002; Eisenhart & DeHaan, 2005; Leech & Goodwin, 2008; Schulman et al., 2006). Por ejemplo, Margaret Eisenhart y Robert DeHaan se adhieren a los principios que definen la investigación científica, formulados por el *National Research Council* de Estados Unidos:

1. Plantear preguntas importantes que puedan investigarse a través de métodos empíricos.
2. Vincular la investigación con la teoría relevante (modelos teóricos, conceptos, enfoques).
3. Formular una cadena de razonamiento explícita, desde el diseño hasta los resultados.
4. Reproducir el proceso de recolección y análisis, así como generalizar los resultados.
5. Hacer pública la investigación para fomentar el escrutinio y la crítica profesional.

Además, ambos autores desarrollan los lineamientos de lo que sería un programa doctoral de investigación educativa con orientación científica, el cual debería tener cuatro elementos básicos:

- a. *Cursos centrales*: Cursos en temas tradicionales para los programas de educación, así como cursos de áreas interdisciplinarias como neurociencia, sociología de la ciencia, etc.
- b. *Experiencia investigadora*: Los estudiantes de doctorado deben involucrarse con todas las etapas de la investigación, a través de una estrecha supervisión y apoyo de un docente investigador; por otra parte, tener la oportunidad para profundizar en una línea de

investigación por sí mismos –bajo la supervisión por parte un profesor del doctorado– y luego comunicar los resultados de la investigación realizada, a través de una tesis, un reporte técnico o un artículo publicado en una revista académica.

c. *Experiencia profesional*: Los estudiantes doctorales deben adquirir experiencia profesional, enseñando en aulas, desarrollando proyectos o formulando políticas educativas.

d. *Colaboración interdisciplinaria*: Además de los cursos y seminarios interdisciplinarios, deben crearse espacios para que los estudiantes de doctorado participen de redes con otros estudiantes e investigadores de otros departamentos vinculados con la educación.

Por su parte, Leech y Goodwin (2008) señalan que en los programas doctorales en Estados Unidos hay una gran variación en los requerimientos para los cursos de metodología de la investigación, ya que si bien es cierto el promedio es de tres cursos durante todo el doctorado, el rango de cursos va de uno a nueve. Dado que el propósito de los cursos de metodología es enseñar a los estudiantes las habilidades y conocimientos sobre la investigación, tres cursos parece ser un requisito mínimo para alcanzar esa meta. Un hecho preocupante, reportado por estos autores, es el bajo porcentaje (23%) de programas doctorales que incluyen un curso de medición, debido a que las destrezas operativas de medición y análisis son fundamentales para diseñar e implementar estudios robustos.

Como puede verse, cada investigador enfatizará aquellas competencias o habilidades de investigación que considere más relevantes, lo cual podría llevar a una lista bastante extensa de las competencias requeridas para elaborar tesis doctorales. Con el fin de llegar a un consenso sobre este polémico tema, el año 2006 la *American Educational Research Association* (AERA) publicó los estándares para presentar resultados de investigación en

ciencias sociales, un documento guía que señala lo que deben incluir los trabajos de investigación educativa. Si bien no detalla una lista de competencias que los estudiantes de doctorado deben haber adquirido para elaborar tesis, sí precisa una serie de elementos y criterios que suponen el manejo avanzado de habilidades de investigación, incluso diferenciando entre estudios cuantitativos y cualitativos. Por ejemplo, en el rubro análisis de datos, las directrices de la AERA (2006, pp. 36-37) señalan:

Con métodos cuantitativos:

Cuando se emplean métodos cuantitativos, los análisis estadísticos son llevados a cabo, informados y luego se discuten los hallazgos obtenidos. Los resultados de los análisis estadísticos típicamente incluyen: a) un índice cuantitativo de la relación entre las variables (o una magnitud) y b) un índice de su incertidumbre.

Mientras que la prueba de significancia estadística tiene una larga historia y es útil en la investigación educativa, los estadísticos han advertido durante mucho tiempo contra el exceso de confianza en la prueba de significancia, que puede llevar a excluir otros métodos de interpretación de los análisis estadísticos. Las pruebas de significancia estadística combinan tanto la magnitud de las relaciones (o estimados) y su incertidumbre dentro del mismo índice cuantitativo. La interpretación de los análisis estadísticos mejora cuando se presentan de forma separada la magnitud de las relaciones (e.g., los tamaños del efecto) y el nivel de incertidumbre. Si bien es cierto pueden realizarse muchos análisis estadísticos dentro de un estudio, usualmente solo una parte de ellos es crítica para los resultados y las interpretaciones. Es importante presentar los resultados de análisis que sean vitales para la interpretación de los hallazgos de tal manera que incluyan la magnitud, así como la significancia estadística de dichos resultados. Los índices cuantitativos de magnitud del efecto (índices de tamaño del efecto) son una manera útil de hacerlo (...)

Con los métodos cualitativos:

Cuando se emplean métodos cualitativos, los análisis tienen lugar durante y después de la recolección de datos. Los primeros análisis pueden contribuir con el posterior recojo de información al, por ejemplo, identificar categorías de hechos, acciones o personas, para un análisis posterior dentro del estudio en curso o en una próxima investigación. Tal como se señaló en la sección Medición y Clasificación, durante las primeras etapas del análisis, los investigadores pueden desarrollar formas de segmentar los datos (e.g., por persona; por acción, actividad, hecho o narrativa; por período) y grupos de categorías importantes o códigos que permitan organizar los segmentos de datos.

Dichas clasificaciones ayudan al investigador a identificar patrones dentro de los datos. Los patrones se refieren a configuraciones de hechos u otras observaciones que ocurren de forma repetida y consistente dentro de un ordenamiento característico. En ocasiones, el análisis busca ofrecer interpretaciones completas y profundas de un texto o registro particular, más que la descripción de patrones de materiales empíricos extendidos o formados por varios componentes. Cualquiera sea el abordaje inicial al análisis de datos, es importante que los investigadores describan de forma completa los procesos empleados de tal forma que otros puedan seguir su lógica de exploración.

Además de demostrar un manejo adecuado de competencias metodológicas, las tesis doctorales deben abordar temas relevantes para el sistema educativo en el cual están insertas las diferentes escuelas o unidades de posgrado. Tal como lo señala José Padrón (2006), no basta con que las tesis doctorales se publiquen como artículos científicos, lo importante es que aporten a la comunidad educativa. Por ejemplo, en un país como Perú, donde la mayoría de estudiantes en edad escolar no comprende lo que lee ni puede resolver operaciones básicas de cálculo aritmético, una tesis sobre métodos de enseñanza efectivos

para mejorar el rendimiento en comprensión lectora o matemática es más relevante que una tesis doctoral sobre la aceptación y uso de teléfonos inteligentes (*smartphones*) por los estudiantes.

2.2.2. Diseños de Investigación

Antes de revisar los principales diseños de investigación en educación es importante comprender los fundamentos filosóficos que sustentan los diferentes estudios en educación. Por ello es necesario describir los enfoques epistemológicos clásicos en los cuales se inscriben los planteamientos específicos para los planes de trabajo que hacen posible la recolección y análisis de datos. En el terreno educativo, la discusión epistemológica ha girado en torno a la vigencia del enfoque cuantitativo (cuyo exponente más destacado en la actualidad es el neo positivismo) y el cualitativo (donde destacan el enfoque socio crítico y la hermenéutica), los cuales son descritos brevemente a continuación.

Neopositivismo.

El planteamiento positivista original buscaba convertir el modelo de las ciencias físicas y naturales en el modelo a seguir por las demás ciencias y disciplinas con aspiraciones científicas, tal como ocurre con las disciplinas sociales, las humanidades y educación. Dicho planteamiento fue criticado por no considerar la naturaleza compleja y relacional de la dinámica social, amplificada por la alta variación del comportamiento humano, lo cual llevó a pensar que la conducta humana no podía reducirse a modelos matemáticos. Efectivamente, dichas críticas al positivismo eran válidas a inicios del siglo XX, debido al estado incipiente de los diseños de investigación empírica, tal como hoy los conocemos. Sin embargo, en las últimas décadas las ciencias biológicas, físicas, químicas y médicas evolucionaron y los métodos de recolección y análisis de datos lograron incorporar elementos contextuales, la dinámica social y la complejidad de la conducta humana, dando

lugar a campos multidisciplinarios, como la epidemiología o la neurociencia, por mencionar algunos.

Actualmente, el enfoque neopositivista (o cuantitativo) trabaja con modelos conceptuales avanzados (e.g., modelos estructurales longitudinales, modelos multinivel o modelos no lineales) que no fueron considerados en el planteamiento positivista original. No obstante, no se trata de modelos finales o acabados: el conocimiento científico sigue en expansión y ello es posible en la medida en que las nuevas teorías y las ya existentes son sometidas a prueba a través de nuevos estudios. En resumen, en el estado actual del conocimiento científico, el neopositivismo también tiene en cuenta la complejidad, la dinámica relacional, los elementos de contexto, dentro de una perspectiva interdisciplinaria que no supone el abandono de la tradición cuantitativa, sino un fortalecimiento del mismo, siempre que la recolección y análisis cumplan con requisitos mínimos de rigurosidad.

Sociocrítica.

Si bien es cierto el enfoque neopositivista responde las preguntas sobre qué tipo de conocimiento científico producir y cómo generarlo, no aborda de forma explícita la pregunta del porqué y para qué; léase, ¿por qué es importante hacer ciencia? Este cuestionamiento del quehacer científico es abordado por el enfoque sociocrítico, el cual plantea que los resultados de la investigación científica deben ser aplicados para combatir la injusticia e inequidad del actual sistema económico y social. En ese sentido, este enfoque destaca la importancia de hacer explícitas las agendas de investigación que establecen qué temas serán investigados y cuáles serán dejados de lado o abordados en estudios futuros.

Aun cuando este enfoque ha llevado a la formulación de concepciones diferentes sobre cómo analizar los hechos estudiados, no ha desarrollado planteamientos metodológicos que reemplacen –de forma exitosa– los diseños de investigación elaborados desde la

perspectiva neopositivista, en todo caso, se trata de un enfoque complementario. Un aporte valioso de la postura sociocrítica es cuestionar la condición *a priori* de los hechos sociales, léase la tendencia a percibirlos como si fueran una consecuencia natural de la evolución de los sistemas económicos. Desde este enfoque el investigador se preguntará porqué hay enfermedades –como el SIDA, por ejemplo– cuya asignación de fondos para el tratamiento y la prevención no guarda relación con la tasa de mortalidad que sí se observa en otras enfermedades, como la tuberculosis. Esto llevará a que el investigador analice porqué ciertas enfermedades se ponen de moda y porque otras pierden interés público. En el campo educativo, el enfoque sociocrítico preguntará porqué los programas de reforma educativa se implementan en función de las modas académicas o los intereses de los grupos de presión, y no por su efectividad para mejorar el rendimiento académico de los estudiantes.

Hermenéutica.

El enfoque hermenéutico también es conocido como enfoque humanístico o fenomenológico. Dado que este enfoque se aproxima al objeto de estudio de manera integral, completa, la pregunta fundamental que busca responder la hermenéutica es: ¿esta interpretación de los hechos es suficiente para comprender el fenómeno estudiado? Dicha visión integradora intenta explicar las distintas facetas del hecho estudiado a fin de determinar si las interpretaciones permiten comprender la complejidad y versatilidad de la realidad analizada, se trate de acciones individuales o de la vida en sociedad.

Al igual que en el enfoque sociocrítico, la hermenéutica no se ha enfocado en desarrollar métodos de recolección y análisis de datos que reemplacen los postulados del neopositivismo, sino en ampliar los criterios para interpretar los resultados obtenidos, es decir, en discutir el significado e importancia de los hallazgos del quehacer científico. Ello ha llevado a que la hermenéutica tenga importantes coincidencias con la a) semiótica, para

entender qué forma los dispositivos lingüísticos contribuyen al conocimiento humano, b) etnografía, ya que se rescatan los aspectos subjetivos del investigador, por ello, entre los métodos preferidos por la hermenéutica destacan la entrevista en profundidad, el estudio de caso, la investigación-acción y la observación participante.

Dado que estas concepciones epistemológicas representan visiones generales sobre cómo ocurren y se registran los hechos materiales y sociales, se requieren métodos y procedimientos específicos para llevar a cabo la recolección y análisis de datos. En otras palabras, se requiere de planes de trabajo, planes operativos sobre cómo llevar a cabo dichas actividades; es decir, se requieren diseños de investigación específicos en educación.

Existen diferentes clasificaciones de los diseños de investigación en educación. A nivel general suele hablarse de diseños cuantitativos, cualitativos o diseños mixtos (Johnson & Christensen, 2007). Sin embargo, dos criterios útiles para clasificar los diseños son el propósito del estudio y el tipo de investigación. De acuerdo con el propósito de la investigación educativa existen cuatro diseños principales: exploratorio, descriptivo, evaluativo y explicativo, cada uno de los cuales forma parte de una tradición epistemológica. Por ejemplo, los diseños exploratorios suelen inscribirse en la perspectiva cualitativa o interpretativa, mientras que los diseños evaluativos o explicativos corresponden al enfoque neopositivista, mientras que el diseño descriptivo puede abordarse desde cualquiera de los ambos enfoques epistemológicos. Según el tipo de investigación, existen los diseños históricos, documentales, de encuesta, correlacionales y cuasi experimentales. Los históricos centrados en el análisis de hechos pasados, los documentales en la revisión de diversos materiales, mientras que los correlacionales usan la correlación y el análisis de regresión.

Especialistas en investigación educativa como Johnson y Christensen (2007) o Butin (2010) han elaborado cuadros comparativos que ayudan a entender las semejanzas y diferencias entre los diseños de investigación antes mencionados (ver tablas 2 y 3). En ese sentido, existe cierto consenso con relación al propósito de la investigación, los supuestos analíticos y las técnicas empleadas para la recolección y análisis de los datos.

Tabla 2

Semejanzas y diferencias entre los principales enfoques de investigación

	Investigación cuantitativa	Investigación mixta	Investigación cualitativa
Método científico	Deductivo o “de arriba abajo”. El investigador prueba las hipótesis y teorías con los datos.	Deductivo o inductivo.	Inductivo o “de abajo arriba”. El investigador genera nuevas teorías e hipótesis.
Propósito de la investigación	Descripción, explicación y predicción.	Propósitos múltiples.	Descripción, exploración y descubrimiento.
Naturaleza de la observación	Trata de estudiar la conducta en situaciones controladas.	Estudia la conducta en varios contextos.	Estudia la conducta humana en ambientes naturales y en su contexto.
Forma de recolección de los datos	Recolección de datos cuantitativos, basada en la medición precisa usando instrumentos de recolección de datos estructurados y validados.	Diferentes formas.	Recolección de datos cualitativos (e.g., entrevistas a profundidad, observación participante, notas de campo y preguntas abiertas).
Naturaleza de los datos	VARIABLES.	Mezcla de variables e imágenes.	Palabras, imágenes y categorías.
Análisis de los datos	Identificación de relaciones estadísticas.	Cuantitativa y cualitativa.	Búsqueda de patrones, temas y características generales.
Resultados	Hallazgos generalizables.	Hallazgos corroborados que son generalizables.	Hallazgos de naturaleza particular. Representación del punto de vista interno.
Forma de presentación de los resultados	Informe estadístico (e.g., incluye correlaciones, coeficientes de regresión, comparación de medias y significancia estadística).	Ecléctico y pragmático.	Informe narrativo con una descripción contextual y citas directas de las respuestas dadas por los participantes de la investigación.

Fuente: Johnson, & Christensen (2007)

Tabla 3

Vínculo entre diseños, preguntas, resultados y marcos de trabajo de las investigaciones

Propósito de la investigación	Pregunta de Investigación	Resultado clave	Marco de trabajo de orientación	Enfoque epistemológico
Exploración	Qué, cómo, por qué	Reflexión Caso o narración	Hipótesis tentativa o diseño inicial	Hermenéutica Sociocrítica
Descripción	¿Qué o cómo?	Reflexión Caso o narración	Categorías	Hermenéutica Sociocrítica
Evaluación	¿Qué tan bien trabaja?	Caso o narración “Mejor práctica”	Análisis de brecha o de necesidades	Neopositivismo
Explicación	¿Por qué?	“Mejor práctica”	Hipótesis formal de investigación	Neopositivismo

Fuente: Butin (2010)

A continuación, se presentan ejemplos del tipo de estudios que se pueden realizar a partir de los diseños de investigación antes descritos:

Investigación con un diseño exploratorio

Percepción de los profesores sobre el aprendizaje de una segunda lengua.

Investigación con un diseño descriptivo

Impacto de la aplicación de un nuevo libro de texto en los estudiantes de sexto de primaria.

Investigación con un diseño explicativo

Factores asociados con el rendimiento en matemáticas de los estudiantes de quinto de secundaria de colegios públicos.

Establecido el marco general de los diseños de investigación en educación, a continuación, se analizará un aspecto clave de los diseños de investigación: validez y confiabilidad de los instrumentos empleados para recolectar los datos.

2.2.3. Validez y Confiabilidad de los Instrumentos de Recolección de Datos

Con relación a las propiedades métricas de los instrumentos de recolección de datos, los investigadores deben demostrar que dichas herramientas tienen altos niveles de validez y confiabilidad, porque ello garantiza que son instrumentos sólidos para respaldar las

hipótesis de cada estudio. No obstante, debe señalarse que cuando se habla de las propiedades métricas lo que en realidad se mide es la validez y confiabilidad de los puntajes obtenidos en cada aplicación del instrumento de colecta de datos. Por esa razón también se dice que la validez y confiabilidad no son atributos de los instrumentos sino propiedades de los puntajes obtenidos. Sin embargo, en aras de facilitar la lectura de esta tesis, de aquí en adelante se hablará de validez y confiabilidad de los instrumentos de recolección, más que de los puntajes.

La validez se refiere al grado en que un instrumento mide lo que en teoría debería medir, mientras que la confiabilidad alude a la precisión en la medición. En cuanto a la manera cómo se integran, se ha señalado que un instrumento puede ser confiable pero no válido, pero para ser válido también tiene que ser confiable. Si bien existen diferentes clasificaciones sobre los tipos de validez, los tres documentos normativos más recientes sobre las propiedades métricas de los instrumentos de recolección de datos (*American Educational Research Association [AERA]*, *American Psychological Association [APA]* & *National Council on Measurement in Education [NCME]*, 1985, 1999 y 2014) coinciden en señalar los tres tipos de validez vigentes: validez de contenido, validez de criterio y validez de constructo. Cabe señalar que existen dos versiones previas de los estándares de medición para pruebas educativas y psicológicas (las de los años 1966 y 1974), pero a fin de cumplir las exigencias de actualidad del presente doctorado se optó por citar los estándares de medición educativa más recientes de las tres entidades rectoras: AERA, APA y NCME.

La validez de contenido se determina a partir del nivel de acuerdo o desacuerdo entre un grupo de expertos sobre la pertinencia de un ítem determinado para medir una dimensión o área previamente definida. El coeficiente V de Aiken y el kappa de Cohen son indicadores que miden el nivel de acuerdo entre los jueces que examinan el contenido de

dichos instrumentos. La validez de criterio se subdivide en dos tipos: validez predictiva y validez concurrente. Para calcular el nivel de validez correspondiente se emplean los análisis de correlación y regresión. Los documentos normativos señalan que la validez de constructo (que incluye la validez convergente y discriminante) es la forma más completa de validez de un instrumento de recolección de datos, por lo que los investigadores deben verificar siempre la validez de constructo de sus instrumentos de recolección (ver Tabla 4).

Tabla 4

Principales métodos para obtener evidencia de validez

Tipo de evidencia	Procedimientos
Evidencia basada en contenido	Estudiar el constructo que se va a medir, examinar el contenido del instrumento y decidir hasta qué punto el contenido del instrumento representa al constructo de forma apropiada. Esta validación suele ser llevada a cabo por expertos.
Evidencia basada en la estructura interna	Primero, determinar cuántas dimensiones o constructos mide el instrumento, para ello se utiliza la técnica estadística llamada <i>análisis factorial</i> . Luego, examinar la <i>homogeneidad</i> de los ítems (para toda la prueba o para sub escalas si es que el instrumento mide más de una dimensión).
Evidencia basada en la relación con otras variables	Relacionar los puntajes de la prueba con un criterio conocido al recolectar evidencia <i>concurrente</i> o <i>predictiva</i> . Correlacionar los puntajes del instrumento de recolección con medidas similares del mismo constructo y medidas de constructos diferentes para obtener evidencia <i>convergente</i> y <i>discriminante</i> . Determinar si los grupos difieren en la prueba de una manera que pueda ser anticipada (e.g., para una escala que mide actitudes liberales, determinar cuánto se diferencian quienes están en contra del aborto de las personas que están a favor del mismo).

Fuente: Johnson, & Christensen (2007)

Para calcular la validez de un instrumento de recolección de datos suelen utilizarse técnicas de reducción de datos como el análisis de componentes principales o enfoques analíticos para la identificación de dimensiones latentes como el análisis factorial confirmatorio. Dada la complejidad de esas herramientas de análisis, ellas suelen estar disponibles en programas de análisis como SPSS o Stata.

En lo que se refiere a la confiabilidad, es una propiedad métrica para determinar el nivel de consistencia o estabilidad en las mediciones. Se clasifica en consistencia externa (con criterios como estabilidad, equivalencia o intercodificador) e interna (con criterios como división por mitades y coherencia entre los ítems). Para la confiabilidad intercodificador, los coeficientes más conocidos son la phi de Scott, el kappa de Cohen, el kappa de Fleiss y el alfa de Krippendorff. Si bien todos ellos emplean un método de cálculo diferente, las magnitudes de dichos coeficientes se expresan en rangos que van desde 0.00 (desacuerdo completo) hasta 1.00 (acuerdo completo). La excepción a este rango es el coeficiente kappa de Cohen que emplea un rango de valores diferente. Cuando la magnitud del coeficiente de confiabilidad intercodificador es superior a 0.70 se considera que el nivel de acuerdo de los evaluadores es confiable y si la magnitud es superior a 0.90 se dice que es altamente confiable.

Desde la perspectiva de la teoría clásica de los tests, los coeficientes más empleados para medir la coherencia entre los ítems son el Kuder-Richardson (para variables dicotómicas) y el alfa de Cronbach (para variables politómicas). De forma similar a la medición del coeficiente de confiabilidad intercodificador, magnitudes superiores a 0.70 para los coeficientes Kuder-Richardson y alfa de Cronbach indican que el ítem o escala evaluados son confiables, mientras que los valores superiores a 0.90 dan cuenta de ítems o escalas de medición altamente confiables. Por el lado de la teoría de respuesta al ítem (TRI), lo que se busca es determinar la probabilidad de responder a un ítem o reactivo de forma específica. Para ello, se emplean modelos logísticos que analizan la relación entre el comportamiento de una persona frente al ítem y el nivel de complejidad de dicho ítem (como un indicador de la probabilidad de responder de forma incorrecta), lo que permite obtener la curva característica del ítem. En ese sentido, los modelos TRI son probabilísticos y no lineales.

En cuanto al cálculo de los coeficientes de confiabilidad, programas como R, SPSS o Stata incluyen módulos especializados para calcular la magnitud de dichos coeficientes.

Para identificar los estudios publicados en revistas de educación que analizaron la validez y confiabilidad de los instrumentos de recolección de datos usados en las tesis doctorales, se emplearon dos estrategias de búsqueda en las bases de datos Scopus y WoS. En Scopus se hizo una búsqueda en los campos resumen (*Abstract*) y título de la revista (*Source title*), mientras que en WoS se combinó la búsqueda en el campo Tema con la categoría temática educación e investigación educacional (*Education & Educational Research*).

Scopus:

(ABS (dissertation AND {reliability | validity}) AND SRCTITLE (education | teaching))

WoS

TEMA: (dissertation {reliability | validity}) Refinado por: Categorías de Web of Science:(Education, Educational research)

La búsqueda en Scopus permitió identificar investigaciones con los términos “tesis doctoral”, validez y confiabilidad en el resumen (n = 14). Los estudios publicados en revistas de educación en ciencias de la salud sí analizan las propiedades métricas de los instrumentos de recolección de datos. En el caso de las revistas de investigación educativa solo se encontró un trabajo que examinó la validez y confiabilidad de un instrumento de recolección diseñado para medir la procrastinación en estudiantes de un doctorado en educación. Para ello, los investigadores emplearon enfoques analíticos derivados de la teoría clásica de los test y la teoría de respuesta al ítem (Johnson et al., 2000). Los demás estudios publicados en revistas de investigación educativa se enfocaron en la enseñanza de la lengua materna o una segunda lengua, clasificando los temas abordados en cada una de ellas (ver Tabla 5).

Tabla 5*Estudios que examinan las propiedades métricas de los instrumentos de recolección*

	Título	Revista	Año	Citas
1	<i>The self-directed learning readiness scale for nursing education revisited: A confirmatory factor analysis</i>	<i>Nurse Education Today</i>	2010	64
2	<i>How to construct and implement script concordance tests: Insights from a systematic review</i>	<i>Medical Education</i>	2012	48
3	<i>Psychometric characteristics of the revised procrastination inventory</i>	<i>Research in Higher Education</i>	2000	21
4	<i>Exploration of nursing doctoral admissions and performance outcomes</i>	<i>Journal of Nursing Education</i>	1996	6
5	<i>Review of doctoral research in English language teaching and learning in Turkey (2009-2013)</i>	<i>Language Teaching</i>	2016	5
6	<i>Review of doctoral research in second language acquisition in Germany (2006-2009)</i>	<i>Language Teaching</i>	2011	4
7	<i>Nursing education research in Finland: review of doctoral dissertations</i>	<i>Nurse Education Today</i>	2016	3
8	<i>Review of doctoral research in second-language acquisition in Australia (2003-2006)</i>	<i>Language Teaching</i>	2008	3
9	<i>Review of doctoral research in second-language teaching and learning in England (2006)</i>	<i>Language Teaching</i>	2009	2
10	<i>Review of doctoral research in language assessment in Canada (2006-2011)</i>	<i>Language Teaching</i>	2013	1
11	<i>Review of doctoral research on second language teaching and learning in Spain (2008-2010)</i>	<i>Language Teaching</i>	2013	1
12	<i>Review of doctoral research in second language acquisition, language teaching and learning in Poland (2006-2010)</i>	<i>Language Teaching</i>	2012	1
13	<i>Review of doctoral research in English language education in the Philippines, Singapore and Malaysia (2007-2010)</i>	<i>Language Teaching</i>	2012	1
14	<i>Review of doctoral research in second-language teaching and learning in the United States (2006-2007)</i>	<i>Language Teaching</i>	2009	1

Fuente: Base de datos Scopus. Fecha de búsqueda: 10 de marzo de 2020.

En WoS, además de los estudios antes mencionados se encontró una investigación que empleó el análisis de contenido para examinar las características de los instrumentos de recolección de datos empleados en las tesis doctorales (Karadağ, 2011). En ese estudio se analizaron 211 tesis doctorales en educación sustentadas en los años 2003 y 2007. Para

recolectar los datos se usó la Escala de Evaluación de la Investigación Educativa, compuesta por 50 ítems medidos en escalas Likert de 11 puntos (0 = No logrado, 10 = Completamente logrado). Los instrumentos de recolección utilizados en dichas tesis doctorales fueron: escalas de medición (63.1%), formatos de entrevistas (20.9%), pruebas de logro (12.4%) y formatos de observación (12.4%).

Las propiedades métricas de los instrumentos de recolección estuvieron por debajo de lo esperado, observándose que la calificación de la validez (2.93 ± 2.13) estuvo por debajo de la valoración de la confiabilidad (4.21 ± 1.86). Con relación a la validez de los instrumentos, los errores más frecuentes fueron: a) dividir las escalas en factores sin hacer análisis factorial, b) dar información insuficiente sobre el análisis factorial, c) interpretar incorrectamente las cargas factoriales de los ítems, d) analizar solo la validez de contenido [criterio de jueces expertos] al diseñar o adaptar la escala de medición. En lo que se refiere a la confiabilidad, las fallas más comunes fueron: a) elegir de forma incorrecta los coeficientes, b) determinar con poca precisión los intervalos de confianza, c) dar información insuficiente sobre los parámetros de confianza, d) borrar un ítem luego de realizar el análisis de confiabilidad.

A manera de síntesis, se puede decir que los estudios que analizaron la validez y confiabilidad de los instrumentos de recolección de datos empleados en las tesis doctorales en educación fueron descriptivos porque utilizaron formatos con categorías predefinidas.

Aun cuando las tesis doctorales suelen incluir los resultados originales de los análisis de validez y confiabilidad, los autores de los estudios previos optaron por analizar el contenido de dichas tesis usando fichas de registro con categorías determinadas de antemano. Por otro lado, si bien es cierto los estudios descriptivos son necesarios para determinar el alcance y complejidad de los temas analizados, resultan insuficientes para explicar en qué condiciones se manifiestan dichas variables o cómo afectan otros factores.

Por todo ello, es importante explorar si validez y confiabilidad de los instrumentos de recolección se relacionan con aspectos vinculados con el desempeño académico o la efectividad de las sesiones de enseñanza y aprendizaje. Dado que las propiedades métricas se expresan en magnitudes numéricas se optó por una métrica para examinar la relación entre dichas propiedades y la efectividad de los programas educativos descritos en investigaciones experimentales y cuasi experimentales. La medición de la efectividad de los programas es necesaria para saber si en verdad mejoraron el aprendizaje de los participantes del estudio. El indicador empleado para medir la efectividad de los programas educativos se conoce como tamaño del efecto (Cohen, 1988; Ellis, 2012).

Por esa razón, en esta tesis se consideró importante determinar si las propiedades métricas de los instrumentos de recolección tenían alguna relación con los programas de enseñanza y aprendizaje reportados en las tesis doctorales. La validez y confiabilidad son atributos fundamentales porque sobre ellos se estructuran las estrategias de análisis para probar las hipótesis de investigación de las tesis doctorales en educación. Para entender el alcance de las hipótesis de este estudio, a continuación, se describen los principales conceptos del estudio.

2.3. Definición de Términos Básicos

Análisis de Validez.

Es un tipo de análisis que permite determinar si los ítems y escalas diseñadas para para medir una variable son apropiados o si deben ser reformulados. Si bien existen diferentes tipos de validez (e.g., convergente, de contenido, de constructo, discriminante, etc.), los coeficientes para analizar la validez suelen expresarse en valores numéricos cuyas magnitudes van de 0.00 a 0.99, donde los valores cercanos a cero indican un bajo nivel de validez y los cercanos a la unidad un alto nivel de validez. A diferencia de la confiabilidad, que es una condición necesaria para los instrumentos de recolección de datos, la validez es

una condición necesaria y suficiente. Es decir, un instrumento de recolección de datos puede ser confiable pero no válido, sin embargo, todo instrumento de recolección que es válido por definición también es confiable.

Análisis de Confiabilidad.

Es un tipo de análisis que permite determinar si los puntajes obtenidos al responder o evaluar un conjunto de ítems (e.g., enunciados, preguntas, etc.) tienen bajos niveles de dispersión; es decir, no se observan grandes variaciones en las magnitudes de los coeficientes que miden la confiabilidad. Los más usados son el análisis de estabilidad en la medición (e.g., test-retest), confiabilidad intercodificador (e.g., alfa de Krippendorff o kappa de Cohen) y el análisis de consistencia interna (e.g., Kuder-Richardson o alfa de Cronbach). Cada uno de esos coeficientes tiene rangos de variación que van de 0.00 a 0.99, donde los valores cercanos a cero indican un bajo nivel de confiabilidad y los valores cercanos a uno un alto nivel. De esta manera se puede determinar si la confiabilidad es baja, moderada o fuerte.

Instrumentos de Recolección de Datos.

Herramientas diseñadas o adaptadas por los investigadores para recoger los datos que les permitirán probar sus hipótesis de investigación. Dado que en educación existen diversos diseños de investigación, la naturaleza de los datos recolectados variará según el tipo de instrumento de recolección empleado (e.g., escala de medición, formato de entrevista, formato de observación, formato de registro, prueba de rendimiento académico, etc.). Dentro de los diseños de investigación experimentales y cuasiexperimentales, los instrumentos de recolección empleados con más frecuencia son las escalas de medición y las pruebas de rendimiento en áreas curriculares como matemática o comunicación.

Tamaño del Efecto.

Medida de la magnitud y dirección (positiva o negativa) de la relación que existe entre dos o más variables dentro de una muestra o población analizadas. Los tamaños de efecto más conocidos son aquellos basados en la diferencia de medias (e.g., d de Cohen, Δ de Glass o g de Hedges), en los coeficientes de asociación (e.g., r de Pearson, ρ de Spearman o τ de Kendall), los coeficientes de regresión (e.g., coeficiente de determinación en modelos de regresión lineal R^2 o coeficiente de dispersión de medias en tres o más grupos f^2) y la comparación entre grupos basada en respuestas dicotómicas (e.g., razón de riesgo [*risk ratio*] o razón de verosimilitud [*odds ratio*]). Cuando se aplican a los programas de enseñanza-aprendizaje, el valor del tamaño del efecto permite medir la efectividad de dichos programas. En ese sentido es una medida que complementa el análisis de la significancia estadística: mientras que la significancia determina si hubo o no un cambio a consecuencia de la intervención educativa, el tamaño del efecto establece cuál es la magnitud del cambio logrado, por eso se dice que es una medida de la efectividad alcanzada.

Tesis Doctorales en Educación.

Producto académico que elaboran quienes concluyen un doctorado en educación. El propósito de la tesis varía según el tipo de doctorado: investigación, profesionalización o especialización. En el primer caso se espera que la tesis doctoral brinde aportes al conocimiento en educación, mientras que en el segundo caso la tesis debe mostrar un manejo experto y actualizado de temas relevantes en el campo educativo. En el caso de los doctorados de especialización, se espera que el producto final presente innovaciones a los diseños instruccionales o programas de enseñanza-aprendizaje.

Capítulo III. Hipótesis y Variables

3.1. Hipótesis

A continuación, se presentan las hipótesis de investigación, que se derivan de los problemas y objetivos detallados en el primer capítulo. En el siguiente capítulo se explican las técnicas de recolección de datos y herramientas analíticas que permitirán poner a prueba estas hipótesis. En el apéndice A se muestran la articulación entre problemas, objetivos e hipótesis, así como la integración entre estos últimos y las definiciones conceptuales y operacionales de las variables analizadas.

3.1.1. Hipótesis General

Existe una relación positiva entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la Universidad Nacional de Educación (UNE) 2005-2015.

3.1.2. Hipótesis Específicas

Primera Hipótesis Específica.

Existe una relación positiva entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.

Segunda Hipótesis Específica.

Existe una relación positiva entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.

Tercera Hipótesis Específica.

El tamaño de la muestra modera la relación entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.

3.2. Variables

3.2.1. Definición Conceptual

Validez del Instrumento.

Es una propiedad métrica de los puntajes obtenidos con un instrumento de recolección de datos. Dicha propiedad permite saber si el instrumento o prueba realmente mide la variable o el constructo que en teoría debe medir. Para que un instrumento de recolección sea realmente válido, la presencia o ausencia del constructo debiera ser la única razón de las respuestas observadas o registradas. Sin embargo, en la práctica esto es casi imposible de lograr. Por esa razón, los procedimientos estadísticos por sí solos no garantizan la validez, los coeficientes de correlación o los modelos estructurales solo proporcionan evidencia parcial de validez. Sin un procedimiento cuidadoso para garantizar la validez, esta propiedad métrica estará ausente (Wilson & Gochyyev, 2013).

Con relación a las evidencias para demostrar que un instrumento mide lo que debe medir, además de tener un sólido respaldo estadístico, la interpretación de los puntajes debe estar acorde con la teoría propuesta. Es decir, dicha teoría debe descartar cualquier otra explicación alternativa que no haya sido contemplada por dicha teoría (AERA et al., 2014).

Confiabilidad del Instrumento.

Es una propiedad métrica de los puntajes obtenidos con un instrumento de recolección de datos. Esta propiedad se expresa en un índice que señala el nivel de consistencia del

puntaje al medir la variable o constructo que debería medir. El tipo de índice a emplear depende del diseño de investigación planteado para probar la hipótesis del estudio.

Cuando los puntajes son confiables, el evaluador o investigador puede asumir que el error de medición es tan pequeño que no hay problema alguno en usar dichos puntajes. Es decir, en los instrumentos con altos niveles de confiabilidad los puntajes observados están muy cerca de los puntajes reales porque el error de medición es muy pequeño. Dado que existen muchas fuentes potenciales para el error de medición, además de aquellas derivadas de las respuestas obtenidas con el instrumento de recolección (e.g., el orden de presentación de los ítems, las condiciones de administración, las actitudes de los evaluadores, etc.) el error de medición siempre estará presente. Por ello, se considera que el error es parte inevitable del proceso de medición y que el evaluador debe hacer todos los esfuerzos posibles por reducirlo (Wilson & Guyyev, 2013).

En cuanto a la interpretación de los puntajes obtenidos con los instrumentos de recolección, el considerar que un puntaje es más confiable que otro depende de las premisas de los evaluadores sobre cuál debe ser el grado de consistencia al examinar el puntaje obtenido en cada administración del instrumento de recolección de datos (AERA et al., 2014). Por esa razón se han establecido rangos de variación para determinar si el valor de los indicadores corresponde a un nivel de baja, mediana o alta confiabilidad.

Tamaño del Efecto.

Un efecto es el resultado obtenido al comparar el cambio que produce un tratamiento sobre dos grupos (e.g., grupo de tratamiento versus grupo control o tratados versus no tratados) o al describir la asociación que existe entre dos variables (e.g., la relación entre la motivación para aprender y el rendimiento académico en matemática o el vínculo entre las condiciones de trabajo y el compromiso con la organización).

Si bien es cierto que los efectos pueden observarse en condiciones controladas (e.g., dentro de un laboratorio experimental o en una muestra seleccionada de forma aleatoria), los tamaños de efecto son magnitudes que existen en el mundo real (Ellis, 2010). En otras palabras, el tamaño del efecto es una medida de la magnitud y dirección (positiva o negativa) de la relación que existe entre dos o más variables dentro de una muestra analizada. Dado que el tamaño del efecto no se ve afectado por el tamaño de la muestra, se considera un complemento a la significancia estadística, ya que el tamaño del efecto se enfoca en la significancia práctica, es decir, en la importancia de la relación observada cuando se manifiesta en el mundo real (Borenstein et al., 2009).

Tamaño de la Muestra.

Cantidad de participantes (e.g., estudiantes, profesores, directores, etc.) que fueron seleccionados para el estudio, a partir de un tipo de muestreo y tomando en cuenta los criterios inclusión y exclusión definidos previamente. En el caso de los diseños de investigación experimentales y cuasi experimentales, el tamaño de la muestra se refiere al número de participantes en el grupo de tratamiento y en el grupo control.

3.2.2. Operacionalización de Variables

Validez del Instrumento.

Esta variable se operacionalizó a dos niveles. A nivel cualitativo se tomó en cuenta si en la tesis se incluyó información sobre el tipo de análisis de validez usado para examinar la variable correspondiente. Cuando no se obtuvo información sobre el tipo de validez se registró “No indica”, caso contrario se registró “Opinión de expertos” para indicar que se usó la validez de contenido por criterio de jueces expertos.

A nivel cuantitativo se registró el nivel promedio de acuerdo entre los jueces que evaluaron la validez del instrumento. Dicho acuerdo se registró como un valor decimal. Por ejemplo, si el acuerdo promedio entre los jueces fue 85.73%, dicho valor se convirtió

en 0.8573. Como la validación por jueces expertos es un requisito para las tesis sustentadas en la Escuela de Posgrado de la UNE, en caso la tesis doctoral tuviera los resultados de la validez de constructo (e.g., porcentaje de varianza explicada o número de componentes extraídos), el valor que se registró para la variable validez fue el acuerdo promedio entre los jueces, expresado como valor decimal.

Confiabilidad del Instrumento.

Esta variable se operacionalizó a dos niveles. A nivel cualitativo se registró el tipo de análisis de confiabilidad empleado con la variable: a) alfa de Cronbach, b) correlación de Pearson, c) Küder-Richardson, d) mitades-partidas, e) test-retest. A nivel cuantitativo se tomaron en cuenta dos indicadores: la magnitud del coeficiente –medido en un rango que va de 0.00 a 0.99– y el número de ítems usados para medir la variable. Este último dato es importante porque el coeficiente de correlación es sensible al número de ítems.

Tamaño del Efecto.

Esta variable se operacionalizó a nivel cuantitativo. Para calcular la magnitud del tamaño del efecto únicamente se consideraron las tesis doctorales con diseño experimental o cuasi experimental. De dichas tesis se extrajeron los datos para calcular el indicador de tamaño del efecto conocido como la d de Cohen. $d = (\text{media del grupo experimental} - \text{media del grupo control}) / \text{desviación estándar agrupada}$.

A partir de ese indicador se calculó el indicador de tamaño del efecto g de Hedges, que fue el indicador empleado para el contraste de las hipótesis de investigación.

Tamaño de la Muestra.

Valor obtenido al sumar el número personas (e.g., estudiantes, profesores, directores, etc.) que formaron parte del grupo experimental y del grupo control. Los valores que corresponden al número de participantes por cada grupo se expresaron como números enteros. Estos valores también se emplearon para calcular el tamaño del efecto.

Capítulo IV. Metodología

Para esta investigación se analizaron todas las tesis doctorales de la Escuela de Posgrado de la UNE sustentadas entre los años 2005 y 2015, registradas en el catálogo de tesis de la biblioteca de posgrado. Para el estudio solo se consideraron las tesis con diseño (cuasi) experimental que analizaron la efectividad de los programas de enseñanza-aprendizaje orientados a los estudiantes educación superior.

4.1. Enfoque de Investigación

Estudio con enfoque cuantitativo ya que busca medir la efectividad de los programas de enseñanza-aprendizaje en educación superior, para saber cuáles tienen un efecto directo en la mejora del rendimiento académico de los estudiantes universitarios. En ese sentido, el estudio se enmarca dentro de la tradición epistemológica neopositivista ya que aspira a formular modelos matemáticos basados en hechos observables y replicables. En el estado actual del conocimiento científico, el neopositivismo también considera la complejidad y diversidad, dentro de una perspectiva interdisciplinaria que garantiza que los criterios de recolección y análisis de datos cumplen los requisitos mínimos de rigurosidad, tales como la validez, confiabilidad, generalizabilidad y operacionalizabilidad.

4.2. Tipo de Investigación

Es un estudio descriptivo con alcance explicativo y corresponde al tipo reseña de la literatura, subtipo revisión sistemática. Este tipo de revisiones se refiere a evaluaciones críticas del material académico que ya ha sido divulgado, como es el caso de las tesis doctorales que ya fueron sustentadas y pueden ser revisadas por cualquier investigador, desde la biblioteca de Escuela de Posgrado de la UNE o el repositorio institucional de la universidad, que a la fecha alberga más de 670 tesis doctorales. El estudio incluye un componente de meta análisis porque los resultados de cada estudio se combinaron de

forma estadística para obtener un indicador agregado que sintetice los hallazgos reportados.

4.3. Diseño de Investigación

En esta investigación se combinó el tamaño del efecto de cada programa de enseñanza aprendizaje diseñado para mejorar el rendimiento académico de los estudiantes peruanos de educación superior. Para efectuar dicha combinación, solo se incluyeron aquellos estudios que tuviesen un diseño (cuasi) experimental, lo que permitió obtener un tamaño del efecto ponderado. Por la naturaleza del estudio, el meta análisis es de tipo explicativo ya que se incluyó un componente de meta regresión para explorar cómo influyen los criterios métricos de los instrumentos de recolección sobre el tamaño del efecto ponderado.

4.4. Método

Para estudiar el efecto de los programas de enseñanza-aprendizaje en la mejora del rendimiento académico de los estudiantes universitarios se consultaron las tesis doctorales del periodo 2005-2015 disponibles en la biblioteca de la Escuela de Posgrado de la UNE y en el repositorio institucional de la universidad. Para extraer los datos de cada tesis doctoral se usaron 13 campos de la hoja de registro detallada en el apéndice B.

Los 13 campos usados para el recojo de datos de las tesis doctorales fueron: 1) apellidos del autor, 2) año de sustentación, 3) análisis de validez, 4) tipo de análisis de confiabilidad, 5) magnitud del coeficiente de confiabilidad, 6) cantidad de ítems para medir la variable, 7) tamaño de la muestra, 8) media de la variable analizada para el grupo experimental, 9) desviación estándar de la variable analizada para el grupo experimental, 10) número de participantes en el grupo experimental, 11) media de la variable analizada para el grupo control, 12) desviación estándar de la variable analizada para el grupo control, 13) número de participantes en el grupo control.

La información de los últimos seis campos permitió calcular los dos indicadores del tamaño del efecto: d de Cohen y g de Hedges. Para determinar si el efecto fue grande, mediano o pequeño, se tomaron en cuenta los umbrales definidos por Jacob Cohen (1988), que fueron adaptados por Robert Slavin (2008) para el campo educativo. Como indicador complementario a la d se empleó la g de Hedges, ya que este estadístico toma en cuenta el tamaño de la muestra a la hora de calcular el tamaño de efecto agregado, generando de esta forma un indicador ponderado.

Un elemento final a tener en cuenta fue el criterio empleado a la hora de extraer los datos de las tesis doctorales y registrarlos en la plantilla creada para dicho propósito (ver Apéndice B). Aun cuando algunas tesis doctorales que examinaron el efecto de los programas de enseñanza-aprendizaje también analizaron el impacto de otras variables (e.g., la motivación del estudiante, sus hábitos de estudio, sus habilidades sociales, etc.), sólo se registraron los datos correspondientes a la validez y confiabilidad de los instrumentos de recolección de datos que midieron las variables que tuvieron un efecto en el aprendizaje.

Se tomó esta decisión porque incluir todos los resultados originales que fueron reportados en las tesis doctorales no habría aportado al propósito de esta investigación. Una vez calculados los valores del tamaño del efecto, las magnitudes obtenidas del tamaño del efecto ponderado se compararon con los hallazgos de otras investigaciones que también han analizado el impacto de las intervenciones educativas, según lo reportado en tesis doctorales de educación (Hattie, 2009, 2012; Pigott et al., 2013; Polanin et al., 2016).

4.5. Población y Muestra

El corpus documental estuvo formado por todas las tesis doctorales de la Escuela de Posgrado de la UNE, sustentadas durante 2005 y 2015. Dado que se trabajó con todas las tesis sustentadas en ese periodo, se trata de un estudio censal, es decir la cobertura

documental fue del 100%. Con relación a los criterios de selección, se consideraron aquellas tesis que cumplieron tres criterios: a) el programa de enseñanza aprendizaje se enfoca en estrategias didácticas, en cualquiera de sus variantes; b) los participantes del estudio son estudiantes de educación superior, procedentes de universidades (a nivel de pre o posgrado) o de institutos tecnológicos; c) los estudios tienen un diseño (cuasi) experimental.

En cuanto a la cobertura temporal se trabajó con el periodo 2005-2015 porque se consideró que un periodo no menor a 10 años es adecuado para examinar la evolución de los métodos de recolección y análisis de datos enseñados en un programa doctoral en ciencia de la educación. Dado que el aprendizaje de dichos métodos requiere un proceso de maduración, ya sea para precisar los alcances de las técnicas analíticas o incorporar las nuevas métricas que pudieran surgir, un periodo no menor a diez años se consideró suficiente para dar cuenta de la maduración de la enseñanza de dichos métodos de recolección y análisis de datos.

4.6. Técnicas de Recolección de Datos

Para la extracción de datos de las tesis doctorales se trabajó con una ficha de recojo de datos elaborado por el autor (ver Apéndice B), la cual es una adaptación del diccionario de códigos para revisiones sistemáticas en investigación educativa propuesto por Olusola Adesope y su equipo de colaboradores el año 2010 (ver Apéndice C). Se adaptó dicha ficha de recojo de datos porque es uno de los pocos estudios en los que se reportó el coeficiente de confiabilidad intercodificador en dos importantes etapas previas al cálculo del tamaño del efecto: la selección de los estudios (κ de Cohen = 0.88) y la codificación de las diferentes secciones del diseño de investigación (α de Krippendorff = 0.92). Para conocer cómo estuvo organizada la ficha de recojo de datos elaborada por Adesope y colaboradores (2010), al llevar a cabo el meta análisis aplicado al ámbito educativo, al

final de esta tesis se incluye la ficha de datos usada por los autores antes mencionados (ver Apéndice C).

El esquema de codificación del presente estudio doctoral tiene dos partes: la primera incluye datos referidos al diseño de investigación (e.g., tipo de diseño, evidencias de validez y confiabilidad, tamaño de la muestra) y la segunda registra los resultados de la tesis (e.g., participantes en el grupo experimental y control, media y desviación estándar del resultado observado en ambos grupos después de la intervención). Los datos de la segunda parte de la ficha de registro se usaron para calcular el tamaño de efecto individual y agregado.

La información se registró en un archivo Excel, que luego fue importado en el programa estadístico R para Windows, versión 3.6.0. Se eligió dicho software de análisis estadístico por la flexibilidad que ofrece para actualizar los módulos de análisis ya que a diferencia de programas como SPSS o Stata la actualización no requiere pago de licencia.

4.7. Tratamiento Estadístico

Cálculo e Interpretación de los Tamaños del Efecto.

Se calcularon los indicadores de tamaño del efecto asociados con cada programa de enseñanza-aprendizaje (d de Cohen y g de Hedges) y el tamaño de efecto promedio. Se definieron los siguientes puntos de corte: igual o mayor a 0.20, entre 0.19 y 0.11, igual o menor a 0.10 para los tamaños de efecto grandes, medianos y pequeños, respectivamente (Slavin, 2008). Para determinar la significancia de cada efecto se obtuvo el p -valor y para medir su dispersión se computó el intervalo de confianza al 95%. Para probar que la distribución de los tamaños de efecto era homogénea se estimó el estadístico Q , que mide hasta qué punto los tamaños del efecto observados son diferentes de los efectos esperados (Borenstein et al., 2009; Ellis, 2010).

Contraste de la Hipótesis General y las Hipótesis Específicas.

Para probar las hipótesis de investigación se emplearon modelos de regresión lineal: múltiple (para la hipótesis general) y bivariada y con efecto moderador (para las hipótesis específicas). En este estudio la variable a ser predicha fue el tamaño del efecto del programa de enseñanza-aprendizaje y las variables predictoras fueron la validez y confiabilidad del instrumento de recolección usado para medir la variable en que se enfocó el programa de enseñanza-aprendizaje, así como el efecto moderador del tamaño de la muestra. A fin de determinar la significancia estadística de cada modelo de regresión se verificó que el p-valor del coeficiente F de Fisher fuese menor a 0.05 y que el p-valor de cada coeficiente de regresión fuese menor a 0.05. Antes de generar cada modelo de regresión, se obtuvieron matrices de correlación y gráficos de dispersión.

Dado que la primera y segunda hipótesis específica también se probaron usando la regresión bivariada, además de calcular el p-valor de cada variable se obtuvo el intervalo de confianza del coeficiente beta asociado a cada variable. Si dicho intervalo de confianza incluía el valor cero, se consideraría que la validez o confiabilidad no explicaron el tamaño del efecto de los programas de enseñanza-aprendizaje.

En cuanto a la tercera hipótesis específica, se analizó el efecto moderador del tamaño de la muestra sobre la validez y confiabilidad al predecir el tamaño del efecto de los programas de enseñanza-aprendizaje. Para verificar la significancia estadística del modelo de regresión con efecto de moderación se verificó que el p-valor del coeficiente F de Fisher, los coeficientes de regresión obtenidos al validar la hipótesis general y los coeficientes del efecto moderador fuesen menores a 0.05.

Para calcular el tamaño del efecto se utilizó el paquete meta, versión 4.9-5, para R, mientras que para los modelos de regresión se empleó el paquete lm para R.

Capítulo V. Resultados

5.1. Validez y Confiabilidad del Instrumento

Dado que no se analizó el contenido de las tesis doctorales ni se asignó un código (tal como ocurre con el análisis de contenido), sino que se extrajeron datos y se registraron en una ficha, el instrumento se validó a través de la opinión de expertos en metodología de investigación. Para ello, cada uno de los cinco expertos (todos ellos, profesores de metodología en maestría y doctorado) recibió la ficha de validación junto con la matriz de consistencia y la ficha de registro de datos. Cada experto verificó que hubiera correspondencia entre las dimensiones de la ficha y los objetivos de la investigación. Producto de ello, asignaron una calificación en forma de porcentaje, tomando en cuenta una escala de 0% a 100%, donde 0 equivale a muy deficiente y 100 a excelente. Los expertos asignaron una valoración global de 92%, que fluctuó entre 81% y 100%, según la dimensión evaluada (ver Tabla 6). Para revisar las fichas de validación que emplearon los expertos puede consultarse el Apéndice D.

Tabla 6

Resultados de la validación según la opinión de los expertos

Dimensión	Experto 1	Experto 2	Experto 3	Experto 4	Experto 5
01. Tipo de diseño	[81-85 %]	[91-95%]	[86-90 %]	[81-85 %]	[91-95 %]
02. Instrumento	[86-90 %]	[96-100 %]	[81-85 %]	[81-85 %]	[91-95 %]
03. Muestreo	[81-85 %]	[91-95 %]	[81-85 %]	[81-85 %]	[96-100 %]
04. Muestra	[91-95 %]	[91-95 %]	[96-100 %]	[81-85 %]	[91-95 %]
05. Validez	[96-100%]	[91-95 %]	[96-100 %]	[81-85 %]	[86-90 %]
06. Confiabilidad	[96-100%]	[91-95 %]	[96-100 %]	[81-85 %]	[91-95 %]
07. Asociación	[86-90 %]	[91-95 %]	[96-100 %]	[81-85 %]	[96-100 %]
08. Promedio	[96-100 %]	[91-95 %]	[96-100 %]	[81-85 %]	[86-90 %]
09. Desviación estándar	[96-100 %]	[91-95 %]	[96-100 %]	[81-85 %]	[91-95 %]
10. Tamaño por grupo	[96-100 %]	[91-95 %]	[96-100 %]	[81-85 %]	[91-95 %]

Fuente: Elaborado por el autor con los resultados de la validación.

Dado que las hipótesis de investigación analizan los tamaños del efecto de cada programa de enseñanza-aprendizaje orientados a estudiantes peruanos de educación superior y su relación con tres covariadas (validez, confiabilidad y tamaño de la muestra),

para medir la confiabilidad de los efectos calculados se examinaron los intervalos de confianza de cada uno. Los resultados de este análisis se detallan en el acápite “Tamaño del efecto en los programas de enseñanza-aprendizaje” de la siguiente sección.

5.2. Presentación y Análisis de los Resultados

Luego de revisar las tesis disponibles en la biblioteca de la Escuela de Posgrado UNE y el repositorio institucional se identificaron 54 tesis doctorales que cumplieron los criterios de inclusión detallados en el capítulo anterior (dos sustentadas en el año 2005, una el 2006, tres sustentadas en 2007, diez en 2008, ocho en 2009, ocho en 2010, cuatro en 2011, cinco en el año 2013, una el 2014, 12 en el año 2015). Dichas tesis reportaron 153 efectos.

5.2.1. Análisis de Validez Empleados en las Tesis Doctorales

Con relación al tipo de validez empleado, en el 83% de las tesis ($n = 45$) los autores usaron la validación por criterio de expertos, en el 2% de las tesis ($n = 1$) se validaron los instrumentos empleando la validez predictiva y en el 15% no se indicó el tipo de validación empleada ($n = 8$). A pesar que en algunas tesis se reportó haber empleado el análisis de componentes principales, los tamaños de muestra asociados eran tan pequeños ($n < 40$) y carecían de indicadores estadísticos, que fueron asignadas a la validación por juicio de expertos. El 61% de las tesis ($n = 33$) reportó el resultado de la validación por criterio de expertos como un porcentaje, lo que permitió probar la hipótesis general con dichas tesis.

5.2.2. Análisis de Confiabilidad Empleados en las Tesis Doctorales

En cuanto a la confiabilidad del instrumento de recolección de datos, se registró el resultado del análisis de confiabilidad para cada efecto obtenido al implementar el programa de enseñanza-aprendizaje. De los 153 efectos reportados, en el 37% de efectos ($n = 56$) se usó el coeficiente alfa de Cronbach, en el 15% de efectos ($n = 23$) el coeficiente Kuder-Richardson, en el 8% ($n = 12$) el test-retest, en el 4% ($n = 6$) el coeficiente de

mitades partidas y en el 2% ($n = 3$) el coeficiente de Pearson, mientras que en el 34% de efectos ($n = 53$) no se mencionó el coeficiente utilizado para examinar la confiabilidad del instrumento usado para medir el efecto obtenido al implementar el programa de enseñanza-aprendizaje. Si bien no se reportó el análisis de confiabilidad para la tercera parte de los efectos, en 76 de ellos sí se reportó la magnitud del coeficiente de confiabilidad (0.78 ± 0.10), la cual mostró una concentración de valores mayores a 0.7 (ver Figura 1). Sobre el número de ítems de cada instrumento (17.30 ± 7.99), se observó una moderada dispersión (ver Figura 2).

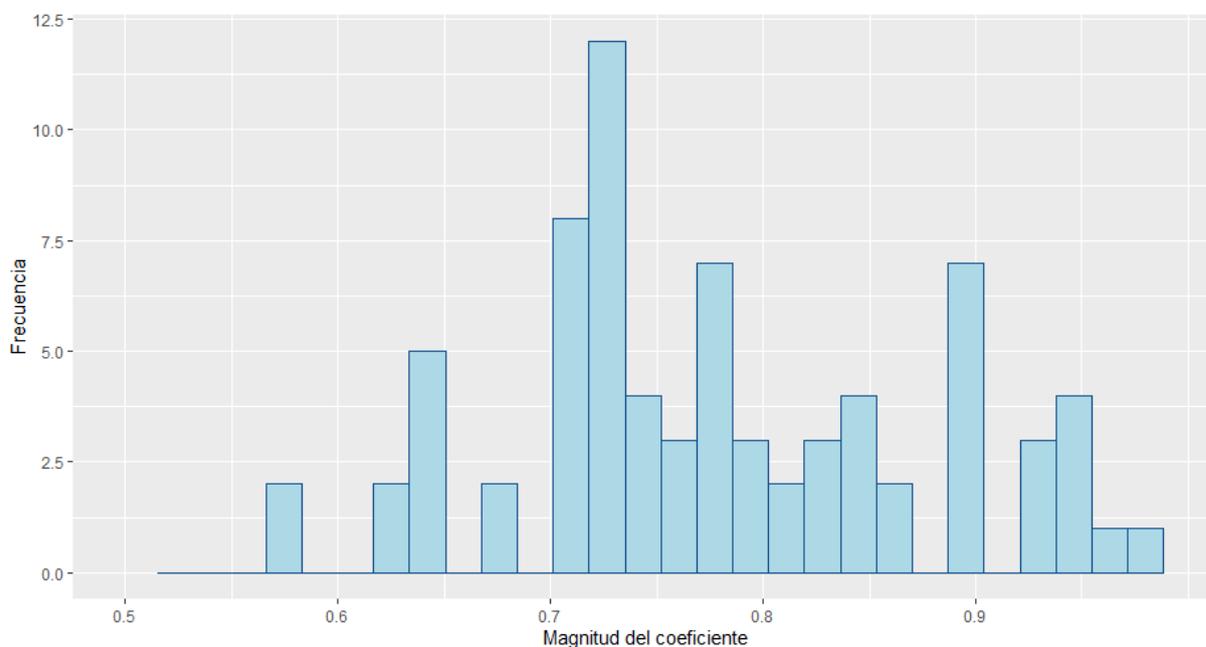


Figura 1. Histograma de la magnitud de los coeficientes de confiabilidad

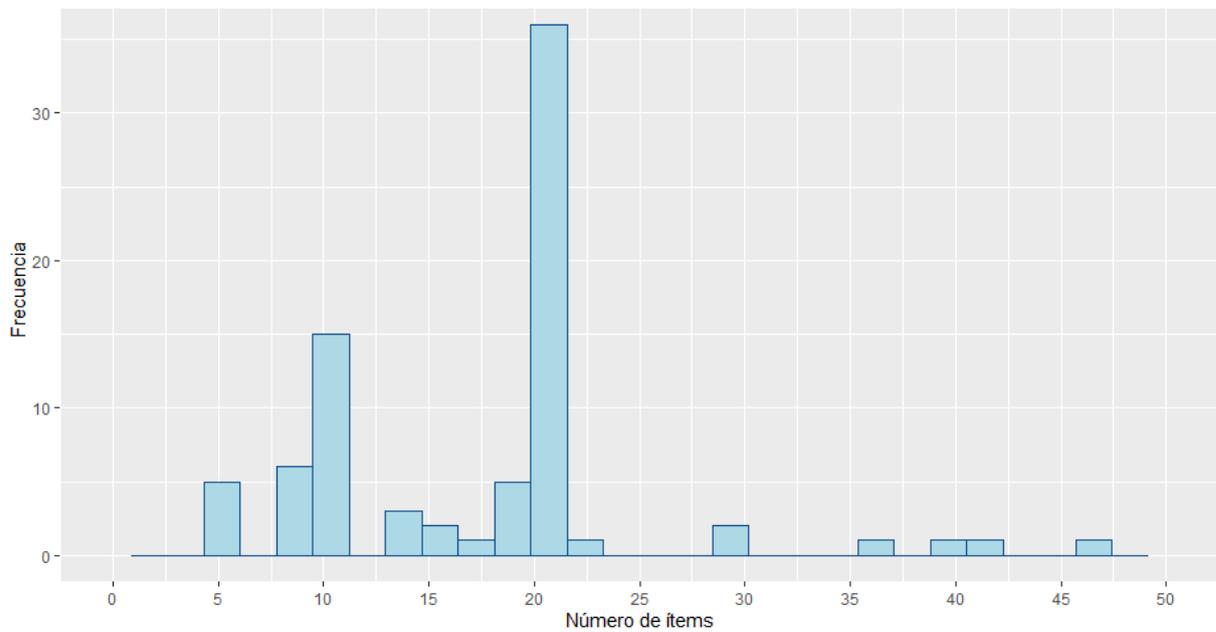


Figura 2. Histograma de frecuencias de la cantidad de ítems incluidos en cada escala

En 20 de los efectos analizados, la magnitud del coeficiente de confiabilidad fue igual o mayor a 0.90, incluso uno de ellos alcanzó el valor de 0.99. La mayoría de las variables se midieron con 20 ítems, pero en ciertos casos dichas variables fueron medidas con 40, 42 y 47 ítems cada una (ver Tabla 7).

Tabla 7

Número de ítems usados para medir cada variable

# ítems	Frecuencia	Porcentaje (%)
No indica	73	47.71
20	31	20.26
10	13	08.50
21	5	03.27
19	5	03.27
8	4	02.61
5	4	02.61
30	2	01.31
14	2	01.31
11	2	01.31
9	2	01.31
47	1	00.65
42	1	00.65
40	1	00.65
36	1	00.65
22	1	00.65

17	1	00.65
16	1	00.65
15	1	00.65
13	1	00.65
6	1	00.65

Fuente: Elaborado por el autor con los resultados del estudio.

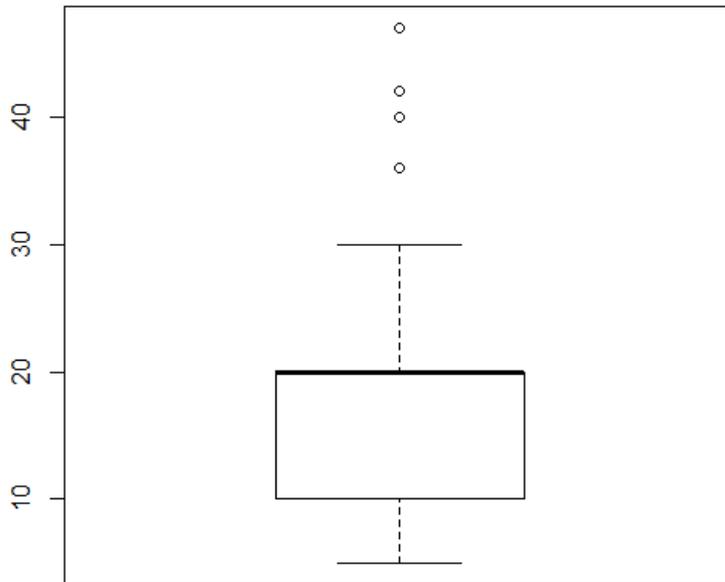


Figura 3. Gráfico de cajas del número de ítems para medir cada variable

En lo que se refiere al tamaño de la muestra se encontró un alto grado de variabilidad (65.67 ± 44.16 , intervalo de confianza al 95% [58.61 – 72.72]). Probablemente por el tipo de diseño empleado predominaron las muestras menores a 60 participantes, observándose la presencia de algunos valores atípicos, tal como se aprecia en el histograma de frecuencias y el gráfico de cajas (ver figuras 4 y 5).

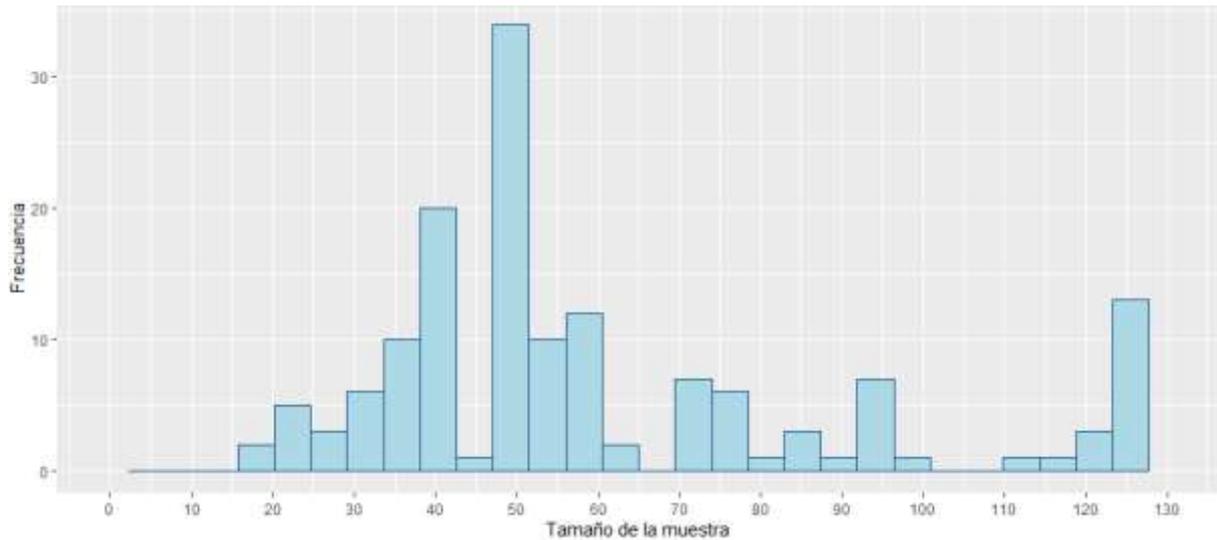


Figura 4. Histograma de frecuencias del número de participantes para medir cada efecto

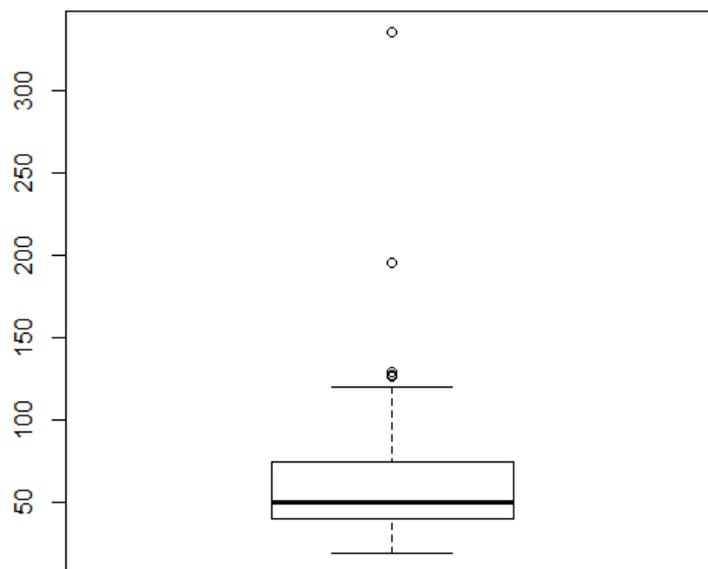


Figura 5. Gráfico de cajas del número de participantes para medir cada efecto

5.2.3. Tamaño del Efecto de los Programas de Enseñanza-Aprendizaje

Las variables empleadas para calcular el tamaño de efecto de cada programa mostraron un moderado nivel de dispersión, según se observó en las medidas de tendencia central y de dispersión. Cabe agregar que se detectaron valores atípicos (ver Tabla 8).

Tabla 8

Datos descriptivos de variables empleadas para calcular el tamaño del efecto (n = 150)

Variable	Media	D.E.	Mediana	Mínimo	Máximo
Media del grupo experimental	68.20	41.02	66	1	141
D. E. del grupo experimental	71.53	41.85	71	1	144
Tamaño del grupo experimental	33.42	23.25	25	8	167
Media del grupo control	68.54	40.46	69	1	139
D. E. del grupo control	68.47	40.74	68	1	140
Tamaño del grupo control	32.12	21.38	27	10	168

Fuente: Creada por el autor con los resultados del estudio.

Nota: D.E. = Desviación estándar. Tres efectos no incluyeron estadísticas descriptivas.

La magnitud de los tamaños de efecto calculados estuvo muy por encima de los umbrales conocidos, ya que el tamaño de efecto promedio fue casi diez veces más grande que los puntos de corte establecidos ($g \geq 0.20$) para un efecto alto: $g = 1.2487$, $p < 0.001$, IC al 95% [1.2008 – 1.2966] para el modelo de efectos fijos y $g = 2.0705$, $p < 0.001$, IC al 95% [1.8421 – 2.2990] para el modelo de efectos aleatorios. En efecto, en el capítulo anterior se estableció que un tamaño de efecto igual o mayor a 0.20 equivalía a un efecto grande (o moderadamente alto), ya que los meta-análisis publicados en revistas emblemáticas de investigación educativa como el *Review of Educational Research* reportan tamaños de efecto promedio que varían entre 0.10 y 0.50. Sin embargo, en las tesis doctorales analizadas se reportaron tamaños del efecto que fueron ocho o nueve veces más grandes (léase un crecimiento de 800% o 900%) que los umbrales fijados en este estudio. Para graficar la magnitud de estos coeficientes se seleccionaron los diez tamaños del efecto con los valores más altos (ver Tabla 9).

Tabla 9

Los diez tamaños del efecto con los valores más altos

N°	g de Hedges	I.C. al 95%	
		inferior	superior
1	13.5460	11.8113	15.2808
2	11.6446	9.8126	13.4766
3	9.8908	8.6018	11.1797
4	9.0203	7.0538	10.9868
5	8.6735	6.7762	10.5707
6	7.9040	6.6217	9.1864
7	7.7502	5.1575	10.3430

8	7.4932	5.8295	9.1569
9	6.2205	4.7710	7.6700
10	5.7796	4.9749	6.5843

Fuente: Creada por el autor con los resultados del estudio.

Nota: I.C. = Intervalo de confianza.

Por todo ello se consideró necesario obtener medidas que fuesen más comparables entre sí. Para ello, se calculó el logaritmo natural de las tres variables asociadas a cada grupo (media, desviación estándar y tamaño de muestra), ya que este procedimiento permite obtener distribuciones más comparables (Rodríguez-Barranco et al., 2017).

Cuando la transformación logarítmica de la desviación estándar generó un valor negativo, el efecto fue omitido de los análisis posteriores. Los valores transformados mostraron una dispersión mucho menor que los valores originales (ver Tabla 10) y fueron los que se usaron en análisis posteriores.

Tabla 10

Datos descriptivos de las variables luego de la transformación logarítmica (n = 122)

Variable	Media	D.E.	Mediana	Mínimo	Máximo
Media del grupo experimental	2.78	0.76	2.74	1.00	5.18
D. E. del grupo experimental	0.71	0.72	0.64	-1.17	2.73
Tamaño del grupo experimental	3.35	0.53	3.22	2.08	5.12
Media del grupo control	2.42	0.94	2.46	-0.36	5.09
D. E. del grupo control	0.75	0.72	0.73	-1.47	2.73
Tamaño del grupo control	3.33	0.49	3.30	2.30	5.12

Fuente: Creada por el autor con los resultados del estudio.

Nota: D.E. = Desviación estándar.

La magnitud de los tamaños del efecto luego de la transformación logarítmica mostró valores más acordes con los valores reportados en la literatura académica ($g = 0.3264$, $p < 0.001$, IC al 95% [0.1824 – 0.4703] para los modelos de efectos fijos y aleatorios). Al ordenarlos de forma decreciente se encontraron efectos no tan atípicos como lo observado con los valores originales (ver Tabla 11). En el apéndice D se muestra la magnitud del tamaño del efecto y el intervalo de confianza al 95% de todas las intervenciones, luego de

la transformación logarítmica. A fin de ver como se distribuyeron los tamaños de efecto se generaron gráficos de floresta de los efectos reportados según el año en que se reportaron los programas de enseñanza-aprendizaje en las tesis doctorales (ver figuras 6-13).

Tabla 11

Los veinte tamaños del efecto transformados con los valores más altos

N°	g de Hedges	I.C. al 95%	
		inferior	superior
1	3.9521	[0.1531	7.7510]
2	3.6037	[0.0819	7.1254]
3	3.3142	[0.0189	6.6095]
4	3.1784	[0.5016	5.8552]
5	2.9837	[-0.0584	6.0258]
6	2.1813	[-0.2802	4.6428]
7	2.1016	[-0.3059	4.5090]
8	1.8421	[-0.2877	3.9719]
9	1.8417	[-0.2878	3.9712]
10	1.7411	[-0.4333	3.9154]
11	1.5830	[-0.4961	3.6620]
12	1.4478	[-0.4667	3.3623]
13	1.0785	[-0.7346	2.8917]
14	1.0529	[-0.8451	2.9510]
15	0.9750	[-0.8836	2.8337]
16	0.9639	[-0.7254	2.6532]
17	0.9563	[-0.8170	2.7295]
18	0.9418	[-0.7009	2.5844]
19	0.9149	[-0.9511	2.7810]
20	0.8646	[-0.9646	2.6938]

Fuente: Creada por el autor con los resultados del estudio.

Nota: I.C. = Intervalo de confianza.

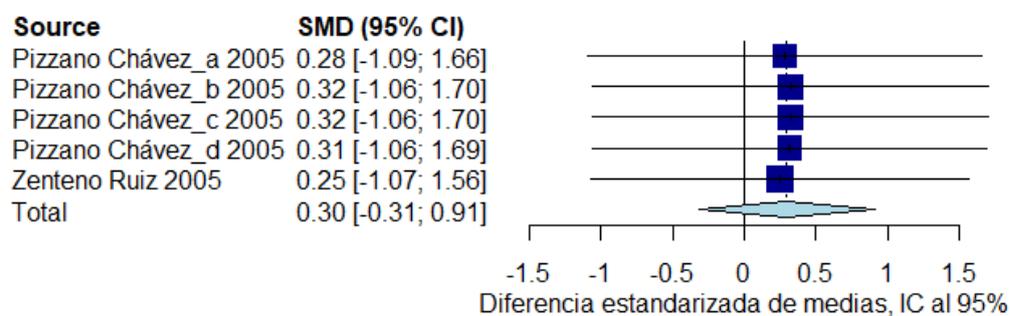


Figura 6. Gráfico de floresta de los tamaños del efecto reportados el año 2005

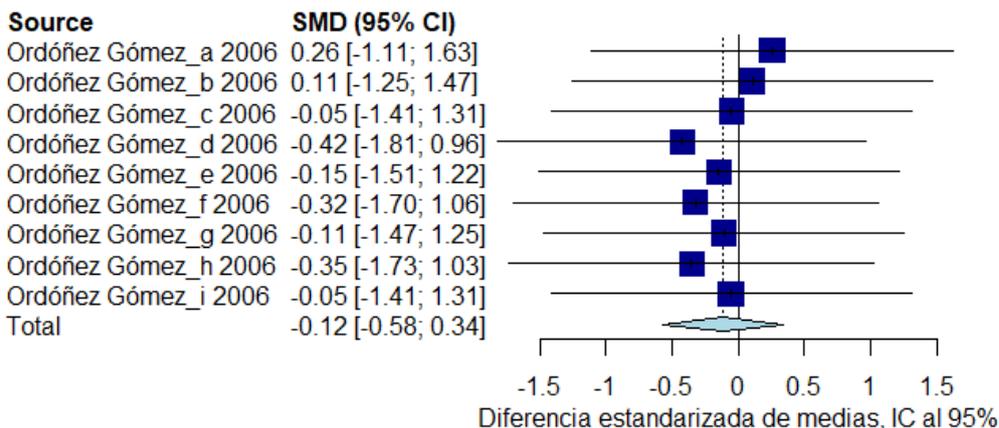


Figura 7. Gráfico de floresta de los tamaños del efecto reportados el año 2006

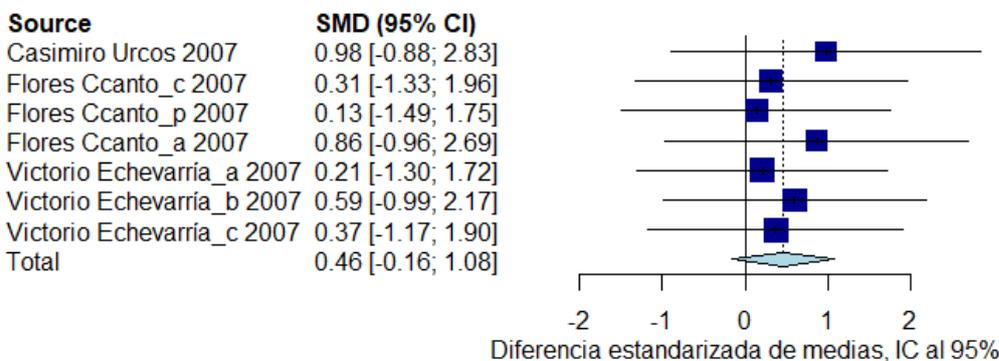


Figura 8. Gráfico de floresta de los tamaños del efecto reportados el año 2007

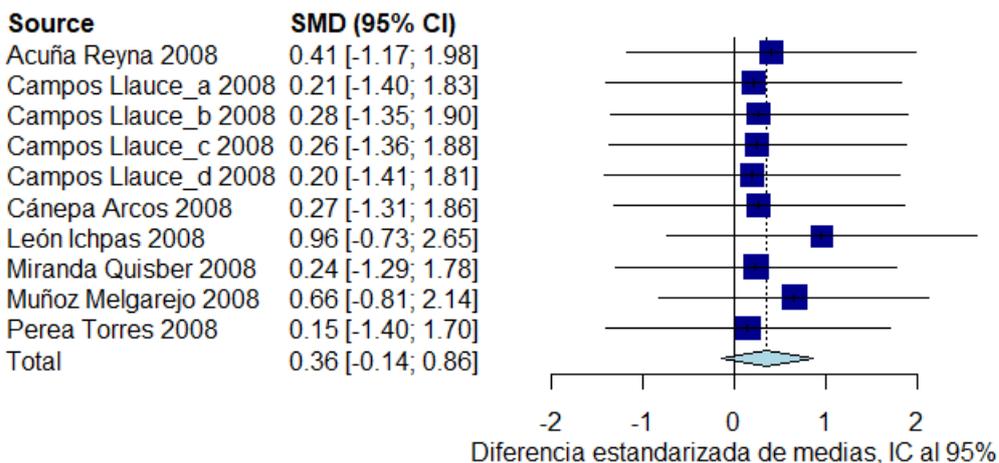


Figura 9. Gráfico de floresta de los tamaños del efecto reportados el año 2008

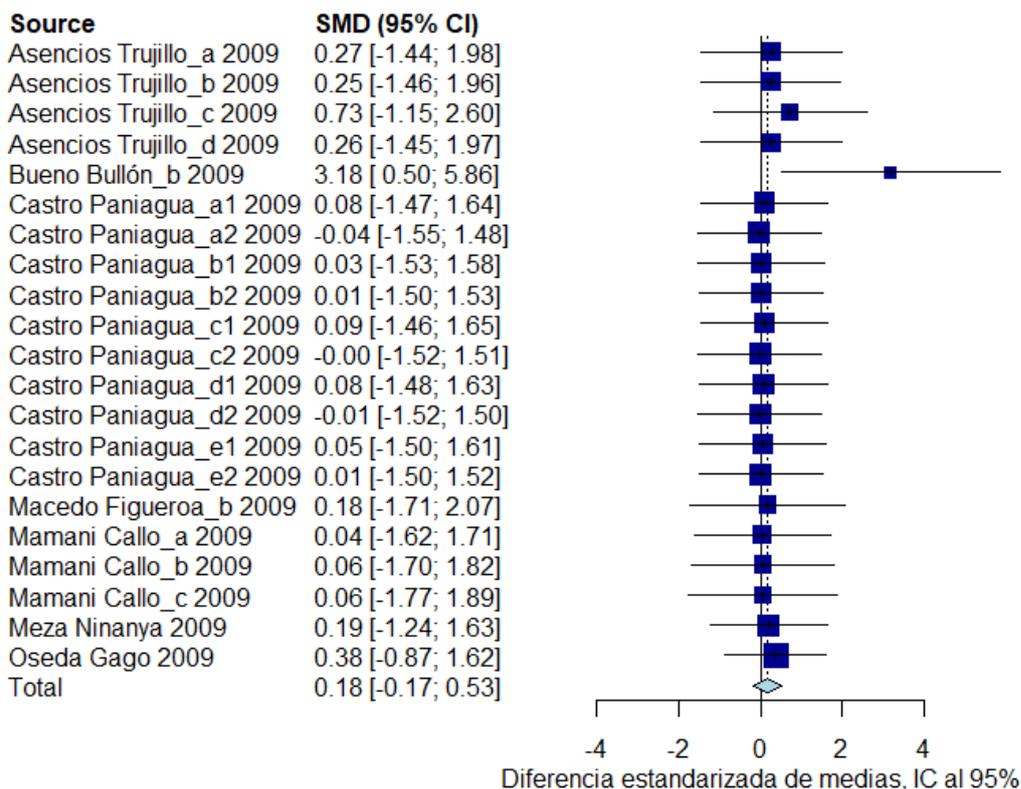


Figura 10. Gráfico de floresta de los tamaños del efecto reportados el año 2009

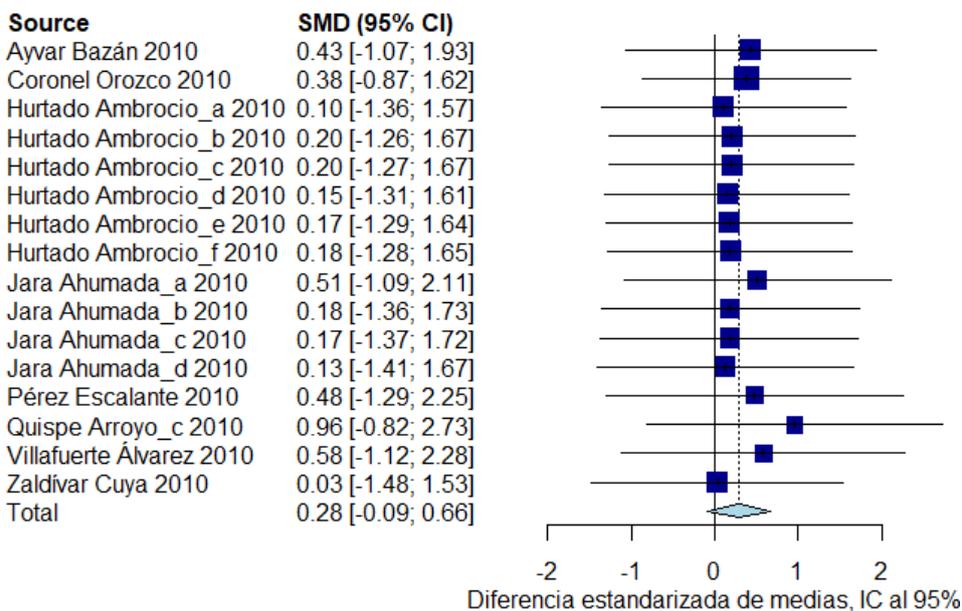


Figura 11. Gráfico de floresta de los tamaños del efecto reportados el año 2010

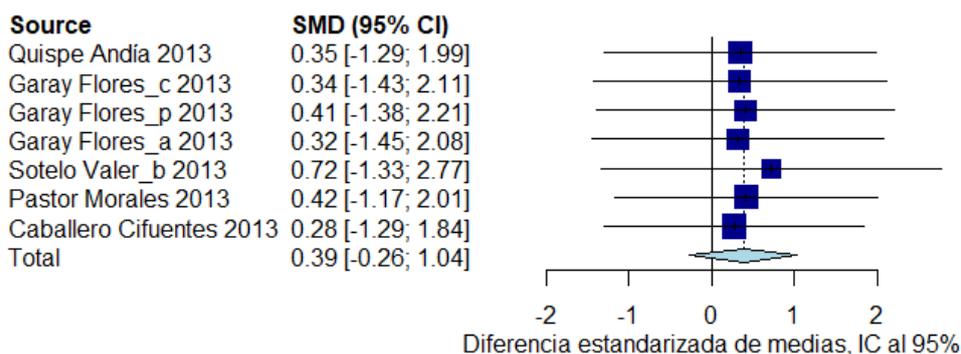


Figura 12. Gráfico de floresta de los tamaños del efecto reportados el año 2013

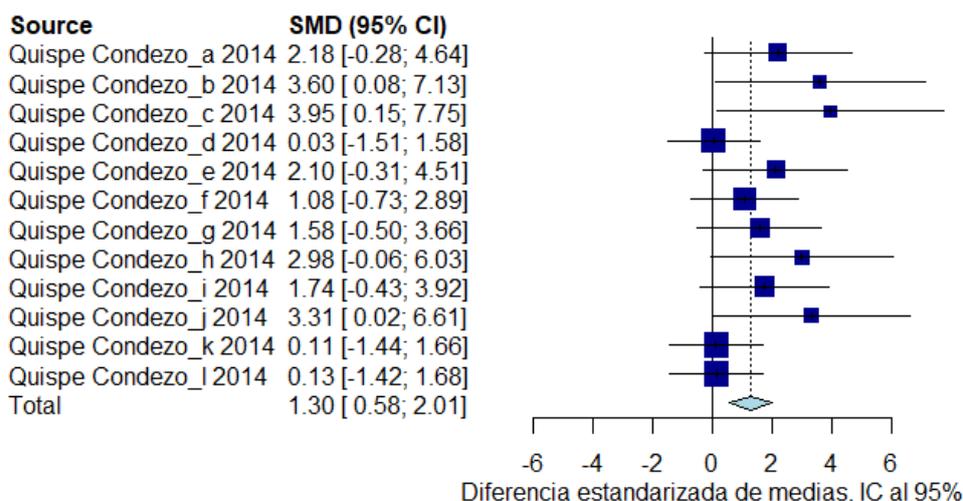


Figura 13. Gráfico de floresta de los tamaños del efecto reportados el año 2014

5.2.4. Contraste de las Hipótesis de Investigación

Para probar cada hipótesis de investigación se siguió el siguiente procedimiento: 1) se formularon las hipótesis nula y alterna, 2) se definió la regla de decisión para determinar si se aceptaba o rechazaba la hipótesis, 3) se realizó un análisis preliminar para evaluar si se seguía adelante con el contraste de hipótesis, 4) se especificó el tipo de modelo de regresión lineal para validar la hipótesis, 5) se aceptó o rechazó la hipótesis de investigación a partir de los resultados obtenidos con el análisis de regresión.

Contraste de la Hipótesis General.

Formulación de las Hipótesis

Hipótesis Nula: No existe una relación positiva entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la Universidad Nacional de Educación (UNE) 2005-2015.

Hipótesis Alternativa: Existe una relación positiva entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la Universidad Nacional de Educación (UNE) 2005-2015.

Definición de la Regla de Decisión

Si el p-valor del coeficiente F de Fisher y los coeficientes de regresión son menores a 0.05 de nivel de significancia se rechaza la hipótesis nula y se acepta la hipótesis alternativa, mientras que si el p-valor > 0.05 de nivel de significancia se acepta la hipótesis nula y se rechaza la hipótesis alternativa.

Análisis Preliminar al Contraste de Hipótesis

Se obtuvo una matriz de correlación, gráfica y tabular, de las variables incluidas en la hipótesis general para ver si procedía el análisis de regresión (ver Figura 14 y Tabla 12).

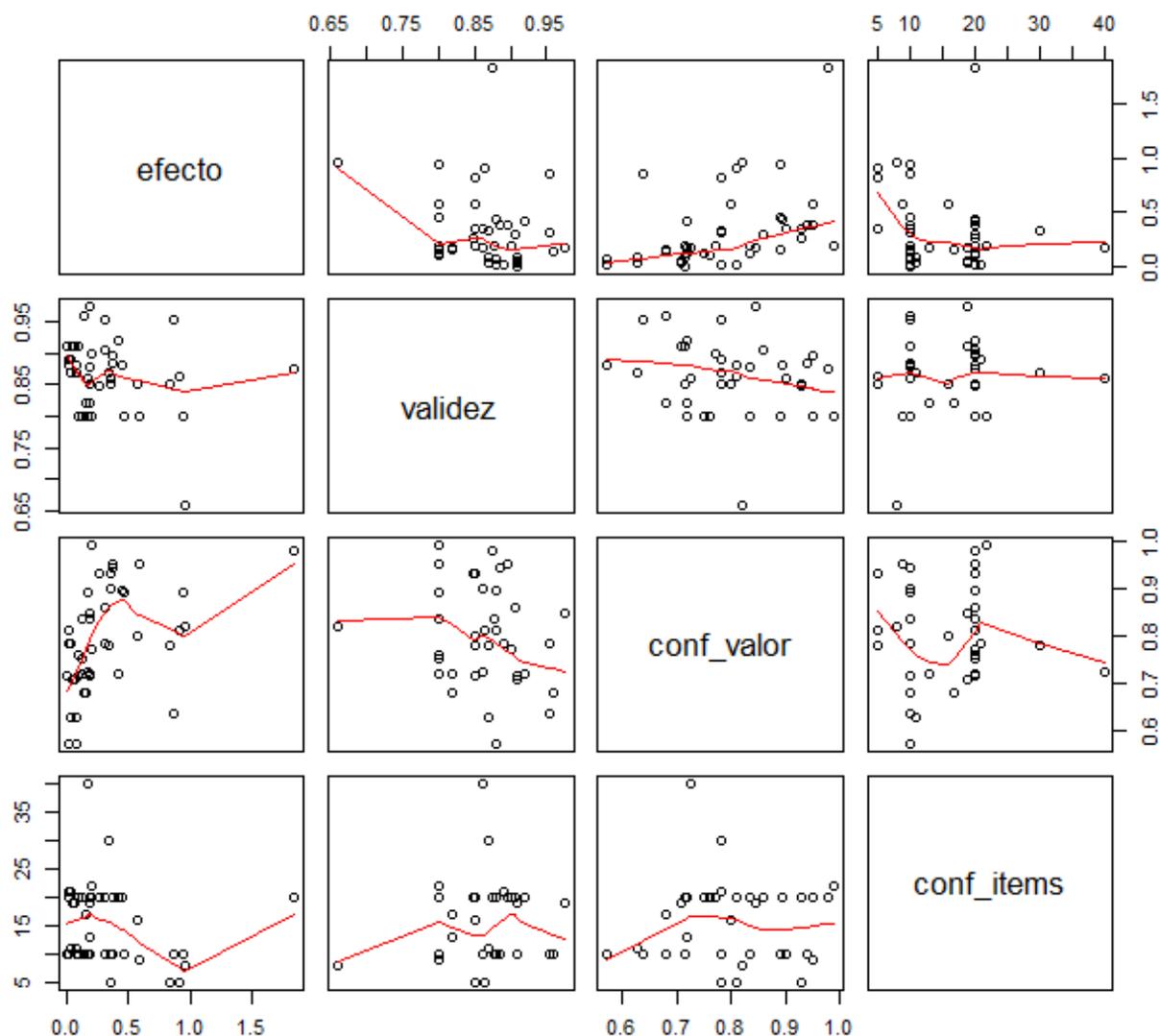


Figura 14. Matriz de correlación de las variables incluidas en la hipótesis general (n = 45)
 Nota: efecto = tamaño del efecto del programa de enseñanza-aprendizaje, validez = Magnitud de la validez de contenido del instrumento de recolección de datos, conf_valor = Magnitud del coeficiente de confiabilidad del instrumento de recolección de datos, conf_items = Número de ítems empleados para medir cada variable.

Tabla 12

Matriz de correlación de las variables incluidas en la hipótesis general (n = 45)

Variable	2.	3.	4.
1. Tamaño del efecto	-0.2041 (0.179)	0.4273 (0.003)	-0.2222 (0.142)
2. Validez de contenido		-0.2710 (0.072)	0.0755 (0.622)
3. Coeficiente de confiabilidad			0.0228 (0.882)
4. Número de ítems por variable			

Fuente: Creada por el autor con los resultados del estudio.

Nota: El p-valor del coeficiente de correlación aparece entre paréntesis.

Técnica Estadística Usada para Validar la Hipótesis

Se empleó la regresión lineal múltiple para probar la hipótesis general (ver Tabla 13).

El coeficiente de determinación R^2 ajustado = 0.186 y el F de Fisher = 4.36 ($p = 0.009$).

Tabla 13

Coefficientes del modelo de regresión lineal para probar la hipótesis general (n = 45)

Variable	Coeficiente	Error estándar	t valor	p-valor
1. Intercepto	-0.1628	0.9292	-0.18	0.8618
2. Validez de contenido	-0.4691	0.8820	-0.53	0.5977
3. Coeficiente de confiabilidad**	1.3504	0.4635	2.91	0.0058
4. Número de ítems por variable	-0.0116	0.0070	-1.65	0.1056

Fuente: Creada por el autor con los resultados del estudio.

Nota: * = significancia al 0.05, ** = significancia al 0.01, *** para significancia al 0.001.

Interpretación de los Resultados de la Prueba de Hipótesis

De acuerdo con los resultados del análisis de regresión lineal múltiple, la hipótesis general se acepta parcialmente ya que el modelo obtuvo significancia estadística ($F = 4.36$, $p = 0.009$), pero solo el coeficiente de confiabilidad logró significancia ($\beta = 1.350$, $p = 0.006$). Es decir, cada incremento de un punto del coeficiente de confiabilidad implica un aumento de 1.350 puntos del logaritmo del tamaño del efecto del programa de enseñanza-aprendizaje.

Contraste de la Primera Hipótesis Específica.

Formulación de las Hipótesis

Hipótesis Nula: No existe una relación positiva entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.

Hipótesis Alterna: Existe una relación positiva entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.

Definición de la Regla de Decisión

Si el p-valor del coeficiente de regresión es menor a 0.05 de nivel de significancia se rechaza la hipótesis nula y se acepta la hipótesis alterna, mientras que si el p-valor > 0.05 de nivel de significancia se acepta la hipótesis nula y se rechaza la hipótesis alterna.

Análisis Preliminar al Contraste de Hipótesis

Se obtuvo un gráfico de dispersión de las variables incluidas en la primera hipótesis específica y se calculó la correlación entre ellas para ver si procedía el análisis de regresión (ver Figura 15). La correlación entre el tamaño del efecto y la validez del instrumento de recolección fue -0.2041 ($p = 0.179$), lo cual revela una relación negativa que se corroboró al llevar a cabo el análisis de regresión lineal bivariada.

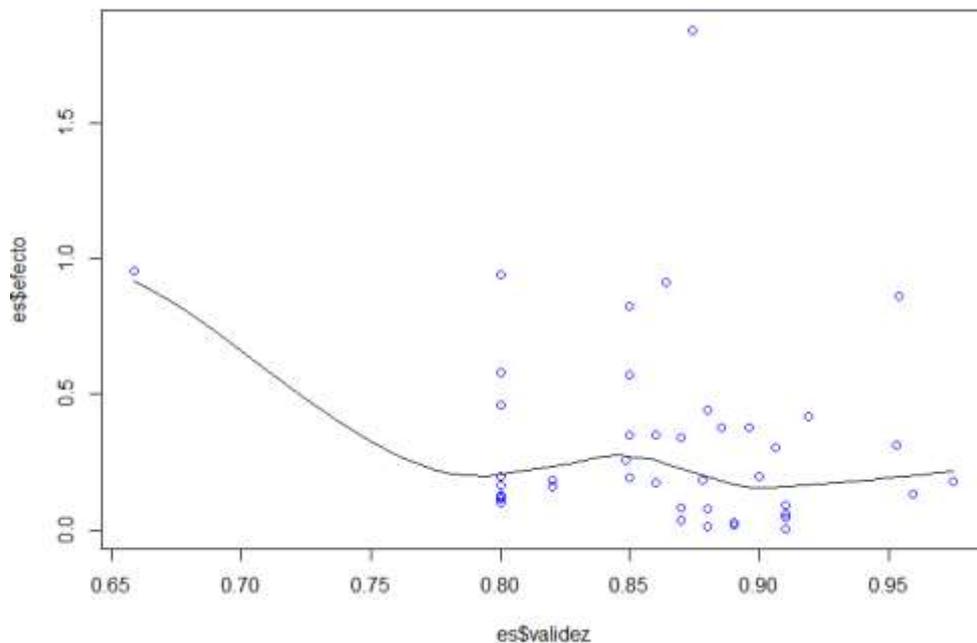


Figura 15. Gráfico de dispersión entre el tamaño del efecto y la validez del instrumento

Técnica Estadística Usada para Validar la Hipótesis

Se empleó la regresión lineal bivariada para probar la hipótesis específica (ver Tabla 14). El coeficiente de determinación R^2 ajustado = 0.019 y el F de Fisher = 1.87 ($p = 0.179$).

Tabla 14

Coefficiente del modelo de regresión lineal para probar la primera hipótesis específica

Variable	Coefficiente	Error estándar	t valor	p-valor
1. Intercepto	1.414	0.804	1.76	0.086
2. Validez de contenido	-1.270	0.929	-1.37	0.179

Fuente: Creada por el autor con los resultados del estudio.

Nota: * = significancia al 0.05, ** = significancia al 0.01, *** para significancia al 0.001.

Interpretación de los Resultados de la Prueba de Hipótesis

A partir de los resultados del análisis de regresión lineal, la primera hipótesis específica se rechaza ya que el modelo no obtuvo significancia estadística ($F = 1.87$, $p = 0.179$) y el coeficiente de validez tampoco logró significancia ($\beta = -1.270$, $p = 0.179$).

Contraste de la Segunda Hipótesis Específica.

Formulación de las Hipótesis

Hipótesis Nula: No existe una relación positiva entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.

Hipótesis Alternativa: Existe una relación positiva entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.

Definición de la Regla de Decisión

Si el p-valor del coeficiente de regresión es menor a 0.05 de nivel de significancia se rechaza la hipótesis nula y se acepta la hipótesis alternativa, mientras que si el p-valor > 0.05 de nivel de significancia se acepta la hipótesis nula y se rechaza la hipótesis alternativa.

Análisis Preliminar al Contraste de Hipótesis

Se obtuvo un gráfico de dispersión de las variables incluidas en la segunda hipótesis específica y se calculó la correlación entre ellas para ver si procedía el análisis de regresión

(ver Figura 16). La correlación entre el tamaño del efecto y el coeficiente de confiabilidad fue 0.4273 ($p = 0.003$).

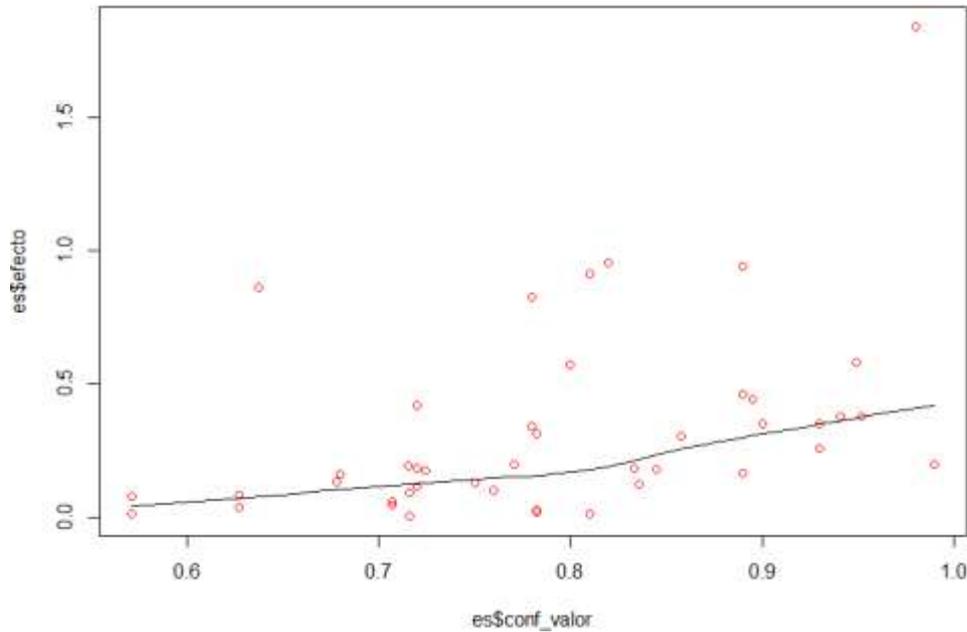


Figura 16. Gráfico de dispersión entre el tamaño del efecto y el coeficiente de confiabilidad

Técnica Estadística Usada para Validar la Hipótesis

Se empleó la regresión lineal bivariada para probar la hipótesis específica (ver Tabla 15). El coeficiente de determinación R^2 ajustado = 0.164 y el F de Fisher = 9.61 ($p = 0.003$).

Tabla 15

Coefficiente del modelo de regresión lineal para probar la segunda hipótesis específica

Variable	Coefficiente	Error estándar	t valor	p -valor
1. Intercepto*	-0.788	0.360	-2.19	0.0339
2. Coeficiente de confiabilidad**	1.401	0.452	3.10	0.0034

Fuente: Creada por el autor con los resultados del estudio.

Nota: * = significancia al 0.05, ** = significancia al 0.01, *** para significancia al 0.001.

Interpretación de los Resultados de la Prueba de Hipótesis

A partir de los resultados del análisis de regresión lineal, la segunda hipótesis específica se acepta ya que el modelo de regresión obtuvo significancia estadística ($F = 9.61$, $p = 0.003$), así como el coeficiente de confiabilidad ($\beta = 1.401$, $p = 0.003$).

Contraste de la Tercera Hipótesis Específica.

Formulación de las Hipótesis

Hipótesis Nula: El tamaño de la muestra no modera la relación entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.

Hipótesis Alternativa: El tamaño de la muestra modera la relación entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.

Definición de la Regla de Decisión

Si el p-valor del coeficiente de regresión es menor a 0.05 de nivel de significancia se rechaza la hipótesis nula y se acepta la hipótesis alternativa, mientras que si el p-valor > 0.05 de nivel de significancia se acepta la hipótesis nula y se rechaza la hipótesis alternativa.

Análisis Preliminar al Contraste de Hipótesis

Se obtuvo una matriz de correlación, gráfica y tabular, de las variables incluidas en la hipótesis específica para ver si procedía el análisis de regresión con efecto de moderación (ver Figura 17 y Tabla 16).

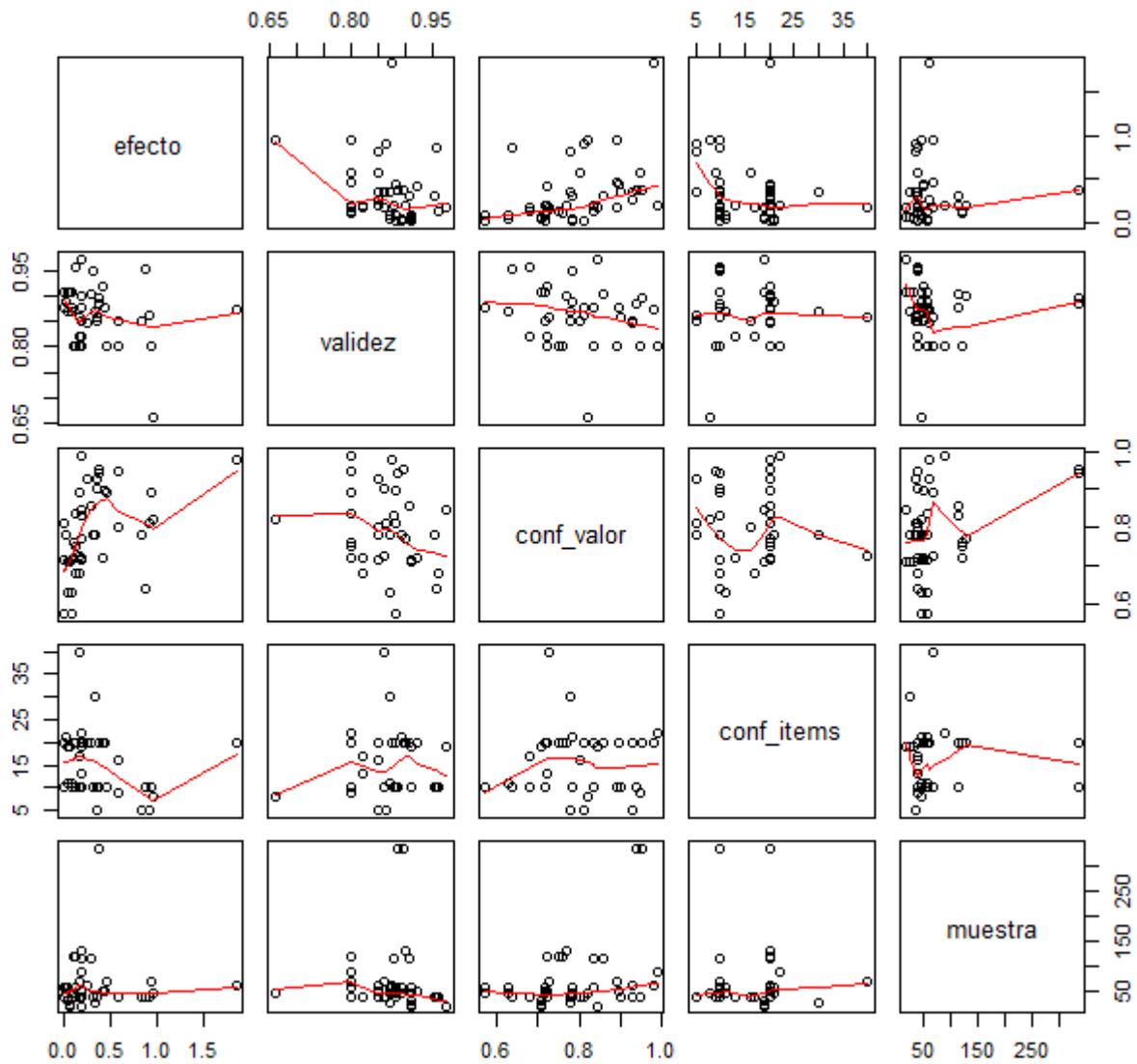


Figura 17. Matriz de correlación de las variables incluidas en la tercera hipótesis específica

Nota: efecto = Tamaño del efecto del programa de enseñanza-aprendizaje, validez = Magnitud de la validez de contenido del instrumento de recolección de datos, conf_valor = Magnitud del coeficiente de confiabilidad del instrumento de recolección de datos, conf_items = Número de ítems empleados para medir cada variable, muestra = Tamaño de la muestra.

Tabla 16

Matriz de correlación de las variables incluidas en la tercera hipótesis específica

Variable	2.	3.	4.	5.
1. Tamaño del efecto	-0.2041 (0.179)	0.4273 (0.003)	-0.2222 (0.142)	-0.0128 (0.933)
2. Validez de contenido		-0.2710 (0.072)	0.0755 (0.622)	-0.0386 (0.801)
3. Coeficiente de confiabilidad			0.0228 (0.882)	0.3419 (0.022)

4. Número de ítems por variable	0.0706 (0.645)
---------------------------------	-------------------

5. Tamaño de la muestra

Fuente: Creada por el autor con los resultados del estudio.

Nota: El p-valor del coeficiente de correlación aparece entre paréntesis.

Técnica Estadística Usada para Validar la Hipótesis

Se empleó la regresión lineal múltiple para probar la tercera hipótesis específica (ver Tabla 17). El coeficiente R^2 ajustado = 0.162 y el F de Fisher = 3.83 ($p = 0.017$).

Tabla 17

Coefficientes del modelo de regresión lineal para probar la tercera hipótesis específica

Variable	Coefficiente	Error estándar	t valor	p-valor
1. Intercepto***	0.30058	0.05425	5.54	0.0000
2. Coeficiente de confiabilidad**	1.68934	0.50021	3.38	0.0016
3. Tamaño de la muestra	-0.00184	0.00156	-1.18	0.2440
4. Efecto de moderación	0.00682	0.01065	0.64	0.5256

Fuente: Creada por el autor con los resultados del estudio.

Nota: * = significancia al 0.05, ** = significancia al 0.01, *** para significancia al 0.001.

Interpretación de los Resultados de la Prueba de Hipótesis

De acuerdo con los resultados del análisis de regresión lineal con efecto moderador, la tercera hipótesis específica se rechaza. Aunque el modelo ($F = 4.36$, $p = 0.009$) y el coeficiente de confiabilidad ($\beta = 1.6983$, $p = 0.0016$) lograron significancia estadística, ni el tamaño de muestra ni el efecto moderador alcanzaron significancia estadística.

5.3. Discusión

Con relación a los tamaños del efecto originales reportados en las tesis doctorales analizadas, la magnitud de ellos reveló una efectividad cinco o seis veces mayor (léase, un incremento de 500% o 600%) que los tamaños del efecto de las tesis doctorales en educación reportados en la literatura académica (Pigott et al., 2013; Polanin et al., 2016). En cuanto a las hipótesis de investigación, de las cuatro planteadas, la hipótesis general se aceptó de forma parcial y la segunda hipótesis específica sí se aceptó, mientras que la primera y tercera hipótesis específicas se rechazaron porque ni los coeficientes ni el modelo de regresión alcanzaron significancia estadística.

Dicha magnitud de 500% o 600% de efectividad de los programas de enseñanza-aprendizaje, comparada con los hallazgos previos, puede deberse a que los programas implementados en los centros de educación superior de Perú tienen un altísimo nivel de efectividad o ello podría explicarse por la presencia de determinados sesgos durante la recolección y análisis de los datos. Para analizar la verosimilitud del argumento de la alta efectividad de los programas de enseñanza-aprendizaje implementados en Perú fue necesario verificar algunos supuestos. Por ejemplo, tamaños de muestra medianos/grandes o un número estándar de ítems para medir las variables. Sin embargo, al revisar los resultados del estudio se constató que la mayoría de las tesis con diseño (cuasi) experimental no tuvieron el tamaño de muestra ($n \geq 80$) requerido para detectar tamaños del efecto reales (Ellis, 2012; Hallahan & Rosenthal, 1996; Maxwell et al., 2008). Por otro lado, menos del 5% de las tesis analizadas emplearon seis o menos ítems para medir cada variable, a pesar que los estándares de medición de pruebas educativas y psicológicas señalan que el nivel de confiabilidad crece con el número de ítems empleados (AERA et al., 1999; 2014).

Dado que no hay evidencia empírica para sostener el argumento de la alta efectividad de los programas de enseñanza-aprendizaje implementados en los centros de educación superior de Perú, a continuación, se explica por qué dicha efectividad sería resultado de la presencia de algunos sesgos en la recolección y análisis de datos. En ese sentido, la elevada magnitud de los tamaños del efecto obtenidos al analizar las tesis doctorales sería un producto del método de investigación empleado, no una consecuencia de haber implementado los programas de enseñanza-aprendizaje. Este crecimiento artificial se conoce como inflación de los tamaños del efecto (Ioannidis, 2008). Esta posible inflación de los tamaños del efecto se explicaría por diversas razones. En primer lugar, está el tamaño de las muestras empleadas (léase, el número de participantes en el grupo experimental y en el grupo control), porque los estudios con muestras pequeñas ($n < 40$ en cada grupo) suelen detectar efectos grandes, mientras que los estudios con muestras grandes por lo general reportan efectos más pequeños. Es decir, a medida que aumenta la muestra, el efecto se reduce.

Estudios previos han detectado la inflación de los tamaños del efecto en la investigación clínica altamente citada (Ioannidis, 2005), los ensayos clínicos sobre intervenciones en salud mental (Trikalinos et al., 2004) y problemas cardiacos (Krum y Tonkin, 2003), la investigación sobre ecología y evolución (Jennions & Moeller, 2002), así como en la investigación psicológica (Leimu & Koricheva, 2004), entre otras áreas. Si bien en la investigación educativa no existe tanta evidencia de inflación de los tamaños del efecto, como en las áreas antes señaladas, es muy probable que en las tesis doctorales analizadas se haya producido una inflación de las magnitudes de los tamaños de dichos efectos, ya que incluso con la transformación logarítmica, hubo varios efectos mayores a uno, lo cual es inusual en la investigación educativa.

En segundo lugar, existiría un posible sesgo de medición, ya que no se tiene certeza sobre el nivel real de confiabilidad de los instrumentos de recolección de datos empleados en las tesis doctorales. Si bien los autores de las tesis siguieron el procedimiento formal para calcular el coeficiente de confiabilidad usando la opción del programa estadístico de su preferencia (e.g., el programa SPSS), no tomaron en cuenta los supuestos teóricos de dichos coeficientes, ni ofrecieron métricas complementarias (por ejemplo, medir la unidimensionalidad de los ítems que forman la escala o emplear indicadores alternativos como el índice de fiabilidad compuesta o el coeficiente de confiabilidad congénica).

En tercer lugar, es posible que se haya producido un sesgo de selección durante el muestreo porque los investigadores fueron quienes crearon o implementaron los programas de enseñanza-aprendizaje reportados en las tesis doctorales. En ese sentido no se trataría de muestras representativas dado que los investigadores seleccionaron a sus estudiantes para formar parte del grupo experimental y del grupo control. Es decir, los investigadores no capacitaron a otros profesores para que implementaran el programa educativo, sino que ellos mismos lo implementaron y evaluaron, asumiendo un doble rol de diseñador/implementador y evaluador. De esta forma, los profesores que diseñaron o implementaron el programa lo hicieron de forma tan eficiente y con tanto entusiasmo que los cambios observados se explicarían por la alta motivación que tuvieron dichos docentes mientras implementaron el programa que ellos diseñaron. Es decir, el interés por demostrar la efectividad del método propuesto hizo que su desempeño fuera tan alto que lograron una efectividad muy por encima de los efectos reportados en la literatura académica. Estudios previos han reportado de qué manera las expectativas previas de los profesores producen el aprendizaje esperado en los estudiantes (Jussim & Harber, 2005; Riley & Ungerleider, 2012; Rist, 2000).

Otro argumento que refuerza la hipótesis del doble rol de diseñador y evaluador tiene que ver con el tipo de evaluación empleada por los autores de las tesis, quienes en casi todos los casos fueron quienes diseñaron los instrumentos de medición e implementaron el programa de enseñanza-aprendizaje. La revisión del número de ítems para medir las variables reveló que la mayoría tenía 20 ítems, muchos de ellos agrupados en tres dimensiones: conceptual, procedimental y actitudinal. Dicha estructura y la revisión del contenido de los ítems mostró que lo que varios autores hicieron fue convertir el examen final de su curso en un instrumento de recolección de datos. Por ello, cuando midieron los logros de aprendizaje de los miembros de los grupos experimental y control, encontraron diferencias tan grandes entre uno y otro grupo, ya que dichos instrumentos de recolección fueron diseñados específicamente para ver cuánto habían aprendido los estudiantes con el nuevo método.

En cuanto a los resultados del contraste de hipótesis, como se señaló previamente, solo se aceptó una hipótesis, una segunda se aceptó parcialmente y dos se rechazaron, de acuerdo con las reglas de decisión definidas de antemano. De las dos variables predictoras, solo la magnitud del coeficiente de confiabilidad estuvo asociada con el tamaño del efecto del programa de enseñanza-aprendizaje, probablemente porque dicha variable sí se calculó a partir de los puntajes obtenidos al aplicar el instrumento de recolección de datos, mientras que la medición de la validez reflejó la opinión de los jueces expertos antes de la aplicación del instrumento. Es decir, la validez no reflejó una propiedad métrica en sentido estricto. En lo que se refiere a la confiabilidad, solo la magnitud del coeficiente de confiabilidad estuvo relacionada con el tamaño del efecto de los programas de enseñanza-aprendizaje, pero no la cantidad de ítems para medir cada variable, un hallazgo inesperado dado que dicho elemento suele tomarse en cuenta al calcular el coeficiente de confiabilidad.

Al examinar la magnitud de los coeficientes de confiabilidad y el número de ítems para medir cada variable se detectó una posible inflación de dichos coeficientes ya que los autores de las tesis emplearon más de diez o 20 ítems (incluso en una tesis se reportó una variable medida con 47 ítems) para medir una variable. La posible inflación se debe a que los coeficientes alfa de Cronbach y Kuder-Richardson son sensibles al número de ítems empleados para medir cada variable, porque a medida que aumenta el número de ítems crece también la magnitud del coeficiente de confiabilidad. Por esa razón, en la actualidad los investigadores prefieren medir las variables con la cantidad mínima necesaria de ítems, es decir, entre tres y cinco ítems por variable. En ese sentido, variables medidas con más de seis ítems podrían estar inflando la correlación entre ellos, generando una confiabilidad producto de los instrumentos de recolección de datos empleados y no de la efectividad real del programa de enseñanza-aprendizaje.

Con relación al tamaño de la muestra, dado que es una variable vinculada con el cálculo de la potencia estadística, se esperaba que también estuviera asociada con el tamaño del efecto. El hecho que no se corroborara dicha relación puede deberse a que los tamaños de muestra empleados en las tesis doctorales no tuvieron la potencia estadística suficiente para detectar efectos reales porque la mayoría de las tesis trabajó con muestras iguales o menores a 60 participantes (30 para el grupo experimental y 30 para el grupo control).

Luego de discutir los resultados reportados en esta investigación, a continuación, se describen y analizan algunos elementos a tener en cuenta en futuros estudios, con el fin de garantizar que los hallazgos presentados en esta tesis doctoral puedan replicarse al emplear el procedimiento descrito en el capítulo cuatro.

5.4. Limitaciones

A pesar del rigor con que se llevó a cabo esta investigación, se identificaron tres limitaciones: disponibilidad de datos, alcance de los programas de enseñanza-aprendizaje y control del sesgo de selección. La primera limitación se refiere a que no todas las tesis doctorales reportaron los resultados del análisis de validez y confiabilidad de los instrumentos de recolección de datos empleados, a pesar que todas las tesis se realizaron en base a un diseño (cuasi) experimental. En muchos casos los autores de las tesis solo mencionaron que validaron su instrumento a través de jueces o usaron el coeficiente alfa de Cronbach o Kuder-Richardson, pero no reportaron el grado de acuerdo de los jueces expertos o la magnitud de dichos coeficientes de confiabilidad. Por tal motivo, si bien las tesis doctorales analizadas reportaron 153 efectos, para probar las hipótesis solo se consideraron 45 efectos, porque solo se tuvo datos completos para calcular aquellos tamaños del efecto.

Ello no quiere decir que los instrumentos de recolección no sean válidos ni confiables, sino que luego de extraer los datos de las tesis doctorales, no se contó con suficiente información para probar si en verdad tenían adecuadas propiedades métricas. En el caso de la validez, esto impidió detectar si la validez por criterio de jueces expertos tenía relación con la efectividad de los programas de enseñanza-aprendizaje implementados en los centros de educación superior de Perú. Frente a esta limitación se puede replicar este estudio incluyendo las tesis doctorales sustentadas durante 2005-2015 y que tuvieron un diseño correlacional. Para ello sería necesario transformar el coeficiente de correlación reportado en la tesis doctoral en una diferencia estandarizada de medias, siguiendo los parámetros establecidos para dicha transformación (Borenstein et al., 2009; Fritz et al., 2012; McGrath & Meyer, 2006), los que luego serían comparados con las magnitudes de la d de Cohen o g de Hedges.

La segunda limitación se relaciona con el número de programas de enseñanza-aprendizaje analizados, ya que se seleccionaron las tesis doctorales sustentadas entre los años 2005 y 2015. A fin de ver si estos hallazgos pueden generalizarse a otras tesis doctorales con diseños (cuasi) experimentales y que reportaron la implementación de programas de enseñanza-aprendizaje en centros de educación superior, en un futuro estudio podría replicarse el procedimiento presentado con tesis de similar diseño de investigación que se sustentaron entre los años 2016-2020. De esta forma, podría verse si los resultados aquí presentados pueden generalizarse a tesis doctorales similares, o podría detectarse una tendencia diferente que amerite una nueva investigación.

La tercera limitación se refiere a cómo evitar el sesgo de auto selección para reducir la inflación del tamaño del efecto. Dado que la efectividad de los programas de enseñanza-aprendizaje dependen de una cuidadosa implementación (por ejemplo, garantizando que las sesiones de aprendizaje, los materiales didácticos y el diseño instruccional en su conjunto son similares al programa original), el diseñador del programa deberá capacitar a otros docentes para que sean ellos quienes lo implementen. Además, para evaluar la efectividad de dichos programas tendrá que diseñar instrumentos de recolección de datos que no incrementen de forma artificial sus propiedades métricas; por ejemplo, limitando el número de ítems para medir cada variable (Agbo, 2010; Wilson & Gochyyev, 2013). Por otro lado, para la fase de implementación, el investigador tendrá que seleccionar aleatoriamente a los participantes del grupo experimental y control, de tal manera que el tamaño de la muestra tenga la suficiente potencia estadística para detectar efectos reales (Westfall et al., 2014). De esta manera, se lograría reducir el sesgo de selección y mejorar la efectividad de los programas de enseñanza-aprendizaje implementados en los centros de educación superior del Perú.

Conclusiones

Una de las razones de ser de los programas doctorales es generar teorías y modelos que aporten a la expansión del conocimiento de cada disciplina. A su vez, dichos modelos teóricos deben sustentarse en rigurosos procedimientos de recolección y análisis de datos, respaldados en adecuadas bases teóricas y conceptuales. Si bien los requerimientos teóricos y metodológicos requieren tiempo para madurar, el fortalecimiento de los procedimientos de recolección y análisis de datos es una meta no tan desafiante como la consolidación de los modelos teóricos. Por esa razón, esta investigación se enfocó en la efectividad de los programas de enseñanza-aprendizaje (medida a través del indicador conocido como tamaño del efecto) y su relación con las propiedades métricas de los instrumentos de recolección.

Para ello se formularon hipótesis que fueron probadas a través de la estrategia analítica diseñada para este estudio. Dicho contraste de hipótesis mostró que la hipótesis general se aceptó de manera parcial, ya que de las dos propiedades métricas solo la confiabilidad del instrumento de recolección de datos se relaciona con la efectividad de los programas de enseñanza-aprendizaje implementados en los centros de educación superior. Por otro lado, de las tres hipótesis específicas, solo una de ellas se aceptó.

Los resultados de esta investigación contribuyen al conocimiento científico ya que proporcionan evidencia empírica que muestra que la inflación de los tamaños del efecto también está presente en las investigaciones educativas, lo que corrobora un hecho reportado previamente en la literatura académica (Ioannidis, 2008, 2015). Habiéndose corroborado el aumento artificial de la efectividad de los programas de enseñanza-aprendizaje descritos en las tesis doctorales analizadas, el siguiente paso es identificar las posibles causas a fin de controlar o reducir la inflación de los tamaños del efecto.

El análisis y discusión de los resultados mostró que, si bien los tamaños del efecto obtenidos estuvieron muy por encima de lo reportado en la literatura académica, ello puede explicarse por la presencia de un sesgo de selección, el crecimiento artificial de la confiabilidad de los instrumentos de recolección y los tamaños de muestra sin la suficiente potencia estadística para detectar tamaños de efecto reales. En la medida que se tomen en cuenta las limitaciones señaladas al replicarse este enfoque analítico en estudios futuros, podrá tenerse una idea más precisa del nivel real de efectividad de los programas de enseñanza-aprendizaje reportados en las tesis doctorales de la Escuela de Posgrado de la UNE y si los resultados aquí presentados pueden generalizarse a otras tesis similares.

Recomendaciones

Dado que el estudio se enfocó en la metodología de la investigación empleada en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE, las dos recomendaciones van en esa dirección. La primera se refiere a replantear el contenido de los cursos de “Tesis II-III” (ver Tabla 18), a fin que los estudiantes doctorales comprendan los supuestos teóricos de las herramientas usadas para analizar la validez y confiabilidad de los instrumentos de recolección de datos empleados en cada estudio. Ello incluye el procedimiento para crear pruebas o analizar las diferencias entre los indicadores basados en la Teoría Clásica de los Test (TCT) y aquellos derivados de la Teoría de Respuesta al Ítem (TRI).

La segunda recomendación tiene que ver con el énfasis que debe ponerse en cómo identificar y evitar los sesgos en la investigación doctoral. Dado que el propósito de los estudios doctorales es producir trabajos de alto rigor académico que contribuyan al crecimiento del conocimiento teórico de la disciplina, ello solo será posible si los investigadores han tomado las precauciones necesarias para evitar la aparición de sesgos, a fin de lograr que las ciencias de la educación brinden nuevos aportes teóricos que sean verificados y verificables.

Tabla 18

Propuesta de temas a incluirse en los cursos de Tesis

Curso	Temas
Tesis II	Diseño y validación de instrumentos: Medición de constructos Mapa del constructo, diseño de ítems, resultados y modelo de medición Análisis de validez Validez de contenido, discriminante, convergente, de constructo Análisis de confiabilidad Consistencia interna, consistencia temporal, confiabilidad intercodificador Teoría clásica de los tests (TCT) Error estándar de medición, estimación del puntaje real, análisis de ítems Teoría de respuesta al ítem (TRI) Modelos de respuesta de ítems dicotómicos y politómicos, modelos Rasch

-
- Tesis III Métodos de análisis I:
- Regresión lineal múltiple
 - Interpretación de los coeficientes y modelos, supuestos y bondad de ajuste
 - Regresión logística
 - Interpretación de los coeficientes y modelos, estimación de parámetros
 - Interacción de variables
 - Efectos de medicación y moderación de variables
 - Análisis de conglomerados
 - Algoritmos de agrupamiento, representaciones visuales
 - Análisis factorial exploratorio
 - Extracción y rotación de factores, varianza explicada y carga factorial
 - Análisis factorial confirmatorio
 - Especificación del modelo teórico, modelo de medición y estructural
 - Introducción al análisis longitudinal de datos
 - Medidas repetidas, modelos de crecimiento, datos omitidos
- Tesis IV Métodos de análisis II:
- Modelamiento multinivel
 - Estructura jerárquica de los datos
 - Modelos de efectos fijos y aleatorios
 - Evaluación de la bondad de ajuste y adecuación del modelo
 - Modelamiento con ecuaciones estructurales
 - Modelos basados en covarianza y en mínimos cuadrados parciales
 - Evaluación del modelo de medición y modelo estructural
 - Criterios de bondad de ajuste del modelo
-

Fuente: Adaptado de Teo (2013).

Nota: Para garantizar que los estudiantes doctorales puedan practicar lo aprendido se utilizarían paquetes disponibles para el programa R y las bases de datos educativas de acceso gratuito (e.g, ECE, ERCE, TERCE, PISA, TIMSS, PIRLS)

Referencias

- Adesope, O.; Lavin, T.; Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80*(2), 207-245. doi: 10.3102/0034654310368803.
- Abgo, A. A. (2010). Cronbach's alpha: Review of limitations and associated recommendations. *Journal of Psychology in Africa, 20*(2), 233-2340. doi: 10.1080/14330237.2010.10820371.
- American Educational Research Association (2006). Standards for reporting on empirical Social Science research in AERA publications. *Educational Researcher, 35*(6), 33-40. doi: 10.3102/0013189X035006033.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1985). *Standards for educational and psychological tests*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological tests*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Anderson, G. (2002). Reflecting on research for doctoral students in Education. *Educational Researcher, 31*(7), 22-25. doi: 10.3102/0013189X031007022.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to Meta-Analysis*. West Sussex: John Wiley & Sons.
- Bustanza Choque, A. (2014). *La productividad científica de las tesis doctorales en ciencias de la educación, sustentadas en la escuela de posgrado de la Universidad Nacional de*

- Educación, Enrique Guzmán y Valle, período 2000-2010* (Tesis doctoral). Universidad Nacional de Educación, Lima, Perú. Recuperado de <http://repositorio.une.edu.pe/handle/UNE/903>.
- Butin, D. (2010). *The education dissertation: A guide for practitioner scholars*. Thousand Oaks, CA: Sage.
- Capraro, R., & Thompson, B. (2008). The educational researcher defined: What will future researchers be trained to do? *The Journal of Educational Research*, 101(4), 247-253. doi: 10.3200/JOER.101.4.247-253.
- Cencia Crispín, O. (2017). *Tesis doctorales en educación en universidades públicas de la región central del Perú 2010-2015* (Tesis doctoral). Universidad Nacional del Centro del Perú, Huancayo, Perú. Recuperado de <http://repositorio.uncp.edu.pe/handle/UNCP/4151>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2a. ed. Hillsdale, NJ: Lawrence Erlbaum.
- Cooper, H. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, 1(1), 104-126.
- Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach*. 4a. ed. Los Angeles, CA: Sage.
- Cruz Neyra, L., Toledo Espinoza, E., Mendoza Ramírez, A., & Jáuregui Quispe, A. (2010). *Análisis metodológico de las tesis de doctorado de la Universidad Nacional de Educación Enrique Guzmán y Valle*. Informe final de investigación. UNE: Instituto de Investigación.
- Davies, R. S., Howell, S. L., & Petrie, J. A. (2010). A review of trends in distance education scholarship at research universities in North America, 1998-2007.

- International Review of Research in Open and Distance Learning*, 11(3), 42-56. doi: 10.19173/irrodl.v11i3.876.
- Díaz Bazo, C., & Sime Poma, L. (2016). Las tesis de doctorado en educación en el Perú: Un perfil de la producción académica en el campo educativo. *Revista Peruana de Investigación Educativa*, 8, 5-40.
- Durak, G., Cankaya, S., Yunkul, E., & Misirli, Z. A. (2018). A content analysis of dissertations in the field of educational technology: The case of Turkey. *Turkish Online Journal of Distance Education*, 19(2), 128-148.
- Eisenhart, M., & DeHaan, R. (2005). Doctoral preparation of scientifically based education researchers. *Educational Researcher*, 34(4), 3-13. doi: 10.3102/0013189X034004003.
- Ellis, P. D. (2012). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge: Cambridge University Press.
- Ferreira-Villa, C., Pascual-García, L., & Pol-Asmarats, C. (2013). La producción española en tesis doctorales sobre orientación en la base de datos Teseo (2001-2012). *Revista Electrónica de Investigación y Evaluación Educativa*, 19(1), art. 4. doi: 10.7203/relieve.19.1.246.
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology General*, 141(1), 2-18. doi: 10.1037/a0024338.
- Granda Sandoval, A. (2013). *Doctorados: garantía para el desarrollo sostenible del Perú*. Lima: Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica.
- Hallahan, M. & Rosenthal, R. (1996). Statistical power: Concepts, procedures, and applications. *Behaviour Research and Therapy*, 34(5-6), 489-499. doi: 10.1016/0005-7967(95)00082-8.

- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Londres: Routledge.
- Ioannidis, J. P. (2008). Why most discovered true associations are inflated. *Epidemiology*, *19*(5), 640-648. doi: 10.1097/EDE.0b013e318813e7.
- Ioannidis, J. P. (2015). Contradicted and initially stronger effects in highly cited clinical research. *Journal of the American Medical Association*, *294*(2), 218-228. doi: 10.1001/jama.294.2.218.
- Jennions, M. D. & Moeller, A. P. (2002). Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution. *Proceedings of the Royal Society, London B*, *269*(1486), 43-48. doi: 10.1098/rspb.2001.1832.
- Johnson, B., & Christensen, L. (2007). *Educational research: Quantitative, qualitative, and mixed approaches*. 3a. ed. Thousand Oaks, CA: Sage.
- Johnson, E. M., Green, K. E., & Kluever, R. C. (2000). Psychometric characteristics of the Revised Procastination Inventory. *Research in Higher Education*, *41*(2), 269-270. doi: 10.1023/A:1007051423054.
- Jussim, L. & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, *9*(2), 131-155. doi: 10.1207/s15327957pspr0902_3.
- Karadağ, E. (2011). Instruments used in doctoral dissertations in educational sciences in Turkey: Quality of research and analytical errors. *Kuram ve Uygulamada Eğitim Bilimleri (Educational Sciences: Theory and Practice)*, *11*(1), 330-334.
- Krum, H. & Tonkin, A. (2003). Why do phase III trials of promising heart failure drugs often fail? The contribution of “regression to the truth”. *Journal of Cardiac Failure*, *9*(5), 364-367. doi: 10.1054/S1071-9164(03)00018-6.

- Leech, N., & Goodwin, L. (2008). Building a methodological foundation: Doctoral-level methods courses in colleges of education. *Research in the Schools, 15*(1), 1-8.
- Leimu, R. & Koricheva, J. (2004). Cumulative meta-analysis: a new tool for detection of temporal trends and publications bias in ecology. *Proceedings of the Royal Society, London B, 271*(1551), 1961-1966. doi: 10.1098/rspb.2004.2828.
- MacGrath, R. E. & Meyer, G. J. (2006). When effect sizes disagree: The case of *r* and *d*. *Psychological Methods, 11*(4), 386-401. doi: 10.1037/1082-989X.11.4.386.
- Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology, 59*, 537-563. doi: 10.1146/annurev.psych.59.103006.093735.
- Mazzi Huaycucho, V. (2017). Impacto de Amauta en la prensa minera de Morococha (1926-1930). *Utopía y Praxis Latinoamericana, 22*(77), 89-99.
- Mazzi Huaycucho, V. (2003). *La educación proletaria en los centros escolares obreros de Morococha: 1924-1930* (Tesis doctoral). Universidad Nacional de Educación Enrique Guzmán y Valle, Perú. Recuperado de <http://repositorio.une.edu.pe/handle/UNE/589>.
- Padrón, J. (2006). Acerca de la investigación educativa del nivel doctoral (testimonios de un tutor). En: Sánchez, J. (Comp.) *Investigación educativa: un compromiso para investigar y aprender con otros*. Barquisimeto: Consejo Nacional de Universidades / Oficina de Planificación del Sector Universitario / Universidad Nacional Experimental Simón Rodríguez, pp. 71-94. Recuperado de <http://padron.entretemas.com/TestimonioDeUnTutor/Index.htm>.
- Pigott, T.D., Valentine, J.C., Polanin, J.R., Williams, R.T., & Canada, D.D. (2013). Outcome-reporting bias in education research. *Educational Researcher, 42*(8), 424-432. doi: 10.3102/0013189X13507104.

- Polanin, J.R., Tanner-Smith, E.E., & Hennessy, E.A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research, 86*(1), 207-236. doi: 10.3102/0034654315582067.
- Richards, J., Dykemann, C., & Bender, S. (2016). Historical trends in counsellor education dissertations. *British Journal of Guidance & Counselling, 44*(5), 550-561. doi: 10.1080/03069885.2016.1213373.
- Riley, T. & Ungerleider, C. (2012). Self-fulfilling prophecy: How teachers' attributions, expectations, and stereotypes influence the learning opportunities afforded aboriginal students. *Canadian Journal of Education, 35*(2), 303-333.
- Rist, R. C. (2000). Student social class and teacher expectations: The self-fulfilling prophecy in Ghetto education. *Harvard Educational Review, 70*(3), 266-301.
- Rodríguez-Barranco, M., Tobías, A., Redondo, D., Molina-Portillo, E., & Sánchez, M. J. (2017). Standardizing effect size from linear regression models with log-transformed variables for meta-analysis. *BMC Medical Research Methodology, 17*, 44. doi: 10.1186/s12874-017-0322-8.
- Schulman, L. *et al.* (2006). Reclaiming education's doctorates: A critique and a proposal. *Educational Researcher, 35*(3), 25-32. doi: 10.3102/0013189X035003025.
- Slavin, R. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*(9), 5-11.
- Slavin, R. (2008). Perspectives on evidence-based research in education: What works? Issues in synthesizing educational program evaluations. *Educational Researcher, 37*(1), 5-14. doi: 10.3102/0013189X08314117.
- Torralbo Rodríguez, M., Vallejo Ruiz, M., Fernández Cano, A., & Rico Romero, L. (2004). Análisis metodológico de la producción española de tesis doctorales en educación matemática (1976-1998). *Revista Electrónica de Investigación y Evaluación*

Educativa, 10(1), 41-59. Recuperado de

http://www.uv.es/RELIEVE/v10n1/RELIEVEv10n1_3.htm.

Trikalinos, T. A., Churchill, R., Ferri, M., et al. (2004). Effect sizes in cumulative meta-analysis of mental health randomized trials evolved over time. *Journal of Clinical Epidemiology*, 57(11), 1124-1130. doi: 10.1016/j.clinepi.2004.02.018.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology General*, 143(5), 2020-2045. doi: 10.1037/xge0000014.

Wisn, M. & Gochyyev, P. (2013). Psychometrics. En: Teo, T. (Ed.). *Handbook of quantitative methods for educational research*. Rotterdam, Holanda: Sense, pp. 3-30.

Apéndices

Apéndice A. Matriz de Consistencia

Problemas	Objetivos	Hipótesis
<p>Problema general: ¿Qué relación existe entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la Universidad Nacional de Educación (UNE) 2005-2015?</p>	<p>Objetivo general: Establecer la relación que existe entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la Universidad Nacional de Educación (UNE) 2005-2015.</p>	<p>Hipótesis general: Existe una relación positiva entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la Universidad Nacional de Educación (UNE) 2005-2015.</p>
<p>Problema específico 1: ¿Qué relación existe entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015?</p>	<p>Objetivo específico 1: Establecer la relación que existe entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.</p>	<p>Primera hipótesis específica: Existe una relación positiva entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.</p>
<p>Problema específico 2: ¿Qué relación existe entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015?</p>	<p>Objetivo específico 2: Establecer la relación que existe entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.</p>	<p>Segunda hipótesis específica: Existe una relación positiva entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.</p>
<p>Problema específico 3: ¿El tamaño de la muestra modera la relación entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015?</p>	<p>Objetivo específico 3: Determinar si el tamaño de la muestra modera la relación entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.</p>	<p>Tercera hipótesis específica: El tamaño de la muestra modera la relación entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.</p>

...

Hipótesis	Definición conceptual	Definición operacional
<p>Hipótesis general: Existe una relación positiva entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la Universidad Nacional de Educación (UNE) 2005-2015.</p>	<p>Tamaño del efecto: Medida de la magnitud y dirección (positiva o negativa) de la relación que existe entre dos o más variables dentro de una muestra analizada. Los tamaños de efecto más conocidos son aquellos basados en la diferencia de medias, en los coeficientes de asociación, los coeficientes de regresión y la tasa de probabilidad</p>	<p>Tamaño de efecto: Para diseños (cuasi) experimentales: $TE (d) = \frac{\text{media del grupo experimental} - \text{media del grupo control}}{\text{desviación estándar agrupada}}$ Los indicadores más conocidos son la <i>d</i> de Cohen y la <i>g</i> de Hedges.</p>
<p>Primera hipótesis específica: Existe una relación positiva entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.</p>	<p>Validez del instrumento Es una propiedad de los datos obtenidos con los instrumentos de recolección, la cual permite saber en qué medida dichos instrumentos realmente miden aquello para lo que fueron diseñados.</p>	<p>Validez del instrumento Grado de acuerdo promedio de los jueces sobre la pertinencia del contenido del instrumento, expresado en numeración decimal. De esta forma, un acuerdo promedio de 85% se registró como 0.85.</p>
<p>Segunda hipótesis específica: Existe una relación positiva entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la confiabilidad de los instrumentos de recolección de datos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.</p>	<p>Confiabilidad del instrumento Es una propiedad de los datos obtenidos con los instrumentos de recolección que permite saber con qué precisión o con grado de estabilidad dichos instrumentos miden aquello que debieran medir.</p>	<p>Confiabilidad del instrumento Tipo de coeficiente (e.g., Alfa de Cronbach, Küder-Richardson, mitades partidas, retest) y magnitud del mismo en un rango que va de 0.001 a 0.999.</p>
<p>Tercera hipótesis específica: El tamaño de la muestra modera la relación entre el tamaño del efecto de los programas de enseñanza-aprendizaje y la validez y confiabilidad de los instrumentos en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015.</p>	<p>Tamaño de la muestra: Cantidad de personas que fueron seleccionadas para el estudio, a partir de un tipo de muestreo y tomando en cuenta los criterios inclusión y exclusión previamente definidos para cada investigación.</p>	<p>Tamaño de la muestra: Dado que solo se consideró diseños (cuasi) experimentales, el tamaño de la muestra fue igual al número de participantes en total (tanto en el grupo experimental como en el grupo control).</p>

Nombre de la variable o atributo	Opciones de respuesta (Operacionalización)
Promedio del grupo de comparación 1	Promedio al aplicarse el instrumento de recolección
D.E. del grupo de comparación 1	Desviación estándar del promedio logrado por el grupo
Tamaño del grupo de comparación 1	Número de participantes en el grupo
Promedio del grupo de comparación 2	Promedio al aplicarse el instrumento de recolección
D.E. de grupo de comparación 2	Desviación estándar del promedio logrado por el grupo
Tamaño del grupo de comparación 2	Número de participantes en el grupo
Promedio del grupo experimental	Promedio al aplicarse el instrumento de recolección
D.E. del grupo experimental	Desviación estándar del promedio logrado por el grupo
Tamaño del grupo experimental	Número de participantes en el grupo
Promedio del grupo control	Promedio al aplicarse el instrumento de recolección
D.E. del grupo control	Desviación estándar del promedio logrado por el grupo
Tamaño del grupo control	Número de participantes en el grupo
Tamaño del grupo del análisis correlación	Número de participantes en el estudio de correlación
Valor del coeficiente de correlación	Valor obtenido por el coeficiente en el rango [-1, +1]
Coeficiente Chi cuadrado	Valor obtenido por el coeficiente de asociación χ^2
Grados de libertad del Chi cuadrado	Grados de libertad

Fuente: Elaborado por el autor, en base a la ficha de datos elaborada por Adesopa et al. (2010)

Apéndice C. Ficha Técnica del Instrumento Original de Recolección de Datos

Categorías	Variables codificadas
Identificación del estudio	Autor(es) Título del artículo Fecha de publicación Revista, volumen, número y páginas Afiliación institucional de cada autor
Características y resultados medidos	Foco del estudio Prácticas instruccionales del inglés como segunda lengua Beneficios cognitivos del bilingüismo País donde se realizó el estudio Australia/Nueva Zelanda, Canadá, Estados Unidos, Reino Unido, otro Variables medidas Habilidades para resolver problemas Pensamiento creativo y divergente Flexibilidad cognitiva Representación abstracta y habilidades de razonamiento Estrategias de aprendizaje Conciencia metalingüística Habilidades metacognitivas Memoria operativa Rango de edad de los participantes Grados K-3 (5-8 años de edad) Grados 4-7 (9-12 años de edad) Grados 8-12 (13-18 años de edad) Rango longitudinal Ubicación del estudio Clase, ambiente externo, laboratorio, otro
Preguntas de investigación	¿Cuál es la pregunta amplia que el investigador busca abordar? ¿Cuál es la pregunta de investigación planteada por el autor del estudio? ¿Cuál es la hipótesis de investigador planteada por el autor del estudio? ¿Cuál es la base empírica o teórica de esta investigación?
Diseños y métodos	Variable incluidas Variables independientes (listar las variables moderadoras) Variables dependientes (listar las variables a ser explicadas) Medidas empleadas Prueba estandarizada (indicar el nombre de la prueba utilizada) Prueba elaborada por el profesor (describir la prueba) Observación (especificar la ficha de registro)
Grupos y aleatorización	Diseño de investigación No aleatorizado con grupos de tratamiento y control (cómo se asignó) Medidas repetidas (la misma muestra de persona se mide repetidamente) Forma en que los grupos difieren Señalado de forma explícita o implícita, no aplicable, no señalado Número de grupos Uno, dos, tres, cuatro o más, otro, no aplica
Estrategia de muestreo	Los autores tratan de obtener resultados representativos de una población Marco muestral del cual se seleccionan a los participantes Método empleado para seleccionar los participantes a partir del marco

Características de la muestra	Número total de participantes en el estudio (la muestra real) Sexo de los participantes en la muestra real Status socioeconómico de los participantes en la muestra real Grupo étnica de los participantes en la muestra real Información sobre necesidades de educación especial de los participantes Características regionales de los individuos/grupos en la muestra real Autores describen estrategias usadas para controlar los sesgos Información (o características) adicional(es) sobre la muestra Número de participantes que abandonaron antes de terminar el estudio
Reclutamiento y consentimiento	Métodos usados para reclutar a los participantes Carta de invitación, llamada telefónica, no aplicable, no especifica, otro Incentivos proporcionados para reclutar a los participantes Señalado de forma explícita o implícita, no aplicable, no señalado De quien se buscó el consentimiento Participantes, padres, otros, no se buscó, no se especifica
Recolección de datos	Métodos usados para recolectar los datos Experimental, evaluación basada en el currículo, grupo focal, entrevista grupal, entrevista uno a uno, observación, cuestionario de autorreporte, exámenes, prueba clínica, prueba práctica, prueba psicológica, escenario hipotético, registro escolar, datos secundarios, no se especifica, otro Quien recolectó los datos Investigador, director, profesor, padres, estudiantes, funcionarios, otro
Análisis de datos	Método de análisis empleados en el estudio Descriptivo, correlación, diferencias de grupo (e.g., prueba <i>t</i> , ANOVA, etc.), análisis de la curva de crecimiento y modelamiento multinivel, modelamiento con ecuaciones estructurales, análisis de ruta, regresión, curva de crecimiento latente, otro
Resultados y conclusión	Resultados reportados por el autor del estudio Media de los grupos, desviación estándar, número de participantes Tamaño de efecto estimado, error estándar del tamaño de efecto Estadístico <i>t</i> o <i>F</i> Nivel de significancia Peso inverso de la varianza
Estudios analizados	63 estudios (a partir de 39 artículos de revista)
Muestra agregada	6,022 participantes
Confiabilidad intercodificador	α de Krippendorff = 0.92

Fuente: Adesopa et al. (2010)

Apéndice D. Validación del Instrumento de Recolección de Datos



UNIVERSIDAD NACIONAL DE EDUCACIÓN
Enrique Guzmán y Valle
Alma Máter del Magisterio Nacional
ESCUELA DE POSGRADO

FICHA DE VALIDACIÓN DEL INSTRUMENTO

DATOS GENERALES	
Nombres y apellidos del experto	Sergio Chión Chacón Grado académico: Doctor
Cargo e institución donde labora	Profesor investigador en CENTRUM Católica Graduate Business School (CCGBS)
Título de la investigación	Metodología de la investigación en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015
Autor del instrumento	Carlos Manuel Vilchez Román
Nombre del instrumento	Ficha de datos del diseño de investigación

Indicador	Criterios	Deficiente 0-20%				Regular 21-40%				Bueno 41-60%				Muy bueno 61-80%				Excelente 81-100%			
		0-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65	66-70	71-75	76-80	81-85	86-90	91-95	96-100
1. Tipo de diseño	Cuasi experimental, correlacional, descriptivo, histórico, filosófico, caso																	✓			
2. Instrumento	Cuestionario, prueba estandarizada, currículo, prueba ad-hoc, entrevista																		✓		
3. Muestreo	Bola de nieve, cuota, intencional, aleatorio simple, estratificado, conglomerado																	✓			
4. Muestra	Número de participantes																			✓	
5. Validez	Aparente, Constructo, Jueces, Predictiva, N.A.																				✓
6. Confiabilidad	Consistencia interna, test-retest, intercodificador																				✓
7. Asociación	χ^2 , r de Pearson, ρ																		✓		
8. Promedio	Promedio de cada grupo																				✓
9. Variación	D.E. del promedio grupal																				✓
10. Tamaño	Participantes por grupo																				✓

Nota: El detalle sobre los ítems de la columna **Indicador** se brindan en el Apéndice B. Instrumento de recolección de datos.

Promedio de valoración 95 %
 Opinión de aplicabilidad Es aplicable No es aplicable
 Lugar y fecha Lima, 15 de mayo de 2019

Sergio Chión Chacón

 Firma del experto evaluador
 DNI/CE: 08240294



UNIVERSIDAD NACIONAL DE EDUCACIÓN
Enrique Guzmán y Valle
Alma Máter del Magisterio Nacional
ESCUELA DE POSGRADO

FICHA DE VALIDACIÓN DEL INSTRUMENTO

DATOS GENERALES		
Nombres y apellidos del experto	Rubén Guevara Mónica	Grado académico: Doctor
Cargo e institución donde labora	Profesor en CENTRUM Católica Graduate Business School (CCGBS)	
Título de la investigación	Metodología de la investigación en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015	
Autor del instrumento	Carlos Manuel Vilchez Román	
Nombre del instrumento	Ficha de datos del diseño de investigación	

Indicador	Criterios	Deficiente 0-20%				Regular 21-40%				Bueno 41-60%				Muy bueno 61-80%				Excelente 81-100%			
		0-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65	66-70	71-75	76-80	81-85	86-90	91-95	96-100
1. Tipo de diseño	Cuasi experimental, correlacional, descriptivo, histórico, filosófico, caso																			X	
2. Instrumento	Cuestionario, prueba estandarizada, currículo, prueba ad-hoc, entrevista																				X
3. Muestreo	Bola de nieve, cuota, intencional, aleatorio simple, estratificado, conglomerado																			X	
4. Muestra	Número de participantes																			X	
5. Validez	Aparente, Contenido, Constructo, Predictiva																			X	
6. Confiabilidad	Consistencia interna, test-retest, intercodificador																			X	
7. Asociación	χ^2 , r de Pearson, p																			X	
8. Promedio	Promedio de cada grupo																			X	
9. Desv. Estand.	D.E. del promedio grupal																				X
10. Tamaño	Participantes por grupo																			X	

Nota: El detalle sobre los ítems de la columna Indicador se brindan en el Apéndice B: Instrumento de recolección de datos.

Promedio de valoración 95 %

Opinión de aplicabilidad Es aplicable [] No es aplicable

Lugar y fecha Lima, 15 de mayo de 2019

Rubén Guevara
 Firma del experto evaluador

DNI/CE: 48642063



UNIVERSIDAD NACIONAL DE EDUCACIÓN
Enrique Guzmán y Valle
Alma Máter del Magisterio Nacional
ESCUELA DE POSGRADO

FICHA DE VALIDACIÓN DEL INSTRUMENTO

DATOS GENERALES	
Nombres y apellidos del experto	Alexander Jean-Luc Coad Grado académico: Doctor
Cargo e institución donde labora	Profesor investigador en CENTRUM Católica Graduate Business School (CCGBS)
Título de la investigación	Metodología de la investigación en las tesis doctorales sustentadas en la Escuela de Posgrado de la UNE 2005-2015
Autor del instrumento	Carlos Manuel Vilchez Román
Nombre del instrumento	Ficha de datos del diseño de investigación

Indicador	Criterios	Deficiente 0-20%				Regular 21-40%				Bueno 41-60%				Muy bueno 61-80%				Excelente 81-100%			
		0-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65	66-70	71-75	76-80	81-85	86-90	91-95	96-100
1. Tipo de diseño	Cuasi experimental, correlacional, descriptivo, histórico, filosófico, caso																			X	
2. Instrumento	Cuestionario, prueba estandarizada, currículo, prueba ad-hoc, entrevista																			X	
3. Muestreo	Bola de nieve, cuota, intencional, aleatorio simple, estratificado, conglomerado																				X
4. Muestra	Número de participantes																			X	
5. Validez	Aparente, Contenido, Constructo, Predictiva																	X			
6. Confiabilidad	Consistencia interna, test-retest, intercodificador																		X		
7. Asociación	χ^2 , r de Pearson, p																				X
8. Promedio	Promedio de cada grupo																	X			
9. Desv. Estand.	D.E. del promedio grupal																		X		
10. Tamaño	Participantes por grupo																		X		

Nota: El detalle sobre los ítems de la columna **Indicador** se brindan en el Apéndice B: Instrumento de recolección de datos.

Promedio de valoración 93 %
 Opinión de aplicabilidad Es aplicable [] No es aplicable
 Lugar y fecha Lima, 3/6 de mayo de 2019

.....
 Firma del experto evaluador
 DNI/CE: 001544887

Apéndice E. Datos Registrados de las Tesis Doctorales

autor	año	validez	conf_tipo	conf_valor	conf_it	muestra	exp_m	exp_de	exp_n	cont_m	cont_de	cont_n
Pizzano Chávez_a	2005	No indica	No indica	.	20	126	60.080	1.670	63	49.240	2.040	63
Pizzano Chávez_b	2005	No indica	No indica	.	20	126	134.660	4.510	63	85.740	2.330	63
Pizzano Chávez_c	2005	No indica	No indica	.	20	126	52.070	2.030	63	40.790	1.860	63
Pizzano Chávez_d	2005	No indica	No indica	.	20	126	92.140	1.870	63	73.180	1.940	63
Zenteno Ruiz	2005	Opinión de expertos	Alfa	.	20	195	13.790	3.017	127	10.570	2.121	68
Ordóñez Gómez_a	2006	Predictiva	Pearson	0.891	15	127	29.000	1.900	64	21.120	3.990	63
Ordóñez Gómez_b	2006	Predictiva	No indica	.	.	127	26.310	3.820	64	22.050	4.400	63
Ordóñez Gómez_c	2006	Predictiva	No indica	.	.	127	34.260	10.250	64	39.060	9.040	63
Ordóñez Gómez_d	2006	Predictiva	No indica	.	.	127	10.320	2.580	64	16.220	2.500	63
Ordóñez Gómez_e	2006	Predictiva	No indica	.	.	127	13.180	4.150	64	16.110	2.530	63
Ordóñez Gómez_f	2006	Predictiva	No indica	.	.	127	11.250	2.850	64	16.580	2.920	63
Ordóñez Gómez_g	2006	Predictiva	No indica	.	.	127	14.400	4.620	64	16.930	2.900	63
Ordóñez Gómez_h	2006	Predictiva	No indica	.	.	127	4.670	1.550	64	6.450	2.850	63
Ordóñez Gómez_i	2006	Predictiva	No indica	.	.	127	6.040	2.560	64	6.390	2.530	63
Casimiro Urcos	2007	Opinión de expertos	Alfa	0.940	20	40	14.200	1.110	20	11.700	1.220	20
Flores Ccanto_c	2007	0.9528	Alfa	0.782	10	38	13.890	1.823	19	9.890	2.865	19
Flores Ccanto_p	2007	0.9594	Alfa	0.678	10	38	34.000	3.512	19	27.370	3.730	19
Flores Ccanto_a	2007	0.9542	Alfa	0.637	10	38	10.740	3.212	19	2.420	4.682	19
Victorio Echevarría_a	2007	Opinión de expertos	Kuder-Richardson	.	.	60	15.900	1.937	30	13.6000	1.7880	30
Victorio Echevarría_b	2007	Opinión de expertos	Kuder-Richardson	.	.	60	15.370	1.661	30	11.0700	1.5070	30
Victorio Echevarría_c	2007	Opinión de expertos	Kuder-Richardson	.	.	60	15.100	1.912	30	11.8600	1.5350	30
Acuña Reyna	2008	Opinión de expertos	Kuder-Richardson	0.891	9	52	16.080	1.215	26	11.700	2.430	26
Campos Llauce_a	2008	Opinión de expertos	Test-retest	.	42	40	29.350	5.700	20	18.300	6.320	20
Campos Llauce_b	2008	Opinión de expertos	No indica	.	.	40	44.400	8.350	20	22.050	6.900	20

Campos Llauce_c	2008	Opinión de expertos	No indica	.	.	40	29.950	6.180	20	16.650	6.010	20
Campos Llauce_d	2008	Opinión de expertos	No indica	.	.	40	39.300	7.360	20	23.750	7.390	20
Cánepa Arcos	2008	Opinión de expertos	Pearson	0.854	47	46	78.541	5.552	23	44.494	5.201	23
Fernández Macedo	2008	Opinión de expertos	Mitades partidas	0.800	20	40	13.550	1.432	20	10.450	0.887	20
León Ichpas	2008	No indica	No indica	.	20	62	17.320	1.628	34	9.540	1.732	28
Miranda Quisber	2008	Opinión de expertos	No indica	.	.	56	13.180	1.982	28	10.540	2.301	28
Muñoz Melgarejo	2008	Opinión de expertos	No indica	0.840	30	100	15.380	1.380	50	12.740	1.140	50
Muñoz Salazar	2008	0.8500	Alfa	0.851	36	82	15.519	0.997	39	13.734	1.180	43
Palacios Villanes_a	2008	0.9556	Test-retest	.	21	86	16.200	2.700	49	7.350	0.490	37
Palacios Villanes_b	2008	0.9556	No indica	.	21	86	19.900	1.960	49	7.620	0.640	37
Palacios Villanes_c	2008	0.9556	No indica	.	21	86	20.490	1.500	49	7.050	0.230	37
Perea Torres	2008	No indica	Alfa	0.730	6	50	81.080	3.650	25	63.600	3.690	25
Asencios Trujillo_a	2009	Opinión de expertos	Alfa	0.636	20	30	18.067	3.011	15	13.333	1.667	15
Asencios Trujillo_b	2009	Opinión de expertos	Alfa	0.636	20	30	17.267	2.865	15	13.533	1.125	15
Asencios Trujillo_c	2009	Opinión de expertos	Alfa	0.636	20	30	19.000	1.690	15	13.200	1.146	15
Asencios Trujillo_d	2009	Opinión de expertos	Alfa	0.636	20	30	17.333	2.992	15	12.933	1.624	15
Bueno Bullón_a	2009	No indica	No indica	.	.	96	4.460	0.930	58	2.920	1.200	38
Bueno Bullón_b	2009	No indica	No indica	.	.	96	3.810	1.058	58	3.080	1.061	38
Bueno Bullón_c	2009	No indica	No indica	.	.	96	3.914	0.915	58	2.895	0.882	38
Bueno Bullón_d	2009	No indica	No indica	.	.	96	3.741	0.975	58	3.000	0.858	38
Bueno Bullón_e	2009	No indica	No indica	.	.	96	4.103	0.712	58	3.000	0.725	38
Bueno Bullón_f	2009	No indica	No indica	.	.	96	4.121	0.672	58	2.842	0.629	38
Bueno Bullón_g	2009	No indica	No indica	.	.	96	2.724	0.518	58	1.711	0.685	38
Castro Paniagua_a1	2009	0.8700	Test-retest	0.627	11	48	31.958	3.169	24	28.708	2.579	24
Castro Paniagua_a2	2009	0.8700	Test-retest	0.627	11	58	28.536	4.443	28	30.400	4.288	30
Castro Paniagua_b1	2009	0.8900	Test-retest	0.782	21	48	56.792	9.686	24	53.167	6.141	24
Castro Paniagua_b2	2009	0.8900	Test-retest	0.782	21	58	54.607	5.763	28	52.733	9.082	30
Castro Paniagua_c1	2009	0.9100	Test-retest	0.716	10	48	30.583	6.128	24	25.500	3.934	24

Castro Paniagua_c2	2009	0.9100	Test-retest	0.716	10	58	27.000	4.698	28	27.200	3.699	30
Castro Paniagua_d1	2009	0.8800	Test-retest	0.571	10	48	29.250	4.445	24	25.417	4.221	24
Castro Paniagua_d2	2009	0.8800	Test-retest	0.571	10	58	27.357	3.955	28	27.933	3.629	30
Castro Paniagua_e1	2009	No indica	Test-retest	0.710	20	48	55.958	8.317	24	49.083	7.729	24
Castro Paniagua_e2	2009	No indica	Test-retest	0.710	20	58	53.286	7.630	28	52.467	6.730	30
Feria Macizo_a	2009	0.7590	No indica	.	.	40	4.800	0.700	20	2.600	0.500	20
Feria Macizo_b	2009	0.7590	No indica	.	.	40	4.800	0.900	20	2.600	0.800	20
Feria Macizo_c	2009	0.7590	No indica	.	.	40	5.000	0.800	20	2.600	0.700	20
Macedo Figueroa_a	2009	0.9750	Alfa	0.845	19	41	16.420	0.902	19	14.360	3.402	22
Macedo Figueroa_b	2009	0.9750	Alfa	0.845	19	19	14.000	1.500	8	10.090	4.780	11
Mamani Callo_a	2009	0.9100	Kuder-Richardson	0.707	19	32	43.500	3.055	16	40.810	3.080	16
Mamani Callo_b	2009	0.9100	Kuder-Richardson	0.707	19	24	44.080	2.644	12	39.910	4.350	12
Mamani Callo_c	2009	0.9100	Kuder-Richardson	0.707	19	20	45.200	2.097	10	41.400	3.530	10
Meza Ninanya	2009	0.8000	Alfa	0.990	22	88	74.240	6.310	41	45.280	11.690	47
Oseda Gago	2009	0.8960	Alfa	0.952	20	335	13.070	2.240	167	9.530	2.030	168
Ayvar Bazán	2010	0.9500	No indica	.	.	72	15.470	1.810	36	12.080	1.440	36
Coronel Orozco	2010	0.8852	Kuder-Richardson	0.941	10	335	13.070	2.240	167	9.530	2.030	168
Hurtado Ambrocio_a	2010	No indica	Alfa	0.720	.	74	23.650	1.650	37	21.140	3.270	37
Hurtado Ambrocio_b	2010	No indica	Alfa	0.720	.	74	13.950	1.370	37	11.970	2.320	37
Hurtado Ambrocio_c	2010	No indica	Alfa	0.720	.	74	11.650	1.860	37	9.650	2.600	37
Hurtado Ambrocio_d	2010	No indica	Alfa	0.720	.	74	16.410	2.060	37	13.970	2.880	37
Hurtado Ambrocio_e	2010	No indica	Alfa	0.720	.	74	14.620	1.480	37	12.840	2.280	37
Hurtado Ambrocio_f	2010	No indica	Alfa	0.720	.	74	14.050	1.470	37	11.920	2.720	37
Jara Ahumada_a	2010	No indica	No indica	.	.	52	16.150	1.405	26	12.190	1.721	26
Jara Ahumada_b	2010	No indica	No indica	.	.	52	44.580	4.254	26	30.440	7.147	26
Jara Ahumada_c	2010	No indica	No indica	.	.	52	21.190	4.673	26	15.540	3.982	26
Jara Ahumada_d	2010	No indica	No indica	.	.	52	16.440	2.022	26	14.040	3.470	26
Pérez Escalante	2010	No indica	No indica	.	.	30	15.100	1.248	15	12.000	1.600	15

Quispe Arroyo_a	2010	0.6587	Kuder-Richardson	0.740	8	48	15.300	0.400	24	9.500	0.800	24
Quispe Arroyo_b	2010	0.6587	Kuder-Richardson	0.750	14	48	25.800	1.500	24	15.200	0.800	24
Quispe Arroyo_c	2010	0.6587	Alfa	0.820	8	48	28.600	2.700	24	10.500	2.000	24
Quispe Arroyo_d	2010	0.6587	Kuder-Richardson	0.760	8	48			24			24
Quispe Arroyo_e	2010	0.6587	Kuder-Richardson	0.750	14	48			24			24
Quispe Arroyo_f	2010	0.6587	Alfa	0.770	8	48			24			24
Villafuerte Álvarez	2010	0.8000	Kuder-Richardson	0.949	9	40	5.950	1.638	20	3.600	2.326	20
Zaldívar Cuya	2010	Opinión de expertos	Alfa	.	.	60	178.000	15.330	30	162.670	15.330	30
Pujay Cristóbal	2011	0.9000	Alfa	0.771	20	129	12.360	2.704	55	10.070	2.284	74
Mitma Mamani_a	2011	0.8130	Alfa	0.959	20	34	5.236	0.539	17	4.361	0.516	17
Mitma Mamani_b	2011	0.8130	No indica	.	.	34	8.750	0.691	17	7.042	0.929	17
Mitma Mamani_c	2011	0.8130	No indica	.	.	34	3.652	0.311	17	3.181	0.506	17
Aramburú Ocaña	2011	0.8484	Pearson	0.930	20	62	15.346	2.399	26	11.527	2.624	36
Gamarra Astuhamán	2011	0.9064	Mitades partidas	0.858	20	115	15.080	2.174	53	11.420	2.270	52
Quispe Andía	2013	0.8600	Kuder-Richardson	0.900	10	40	14.200	1.881	20	10.300	2.270	20
Garay Flores_c	2013	0.8700	Kuder-Richardson	0.780	30	27	8.500	1.910	14	6.850	1.140	13
Garay Flores_p	2013	0.8700	No indica	.	.	27	8.430	1.220	14	6.690	1.750	13
Garay Flores_a	2013	0.8700	No indica	.	.	27	9.140	1.290	14	7.920	1.500	13
Sotelo Valer_a	2013	0.8640	Mitades partidas	0.930	20	23	16.580	0.793	12	10.640	0.674	11
Sotelo Valer_b	2013	0.8640	No indica	.	.	23	15.800	1.749	12	10.450	1.128	11
Sotelo Valer_c	2013	0.8640	No indica	.	.	23	16.080	0.996	12	10.180	0.982	11
Sotelo Valer_d	2013	0.8640	No indica	.	.	23	16.500	0.674	12	11.090	1.136	11
Pastor Morales	2013	0.9190	Kuder-Richardson	0.720	20	50	17.520	1.558	25	12.760	2.127	25
Caballero Cifuentes	2013	Opinión de expertos	No indica	.	.	50	16.080	2.040	25	12.360	2.307	25
Quispe Condezo_a	2014	0.9400	Alfa	.	.	50	7.260	1.054	23	3.110	1.553	27
Quispe Condezo_b	2014	0.9400	Alfa	.	.	50	7.870	1.014	23	2.590	1.421	27
Quispe Condezo_c	2014	0.9400	Alfa	.	.	50	7.650	1.112	23	2.440	1.368	27
Quispe Condezo_d	2014	0.9400	Alfa	.	.	50	51.520	9.361	23	46.930	9.131	27

Quispe Condezo_e	2014	0.9400	Alfa	.	.	50	7.090	2.429	23	1.150	1.634	27
Quispe Condezo_f	2014	0.9400	Alfa	.	.	50	6.650	2.673	23	2.330	1.776	27
Quispe Condezo_g	2014	0.9400	Alfa	.	.	50	6.830	2.146	23	1.810	1.841	27
Quispe Condezo_h	2014	0.9400	Alfa	.	.	50	7.390	2.463	23	0.700	1.265	27
Quispe Condezo_i	2014	0.9400	Alfa	.	.	50	6.570	2.107	23	1.670	1.710	27
Quispe Condezo_j	2014	0.9400	Alfa	.	.	50	7.960	1.894	23	0.850	1.586	27
Quispe Condezo_k	2014	0.9400	Alfa	.	.	50	67.040	10.283	23	49.440	7.807	27
Quispe Condezo_l	2014	0.9400	Alfa	.	.	50	80.700	15.086	23	57.190	5.589	27
Chang Landa_a	2015	0.8740	Kuder-Richardson	0.980	20	60	8.000	1.200	30	4.000	1.500	30
Chang Landa_b	2015	0.8740	No indica	.	.	60	9.000	1.100	30	5.000	1.600	30
Chang Landa_c	2015	0.8740	No indica	.	.	60	17.000	1.100	30	11.000	1.300	30
Del Águila	2015	0.8000	Alfa	0.836	20	56	12.340	2.351	35	11.000	1.975	21
Flores Flores_a	2015	0.8500	Alfa	0.715	20	50	15.680	2.286	25	12.880	2.386	25
Flores Flores_b	2015	0.8500	Alfa	.	.	50	15.600	2.708	25	13.120	2.315	25
Flores Flores_c	2015	0.8500	Alfa	.	.	50	15.680	1.796	25	13.840	2.075	25
Flores Flores_d	2015	0.8500	Alfa	.	.	50	15.640	1.551	25	12.880	2.386	25
Flores Salinas_a	2015	0.8000	Alfa	0.890	10	70	11.579	1.132	35	10.452	1.040	35
Flores Salinas_b	2015	0.8000	Alfa	0.890	10	70	11.586	1.134	35	10.129	1.386	35
Flores Salinas_c	2015	0.8000	Alfa	0.890	10	70	11.571	1.669	35	10.771	1.128	35
Gutiérrez López_a	2015	0.8800	Alfa	0.895	20	50	14.040	1.241	25	11.000	1.826	25
Gutiérrez López_b	2015	0.8800	Alfa	.	.	50	14.360	1.221	25	10.520	2.064	25
Gutiérrez López_c	2015	0.8800	Alfa	.	.	50	15.040	1.837	25	11.040	1.925	25
Gutiérrez López_d	2015	0.8800	Alfa	.	.	50	15.680	1.376	25	10.000	2.041	25
Hilares Soria_a	2015	0.8500	Alfa	0.800	16	40	14.550	1.849	20	8.250	2.552	20
Hilares Soria_b	2015	0.8500	No indica	.	.	40	8.450	1.572	20	5.250	2.074	20
Hilares Soria_c	2015	0.8500	No indica	.	.	40	6.100	1.518	20	3.000	1.892	20
Huamán Valle_a	2015	0.8200	Kuder-Richardson	0.680	17	40	15.158	1.675	20	13.000	2.582	20
Huamán Valle_b	2015	0.8200	Alfa	0.720	13	40	33.263	3.347	20	24.211	4.614	20

Núñez Vara	2015	0.8780	Alfa	0.833	10	114	16.141	3.468	64	11.880	5.102	50
Pantoja Carhuavilca_a	2015	0.9321	Kuder-Richardson	0.792	20	56	8.286	0.535	28	5.179	2.611	28
Pantoja Carhuavilca_b	2015	0.9321	No indica	.	.	56	7.679	0.612	28	2.714	1.922	28
Pantoja Carhuavilca_c	2015	0.9321	No indica	.	.	56	41.964	3.873	28	27.321	3.278	28
Rivera Fabián_a	2015	0.8000	Mitades partidas	0.750	20	120	13.170	3.530	60	10.750	4.320	60
Rivera Fabián_b	2015	0.8000	Mitades partidas	0.720	20	120	13.000	3.650	60	10.830	4.320	60
Rivera Fabián_c	2015	0.8000	Mitades partidas	0.760	20	120	13.080	3.710	60	11.170	4.280	60
Salcedo Meza_a	2015	0.8500	Kuder-Richardson	0.930	5	37	17.160	1.950	18	12.310	2.290	19
Salcedo Meza_b	2015	0.8500	Kuder-Richardson	0.780	5	37	17.380	1.540	18	11.840	1.340	19
Salcedo Meza_c	2015	0.8640	Kuder-Richardson	0.810	5	37	17.720	1.600	18	12.000	1.100	19
Salcedo Meza_d	2015	0.8640	Kuder-Richardson	0.720	5	37	17.560	1.690	18	12.050	0.910	19
Soria Pérez_a	2015	0.8800	Alfa	0.754	20	40	16.100	1.252	20	13.700	0.801	20
Soria Pérez_b	2015	0.8800	Alfa	0.810	20	40	16.150	1.226	20	16.100	1.252	20
Soria Pérez_c	2015	0.8800	No indica	.	.	40	16.850	0.813	20	15.450	0.686	20
Zambrano Arce_a	2015	0.8600	Alfa	0.724	40	70	13.487	1.502	37	11.303	3.235	30
Zambrano Arce_b	2015	0.8600	No indica	.	.	70	13.568	2.500	37	10.849	3.483	30
Zambrano Arce_c	2015	0.8600	No indica	.	.	70	12.865	1.946	37	11.242	3.491	30

Nota: autor = efecto reportado por el autor de la tesis doctoral (cuando la tesis reportó más de un efecto, cada efecto se identificó con una letra luego del apellido materno), año = año de sustentación de la tesis doctoral, validez = tipo de análisis de validez (cuando se reportó o cálculo el grado de acuerdo entre los jueces expertos, este se registró como un número decimal), conf_tipo = tipo de análisis de confiabilidad, conf_valor = magnitud del coeficiente de confiabilidad, conf_it = número de ítems para medir la variable, muestra = total de participantes del grupo experimental y grupo control, exp_m = media del grupo experimental en la variable analizada, exp_de = desviación estándar del grupo grupo experimental en la variable analizada, exp_n = número de participantes en el grupo experimental, cont_m = media del grupo control en la variable analizada, cont_de = desviación estándar del grupo control en la variable analizada, cont_n = número de participantes en el grupo control.

Apéndice F. Tamaños del Efecto e Intervalos de Confianza Originales

id	g	lower	upper	% W(fixed)	% W(random)
1	5.7796	[4.9749;	6.5843]	0.4	0.7
2	13.5460	[11.8113;	15.2808]	0.1	0.5
3	5.7588	[4.9565;	6.5612]	0.4	0.7
4	9.8908	[8.6018;	11.1797]	0.1	0.6
5	1.1709	[0.8538;	1.4879]	2.3	0.7
6	2.5129	[2.0443;	2.9814]	1.0	0.7
7	1.0283	[0.6575;	1.3991]	1.7	0.7
8	-0.4935	[-0.8467;	-0.1402]	1.8	0.7
9	-2.3083	[-2.7601;	-1.8565]	1.1	0.7
10	-0.8459	[-1.2094;	-0.4823]	1.7	0.7
11	-1.8364	[-2.2531;	-1.4197]	1.3	0.7
12	-0.6509	[-1.0081;	-0.2936]	1.8	0.7
13	-0.7729	[-1.1340;	-0.4119]	1.8	0.7
14	-0.1367	[-0.4850;	0.2116]	1.9	0.7
15	2.1009	[1.3140;	2.8879]	0.4	0.7
16	1.6309	[0.8863;	2.3754]	0.4	0.7
17	1.7918	[1.0266;	2.5569]	0.4	0.7
18	2.0288	[1.2310;	2.8266]	0.4	0.7
19	1.2179	[0.6639;	1.7719]	0.7	0.7
20	2.6762	[1.9681;	3.3844]	0.5	0.7
21	1.8445	[1.2340;	2.4549]	0.6	0.7
22	2.2456	[1.5406;	2.9506]	0.5	0.7
23	1.7997	[1.0536;	2.5458]	0.4	0.7
24	2.8600	[1.9546;	3.7655]	0.3	0.7
25	2.1386	[1.3462;	2.9309]	0.4	0.7
26	2.0666	[1.2845;	2.8486]	0.4	0.7
27	6.2205	[4.7710;	7.6700]	0.1	0.6
28	2.5509	[1.6961;	3.4058]	0.3	0.7
29	4.5848	[3.6124;	5.5572]	0.2	0.6
30	1.2122	[0.6390;	1.7855]	0.7	0.7
31	2.0698	[1.5806;	2.5590]	1.0	0.7
32	1.6120	[1.1103;	2.1138]	0.9	0.7
33	4.2452	[3.4680;	5.0223]	0.4	0.7
34	7.9040	[6.6217;	9.1864]	0.1	0.6
35	11.6446	[9.8126;	13.4766]	0.1	0.5
36	4.6881	[3.5818;	5.7943]	0.2	0.6
37	1.8927	[1.0116;	2.7737]	0.3	0.7
38	1.6693	[0.8222;	2.5164]	0.3	0.7
39	3.9085	[2.6286;	5.1883]	0.1	0.6
40	1.7784	[0.9151;	2.6417]	0.3	0.7
41	1.4624	[1.0020;	1.9228]	1.1	0.7
42	0.6837	[0.2629;	1.1045]	1.3	0.7
43	1.1205	[0.6806;	1.5604]	1.2	0.7
44	0.7898	[0.3651;	1.2145]	1.3	0.7
45	1.5257	[1.0611;	1.9904]	1.1	0.7
46	1.9358	[1.4403;	2.4313]	0.9	0.7

47	1.7049	[1.2275;	2.1824]	1.0	0.7
48	1.1065	[0.4953;	1.7176]	0.6	0.7
49	-0.4214	[-0.9425;	0.0997]	0.8	0.7
50	0.4397	[-0.1335;	1.0129]	0.7	0.7
51	0.2413	[-0.2758;	0.7583]	0.9	0.7
52	0.9710	[0.3699;	1.5720]	0.6	0.7
53	-0.0469	[-0.5620;	0.4682]	0.9	0.7
54	0.8698	[0.2756;	1.4640]	0.6	0.7
55	-0.1499	[-0.6657;	0.3658]	0.9	0.7
56	0.8423	[0.2498;	1.4348]	0.7	0.7
57	0.1126	[-0.4029;	0.6280]	0.9	0.7
58	3.5449	[2.5185;	4.5713]	0.2	0.6
59	2.5324	[1.6805;	3.3843]	0.3	0.7
60	3.1295	[2.1777;	4.0812]	0.3	0.6
61	0.7859	[0.1465;	1.4253]	0.6	0.7
62	0.9854	[0.0090;	1.9617]	0.2	0.6
63	0.8548	[0.1267;	1.5830]	0.4	0.7
64	1.1185	[0.2467;	1.9903]	0.3	0.7
65	1.2536	[0.2757;	2.2314]	0.2	0.6
66	2.9992	[2.3820;	3.6164]	0.6	0.7
67	1.6526	[1.4042;	1.9010]	3.7	0.7
68	2.0505	[1.4742;	2.6267]	0.7	0.7
69	1.6526	[1.4042;	1.9010]	3.7	0.7
70	0.9590	[0.4765;	1.4416]	1.0	0.7
71	1.0284	[0.5420;	1.5149]	1.0	0.7
72	0.8755	[0.3973;	1.3537]	1.0	0.7
73	0.9643	[0.4815;	1.4472]	1.0	0.7
74	0.9164	[0.4361;	1.3967]	1.0	0.7
75	0.9641	[0.4813;	1.4469]	1.0	0.7
76	2.4827	[1.7466;	3.2188]	0.4	0.7
77	2.3680	[1.6472;	3.0889]	0.4	0.7
78	1.2818	[0.6809;	1.8828]	0.6	0.7
79	0.8324	[0.2639;	1.4009]	0.7	0.7
80	2.1021	[1.1868;	3.0175]	0.3	0.6
81	9.0203	[7.0538;	10.9868]	0.1	0.5
82	8.6735	[6.7762;	10.5707]	0.1	0.5
83	7.4932	[5.8295;	9.1569]	0.1	0.5
84	1.1450	[0.4712;	1.8188]	0.5	0.7
85	0.9870	[0.4490;	1.5250]	0.8	0.7
86	0.9211	[0.5540;	1.2883]	1.7	0.7
87	1.6192	[0.8321;	2.4063]	0.4	0.7
88	2.0370	[1.1902;	2.8838]	0.3	0.7
89	1.0950	[0.3680;	1.8220]	0.4	0.7
90	1.4890	[0.9164;	2.0615]	0.7	0.7
91	1.6351	[1.1911;	2.0791]	1.2	0.7
92	1.8337	[1.0832;	2.5842]	0.4	0.7
93	1.0077	[0.1987;	1.8167]	0.4	0.7
94	1.1264	[0.3044;	1.9483]	0.3	0.7
95	0.8482	[0.0546;	1.6418]	0.4	0.7
96	7.7502	[5.1575;	10.3430]	0.0	0.4

97	3.4701	[2.0980;	4.8423]	0.1	0.6
98	5.7479	[3.7482;	7.7476]	0.1	0.5
99	5.6477	[3.6770;	7.6183]	0.1	0.5
100	2.5134	[1.7579;	3.2688]	0.4	0.7
101	1.6815	[1.0294;	2.3336]	0.5	0.7
102	3.0316	[2.1994;	3.8638]	0.3	0.7
103	4.1543	[3.1399;	5.1687]	0.2	0.6
104	4.0792	[3.0776;	5.0808]	0.2	0.6
105	0.4891	[-0.0759;	1.0541]	0.7	0.7
106	2.8699	[2.0620;	3.6778]	0.4	0.7
107	1.9048	[1.2261;	2.5835]	0.5	0.7
108	2.4872	[1.7341;	3.2404]	0.4	0.7
109	3.4480	[2.5508;	4.3453]	0.3	0.7
110	2.5354	[1.7756;	3.2953]	0.4	0.7
111	4.0360	[3.0417;	5.0303]	0.2	0.6
112	1.9192	[1.2388;	2.5995]	0.5	0.7
113	2.1017	[1.3992;	2.8042]	0.5	0.7
114	2.9066	[2.1680;	3.6452]	0.4	0.7
115	2.8756	[2.1411;	3.6100]	0.4	0.7
116	4.9180	[3.8765;	5.9596]	0.2	0.6
117	0.5954	[0.0424;	1.1484]	0.8	0.7
118	1.1795	[0.5751;	1.7840]	0.6	0.7
119	0.9690	[0.3804;	1.5576]	0.7	0.7
120	0.9333	[0.3471;	1.5195]	0.7	0.7
121	1.3500	[0.7309;	1.9692]	0.6	0.7
122	1.0253	[0.5253;	1.5254]	0.9	0.7
123	1.1379	[0.6308;	1.6450]	0.9	0.7
124	0.5554	[0.0774;	1.0334]	1.0	0.7
125	1.9167	[1.2381;	2.5953]	0.5	0.7
126	2.2290	[1.5117;	2.9462]	0.4	0.7
127	2.0926	[1.3926;	2.7925]	0.5	0.7
128	3.2121	[2.3533;	4.0709]	0.3	0.7
129	2.7710	[1.8804;	3.6616]	0.3	0.7
130	1.7044	[0.9703;	2.4385]	0.4	0.7
131	1.7714	[1.0289;	2.5139]	0.4	0.7
132	0.9719	[0.3128;	1.6310]	0.5	0.7
133	2.2012	[1.3998;	3.0026]	0.4	0.7
134	0.9934	[0.6008;	1.3859]	1.5	0.7
135	1.6256	[1.0158;	2.2354]	0.6	0.7
136	3.4325	[2.5904;	4.2745]	0.3	0.7
137	4.0243	[3.0906;	4.9581]	0.3	0.6
138	0.6096	[0.2432;	0.9759]	1.7	0.7
139	0.5392	[0.1747;	0.9037]	1.7	0.7
140	0.4738	[0.1109;	0.8368]	1.7	0.7
141	2.2261	[1.3874;	3.0649]	0.3	0.7
142	3.7626	[2.6499;	4.8753]	0.2	0.6
143	4.0973	[2.9179;	5.2767]	0.2	0.6
144	4.0037	[2.8431;	5.1643]	0.2	0.6
145	2.2382	[1.4314;	3.0451]	0.4	0.7
146	0.0396	[-0.5803;	0.6594]	0.6	0.7

147	1.8243	[1.0750;	2.5735]	0.4	0.7
148	0.8871	[0.3813;	1.3929]	0.9	0.7
149	0.9021	[0.3955;	1.4087]	0.9	0.7
150	0.5842	[0.0920;	1.0764]	0.9	0.7

Apéndice G. Tamaños del Efecto Luego de la Transformación Logarítmica

id	<i>g</i>	lower	upper	%W(fixed)	%W(random)
1	0.2806	[-1.0939;	1.6551]	1.1	1.1
2	0.3236	[-1.0550;	1.7023]	1.1	1.1
3	0.3212	[-1.0572;	1.6996]	1.1	1.1
4	0.3130	[-1.0645;	1.6906]	1.1	1.1
5	0.2466	[-1.0673;	1.5604]	1.2	1.2
6	0.2577	[-1.1134;	1.6289]	1.1	1.1
7	0.1096	[-1.2529;	1.4720]	1.1	1.1
8	0.0507	[-1.4116;	1.3102]	1.1	1.1
9	0.4249	[-1.8143;	0.9644]	1.1	1.1
10	0.1463	[-1.5102;	1.2176]	1.1	1.1
11	0.3207	[-1.6977;	1.0563]	1.1	1.1
12	0.1075	[-1.4699;	1.2548]	1.1	1.1
13	0.3527	[-1.7332;	1.0277]	1.1	1.1
14	0.0528	[-1.4137;	1.3081]	1.1	1.1
15	0.9750	[-0.8836;	2.8337]	0.6	0.6
16	0.3147	[-1.3306;	1.9599]	0.8	0.8
17	0.1338	[-1.4869;	1.7546]	0.8	0.8
18	0.8646	[-0.9646;	2.6938]	0.6	0.6
19	0.2097	[-1.3031;	1.7224]	0.9	0.9
20	0.5942	[-0.9856;	2.1740]	0.8	0.8
21	0.3672	[-1.1656;	1.8999]	0.9	0.9
22	0.4078	[-1.1677;	1.9832]	0.8	0.8
23	0.2107	[-1.4037;	1.8250]	0.8	0.8
24	0.2758	[-1.3477;	1.8993]	0.8	0.8
25	0.2597	[-1.3613;	1.8808]	0.8	0.8
26	0.2015	[-1.4117;	1.8148]	0.8	0.8
27	0.2749	[-1.3102;	1.8600]	0.8	0.8
28	0.9639	[-0.7254;	2.6532]	0.7	0.7
29	0.2434	[-1.2888;	1.7755]	0.9	0.9
30	0.6629	[-0.8140;	2.1398]	1.0	1.0
31	0.1533	[-1.3975;	1.7041]	0.9	0.9
32	0.2699	[-1.4424;	1.9822]	0.7	0.7
33	0.2483	[-1.4597;	1.9563]	0.7	0.7
34	0.7250	[-1.1514;	2.6015]	0.6	0.6
35	0.2637	[-1.4473;	1.9747]	0.7	0.7
36	3.1784	[0.5016;	5.8552]	0.3	0.3
37	0.0831	[-1.4735;	1.6396]	0.9	0.9
38	0.0358	[-1.5468;	1.4752]	0.9	0.9
39	0.0262	[-1.5288;	1.5812]	0.9	0.9
40	0.0146	[-1.4962;	1.5253]	0.9	0.9
41	0.0925	[-1.4645;	1.6495]	0.9	0.9
42	0.0043	[-1.5150;	1.5064]	0.9	0.9
43	0.0783	[-1.4781;	1.6347]	0.9	0.9
44	0.0130	[-1.5238;	1.4977]	0.9	0.9
45	0.0515	[-1.5040;	1.6070]	0.9	0.9
46	0.0065	[-1.5042;	1.5173]	0.9	0.9

47	0.1802	[-1.7080;	2.0685]	0.6	0.6
48	0.0440	[-1.6213;	1.7093]	0.7	0.7
49	0.0577	[-1.7024;	1.8179]	0.7	0.7
50	0.0579	[-1.7714;	1.8872]	0.6	0.6
51	0.1948	[-1.2378;	1.6274]	1.0	1.0
52	0.3772	[-0.8652;	1.6196]	1.3	1.3
53	0.4257	[-1.0749;	1.9263]	0.9	0.9
54	0.3772	[-0.8652;	1.6196]	1.3	1.3
55	0.1047	[-1.3561;	1.5656]	1.0	1.0
56	0.2046	[-1.2624;	1.6716]	1.0	1.0
57	0.1985	[-1.2680;	1.6651]	1.0	1.0
58	0.1509	[-1.3123;	1.6141]	1.0	1.0
59	0.1708	[-1.2937;	1.6353]	1.0	1.0
60	0.1841	[-1.2813;	1.6496]	1.0	1.0
61	0.5118	[-1.0861;	2.1098]	0.8	0.8
62	0.1821	[-1.3615;	1.7257]	0.9	0.9
63	0.1746	[-1.3684;	1.7176]	0.9	0.9
64	0.1287	[-1.4110;	1.6683]	0.9	0.9
65	0.4773	[-1.2929;	2.2474]	0.7	0.7
66	0.9563	[-0.8170;	2.7295]	0.7	0.7
67	0.5810	[-1.1162;	2.2783]	0.7	0.7
68	0.0276	[-1.4756;	1.5307]	0.9	0.9
69	0.1972	[-1.1696;	1.5641]	1.1	1.1
70	0.2591	[-1.2560;	1.7742]	0.9	0.9
71	0.3022	[-1.1063;	1.7106]	1	1
72	0.3508	[-1.2862;	1.9878]	0.8	0.8
73	0.3413	[-1.4280;	2.1106]	0.7	0.7
74	0.4145	[-1.3784;	2.2073]	0.6	0.6
75	0.3173	[-1.4451;	2.0798]	0.7	0.7
76	0.7213	[-1.3300;	2.7725]	0.5	0.5
77	0.4205	[-1.1678;	2.0089]	0.8	0.8
78	0.2780	[-1.2860;	1.8421]	0.8	0.8
79	2.1813	[-0.2802;	4.6428]	0.3	0.3
80	3.6037	[0.0819;	7.1254]	0.2	0.2
81	3.9521	[0.1531;	7.7510]	0.1	0.1
82	0.0344	[-1.5120;	1.5809]	0.9	0.9
83	2.1016	[-0.3059;	4.5090]	0.4	0.4
84	1.0785	[-0.7346;	2.8917]	0.6	0.6
85	1.5830	[-0.4961;	3.6620]	0.5	0.5
86	2.9837	[-0.0584;	6.0258]	0.2	0.2
87	1.7411	[-0.4333;	3.9154]	0.4	0.4
88	3.3142	[0.0189;	6.6095]	0.2	0.2
89	0.1140	[-1.4354;	1.6634]	0.9	0.9
90	0.1253	[-1.4248;	1.6754]	0.9	0.9
91	1.8417	[-0.2878;	3.9712]	0.5	0.5
92	1.4478	[-0.4667;	3.3623]	0.6	0.6
93	1.8421	[-0.2877;	3.9719]	0.5	0.5
94	0.1216	[-1.4123;	1.6555]	0.9	0.9
95	0.1903	[-1.3636;	1.7442]	0.9	0.9
96	0.1543	[-1.3966;	1.7051]	0.9	0.9

97	0.1549	[-1.3960;	1.7057]	0.9	0.9
98	0.2314	[-1.3268;	1.7896]	0.9	0.9
99	0.9418	[-0.7009;	2.5844]	0.8	0.8
100	0.4595	[-1.0534;	1.9724]	0.9	0.9
101	0.1628	[-1.3126;	1.6383]	1	1
102	0.4429	[-1.1501;	2.0359]	0.8	0.8
103	0.4806	[-1.1208;	2.0820]	0.8	0.8
104	0.4016	[-1.1829;	1.9862]	0.8	0.8
105	0.6681	[-0.9842;	2.3204]	0.8	0.8
106	0.5726	[-1.1220;	2.2671]	0.7	0.7
107	0.6270	[-1.0855;	2.3395]	0.7	0.7
108	1.0529	[-0.8451;	2.9510]	0.6	0.6
109	0.1608	[-1.4482;	1.7698]	0.8	0.8
110	0.1843	[-1.4270;	1.7956]	0.8	0.8
111	0.1852	[-1.2010;	1.5714]	1.1	1.1
112	0.2798	[-1.2567;	1.8163]	0.9	0.9
113	0.1299	[-1.2428;	1.5025]	1.1	1.1
114	0.1155	[-1.2565;	1.4875]	1.1	1.1
115	0.0996	[-1.2719;	1.4711]	1.1	1.1
116	0.3485	[-1.3118;	2.0089]	0.8	0.8
117	0.8251	[-0.9980;	2.6482]	0.6	0.6
118	0.9149	[-0.9511;	2.7810]	0.6	0.6
119	0.0116	[-1.5899;	1.6130]	0.8	0.8
120	0.1722	[-1.3150;	1.6594]	0.9	0.9
121	0.1732	[-1.3141;	1.6605]	0.9	0.9
122	0.1147	[-1.3690;	1.5985]	0.9	0.9
