



FACULTAD DE PSICOLOGÍA

DESEMPEÑO DOCENTE Y DESEMPEÑO ESTUDIANTIL PERCIBIDOS POR
ESTUDIANTES DE PSICOLOGÍA DE UNA UNIVERSIDAD NACIONAL DE LIMA

Línea de investigación:

Psicología de los procesos básicos y psicología educativa

Tesis para optar el Título Profesional de Licenciado en Psicología con
mención en Psicología Educativa

Autor:

Tirado Malpartida, Jhonatan Juan

Asesor:

Mendoza Huamán, Vicente Eugenio

ORCID: 0000-0003-1946-6831

Jurado:

Figueroa Gonzales, Julio Lorenzo

Hervias Guerra, Edmundo Magno

Ávila Miñán, Mildred

Lima - Perú

2024

DESEMPEÑO DOCENTE Y DESEMPEÑO ESTUDIANTIL PERCIBIDOS POR ESTUDIANTES DE PSICOLOGÍA DE UNA UNIVERSIDAD NACIONAL DE LIMA

INFORME DE ORIGINALIDAD

14%

INDICE DE SIMILITUD

13%

FUENTES DE INTERNET

5%

PUBLICACIONES

3%

TRABAJOS DEL
ESTUDIANTE

FUENTES PRIMARIAS

1	www.researchgate.net Fuente de Internet	1%
2	repositorio.unfv.edu.pe Fuente de Internet	1%
3	datospdf.com Fuente de Internet	1%
4	www.uv.mx Fuente de Internet	1%
5	www.alternativas.me Fuente de Internet	1%
6	Submitted to Universidad Cesar Vallejo Trabajo del estudiante	1%
7	Submitted to Universidad Nacional Federico Villarreal Trabajo del estudiante	<1%
8	es.scribd.com Fuente de Internet	<1%



Universidad Nacional
Federico Villarreal

VRIN | VICERRECTORADO
DE INVESTIGACIÓN

FACULTAD DE PSICOLOGÍA

**DESEMPEÑO DOCENTE Y DESEMPEÑO ESTUDIANTIL PERCIBIDOS POR
ESTUDIANTES DE PSICOLOGÍA DE UNA UNIVERSIDAD NACIONAL DE LIMA**

Línea de Investigación:

Psicología de los procesos básicos y psicología educativa

Tesis para optar el Título Profesional de Licenciado en Psicología con mención en Psicología
Educativa

Autor:

Tirado Malpartida, Jhonatan Juan

ORCID: 0000-0001-5300-6914

Asesor:

Mendoza Huamán, Vicente Eugenio

ORCID: 0000-0003-1946-6831

Jurado:

Figueroa Gonzáles, Julio Lorenzo

Hervias Guerra, Edmundo Magno

Ávila Miñán, Mildred

Lima – Perú

2024

Dedicatoria

A la memoria de mis familiares Delia, Dedicación y Luisa, quienes me criaron con alegría y amor.

A mis padres Segundo y Hortencia; mi familia de Lima, Pasco y La Libertad; y mi pareja Kelly, por apoyarme incondicionalmente y motivarme a ser mejor.

Agradecimientos

A la Facultad de Psicología de la Universidad Nacional Federico Villarreal, por brindarme los espacios para desarrollar mis competencias profesionales, a través de sus cursos y oportunidades extracurriculares.

Al doctor Vicente Mendoza, por sus valiosas recomendaciones en la asesoría de la tesis y sus importantes enseñanzas en los cursos de psicología educativa.

A todos los científicos cuyas obras contribuyen a construir una psicología científica, especialmente a los doctores con los que compartí comunicación: Ricardo Pérez-Almonacid, Karla Acuña, Miriam Jiménez, Emilio Ribes, y dedicado a la memoria del doctor Juan José Irigoyen.

Al doctor Aldo Bazán-Ramírez, que asesoró el presente proyecto, como representante de la Universidad Autónoma del Estado de Morelos, en el marco de una colaboración interinstitucional.

Índice de contenido

Dedicatoria.....	ii
Agradecimientos	iii
Índice de contenido	iv
Índice de tablas	vi
Índice de figuras.....	ix
Resumen.....	x
Abstract.....	xi
I. Introducción	12
1.1. Descripción y formulación del problema.....	12
1.1. Antecedentes	15
1.2. Objetivos	22
1.3. Justificación.....	22
1.4. Hipótesis.....	23
II. Marco teórico	24
2.1. Bases teóricas sobre el tema de investigación.....	24
2.2. Modelos de Interacciones Didácticas.....	31
2.3. Marco conceptual de las interacciones didácticas.....	46
2.4. Factores de las interacciones didácticas	49
2.5. Ámbitos didácticos del desempeños docente y estudiantil	59
III. Método	70
3.1. Tipo de investigación	70
3.2. Ámbito temporal y espacial	70

3.3. Variables	70
3.4. Población y muestra	73
3.5. Instrumentos	74
3.6. Procedimientos	77
3.7. Análisis de datos	78
IV. Resultados	82
4.1. Análisis de ítems	82
4.2. Análisis factorial confirmatorio	86
4.3. Invarianza de la medición	96
4.4. Fiabilidad	102
4.5. Resultados descriptivos	104
4.6. Correlaciones entre variables	109
4.7. Resultados comparativos	110
V. Discusión de resultados	124
VI. Conclusiones	134
VII. Recomendaciones	135
VIII. Referencias	136
IX. Anexos	150

Índice de tablas

Tabla 1 <i>Ámbitos funcionales del desempeño (Carpio et al., 1998)</i>	35
Tabla 2 <i>Propuesta de clasificación de los factores y ámbitos de las interacciones didácticas</i>	48
Tabla 3 <i>Clasificaciones de categorías de desempeño didáctico del docente</i>	56
Tabla 4 <i>Propuesta de correspondencia entre ámbitos didácticos del desempeño docente y desempeño estudiantil</i>	60
Tabla 5 <i>Operacionalización de la variable “Desempeño docente”</i>	70
Tabla 6 <i>Operacionalización de la variable “Desempeño estudiantil”</i>	71
Tabla 7 <i>Distribución de la muestra por salón y tipo de asignatura</i>	73
Tabla 8 <i>Estadísticos descriptivos de los ítems de la variable desempeño docente</i>	82
Tabla 9 <i>Estadísticos descriptivos de los ítems de la variable desempeño estudiantil</i>	84
Tabla 10 <i>Índices de bondad de ajuste para tres modelos de Desempeño docente (n=475) ...</i>	88
Tabla 11 <i>Índices del modelo Bifactor del cuestionario de Desempeño docente</i>	91
Tabla 12 <i>Índices de bondad de ajuste para tres modelos de Desempeño estudiantil (n=475)</i>	92
Tabla 13 <i>Índices del modelo Bifactor del cuestionario de Desempeño estudiantil</i>	95
Tabla 14 <i>Modelos de invarianza de las puntuaciones de Desempeño Docente (n=475) según agrupaciones sociodemográficas en ...</i>	96
Tabla 15 <i>Modelos de invarianza de las puntuaciones de Desempeño Docente (n=475) según agrupaciones socioeducativas en ...</i>	97
Tabla 16 <i>Modelos de invarianza de las puntuaciones de Desempeño Docente (n=475) según agrupaciones de encuesta educativa en ...</i>	99
Tabla 17 <i>Modelos de invarianza de las puntuaciones de Desempeño Estudiantil (n=475) según agrupaciones sociodemográficas en ...</i>	99

Tabla 18 Modelos de invarianza de las puntuaciones de Desempeño Estudiantil (n=475) según agrupaciones socioeducativas en	100
Tabla 19 Modelos de invarianza de las puntuaciones de Desempeño Estudiantil (n=475) según agrupaciones de encuesta educativa en	101
Tabla 20 Coeficientes alfa y alfa ordinal de las dimensiones de desempeño docente.....	102
Tabla 21 Coeficientes alfa y alfa ordinal de las dimensiones de desempeño estudiantil	103
Tabla 22 Características sociodemográficas de los participantes (N= 475).	104
Tabla 23 Estadígrafos para las dimensiones de la variable desempeño docente.....	106
Tabla 24 Estadígrafos para las dimensiones de la variable desempeño estudiantil	107
Tabla 25 Matriz de correlación entre las variables y dimensiones de desempeño docente y desempeño estudiantil percibido.....	109
Tabla 26 Prueba de U de Mann Whitney para el análisis de la variable desempeño docente percibido según sexo.....	111
Tabla 27 Prueba de U de Mann Whitney para el análisis de la variable desempeño estudiantil percibido según sexo	111
Tabla 28 Prueba de U de Mann Whitney para el análisis de la variable desempeño docente percibido según edad	112
Tabla 29 Prueba de U de Mann Whitney para el análisis de la variable desempeño estudiantil percibido según edad	112
Tabla 30 Prueba de U de Mann Whitney para el análisis de la variable desempeño docente percibido según nivel de especialidad	113
Tabla 31 Prueba de U de Mann Whitney para el análisis de la variable desempeño estudiantil percibido según nivel de especialidad	114
Tabla 32 Prueba de U de Mann Whitney para el análisis de la variable desempeño docente percibido según turno de estudios.	115

Tabla 33 <i>Prueba de U de Mann Whitney para el análisis de la variable desempeño estudiantil percibido según turno de estudios.</i>	115
Tabla 34 <i>Prueba H de Kruskall Wallis para el análisis de la variable desempeño docente según eje curricular</i>	116
Tabla 35 <i>Análisis post-hoc mediante la prueba Bonferroni para los grupos de eje curricular de la variable desempeño docente</i>	117
Tabla 36 <i>Prueba H de Kruskall Wallis para el análisis de la variable desempeño estudiantil percibido según eje curricular</i>	118
Tabla 37 <i>Análisis post-hoc mediante la prueba Bonferroni para los grupos de eje curricular de la variable desempeño estudiantil</i>	118
Tabla 38 <i>Prueba H de Kruskall Wallis para el análisis de la variable desempeño docente según ciclo de estudios</i>	119
Tabla 39 <i>Análisis post-hoc mediante la prueba Bonferroni para los grupos de ciclos de estudios de la variable desempeño docente.</i>	120
Tabla 40 <i>Prueba H de Kruskall Wallis para el análisis de la variable desempeño estudiantil percibido según ciclo de estudios.</i>	121
Tabla 41 <i>Análisis post-hoc mediante la prueba Bonferroni para los grupos de ciclos de estudios de la variable desempeño estudiantil</i>	122

Índice de figuras

Figura 1 <i>Esquema representativo de la propuesta de la planeación de la enseñanza del comportamiento inteligente (Ribes, 1990).</i>	32
Figura 2 <i>Representación del lenguaje como medio, instrumento y forma de vida</i>	34
Figura 3 <i>Modelo de interacciones didácticas propuesto por Ibáñez</i>	36
Figura 4 <i>Modelo de las relaciones de enseñar y estudiar estructuradas por el criterio de logro</i>	39
Figura 5 <i>Niveles analíticos de la interacción didáctica</i>	40
Figura 6 <i>Factores de la interacción didáctica</i>	43
Figura 7 <i>Niveles analíticos de la educación</i>	48
Figura 8 <i>Ruta hipotética del desempeño autodidacta a partir del desempeño didáctico</i>	68
Figura 9 <i>Diagrama de senderos del modelo bifactor de desempeño docente</i>	90
Figura 10 <i>Diagrama de senderos del modelo bifactor de desempeño estudiantil</i>	94
Figura 11 <i>Frecuencia y porcentaje de docentes según nivel de desempeño alcanzado</i>	106
Figura 12 <i>Frecuencia y porcentaje de alumnos según nivel de desempeño estudiantil</i>	108

Resumen

Se diseñó una investigación empírica e instrumental con el objetivo de estudiar como los estudiantes perciben las variables desempeño docente y desempeño estudiantil, en una muestra de 475 estudiantes de la carrera de psicología, de una universidad pública en Lima, Perú. El marco teórico fue el modelo de interacciones didácticas, el cual delimita los “ámbitos didácticos” como un conjunto de prácticas en contextos específicos que realizan los docentes y estudiantes para lograr los objetivos de aprendizaje. Los ámbitos didácticos son seis y fueron incluidos como dimensiones para cada variable principal en los instrumentos: Cuestionario de desempeño docente – CDD y Cuestionario de desempeño estudiantil – CDE. Se obtuvieron favorables evidencias de validez y fiabilidad para cada instrumento, en las pruebas de análisis factorial confirmatorio, invarianza de la medición y coeficientes alfa ordinal y omega jerárquico. En los resultados, se halló una correlación directa y grande entre las variables principales ($r_s = 0.596; p < .001$). También se hallaron correlaciones moderadas y grandes entre las dimensiones de las variables principales. Las comparaciones de puntajes por grupo se ejecutaron con las siguientes variables de agrupación género, edad, ciclo de estudios, turno de estudios, especialidad de la carrera y eje curricular del curso. Para desempeño docente evaluado por los alumnos, se hallaron diferencias según el eje curricular, turno y ciclo de estudios. Para el desempeño estudiantil autoevaluado, se hallaron diferencias por edad, especialidad y ciclo de estudios. Los resultados se discuten comparándolos con los antecedentes de esta área de investigación.

Palabras clave: desempeño docente, desempeño estudiantil, ámbitos didácticos, carrera de psicología, invarianza de la medición.

Abstract

An empirical and instrumental research was designed with the objective of studying how students perceive the variables teacher performance and student performance, in a sample of 475 psychology students from a public university in Lima, Peru. The theoretical framework was the model of didactic interactions, which defines "didactic ambits" as a set of practices in specific contexts conducted by teachers and students to achieve learning objectives. There are six didactic ambits, and they were included as dimensions for each main variable in the instruments: Teacher Performance Questionnaire - CDD and Student Performance Questionnaire - CDE. Positive evidence of validity and reliability was obtained for each instrument, in confirmatory factor analysis tests, measurement invariance and ordinal alpha and hierarchical omega coefficients. In the results, a direct and large correlation was found between the main variables ($r_s = 0.596; p < .001$). Moderate and large correlations were also found between the dimensions of the main variables. The comparisons of scores by group were analyzed with the following grouping variables: gender, age, study cycle, study turn, career specialty and curricular axis of the course. For student-assessed teaching performance, differences were found according to curricular axis, turn and study cycle. For student self-evaluated performance, differences were found by age, specialty and study cycle. The results are discussed in comparison with the antecedents of this area of research.

Keywords: teaching performance, student performance, didactic ambits, psychology career, measurement invariance.

I. Introducción

En Perú todavía existen muchos problemas que perjudican la calidad de la educación superior universitaria, impactando en profesores y estudiantes. Para que estos actores educativos desarrollen las habilidades de enseñanza y estudio necesarias para su éxito académico, es relevante investigar sus comportamientos. En la presente tesis se investigó los ámbitos didácticos del desempeño docente y el desempeño estudiantil desde la percepción de los estudiantes de una facultad de psicología de una universidad pública de Lima.

En el capítulo 1 se describen los antecedentes científicos y se brindan todos los lineamientos para ejecutar la investigación, como es los objetivos, descripción del problema, justificación e hipótesis.

En el capítulo 2 se desarrolló el marco teórico, donde se sustenta el modelo de interacciones didácticas y los lineamientos metodológicos para su estudio.

En el capítulo 3 se expuso la metodología para describir, relacionar y comparar los fenómenos estudiados según las variables de agrupación.

En los capítulos 4 y 5 se presentan los resultados psicométricos, los resultados correlacionales y comparativos, así como la discusión respectiva. Finalmente, se brindan las conclusiones y recomendaciones en los capítulos 6 y 7.

1.1. Descripción y formulación del problema

El desempeño de los docentes juega un papel fundamental en la calidad educativa que puede ofrecer una institución, por lo que su investigación se convierte en una necesidad, ya que permite conocer su estado actual y las formas en que se puede mejorar. Desde la psicología interconductual, se han elaborado propuestas interdisciplinarias para estudiar los factores y procesos educativos, denominados como “modelo de interacciones didácticas” (Carpio & Irigoyen, 2005; Ibáñez & Ribes, 2001), donde se considera esencial al concepto de competencia, que cuando se incluye en el análisis del desempeño ideal para el docente, se

define como la tendencia a enseñar de manera efectiva y variada, en un amplio grupo de situaciones educativas, promoviendo que los estudiantes alcancen los logros de aprendizaje (Morales et al., 2016). De acuerdo con esta forma de entender la enseñanza, trasladada al escenario de la educación superior universitaria, se han delimitado seis ámbitos funcionales del desempeño didáctico (Silva et al., 2014): Exploración competencial, Explicitación de criterios, Ilustración, Supervisión de la práctica, Retroalimentación y Evaluación. Si un docente universitario desea alcanzar la competencia de enseñanza, deberá actuar de manera efectiva y variada en cada uno de estos “ámbitos didácticos”.

Respecto al desempeño docente universitario en nuestro país, en la década de los 2010 la educación superior recibía críticas por la baja calidad educativa que brindaba (Yamada et al., 2012) y la baja calidad de enseñanza y formación profesional de los docentes universitarios (Espinoza, 2014). Ante este escenario se produjo la irrupción de la SUNEDU como nuevo ente regulador, estableciendo un nuevo proceso de licenciamiento de universidades y exigiendo un cambio sustancial en los planes de estudio (Tupayachi-Díaz & Lopez-Vega, 2018). Si bien los procesos de licenciamiento mejoraron la baja calidad educativa de las universidades peruanas, todavía se han reportado falencias y debilidades en la enseñanza, formación y producción científica del profesorado universitario (Beltrán et al., 2022; Pérez, 2018).

Un bajo desempeño docente puede impactar negativamente en el desempeño estudiantil (Irigoyen & Jiménez, 2004) y, en el contexto universitario, producir una carencia en sus competencias profesionales y en la obtención de empleo (Yamada et al., 2012). Si bien existen investigaciones sobre el desempeño estudiantil, esta adquiere nuevas características cuando se trata de la etapa universitaria.

Desde el modelo de interacciones didácticas, se conceptualiza al desempeño estudiantil universitario como los comportamientos desplegados por el estudiante o alumno en interacción con el comportamiento del docente y la tarea disciplinaria, para alcanzar los criterios de logro.

Cuando un estudiante afronta con éxito las etapas planteadas en el plan de estudios, se puede hablar de “competencias de estudio” (Morales et al., 2016). Estas competencias se pueden estudiar desde diferentes categorías de análisis, como son las modalidades del lenguaje (Morales et al., 2013), los ámbitos del desempeño científico (Acuña et al., 2013), desempeño tecnológico-técnico (Silva et al., 2016) y el desempeño didáctico (Morales et al., 2017). La interacción sincrónica docente-estudiante también ocurre en los seis ámbitos didácticos nombrados, y que en un primer análisis se les nombro como: Precurrentes y potenciales habilidades, Identificación de criterios, Participación, Práctica pertinente, Mejoramiento y Desempeño resultante. Estos ámbitos se corresponden como pares (en el orden de mención) con los ámbitos didácticos para el docente.

En la presente investigación, se analizará la relación entre el desempeño docente y el desempeño estudiantil, mediante cuestionarios de autoinforme sometidos a procesos de adaptación psicométrico. Asimismo, se profundizará en esta relación, a través de los ámbitos didácticos, plasmados como dimensiones en los cuestionarios. Sin embargo, el desempeño estudiantil no se ve totalmente influenciado por el desempeño docente, sino también por otras variables educativas circunscritas a la institución educativa, la comunidad, a sus compañeros e incluso a su propio comportamiento (Montes de Oca et al., 2023; Morales et al., 2021; Morales & Chávez, 2019). Para observar si los puntajes de autoinforme son diferentes por grupos, se han incluido análisis comparativos según las variables de agrupación: género, rango de edad, turno de estudios, ciclo del plan de estudios, especialidad de la carrera y eje curricular. Respecto a las dos últimas variables educativas, se han adecuado a las características de la carrera universitaria de la muestra de estudio: Especialidad de la carrera de psicología (organizacional, educativa, social y clínica) y el eje curricular del profesor evaluado (básica-disciplinar, investigación científica y especialidad profesional) de acuerdo a la diferenciación de las practicas que realiza el psicólogo (Carpio et al., 1998; Silva et al., 2016).

Con base en lo argumentado, se plantea la pregunta principal de investigación:

¿Cuál es la relación entre el desempeño docente y desempeño estudiantil percibidos por estudiantes de psicología de una universidad pública de Lima – 2019?

1.1. Antecedentes

1.2.1. Antecedentes Nacionales

Bazán-Ramírez et al. (2022a) realizaron una investigación para evaluar las relaciones entre variables de desempeño docente (D.D.) y desempeño estudiantil (D.E.). Para lograrlo, propusieron un modelo explicativo de la variable “evaluación-aplicación” (D.E.) que se vería afectado por las siguientes variables: regulado indirectamente por las variables de desempeño docente “explicitación de criterios”, “ilustración”, “monitoreo-supervisión de la práctica”, “retroalimentación” y “evaluación”; y regulado directamente por las variables de desempeño estudiantil “Participación”, “Práctica pertinente” y “retroalimentación-mejora”. La muestra fue de 309 estudiantes de maestría y doctorado en ciencias de la educación, de una universidad estatal ubicada en el sur del Perú. Los instrumentos de medición fueron dos: 1) Una versión reeditada del Cuestionario de desempeño docente de Bazán-Ramírez et al. (2021), donde se agregó la dimensión “Supervisión de la práctica” (Bazán-Ramírez, et al., 2022b) pero no se incluyó la dimensión “Exploración de competencias” en el modelo explicativo propuesto, conformando 20 ítems el cuestionario final usado. 2) También se usó una versión reeditada del Cuestionario de desempeño estudiantil de Bazán-Ramírez et al. (2021), donde se agregó la dimensión “Práctica pertinente” (Bazán-Ramírez, et al., 2022b) pero no se incluyeron las dimensiones “Comportamientos de aprendizaje precurrentes” e “Identificación de criterios” en el modelo explicativo propuesto, por lo que se emplearon cuatro dimensiones y 18 ítems. Ambos instrumentos superaron las pruebas de AFC y AFC-MG.

Para el análisis del modelo explicativo se usó el estimador máxima verosimilitud robusto (MLM), obteniendo satisfactorios índices de ajuste (χ^2 escalado de Satorra-Bentler

(539) = 859.032, $p = .000$; IFI = 0,931; CFI = 0,931; TLI = 0,923; RMSEA = 0,044; RMSEA-IC = 0,039 - 0,049; SRMR = 0,065) para el primer modelo. De acuerdo a los resultados, las variables “explicitación de criterios” e “ilustración”, en combinación, afectan la variabilidad de “Participación” en un 41%. La variable “monitoreo-supervisión de la práctica” afecta la variabilidad de “Práctica pertinente” en un 56%. La “retroalimentación” y “evaluación” afectan la variabilidad de “retroalimentación-mejora” en un 38%. En cuanto a la variable “evaluación-aplicación”, esta es explicada, en 46%, de manera directa por las restantes tres variables de desempeño estudiantil. Finalmente, los autores realizaron un análisis comparativo entre los estudiantes de maestría y doctorado, concluyendo que no existen diferencias estadísticamente significativas, ni diferencias descriptivas (ES), en ninguna de las variables de desempeño docente ni desempeño estudiantil.

Bazán-Ramírez et al. (2022b), investigadores de la Universidad Nacional Federico Villarreal y de la Universidad San Cristóbal de Huamanga, realizaron un estudio con el objetivo de comparar, según el género y el tipo de posgrado, los niveles en las dimensiones de desempeño docente y desempeño estudiantil, desde la perspectiva del estudiante. La muestra estuvo conformada de 310 alumnos de maestría y doctorado en ciencias de la educación, de una universidad estatal de Perú. Los instrumentos fueron aplicados durante junio de 2020. Los instrumentos de evaluación fueron los Cuestionarios de Desempeño Docente y Desempeño Estudiantil, en la versión adaptada por Bazán-Ramírez, et al. (2022b) que incluye las dimensiones “Supervisión de la práctica” en la variable desempeño docente, y “Práctica pertinente” en la variable desempeño estudiantil. Son 6 dimensiones para cada variable y cada dimensión contiene cuatro ítems, conformando 24 ítems por cada cuestionario. Como objetivo previo, aplicaron el AFC para cada cuestionario, obteniendo adecuados índices de ajuste para el cuestionario de desempeño docente (GFI = 0,98; AGFI = 0,98; CFI = 1,11; RMSEA = 0,04; RMR = 0,043; NFI = 0,98; NNFI = 1,11) y para un modelo corregido de desempeño estudiantil

(GFI = 0,98; AGFI = 0,97; CFI = 1,00; RMSEA = 0,04; RMR = 0,03; NFI = 0,98; NNFI = 0,99) de 22 ítems.

En los resultados comparativos, entre varones y mujeres no se hallaron diferencias estadísticamente significativas en 10 dimensiones de desempeño docente y desempeño estudiantil. Únicamente se hallaron diferencias en la dimensión “ilustración”, donde los varones calificaron más alto a sus docentes, y en la dimensión “aplicación”, donde las mujeres autocalificaron más alto su desempeño. En cuanto a la comparación entre estudiantes de maestría y doctorado, no se hallaron diferencias estadísticamente significativas en ninguna de las 12 dimensiones.

Bazán-Ramírez et al. (2021) buscaron obtener evidencias de la estructura interna e invarianza de la medición de la Escala de Desempeño Docente para carreras de psicología, y hacer comparaciones de puntajes según género, edad y nivel académico. La muestra fue de 316 alumnos de pregrado, de la licenciatura en psicología, provenientes de dos universidades públicas en Lima, Perú. Las variables de control fueron género, edad y etapa académica. El instrumento de medición fue la Escala de Desempeño docente, que es una versión producida a partir del cuestionario de Bazán-Ramírez y Velarde (2021) y la presente investigación. Cuenta con 18 ítems y evalúa cinco ámbitos de desempeño docente. En los resultados psicométricos, se realizó el AFC a partir de una matriz de correlaciones policóricas, mostrando adecuados índices de ajuste ($\chi^2 = 229,29$; $gl = 125$; NNFI = 0,94; CFI = 0,95; SRMR = 0,050; RMSEA = 0,052) y altas cargas factoriales de los ítems a sus respectivas dimensiones. Para la invarianza de la medición, los autores aplicaron el procedimiento de cuatro restricciones secuenciales, considerando que el χ^2 de los modelos no muestren diferencias estadísticamente significativas ($p \geq .05$) y las diferencias de CFI ($\Delta CFI \leq -0.01$) y RMSEA ($\Delta RMSA \leq 0.06$) sean inferiores. Las variables género y edad alcanzaron la invarianza estricta (misma varianza de error entre

grupos) y la variable etapa académica alcanzó la invarianza fuerte (mismos interceptos de ítems entre grupos).

En los resultados comparativos, no se obtuvo diferencias estadísticamente significativas en los puntajes de desempeño docente según la variable género ($p = .358$; $\eta^2 = 0.003$). Si se obtuvieron diferencias estadísticamente significativas en las agrupaciones edad ($p = .027$; $\eta^2 = 0.016$) y etapa académica ($p < .001$; $\eta^2 = 0.102$).

Dominguez-Lara y Fernández-Arata (2019) buscaron identificar los niveles de autoeficacia académica en universitarios peruanos, y compararlas según año de estudios y género. La muestra fue conformada por 730 alumnos de pregrado de la licenciatura de Psicología, de una universidad privada en Lima. Las participantes fueron en su mayoría mujeres (76.6%) y el rango de edades fue de 16 a 47 años. El instrumento de medición fue la Escala de Autoeficacia Percibida Específica de Situaciones Académicas – EAPESA, compuesto de 9 ítems y una escala de respuesta de cuatro alternativas. El coeficiente alfa de la escala fue 0.905.

En la comparación entre varones y mujeres, se obtuvieron diferencias estadísticamente significativas y una d de Cohen = 0.32, donde los varones obtuvieron un mayor puntaje promedio que las mujeres. En la comparación por año de estudios, que estuvo compuesto por grupos de primer, segundo, tercer, cuarto y quinto año, no se hallaron diferencias estadísticamente significativas y la magnitud del efecto omega cuadrado fue insignificante ($\omega^2 = 0.16$). Finalmente, los autores mencionan que, si bien no se hallaron diferencias por año académico, en una comparación descriptiva entre estudiantes de 1er y 5to año, el nivel de autoeficacia académica alto es mayor en los de quinto año y el nivel bajo es mayor en los de primer año.

1.2.2. Antecedentes Internacionales

Abellán-Roselló et al. (2023) investigaron la variable “apoyos docente” conformada por los factores: 1) accesibilidad y cercanía de profesor; 2) apoyo a la autonomía; y 3) apoyo a la comprensión de contenido; con el objetivo de evaluar sus relaciones y comparar las diferencias de puntaje según edad, grado y modalidad de estudios. La muestra fue compuesta de 288 alumnos universitarios provenientes de universidades españolas. El instrumento de evaluación fue la Escala sobre apoyos docentes y estrategias de evitación - MOCSE-EADEE. En los resultados comparativos, se obtuvieron diferencias estadísticamente significativas y magnitudes del efecto pequeñas y moderadas entre los grupos de edad jóvenes (19 a 30 años de edad) y adultos (31 a 59 años de edad), en las tres dimensiones de apoyo docente: 1) $d = .43$; 2) $d = .46$; y 3) $d = .52$. Para la comparación por grado de estudios (entre grado universitario y máster) solo se hallaron diferencias en la dimensión 3) apoyo a la autonomía. Finalmente, en la comparación por modalidad de estudios (presencial y virtual), se hallaron diferencias estadísticamente significativas y magnitudes del efecto moderadas en todas las dimensiones.

Moreno (2023) desarrollo un estudio correlacional simple entre las variables práctica docente y aprendizaje autónomo, en una muestra de 70 estudiantes de enfermería, provenientes de una universidad estatal de Los Ríos, ubicada en Ecuador. Los instrumentos de medición fueron el “Cuestionario de Práctica Docente” y el “Cuestionario de Aprendizaje Autónomo”, creadas por la autora del estudio y sometidas a un proceso de juicio de expertos. En los resultados, se halló una correlación una correlación directa y grande ($r_s = 0.507$; $p < .001$) entre las variables principales del estudio.

Bazán-Ramírez y Velarde (2021) elaboraron un estudio psicométrico y descriptivo con los siguientes objetivos: 1) evaluar de manera exploratoria las características psicométricas de las escalas, 2) describir las percepciones de los estudiantes mexicanos de la licenciatura de psicología sobre el desempeño del docente y su propio desempeño estudiantil, según las

variables de agrupación nivel de estudios y sexo. El diseño fue observacional, de tipo transversal y exploratorio. La muestra estuvo conformada por 124 alumnos de la licenciatura de Psicología, de una universidad estatal ubicada en Morelos, México. Las edades fluctúan entre 19 y 44 años. El muestreo fue aleatorio e incluyó a los alumnos que estudiaban desde el segundo hasta el décimo semestre, clasificando los cursos evaluados en alguna de las tres etapas de la carrera: general, disciplinar y profesional. Los instrumentos usados se agrupan en una batería denominada “Auto reporte de desempeño del docente y del estudiante de psicología”, que es una adaptación en formato abreviado del “Cuestionario de Interacciones Didácticas” – CID (Velarde, 2020). Las propiedades psicométricas de este antecedente se exponen en la sección “3.5 Instrumentos”. Sobre los resultados descriptivos, los autores reportaron la media y desviación estándar según ámbito de desempeño y agrupación, por lo que las comparaciones están basadas en la observación de los autores, los cuales reportan que los ámbitos “explicitación de criterios”, “ilustración”, “identificación de criterios” y “trabajos para realimentación” obtuvieron las mediciones más elevadas. En cuanto a los grupos, hubo diferencias en el nivel de medición de los desempeños de los docentes, dependiendo del nivel de estudios, siendo el grupo de nivel disciplinar los que mayores puntajes otorgaron en los cinco ámbitos de desempeño docente y en tres ámbitos de desempeño estudiantil. El grupo de nivel general obtuvo las valoraciones más bajas, en los cinco ámbitos de desempeño docente y en tres de desempeño estudiantil. En cuanto a la variable sexo, las mujeres obtuvieron mayores valoraciones en “Exploración competencial” y “Evaluación” para desempeño docente y “Identificación de criterios” y “Trabajos para Realimentación” en cuanto a desempeño estudiantil. Los hombres obtuvieron mayores niveles en “Explicitación de criterios”, “Ilustración” y “Realimentación” para desempeño docente y “Competencias precurrentes”, “Ajuste a Ilustración” y “Evaluación / Participación” para desempeño estudiantil. La principal

limitación del estudio es el no uso de pruebas inferenciales o pruebas de magnitud del efecto para la comparación de los grupos.

Morales y Chávez (2019) investigaron la satisfacción académica en universitarios, en función de variables educativas y personales. Si bien la satisfacción académica es una variable que abarca muchos aspectos (infraestructura, beneficios, etc.) que ofrece el programa educativo y la evaluación que hace el estudiante según sus características personales, dentro de esa evaluación se considera relevante la opinión sobre la enseñanza del docente y la percepción que tiene el propio estudiante sobre su eficacia para estudiar, por lo que se puede decir que esta variable está relacionada a la investigación actual. La muestra fue compuesta de 535 alumnos de la carrera de Psicología, provenientes de la Universidad Michoacana de San Nicolás de Hidalgo. El rango de edad fue entre los 18 a los 23 años y el 73% de la muestra son mujeres. La distribución de frecuencias para cada semestre fue similar: 2do(26.8%), 4to(26%), 6to(24%), y 8vo(23.2%). El instrumento de medición fue la Escala de Satisfacción Académica con el ambiente pedagógico, compuesto de 8 ítems y de estructura unidimensional. Los instrumentos para evaluar las variables personales fueron: La Escala de Autoeficacia percibida en situaciones académicas – EAPESA; y el Inventario Maslach de Burnout para estudiantes – MBI-SS. En los resultados de esta investigación, se halló una correlación entre satisfacción académica y autoeficacia académica ($r = 0.414$; $p < .01$). En la comparación de satisfacción académica según los cuatro grupos de semestre académico, se hallaron diferencias estadísticamente significativas ($F = 4.97$; $p = .002$), en favor del grupo de octavo ciclo. En la comparación por turno de estudios, se hallaron diferencias estadísticamente significativas ($t = 2.10$; $p = .036$), a favor del el turno matutino ($M = 26.6$; $DE = 4.25$), aunque la distancia de puntaje no es tan lejana con el turno vespertino ($M = 25.4$; $DE = 4.19$).

1.2. Objetivos

1.3.1. Objetivo General

Determinar la relación entre el desempeño docente y desempeño estudiantil percibidos por estudiantes de psicología de una universidad estatal de Lima – 2019.

1.3.2. Objetivos Específicos

- a. Identificar las fuentes de evidencias de fiabilidad, contenido del test, estructura interna global y por grupos, de los Cuestionarios sobre desempeño docente (CDD) y desempeño estudiantil (CDE) en estudiantes de la carrera de psicología de una universidad pública de Lima.
- b. Identificar la relación entre las dimensiones del desempeño docente y desempeño estudiantil percibidos en estudiantes de la carrera de psicología de una universidad pública de Lima.
- c. Identificar las diferencias grupales en la variable desempeño docente, según las variables de control, en estudiantes de la carrera de psicología de una universidad pública de Lima.
- d. Identificar las diferencias grupales en la variable desempeño estudiantil, según las variables de control, en estudiantes de la carrera de psicología de una universidad pública de Lima.

1.3. Justificación

Este proyecto se justifica en 3 aspectos: teórico, metodológico y social-educativa. En el primer aspecto, para describir los desempeños de docentes y estudiantes, se usan los ámbitos didácticos en formato de dimensiones plasmadas en cuestionarios, bajo el sustento del modelo de interacciones didácticas.

Respecto a la metodología, el estudio aplico cuestionarios sometidos a los procesos de obtención de evidencias de fiabilidad y validez requeridos por los estándares para pruebas educativas y psicológicas. En cuanto a los análisis comparativos, se añadió los tamaños del efecto como medida descriptiva más precisa para evaluar las diferencias o semejanzas entre los grupos. Los tamaños del efecto son medidas requeridas por la APA.

Sobre la justificación para el ámbito educativo: Los instrumentos obtuvieron evidencias de fiabilidad y validez para su uso en estudiantes universitarios de psicología. Así mismo, los resultados de la investigación actual pueden ser útiles para la evaluación de la calidad educativa ofrecida por la universidad. De los resultados actuales, se pueden formular programas de capacitación que incidan en las dimensiones donde los docentes obtienen bajos puntajes.

1.4. Hipótesis

1.5.1. Hipótesis General

Existe una correlación positiva entre el desempeño docente y desempeño estudiantil percibidos por estudiantes de psicología de una universidad pública de Lima, 2019.

1.5.2. Hipótesis Específicas

- a. Existen correlaciones entre las dimensiones del desempeño docente y desempeño estudiantil percibidos en estudiantes de la carrera de psicología de una universidad pública de Lima.
- b. Existen diferencias grupales en la variable desempeño docente, según las variables de control, en estudiantes de la carrera de psicología de una universidad pública de Lima.
- c. Existen diferencias grupales en la variable desempeño estudiantil, según las variables de control, en estudiantes de la carrera de psicología de una universidad pública de Lima.

II. Marco teórico

El objetivo de esta sección es exponer las bases de los modelos básicos, interdisciplinarios y transdisciplinarios que servirán para el estudio de las interacciones educativas.

2.1. Bases teóricas sobre el tema de investigación

2.1.1. *Reseña del programa de investigación*

Jacob Robert Kantor es el fundador de la “psicología interconductual”, un modelo metasistémico sobre un proyecto de ciencia psicológica o de la psicología, proponiendo como base el modelo teórico de campo (Kantor, 1978). Se postula que el objeto de estudio debe ser la conducta, pero redefinido conceptualmente como “interconducta”, para enfatizar la funcionalidad interactiva de las relaciones entre el organismo y el ambiente (Kantor & Smith, 2015). A su vez, esta “interconducta” se esquematiza mediante un modelo de campo psicológico, una construcción teórica cuyo objetivo es definir los eventos psicológicos como factores interrelacionados de distintas lógicas explicativas: la interacción, la possibilitación y la probabilización (Kantor & Smith, 2015). Kantor fue un epistemólogo que abordó diversos problemas disciplinares de la psicología, como las raíces filosóficas de los diversos proyectos científicos de disciplina psicológica, los niveles de estudio intradisciplinarios de la psicología, las relaciones multidisciplinarias de la psicología con otras ciencias, y las formas de aplicación tecnológica del conocimiento psicológico. Después de su retiro como docente universitario, Kantor es invitado a México por el grupo Xalapa (pioneros de la investigación científica psicológica en ese país), donde conoce a Emilio Ribes Iñesta, quien luego de unos años adopta los planteamientos metateóricos de Kantor y desarrolla un programa de investigación experimental sobre el comportamiento animal y humano, denominado “Teoría de la conducta” (Pérez-Almonacid & Gómez, 2014).

Emilio Ribes y Francisco López Valadez escribieron en 1983 “Teoría de la conducta: Un análisis de campo y paramétrico”, un libro que busco tres objetivos: 1) Argumentar una crítica hacia la psicología conductista operante, aduciendo la insuficiencia del análisis atomista y causalista de la teoría del condicionamiento; 2) Reemplazar la teoría del condicionamiento por una teoría de la conducta, mediante un análisis de campo y paramétrico, basados en la psicología interconductual de J. R. Kantor; y 3) Formular una taxonomía funcional de la conducta, basadas en los conceptos de mediación y desligamiento funcional (Ribes & Lopez, 1985). Este libro fue el corazón del programa de investigación de la psicología interconductual, ya que la gran mayoría de la producción científica la refiere directa o indirectamente. Teoría de la conducta también ha sido la base para la construcción de los primeros modelos interdisciplinarios, como extensiones a los ámbitos sociales de la salud, la educación, la ciencia, etc.(Ribes, 2010). De todos estos ámbitos sociales se destaca la educación por recibir mayores aportes. Se debe mencionar que la amplitud de los estudios empíricos citados y el propio Modelo de Interacciones Didácticas se fundamentan en “Teoría de la conducta”(Ribes & Lopez, 1985).

El programa de investigación de la psicología interconductual se puede dividir en cuatro niveles del desarrollo y extensión de toda ciencia: disciplinar, transdisciplinar, interdisciplinar y multidisciplinar. Se explicará brevemente cada uno de estos niveles:

- a. Disciplinar: actividades de investigación referentes al objeto de estudio de la psicología, considerando la complejidad y especificidad científica (Kantor, 1978).
- b. Multidisciplinar: mediante la conjugación del conocimiento (trabajo colaborativo) de dos o más disciplinas se genera de un nuevo objeto de estudio, que pierde su especificidad al combinar elementos teóricos y metodologías específicas.

- c. Transdisciplinar: Hace referencia a disciplinas (generalmente matemáticas o tecnológicas) que son usadas por otras disciplinas para solucionar sus problemas internos de investigación (Camacho et al., 2012).
- d. Interdisciplinar: colaboración e intercambio de información entre dos o más disciplinas con fines tecnológicos y/o profesionales, es decir, la solución de problemas de relevancia social.

La presente investigación se ubica en el nivel interdisciplinar, ya que se emplea modelos derivados de “Teoría de la conducta”, en conjunto con otros conocimientos tecnológicos educativos, para identificar, describir, modelar, diagnosticar, cambiar y predecir los problemas del ámbito social de la educación, específicamente la enseñanza superior universitaria en pre – grado de carreras de psicología.

2.1.2. Clasificación del ejercicio de la psicología

Ribes (1982) analizo y reflexiono sobre la práctica profesionista del psicólogo en el contexto del comienzo de la proliferación de áreas laborales nuevas para el ejercicio de la psicología en Hispanoamérica. Identifico las graves falencias de la denominada “psicología aplicada” resumidas en los siguientes puntos: 1) Endebles bases de conocimiento científico, reflejada por ejemplo, en la ausencia de un objeto de estudio consensuado; 2) poca o nula conexión entre la aplicación tecnológica y el conocimiento científico; 3) aplicación de modelos ad hoc pero sin sustento teórico o incluso tecnológico; 4) Conflictos e intrusismo entre disciplinas tradicionales y disciplinas “psicologizadas”; etc.

Ante este escenario, donde el psicólogo puede ser un científico que estudia el comportamiento, y/o también un profesional que aplica técnicas sobre el comportamiento para intervenir en problemas personales y sociales, se han planteado diversos abordajes para solucionar el conflicto disciplinar/profesional. Desde la psicología interconductual, el planteamiento más conocido es el de Ribes (2009) que afirma que la psicología es una

disciplina científica básica en primer lugar, de cual el conocimiento abstracto y analítico generado puede trasladarse a otros escenarios interdisciplinarios siguiendo una serie de consideraciones recomendadas por el autor. Silva et al. (2016) continúan con el planteamiento de Ribes, agregando lineamientos para su enseñanza educativa-universitaria. Para la investigación actual se asumirá el planteamiento de Silva, Ruiz, Aguilar, Canales y Guerrero, que se pasará a explicar a continuación.

El objetivo de la revisión fue proponer una clasificación de tres roles profesionales que puede asumir el psicólogo; delimitar sus ámbitos y competencias; y dar lineamientos para su enseñanza universitaria. Los tres roles o prácticas profesionales son la científica, la tecnológica y la técnica. Para entender la práctica científica, se asume en primer lugar que la psicología es una ciencia básica que tiene por objetivo la indagación y sistematización del conocimiento sobre la interconducta de los organismos, con el fin de comprenderla. Estas prácticas, generalmente englobadas como “investigación científica”, son comportamientos lingüísticamente enmarcados en ámbitos funcionales, que para el estudio de la conducta científica, Ribes (1993) las delimitó en los siguientes juegos de lenguaje: a) identificación de hechos; b) formulación de preguntas pertinentes a los problemas; c) la aparatología; d) la observación; e) la evidencia; y f) las inferencias y las conclusiones. El psicólogo investigador puede ubicar cualquiera de sus tareas y escenarios en alguno de los juegos de lenguaje descritos.

Para la práctica tecnológica y práctica técnica, Silva et al. (2016) realizan algunas distinciones entre la ciencia y la tecnología, reflexionan sobre la naturaleza valorada de todo problema social y su correspondiente dimensión psicológica, y como se puede transferir el conocimiento científico a escenarios de aplicación tecnológica, desde una perspectiva interconductual, mediante la elaboración de modelos de interfase entre ambos tipos de lenguajes. Para el caso de la tecnología, esta se definió como una práctica humana para elaborar productos o sistemas que permitan la transformación, eliminación o creación de objetos o

condiciones particulares concretas con una finalidad específica. Lo último puede entenderse como las demandas sociales valoradas por algún grupo de referencia, que busca la modificación, eliminación o creación de tal objeto o condiciones. ¿De dónde proviene la tecnología? Esta pregunta tiene dos respuestas, en un inicio surgió como sistematización de las técnicas tradicionales u operativas para resolver un problema, para luego, en una época posterior de desarrollo histórico, tomar el conocimiento científico y sintetizarlo de acuerdo a las particularidades del problema. Un ejemplo ilustrativo es la medicina, que surgió como tecnología por la sistematización de las técnicas tradicionales para aliviar las enfermedades, para en una época posterior alimentarse del conocimiento de la biología y elaborar medicinas e intervenciones más efectivas. En este punto, los autores señalan un aspecto crítico: la tecnología no se aplica sola, sino que es aplicada por grupos humanos a circunstancias específicas, y estos grupos se les conoce como “técnicos”.

Se define a la técnica como un procedimiento o conjunto de reglas, normas o protocolos, que aplicado se puede obtener un resultado predicho (Silva et al., 2016). Se predica del técnico pericia, experiencia y habilidad para usar los protocolos y recursos, de acuerdo a un sistema tecnológico de referencia. Para el caso de la psicología en general, desde la segunda mitad del siglo XX se definió al psicólogo como un técnico, que evalúa y modifica el comportamiento de las personas según el contexto y/o área social en el que se encuentre.

Desde la propuesta de Silva et al. (2016), siguiendo a Ribes, se define el rol de un psicólogo técnico como “un experto en el diseño y evaluación de procedimientos de modificación de conducta y se vuelve en un asesor metodológico en la formulación de problemas y planteamiento de soluciones” (p. 232). La forma en la que trabaja el psicólogo tecnólogo no se limita a interactuar directamente con los usuarios que contratan el servicio, sino que también puede trabajar indirectamente, mediante la colaboración con otros profesionales que tratan con la dimensión individual del comportamiento de las personas,

involucradas en escenarios sociales. Un ejemplo de este trabajo indirecto es cuando un psicólogo asesora a profesores en planes para cambiar el comportamiento de alumnos con alguna dificultad, donde es el profesor quien interactúa de manera directa con el alumno.

Los tecnólogos y técnicos deben trabajar de manera colaborativa o directiva con el resto de especialistas profesionales, transfiriendo el conocimiento de la ciencia básica analítica a escenarios sociales, mediante sistemas tecnológicos sintéticos. Para lograrlo, deben crear y usar un lenguaje interfase o modelo puente que facilite la comunicación con los otros especialistas, sin perder la especificidad del análisis del comportamiento. Los roles mencionados han sido sistematizados en Carpio et al., (1998) y actualmente se consideran como los ámbitos funcionales del desempeño tecnológico y técnico del psicólogo, los cuales se presentan a continuación:

- a) Análisis de las demandas, consiste en la recolección de las demandas explícitas y tácitas de parte de los usuarios de los servicios y se contrastan con los sistemas o marcos tecnológicos.
- b) Formulación de problemas y soluciones, donde las demandas de los usuarios, bajo el análisis del sistema o marco tecnológico, se analiza como un problema y se plantean las soluciones posibles, bajo un proceso desde la identificación del problema en el lenguaje ordinario, hasta el análisis funcional o contingencial.
- c) Definición y establecimiento de estrategias y criterios de éxito, se refiere al diseño de estrategias para el diagnóstico e intervención bajo protocolos tecnológicos, así también los medios de verificación para evaluar si fue exitosa la intervención.
- d) Intervención, es el ámbito donde se aplican operacionalmente las estrategias de intervención y su respectiva medición.

- e) Evaluación y seguimiento, donde de acuerdo a las mediciones antes, durante y/o después de la intervención, se evalúa su éxito o fracaso. Esta etapa también incluye la opción de redefinición de los problemas e implementación de nuevas soluciones.
- f) Transferencia tecnológica, es el ámbito donde se transmite, de manera sintética, protocolos o técnicas de aplicación directa en circunstancias específicas, mediante el entrenamiento de técnicas de observación y alteración del comportamiento propio y/o de otros. Las personas a las cuales se transfiere pueden ser los usuarios finales (ejemplo: el caso de un trabajador que quiere tratar su estrés), como también a los profesionales o paraprofesionales que tienen contacto con los usuarios finales (ejemplo: el entrenamiento para administradores de un centro laboral, para que sean ellos quienes “capaciten” a su personal en técnicas para el manejo del estrés).
- g) Investigación tecnológica, donde se abordan los ámbitos descritos anteriormente como “problemas de investigación”, planteando modelos, métodos, evidencias para el diseño de nuestras estrategias de evaluación e intervención tecnológica.

Si se desea diseñar un currículo para la formación de un especialista en psicología tecnológica y técnica, esta deberá considerar una lista de habilidades y competencias enmarcadas en los ámbitos de desempeño descritos anteriormente. Y si se agrega al perfil un perfil de índole científico, deberán considerar las competencias enmarcadas en los ámbitos de desempeño de la práctica científica. Sin embargo, dada la gran cantidad de exigencias competenciales que tienen los diversos escenarios sociales o ámbitos de la sociedad (ejemplo: las instituciones educativas, las organizaciones laborales, etc.), se convierte en un desafío educacional formar psicólogos. Algunos problemas derivados de este desafío son mencionados por los autores, como el eclecticismo en la enseñanza y ejercicio de la psicología, la confusión vocacional de los ingresantes a la carrera y los problemas de inserción laboral (Silva et al., 2016).

El análisis presentado, además de aportar reflexiones al ejercicio de la práctica, sirvió a la presente tesis para clasificar en tres grupos a las asignaturas dictadas por los profesores evaluados mediante la encuesta de desempeño docente y estudiantil:

1. Fundamentos básicos de psicología: Agrupa a las asignaturas que ofrecen una introducción didáctica a los temas principales de la psicología científica.
2. Investigación científica: Agrupa a las asignaturas que entrenan habilidades en los ámbitos de desempeño de la práctica científica y práctica tecnológica.
3. Especialización profesional: Agrupa a las asignaturas que entrenan habilidades en los ámbitos de desempeño de la práctica técnica de la psicología.

2.2. Modelos de Interacciones Didácticas

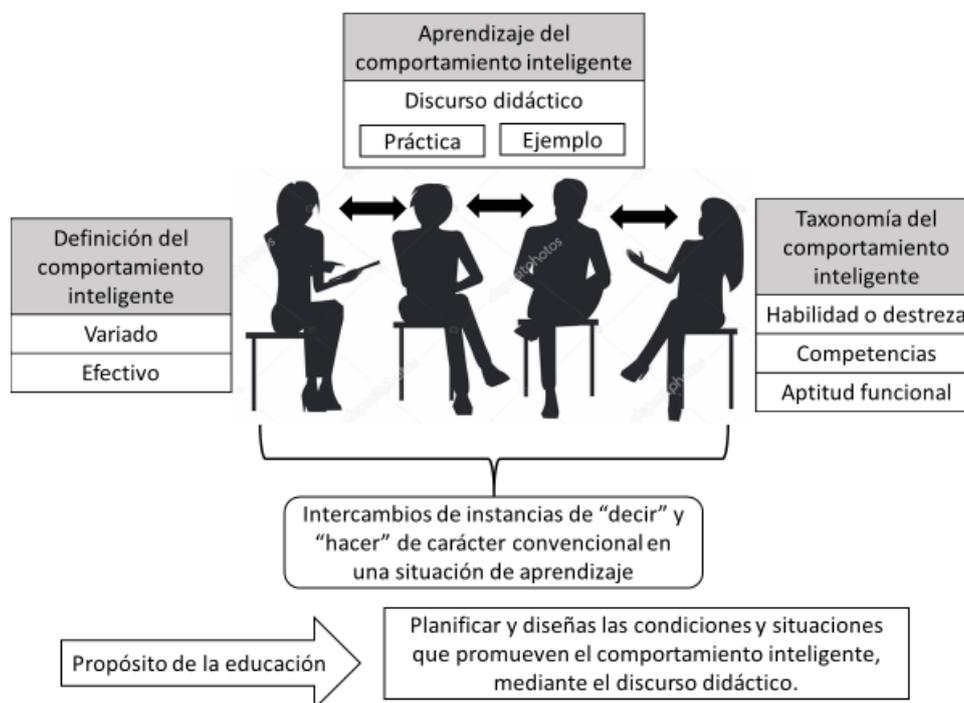
Los inicios del modelo se remontan a los primeros textos de análisis y propuestas de la aplicabilidad del conocimiento de la psicología interconductual al ámbito de la educación, siendo uno de los primeros análisis el tema de la inteligencia y el comportamiento inteligente (Ribes, 1981, 1989). En el libro “Psicología general” (Ribes, 1990), en el capítulo 7 llamado “Aptitudes sustitutivas y planeación del comportamiento inteligente en instituciones educativas” se expone una primera propuesta para la planeación de la enseñanza del comportamiento inteligente mediante la práctica, el ejemplo y el discurso didáctico, logrando como resultado un desempeño variado y efectivo en múltiples situaciones. “Practicar” y “ejemplificar” son términos conocidos en psicología y pedagogía, a diferencia del término “discurso didáctico” que proviene del análisis filosófico hecho por Gilbert Ryle sobre los conceptos mentales y las implicaciones de su uso (Ryle, 2005). Originalmente Ryle creó el concepto a partir de su análisis sobre el intelecto, donde fue usado para diferenciar las potencialidades, propensiones y desempeños intelectuales, aduciendo de estos últimos que “son resultados de las lecciones que se aprenden por lo menos en parte de libros y conferencias, o en general, del discurso didáctico” (p.284). Sus características son:

- a. Está fundamentado en el habla conversacional.
- b. Se da en estilo magistral y estudiado, más que sociable o espontáneo.
- c. Debe ser recordado, imitado y ensayado por el estudiante.
- d. Es impersonal y su aplicación no se limita a circunstancias específicas.
- e. Su propósito es enseñar, instruir, volver competente al estudiante.

Regresando al primer análisis de Ribes, este describe las características del discurso didáctico como una forma de enseñar el comportamiento inteligente que trasciende y supera a las estrategias y/o programas de enseñanza contemporáneos de la educación mexicana del siglo XX. Agrega también que el discurso didáctico promueve primordialmente las aptitudes de tipo sustitutivo referencial y no referencial, ya que su medio de contacto por excelencia es el convencional. Esta última característica es importante destacarla, ya que el autor agrega esta característica, en contraste con la definición original, donde se habla de los desempeños intelectuales en los entornos académicos. Mediante el siguiente esquema se puede representar la propuesta de planeación del comportamiento inteligente.

Figura 1

Esquema representativo de la propuesta de la planeación de la enseñanza del comportamiento inteligente (Ribes, 1990).



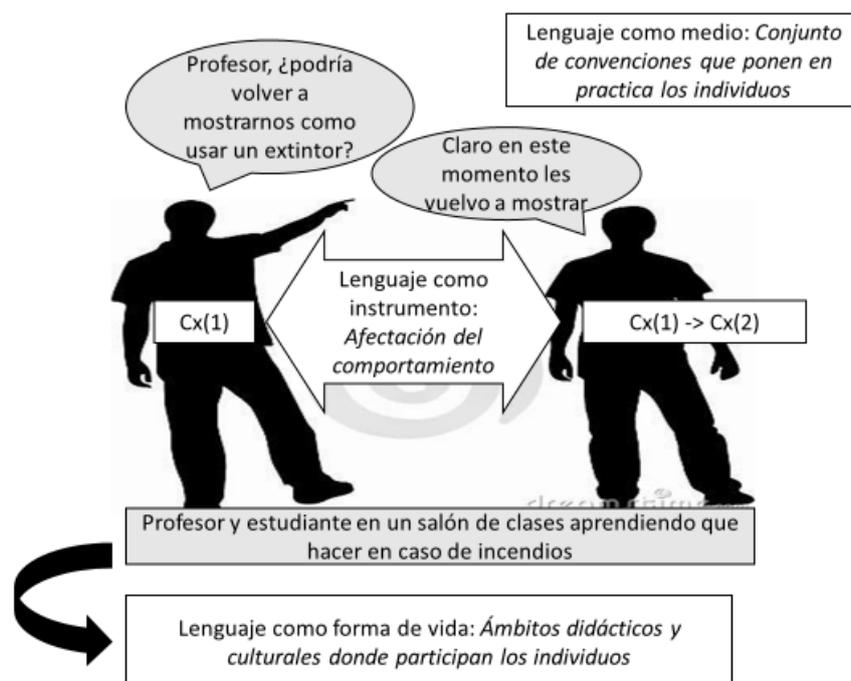
En el último párrafo del capítulo, se menciona por primera vez el concepto de “interacciones didácticas” refiriéndose a que la relación maestro – aprendiz y otras variables educativas deben ser investigadas mediante estudios longitudinales, observacionales y experimentales, para identificar los componentes, condiciones, intercambios y emergencia de las interacciones efectivas y/o comportamientos en un nivel sustitutivo. A partir de este primer análisis, la investigación se direccionó hacia un tema poco abordado por la psicología contemporánea, los cuales son: el estudio psicológico de la práctica científica y su enseñanza. En cuanto al primer tema, en Ribes (1993) y Ribes et al. (1996) se encuentra la fundamentación del modelo psicológico de estudio de la práctica científica individual. A partir de este modelo, se inicia otra línea de investigación sobre la enseñanza de la práctica científica, específicamente sobre la ciencia de la psicología y en el nivel educativo superior universitario (Carpio et al., 1998).

La enseñanza e instrucción de una disciplina científica en el contexto de la educación implica reconocer y diferenciar los distintos escenarios sociales donde se comportarán los individuos y los criterios a los que se ajustarán. Para ello, Ribes et al., (1992) introducen el

término de “juego de lenguaje” como un concepto que describe la red de prácticas convencionales del lenguaje a los cuales los individuos se ajustan. Según Wittgenstein (1999), hablar sobre el lenguaje no significa hablar sobre las morfologías lingüísticas y sus reglas gramaticales, sino sobre un juego de lenguaje, constituido por un grupo de reglas y roles que los individuos practican. Hablar sobre el lenguaje implica una triple dimensionalidad: el lenguaje como medio, como instrumento y como forma de vida que le da sentido al comportamiento humano.

Figura 2

Representación del lenguaje como medio, instrumento y forma de vida



A partir de la noción de “juego de lenguaje” Carpio et al., (1998) tecnifican la noción mediante el concepto de “ámbito funcional del desempeño”, refiriéndose al conjunto de prácticas convencionales que se ubican en un determinado escenario social y cumplen criterios de logro características del escenario, modelando la morfología y función del comportamiento de los individuos. Para ejemplificar “ámbitos funcionales” clasificaron los escenarios de la

práctica científica del psicólogo, la práctica científica tecnológica y la práctica docente en educación superior.

Tabla 1

Ámbitos funcionales del desempeño (Carpio et al., 1998)

Práctica científica del psicólogo	Práctica tecnológica	Práctica docente
Identificación de hechos	Análisis de demandas	Exploración cognoscitiva
Preguntas pertinentes	Formulación de problemas y soluciones	Criterios
Aparatología	Definición de estrategias y criterios de éxito	Ilustración
Observación	Intervención	Práctica
Representación	Evaluación y seguimiento	Evaluación
Inferencia y conclusiones	Investigación tecnológica	
	Transferencia	

Esta pionera clasificación sobre los ámbitos funcionales del desempeño, según el tipo de práctica, abrió las puertas para el desarrollo de las próximas líneas de investigación sobre la enseñanza de la ciencia psicología (Irigoyen et al., 2007; Morales et al., 2010; Padilla et al., 2009) generando evidencia empírica (Irigoyen et al., 2007; Morales et al., 2005; Padilla, 2008).

En las siguientes secciones, se analizará los subprogramas de investigación de los modelos de “Interacciones Didácticas” (abreviado en adelante como: MID) producidos por los grupos de investigación expuestos anteriormente, mediante la respuesta de cada modelo a las siguientes preguntas: ¿Qué son las “interacciones didácticas”? ¿Cuáles son los factores de las “interacciones didácticas”? ¿Cómo se investiga las “interacciones didácticas”? y ¿Cuáles son las críticas al modelo de interacciones didácticas?.

2.2.1. MID de Carlos Ibáñez

El primer modelo por examinar es el liderado por el Dr. Carlos Ibáñez, que surgió a partir del análisis interconductual de los procesos educativos (Ibáñez, 1994; Ibáñez & Ribes, 2001) derivando un modelo de interacciones didácticas cuyo objetivo es identificar los factores y procesos psicológicos fundamentales en el ámbito de la educación institucional. Consideramos los libros “Metodología para la planeación de la educación superior”(Ibáñez, 2007) y “Una propuesta psicopedagógica para el aprendizaje de competencias básicas en educación superior” (Reyes et al., 2009) como las referencias principales del modelo de interacciones didácticas de Carlos Ibáñez y colaboradores.

¿Qué son las “interacciones didácticas”?

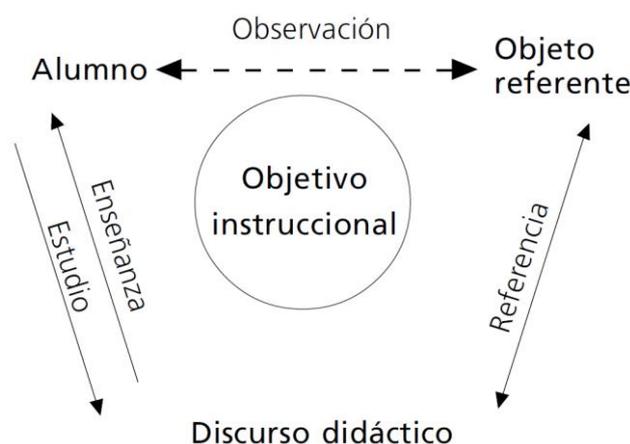
Las interacciones didácticas son las relaciones que se establecen entre los agentes y factores de los procesos educativos durante un episodio instruccional.

¿Cuáles son los factores de las “interacciones didácticas”?

El objetivo instruccional, el alumno, el discurso didáctico y el objeto referente. En el siguiente esquema se muestra como están organizados y que tipos de procesos se establecen a partir de su interacción bajo condiciones idóneas.

Figura 3

Modelo de interacciones didácticas propuesto por Ibáñez



Nota. Figura original de Ibáñez, C. (2007). Un análisis crítico del modelo del triángulo pedagógico: Una propuesta alternativa. *Revista mexicana de investigación educativa*, 12(32), 435-456.

1. El objetivo instruccional enmarca funcionalmente todos los procesos
2. Proceso del discurso didáctico al alumno: Enseñanza
3. Proceso del alumno hacia el discurso didáctico: Estudio
4. Proceso del estudiante con el objeto referente: Observación
5. Proceso del discurso didáctico con el objeto referente: Referencia. También se exponen dos subprocesos del contacto entre la Referencia y el objetivo instruccional: Congruencia y Coherencia.

¿Cómo se investiga las “interacciones didácticas”?

Se debe optar por una estrategia manipulativa de las variables, con diseños experimentales y cuasiexperimentales para determinar las condiciones situacionales que generan competencias. Para ello se necesita de un paradigma metodológico y operacional para el análisis sistemático de las situaciones educativas, segmentadas en dos episodios básicos: Episodio instruccional y episodio de aplicación. Las investigaciones realizadas bajo este modelo han variado los factores, los procesos y otros elementos de los episodios instruccionales y de aplicación, A continuación, se presenta un resumen de estas variables:

1. Dominio
2. Nivel de especialización de la competencia con relación al dominio
3. Nivel de aptitud funcional de la competencia a entrenar: Contextual, Suplementaria, Selectora, Sustitutiva referencial y Sustitutiva no referencial.
4. Objetivo instruccional: Presencia o ausencia, Función en un episodio instruccional.
5. Alumno: Condiciones situacionales y organísmicas.
6. Discurso didáctico: Modalidad lingüística.

7. Objeto referente: Presencia o ausencia, Tipo: Físico (real o sucedáneo) o convencional.
8. Proceso de Estudio: Activo o pasivo.
9. Proceso de Enseñanza: Arreglo particular de la secuencia de presentaciones del discurso didáctico.
6. Proceso de Observación: Activo o pasivo.
7. Proceso de Referencia: previo, simultáneo, posterior del discurso didáctico a la presencia del objeto referente.

¿Cuáles son las críticas al modelo de interacciones didácticas de Carlos Ibáñez?

Los principales cuestionamientos provienen de Morales et al. (2013) los cuales se resumen en los siguientes puntos: Problemas lógicos en la representación de relaciones o procesos entre conceptos funcionales (discurso didáctico) y conceptos morfológicos (alumno, objetos referentes); problemas en el calificativo “didáctico” añadido a algunos factores y procesos antes de cumplir su finalidad “didáctica”; desacuerdo con la definición de “discurso didáctico” como un tipo especial de habla para instruir a otros. Estos cuestionamientos tienen respuestas en Ibáñez (2011). Se debe agregar la carencia de recomendaciones de elaborar investigaciones metodológicas e investigaciones empíricas de estrategia asociativa: estudios comparativos retrospectivos, transversales y longitudinales (Ato et al., 2013). Para el presente trabajo se tomó algunos abordajes conceptuales que este programa de investigación hizo sobre el desempeño del estudiante en las interacciones didácticas.

2.2.2. MID de la FES Iztacala - UNAM

El segundo y tercer modelo guardan bastante correspondencia en la estructura teórica y hay una continuidad de avance entre las investigaciones de ambos grupos. Consideramos que sus líneas surgen a partir de Carpio et al. (1998) y Carpio & Irigoyen (2005). Los presentaremos por separado por propósitos ilustrativos.

El segundo modelo es generado por el grupo de investigación interconductual de la FES Iztacala de la UNAM. Los últimos avances investigativos de su modelo de interacciones didácticas los encontramos en Morales et al. (2013); Morales et al. (2016) y Morales et al. (2017).

¿Qué son las “interacciones didácticas”?

Categoría analítica que aborda funcionalmente de la relación en la que participan un docente y un estudiante, para que este último logre los criterios disciplinariamente esperados. La interacción didáctica se estructura alrededor de un criterio al que hay que ajustarse.

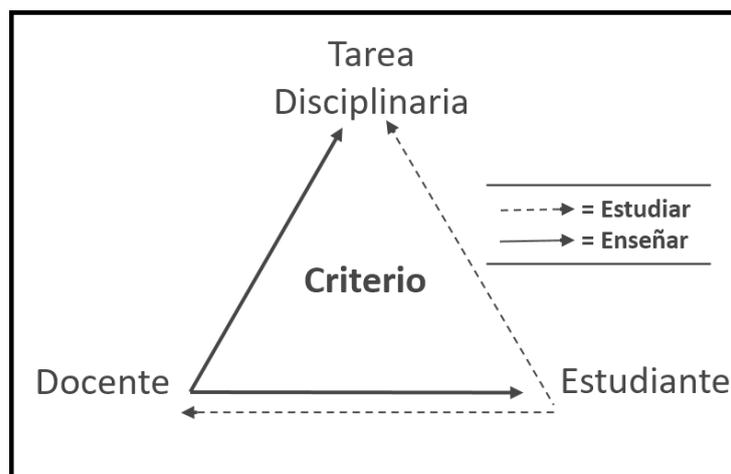
¿Cuáles son los factores de las “interacciones didácticas”?

El criterio de ajuste, el desempeño del docente, el desempeño del estudiante, la tarea disciplinaria, los factores disposicionales (tipo situacional o histórico) y el programa o plan de estudios.

De similar manera al anterior modelo, cuando interactúan estos factores se establecen relaciones de importancia fundamental para la investigación educativa. El orden de los factores y relaciones se muestra en la siguiente figura.

Figura 4

Modelo de las relaciones de enseñar y estudiar estructuradas por el criterio de logro



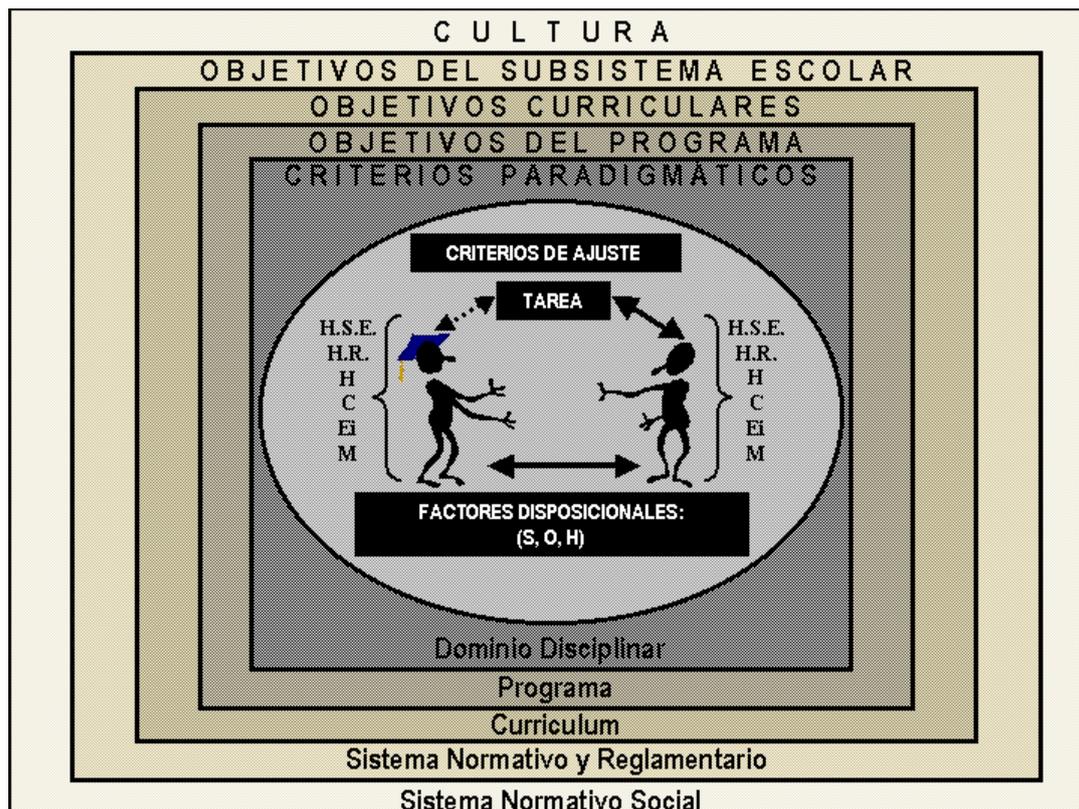
Nota. Figura original de Morales, G., Peña, B., Hernández, A., & Carpio, C. (2017). Competencias didácticas y competencias de estudio: su integración funcional en el aprendizaje de una disciplina. *Alternativas en Psicología*, 37(1), 24-35.

1. Relaciones de enseñar: Compuesto de la relación docente-tarea, expresado en el desempeño en un dominio disciplinar, y la relación docente-estudiante, cuyos desempeños se enmarcan en un dominio didáctico (también denominado ámbito didáctico).
2. Relaciones de estudiar: Compuesto por la relación estudiante-tarea disciplinaria y la relación estudiante - (relación docente-tarea disciplinaria), ambas moduladas por el criterio de logro.
3. Aprendizaje: relación exitosa entre enseñar y estudiar.

Se debe considerar que estas relaciones también se encuentran subordinadas a otros niveles analíticos, como el de las competencias, los cuales se muestran en el siguiente esquema:

Figura 5

Niveles analíticos de la interacción didáctica



Nota. definición de abreviaturas: HSE= Historia Situacional Efectiva, HR= Historia de referencialidad, H= Habilidades, C= Competencias, EI= Estilos interactivos y M= Motivos. Figura original de León, A., Morales, G., Silva, H., & Carpio, C. (2011). Análisis y evaluación del comportamiento docente en el nivel educativo superior. En V. Pacheco & C. Carpio (Eds.), *Observación y métricas en el análisis del comportamiento*. (pp. 84-104). FES-Iztacala UNAM.

Las competencias docentes y estudiantiles se delimitan a partir del criterio de logro, que permite identificar la efectividad e ineffectividad del docente y del estudiante, y el ámbito funcional o dominio específico, que enmarca el análisis del desempeño didáctico y desempeño estudiantil como inseparables y altamente correlacionables. Cuando cruzaron analíticamente los ámbitos funcionales y las interacciones didácticas con las competencias docentes o didácticas y las competencias estudiantiles, obtuvieron 7 ámbitos funcionales (juegos de lenguaje) del desempeño didáctico y 9 ámbitos funcionales del desempeño estudiantil. Hipotéticamente 7 ámbitos didácticos y estudiantiles se correlacionarían. Los restantes son

ámbitos, didácticos o estudiantiles, que se correlacionarían con la tarea disciplinar o algún nivel más grande de expresión del dominio disciplinar.

Finalmente, estas prácticas se dan bajo diferentes modalidades de las interacciones didácticas. El criterio de organización es la forma de la modalidad, clasificada en interacciones didácticas lectoras, escritoras y orales.

¿Cómo se investiga las “interacciones didácticas”?

Consideran los diseños experimentales y observacionales (in-situ) como útiles para la indagación empírica de los factores que probabilizan las competencias de estudio, considerando las variables de orden, secuencia, tipo de criterio, complejidad funcional de la competencia, materiales empleados, tipo de ámbito, modalidad de las interacciones, variabilidad y constancia de los criterios impuestos, de las situaciones y los desempeños estudiantiles

Todas estas variables deben ser investigadas bajo un lenguaje operacional, para ello crean la noción de “episodio didáctico”, modulada por los siguientes factores:

1. El criterio de logro
2. La situación
3. Ilustración del desempeño ideal
4. El desempeño estudiantil
5. La retroalimentación
6. La evaluación

¿Cuáles son las críticas al modelo de interacciones didácticas del grupo de investigación interconductual de la FES Iztacala de la UNAM?

Se pueden plantear cuatro observaciones: La reformulación del concepto de “discurso didáctico” como interacción sustitutiva de contingencias, alejándose del sentido original ryleano (Ibáñez, 2011). Los problemas metodológicos en la evaluación del ajuste congruente

en interacciones didácticas lectoras y escritoras (Peña-Correal et al., 2012). Inconsistencias identificadas en el uso de los conceptos “criterio de ajuste” y “criterio de logro” (De la Sancha & Pérez-Almonacid, 2017). La insuficiencia de análisis al separar el desempeño didáctico del desempeño estudiantil que realizaron en anteriores trabajos y que el mismo grupo reconoció y corrigió (Morales et al., 2017). Como última observación, todavía no hay de investigaciones empíricas de estrategia asociativa: estudios comparativos retrospectivos, transversales y longitudinales (Ato et al., 2013).

2.2.3. MID de la Universidad de Sonora

El tercer modelo contiene pocas variaciones al segundo modelo de interacciones didácticas, lo cual es favorable para el avance y comunicación interdisciplinar. Es desarrollado por el grupo Seminario Interactum, de la Universidad de Sonora, cuyas principales actualizaciones las podemos identificar en Acuña et al. (2013) e Irigoyen et al. (2016).

¿Qué son las “interacciones didácticas”?

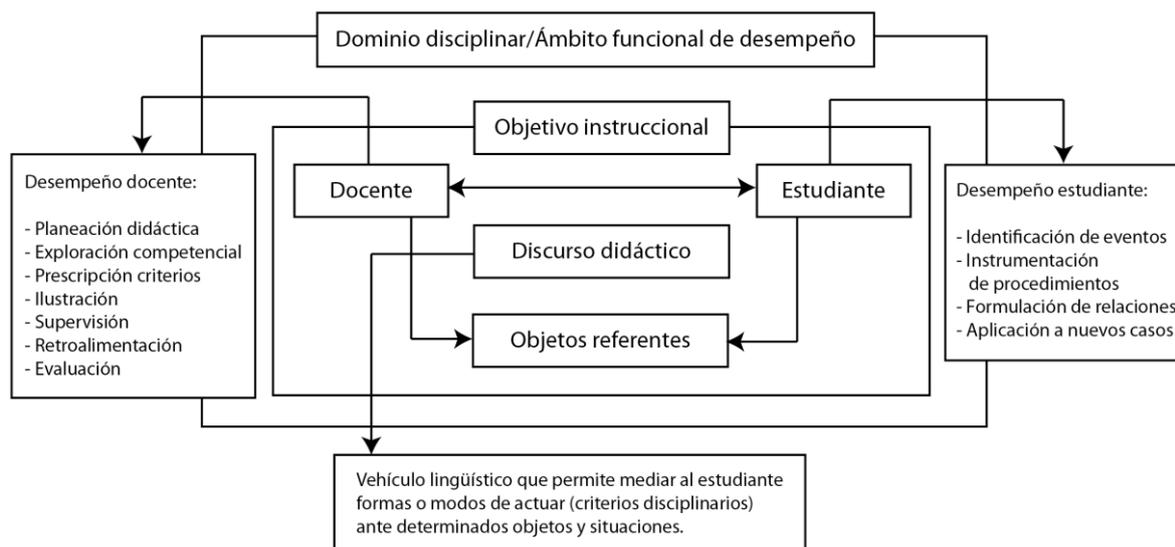
Es el segmento analítico para analizar las interacciones educativas (episodios instruccionales), definida en el intercambio recíproco entre sujetos (docente y estudiante) y objetos o situaciones referentes (materiales de estudio) en condiciones definidas por el ámbito de desempeño (disciplina o profesión).

¿Cuáles son los factores de las “interacciones didácticas”?

El desempeño del docente, el desempeño del estudiante, el objetivo instruccional, los objetos referentes (material de estudio), el dominio disciplinar o ámbito funcional de desempeño.

Figura 6

Factores de la interacción didáctica



Nota. Figura original de Irigoyen, J., Acuña, K., & Jiménez, M. (2011). Interacciones didácticas en educación superior. Algunas consideraciones sobre la evaluación de desempeños. En J. J. Irigoyen, K. Acuña, & M. Jiménez (Eds.), *Evaluación de desempeños académicos* (pp. 73-95). Universidad de Sonora.

El dominio disciplinar o ámbito funcional delimita los requerimientos a lograr en las relaciones entre los factores, los cuales son:

1. Docente – estudiante – objetos referentes: Denominado “discurso didáctico”
2. Docente – objetos referentes
3. Estudiante – objetos referentes: Denominado “compresión”
4. Estudiante – objetos referentes – Docente

¿Cómo se investiga las “interacciones didácticas”?

Los autores han realizado diversas investigaciones descriptivas y experimentales sobre las relaciones de los factores de las interacciones didácticas en el dominio o ámbito funcional de la ciencia, enfatizando en la relación estudiante – objeto referente, derivando un modelo de evaluación del aprendizaje de contenidos científicos, compuesto de cuatro grupos de factores:

1. Factores relacionados con el estudiante: Repertorios lingüísticos básicos y repertorios lingüísticos referidos a la disciplina de conocimiento.

2. Factores relacionados con los objetos referentes: Forma física, sustituta o audiovisual, modalidades de presentación y sus dimensiones,
3. Condiciones necesarias para la interacción: variables situacionales (estado del organismo y características de la situación) e históricas (historia situacional efectiva e historia de referencialidad).
4. Criterio de ajuste: tipo de tarea, nivel de complejidad conductual y modalidad lingüística implicada.

Mediante la interacción de estos factores, se evalúa como el estudiante se ajusta a los criterios de las tareas (objetos referentes) cumpliéndolos de manera efectiva, según el ámbito disciplinar de la ciencia. A esta competencia lograda se ha denominado “comprensión”, diferenciándolo de otras definiciones cognoscitivistas o constructivistas.

¿Cuáles son las críticas al modelo de interacciones didácticas del grupo Seminario Interactum?

Comparten la segunda y tercera críticas expuestas en el anterior modelo. Respecto a la primera observación sobre el tratamiento del concepto de “discurso didáctico”, Irigoyen et al. (2016) reafirman que el discurso didáctico promueve interacciones sustitutivas de contingencias, pero también interacciones situacionales, al redefinir el concepto como un arreglo instruccional que media de modo efectivo las situaciones problema del dominio disciplinar. También agregan procedimientos y pautas para su investigación y aplicación. Como última observación, se ha cuestionado el uso del concepto de “comprensión” como ajuste efectivo a criterios de logro (De la Sancha & Pérez-Almonacid, 2017). A diferencia de los dos anteriores, este grupo si ha generado investigaciones metodológicas e investigaciones empíricas de estrategia asociativa.

Una vez respondida las preguntas principales sobre los tres modelos de interacciones didácticas, se escogieron los MID de la FES Iztacala-UNAM y de la Universidad de Sonora como base para la presente investigación, agregando algunos componentes de otros abordajes,

para establecer una propuesta de investigación. Ante tal desafío nos planteamos la pregunta ¿se pueden comparar modelos teóricos?.

Generalmente se considera a la psicología como una ciencia multiparadigmática o preparadigmática, debido al conglomerado y variación de teorías y métodos para estudiar “lo psicológico”, resultando en diversos programas de investigación que aparentemente son incapaces de comunicarse. Ante este panorama, Gutiérrez (2018) ofrece perspectivas diferentes al problema de la integración de la psicología, hipotetizando en que la pluralidad teórica puede ser resultado de las complejas y múltiples relaciones entre los niveles disciplinares expresadas en las interacciones psicológicas. En sintonía con esta nueva perspectiva, Pérez-Almonacid (2018) propuso una clasificación de la psicología contemporáneo basado en los criterios: naturaleza del objeto de estudio, el propósito epistémico y el alcance disciplinario. Sus implicaciones son ofrecer una orientación para el análisis de las relaciones verticales y horizontales entre categorías supra – teóricas, teóricas y sub – teóricas, detectando cuales serían viables o inviables.

En el caso de los modelos de interacciones didácticas, establecer comparaciones en sus relaciones inter – teóricas es viable porque comparten una misma modalidad dimensional (una psicología no – mediacional, interdisciplinar de corte nomotético e ideográfico), las mismas categorías supra – teóricas (teoría de la conducta), y sus categorías teóricas y sub – teóricas son similares.

2.3. Marco conceptual de las interacciones didácticas

El proceso de enseñanza-aprendizaje ha sido comprendido tradicionalmente como un intercambio comunicativo para transmitir el conocimiento, estudiándose desde perspectivas pedagógicas carentes de bases psicológicas (Carpio & Irigoyen, 2005). Considerando las bases psicológicas de este intercambio, las acciones de los docentes y estudiantes cobran significados funcionales que permiten comprender de mejor manera el fenómeno educativo. En primer

lugar, se debe considerar que estas interacciones ocurren en un ámbito educativo, donde se exigen resultados, curricularmente expresados en desempeños ideales objetivo. En segundo lugar, las acciones de los docentes y estudiantes pueden ser analizadas desde modelos de análisis funcional, donde se hace una distinción entre la morfología de las acciones y la funcionalidad de las conductas, asumiendo un supuesto de sincronidad recíproca entre la conducta del docente y el estudiante. En tercer lugar, ubicar las características ambientales y sociales en las cuales se ven involucrados los actores educativos. Finalmente, asumir un marco teórico naturalista para representar al fenómeno educativo (Ibáñez & Ribes, 2001).

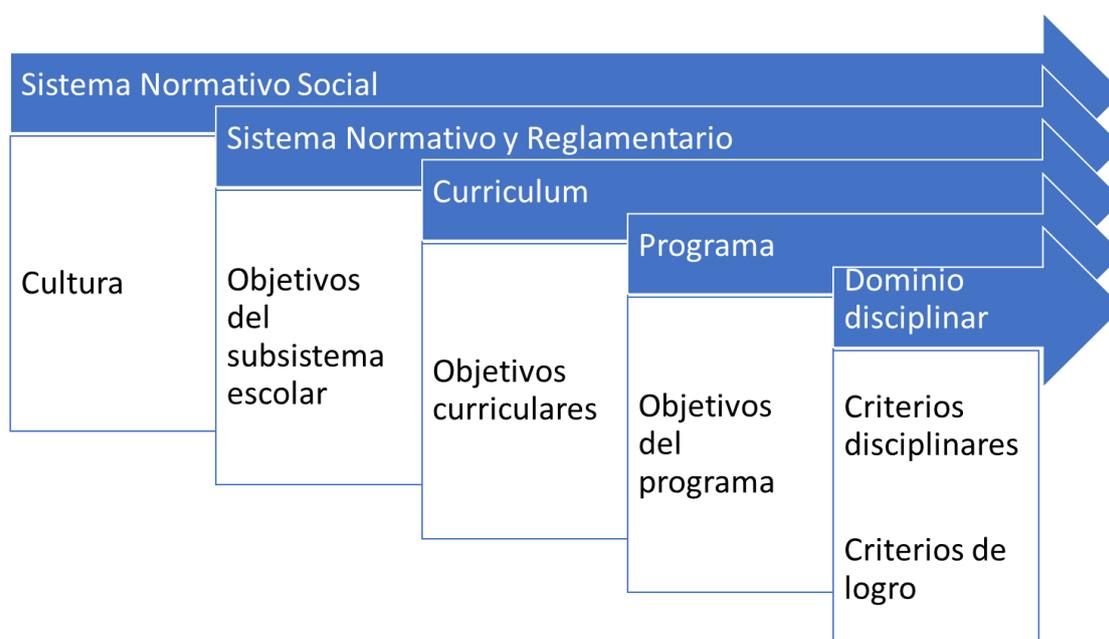
Desde la psicología interconductual, se propone el concepto de interacción didáctica, como segmento analítico para estudiar estas interacciones educativas, descritas anteriormente como “fenómeno educativo” o “proceso de enseñanza-aprendizaje”. En la interacción didáctica participan el docente, el estudiante y la tarea disciplinaria. Sus relaciones o interacciones son determinadas por el requerimiento de logro, que es una especificación para evaluar el éxito o error de un comportamiento (León et al., 2011). El requerimiento de logro se construye a partir de los criterios disciplinares o paradigmáticos pertenecientes a determinados ámbitos o dominios. Por ejemplo, un criterio de la disciplina matemática es la suma de números enteros, los cuales al ser especificados en un requerimiento de logro como “ejecución correcta de suma de números enteros” se pueden idear actividades para enseñarla.

La enseñanza, desde este marco conceptual, se da como circunstancia funcional delimitada por los criterios paradigmáticos y didácticos, para la promoción de los comportamientos disciplinarios, es decir, lo que habrá que “hacerse” y “decirse” cuando se enseña (Morales et al., 2013). Se debe especificar que los criterios paradigmáticos son las convenciones o reglas que comparten los miembros de una comunidad de un dominio disciplinar específico, como por ejemplo la disciplina matemática o la disciplina del entrenamiento físico o corporal.

El plan o programa de estudios reúne a los criterios disciplinarios y los define. Los planes de estudios representativos de las diferentes disciplinas se encuentran articulados mediante el plan curricular, el cual representa de manera general la fase educativa por la cual transita el estudiante, ya sea a nivel de educación básica o educación superior. A nivel nacional o de país, se encuentran los objetivos del subsistema escolar, a partir de los cuales los objetivos curriculares deben ser coherentes y congruentes. Finalmente, el objetivo de la educación formal es formar a los próximos ciudadanos, que se integran en el sistema normativo social (Morales et al., 2013). En la figura 7 se representan estos niveles:

Figura 7

Niveles analíticos de la educación



A partir de esta perspectiva panorámica de las interacciones educativas, se elaboró una tabla que resume de factores de las interacciones didácticas, ubicando también los ámbitos de la práctica didáctica, los cuales se definirán y explicarán en los siguientes subcapítulos.

Tabla 2

Propuesta de clasificación de los factores y ámbitos de las interacciones didácticas

Áreas	Factores y relaciones		
¿Cuáles son los niveles analíticos?	Cultura, objetivos del subsistema escolar, objetivos curriculares, objetivos del programa, criterios paradigmáticos o profesionales		
¿Quién se ajusta a los criterios paradigmáticos?	Docentes, Estudiantes o Programa de instrucción	Estudiantes	
¿Cómo se tipifica el desempeño de quien se ajusta?	Desempeño didáctico	Desempeño estudiantil	
¿Cuáles son los ámbitos funcionales del desempeño didáctico?	Exploración competencial		
	Prescripción de criterios		
	Planeación competencial	Ilustración	Autodirección del estudio
	Supervisión de la práctica		
	Retroalimentación		
¿Cómo se planifica y evalúa el Aprendizaje?	Requerimientos de logro y criterios de pertinencia		Competencias y habilidades desarrolladas
	Evaluación		

2.4. Factores de las interacciones didácticas

En la construcción de los MIDs (modelos de interacciones didácticas) se han usado palabras del lenguaje ordinario relacionadas al área de la educación. Estas palabras han sido redefinidas para: 1- Evitar algunos sentidos ordinario (cotidiano) de uso; 2-Reconceptualizarlas bajo una perspectiva naturalista y evitando, a la vez, perspectivas cognoscitivistas o constructivistas; y 3) Emplear definiciones útiles para la práctica científica y tecnológica. El método para lograr este objetivo ha sido el análisis histórico-conceptual (Ribes, 2018). Para una lectura detallada de esta labor académica recomendamos los libros “Psicología

general”(Ribes, 1990), “Psicología del aprendizaje”(Ribes, 2002), “Psicología y Educación. Aportes desde la teoría de la conducta”(Carpio & Irigoyen, 2005), y “Metodología para la planeación de la educación superior. Una aproximación desde la psicología Interconductual”(Ibáñez, 2007). En la presente síntesis de definiciones, si bien parte de los trabajos iniciales, se afinaran algunas definiciones basados en las reflexiones de los doctores De la Sancha y colaboradores (De la Sancha, 2016; De la Sancha & Pérez-Almonacid, 2017) y el doctor Morales y colaboradores (Morales et al., 2016, 2017, 2021).

2.4.1. Competencia

Se ha reconocido al concepto de “competencia” como fundamental para el análisis de las interacciones educativas, en conjunto con los conceptos de “habilidad”, “criterio de logro” y “aptitud funcional”. Sin embargo, De la Sancha & Pérez-Almonacid (2017) han identificado inconsistencias en el tratamiento de estos conceptos, atribuyéndole características de categorías disposicionales y/o categorías episódicas indistintamente en el desarrollo de las investigaciones conceptuales y empíricas, a pesar de que estas categorías obedecen a lógicas distintas. Sobre la “competencia” y la “habilidad”, estas deben ser consideradas únicamente como categorías disposicionales, ya que hacen referencia a tendencias del comportamiento en situaciones, es decir, sobre la probabilidad de la ocurrencia de comportamientos específicos. Sus definiciones se citan textualmente: “ser competente se predica de la tendencia a satisfacer un requerimiento de logro en formas exitosas pero también variadas porque además tal requerimiento exige tal variabilidad y no puede ser satisfecho desplegando sólo un procedimiento estándar” (p.50).

Aunque sus ocurrencias sean episodios, su identificación exige un análisis relacional o funcional de las variables del episodio, para comprobar la “variabilidad” y el “logro exitoso”. Esto implica que no sería correcto adjetivarlo como categoría episódica, y asignarle una correspondencia con alguno de los contactos funcionales.

2.4.2. Habilidad

Igual tratamiento recibe la “habilidad”, ya que las definiciones previas connotaban una correspondencia rígida entre morfologías de respuesta y estímulo. Se define a la habilidad como la tendencia a satisfacer un requerimiento de logro, considerando lo siguiente:

a) el requerimiento exige suficientemente un procedimiento estándar y relativamente rígido; b) lo que se hace corresponde funcionalmente con ese requerimiento de una forma invariante, rutinaria o estereotipada; y c) tal invarianza se asocia comúnmente a un rango de características morfológicas tanto de los objetos como de las acciones correspondientes (p.49).

2.4.3. Requerimiento de logro

Respecto a los criterios de logro “son requerimientos planteados al interior de los diferentes dominios de vida, culturales o institucionales y de los cuales siempre es posible sancionar éxito o fracaso” (p. 48). Renombrado como “requerimiento de logro” para una mayor distinción de los criterios de ajuste, el requerimiento que un desempeño logra o no logra siempre se da en episodios (es situacional), en los cuales el requerimiento de logro puede plantearse de manera explícita o tácita. A su vez se pueden plantear el concepto de “jerarquía de requerimientos” para hacer referencia a un conjunto de requerimientos de logro organizados y que se pueden satisfacer progresivamente de forma variada y efectiva.

Sobre el “criterio de ajuste”, tiene la función de identificar cuando un tipo de contacto funcional o aptitud se ha actualizado en un episodio. Las aptitudes contextual, suplementaria, selectora, sustitutiva referencial y sustitutiva no referencial especifican la complejidad del episodio (Ribes, 1990). Sin bien el requerimiento de logro y los criterios de ajuste se implican mutuamente, ambos guardan lógicas diferentes, ya que la primera se refiere en episodios que forman partes de dominios, y la segunda se refiere de la actualización del tipo de contacto funcional.

2.4.4. Criterio de pertinencia

De la Sancha & Pérez-Almonacid (2017) reconocen los “criterios de pertinencia” como un hacer pertinente, donde no se predica éxito o fracaso del desempeño, sino si se dio con sentido o es absurdo, en congruencia con un determinado dominio o ámbito funcional. Los criterios de pertinencia se deben considerar en el estudio y promoción de las interacciones didácticas, ya que permite explorar los predicados de comprensión en un episodio conductual.

2.4.5. Ámbito funcional de desempeño

Se define como el conjunto de prácticas (decires y haceres) que se producen y articulan en un determinado contexto específico (Acuña et al., 2013). Se debe explicitar que por “prácticas” se hace referencias a los comportamientos humanos, producidos bajo un medio de contacto convencional, y desplegados en situaciones. El conjunto de prácticas puede ser ejercitada, aprendida, transmitida y generada por personas en circunstancias estructuradas por criterios y valoraciones.

El desarrollo de las formas de conducta de las personas se origina en los ámbitos funcionales de las actividades humanas, y se transmiten culturalmente al interior de un grupo. Esta posible encontrar similitudes en esta descripción con el ejercicio de la enseñanza, donde los expertos imponen situaciones desafiantes (requerimientos de logro) a los novatos, con el objetivo de desarrollar destrezas, que serán los medios para que el novato alcance las habilidades y/o competencias que son delimitadas por el ámbito funcional de desempeño (León et al., 2011).

Para la instrucción y enseñanza de la ciencia y la tecnología, se han delimitado ámbitos funcionales (Carpio et al., 1998; Ribes, 1993) para estos respectivos modos de conocimiento, expuestos en la sección 2.1.2 de la presente tesis. En lo que respecta a la enseñanza a nivel profesional, se han delimitado ámbitos, llamados comúnmente “ámbitos didácticos” o “ámbitos

del desempeño didáctico” (Silva et al., 2014) y se expondrán en la sección 2.5 de la presente tesis.

Es posible delimitar más ámbitos funcionales a partir de los modos de conocimiento (López-Valadez, 2015). Un ejemplo de la construcción de situaciones didácticas, a partir de los ámbitos disciplinares y ámbitos de vida, la podemos encontrar en la propuesta de Ribes (2008).

2.4.6. *Desempeño docente*

Hace referencia a los comportamientos desplegados por quien tiene el rol institucional de enseñar. De acuerdo a las habilidades aprendidas por los estudiantes, se puede evaluar el “desempeño docente” estableciendo estándares de calidad para identificar desempeños excelentes, o también llamado “Competencia didáctica”. Desde un abordaje interconductual (Carpio & Irigoyen, 2005), se han establecido alguno de estos estándares de calidad.

Como requisito indispensable, el docente debe dominar las formas de actuación disciplinaria, agrupados como los criterios aceptados por la comunidad científica, tecnológica o humanística de la que es miembro. Esto quiere decir que: 1) El docente debe poder identificar, elaborar y explicitar los requerimientos de su disciplina, englobando sus respectivas prácticas, teorías, aparatos o situaciones relacionadas; 2) Cumplir con los requerimientos disciplinarios, según su dominio, sea científico (Ribes, 1993) o tecnológica (Carpio, et al. 1998). En palabras más simples, debe ser un experto en su área. Con este respaldo, el docente debe estar en la capacidad de desempeñarse exitosamente en los seis ámbitos didácticos.

En cuanto al rol del docente de educación superior, Silva et al. (2014) plantean las siguientes funciones: a) promueve el ajuste en el desempeño del alumno con los criterios que plantea su propio paradigma educativo; b) Media la interacción del alumno con sus referentes de dominio/disciplina; y c) probabiliza el desarrollo de habilidades estudiantiles.

2.4.6.1. *Problema del discurso didáctico.* En la introducción de la sección 2.2 se describió la introducción del concepto de “discurso didáctico” en el análisis de las

interacciones educativas y su incorporación en los modelos de interacciones didácticas. Sin embargo, en cada modelo este concepto ha sido desarrollado de diferentes formas, como lo analizaron Irigoyen et al. (2016) al revisar los usos de la noción de “discurso didáctico”, siendo definido en los múltiples modelos como: 1) acto lingüístico que media prácticas convencionales; 2) contactos de extensión y transformación; y 3) modalidades y modos lingüísticos. Los autores concluyen su artículo con una tecnificación de la noción, como arreglo instruccional que media efectivamente el ajuste del estudiante a los criterios disciplinares, en mínimamente dos episodios consecuentes. Desde nuestra posición, se identifica un uso indistinto del término como noción y como concepto técnico en los modelos de interacciones didácticas, lo cual representa un obstáculo para investigar y divulgar, más que un aporte al análisis. Se considera que es mejor mantener su definición original ryleana y usarlo como noción para resaltar las características funcionales de las interacciones educativas (Pérez-Almonacid & Gómez, 2014), por lo que no se incluirá en el análisis de esta investigación.

2.4.7. Desempeño estudiantil

Hace referencia a los comportamientos desplegados por quien participa en el programa de estudios y se espera que alcance los objetivos de aprendizaje. Cuando un estudiante es “efectivo” en cumplir los objetivos de aprendizaje de varios cursos sucesivos y en simultáneo, es posible delimitar “competencias de estudio”, como una tendencia a cumplir de manera efectiva y variada los diversos requerimientos de logro que exige el programa de estudios (Morales et al., 2016).

Para alcanzar estas competencias de estudio, lo básico es que el estudiante haya desarrollado destrezas efectivas en los modos lingüísticos observar, escuchar, leer, gesticular, hablar y escribir. La clasificación de los modos lingüísticos ha aportado en el estudio de la transferencia competencial y la habilitación lingüística (Tamayo, 2014), y es una fuente de evidencia experimental que puede alimentar al modelo interdisciplinar de interacciones

didácticas. Desde la UNAM, el Dr. Morales y colaboradores (2013) han planteado una clasificación de las modalidades que puede cobrar la interacción didáctica del docente con el estudiante, que pueden ser de tipo oral, lectora y escrita. Luego, especifican una serie de condiciones, conocimientos y recomendaciones de cada una de estas modalidades, donde se puede destacar que el estudiante, para un aprender efectivo, debe desenvolverse en todas las modalidades didácticas, a través de variadas situaciones y crecientes en complejidad. En contraposición, el uso repetitivo de una secuencia donde el docente hable y el estudiante escuche o escriba no contribuirá a un estudio efectivo (Morales et al., 2013).

Finalmente, las competencias de estudio, en sus diferentes modalidades, se encuentran circunscritas a los ámbitos disciplinares y los ámbitos didácticos, es por ello que estudiar leyendo un texto de matemáticas es funcionalmente diferente a estudiar leyendo un texto de literatura, por lo que la promoción de las competencias de estudio tienen que particularizarse a los ámbitos respectivos. Desde esta perspectiva, carecería de sentido las estrategias de estudio genéricas, abstractas o que pretendan aplicarse a todo (Morales et al., 2016). En cuanto a los ámbitos didácticos, en la presente investigación se ha caracterizado el desempeño estudiantil en los ítems correspondientes a los factores “desde la perspectiva del estudiante” de cada uno de los ámbitos didácticos.

El abordaje de las relaciones contingenciales entre el comportamiento del docente con la tarea disciplinaria y con el comportamiento del estudiante (2 relaciones) puede resumirse en el concepto de “enseñar”. Estas relaciones de enseñanza están subordinadas al requerimiento de logro que el docente impone (Morales et al., 2016).

2.4.8. Perspectiva desde el desempeño didáctico o docente

A partir de la psicología interconductual, se han propuesto clasificaciones del desempeño docente. En la Tabla 3 presentamos tres clasificaciones de ámbitos funcionales del desempeño docente. La primera clasificación (con cinco ámbitos) fue actualizada por el mismo grupo de

investigación, reconsiderando ahora 7 ámbitos. Estos mismos fueron replicados de manera similar por Acuña et al., (2013) e Ibañez & Ribes (2001) formulan descripciones similares al de ámbito didáctico, pero en forma de competencias.

Tabla 3

Clasificaciones de categorías de desempeño didáctico del docente

Ibañez y Ribes (2001)	Acuña, Irigoyen y Jiménez (2013)	Silva, Morales, Pacheco, Camacho, Garduño y Carpio (2014)
-	Planeación didáctica	Planeación didáctica
-	Exploración competencial	Exploración competencial
Expresar claramente ante sus alumnos como se logra el criterio	Prescripción criterios	Explicitación de criterios
Ejemplificar ante los alumnos el desempeño que logra el criterio	Ilustración	Ilustración
Disponer de situaciones necesarias para que el alumno practique “El cómo” se logra el criterio	Supervisión	Práctica supervisada
Informar puntualmente al alumno sobre sus aciertos o errores a lo largo del proceso de aplicación del conocimiento	Retroalimentación	Retroalimentación
-	Evaluación	Evaluación

Fuente: Elaboración propia

2.4.9. *Perspectiva desde el desempeño estudiantil*

Para delimitar las categorías del desempeño estudiantil en correspondencia con el desempeño docente y el objeto referente, se presentarán tres análisis interconductuales.

Un primer análisis lo realizó Ibañez (1999) sobre la conducta de estudio, al definirla como “el contacto interconductual del estudiante con el discurso didáctico” (p.49), este contacto es un factor relevante para el aprendizaje de competencias, entendido como logro de criterios. Para que el estudio sea efectivo, este depende de dos factores principales: la historia experiencial del estudiante y la naturaleza del discurso didáctico. Sobre estos factores se han propuesto sub-factores y se asumen los supuestos básicos de que la conducta de estudio necesariamente es “propositiva” o intencional y los individuos difieren sobre su capacidad de estudio, identificado operacionalmente como la identificación de criterios y ajuste a las contingencias. Los resultados exploratorios de la investigación parecen evidenciar estos supuestos.

Más tarde, Ibañez (2007) extiende su propuesta conceptual y metodológica. Redefine al estudiante como “aprendiz” para hacer énfasis en el logro efectivo de comportarse de acuerdo con los criterios morfológicos y funcionales de la comunidad epistémica a la que se va a insertar. Para lograr cumplir con los criterios, el estudiante “aprendiz” necesariamente debe actuar (en los modos lingüísticos activos) en situaciones reales o funcionalmente equivalentes a los problemas que debe resolver. De esta manera, también se trata de evitar caer en una perspectiva centrada en el docente. Respecto a los episodios instruccionales, Ibañez analiza las tres interacciones didácticas principales, según su modelo: alumno-discurso didáctico - objeto referente (ver figura 1). Solo presentaremos dos para los propósitos de nuestro análisis.

Sobre la interacción alumno-discurso didáctico: Primero se destaca la relación de “enseñanza” desde el discurso didáctico al alumno, pero también se debe tomar en cuenta la relación de “estudio” desde el alumno al discurso didáctico, definido como proceso correlativo

donde el aprendiz hace contacto con los productos lingüísticos del maestro. Morfológicamente, se le identifica como una reactividad lingüística que depende de la modalidad del discurso didáctico. No se delimitan ámbitos didácticos o situaciones convencionalmente específicas para esta interacción, más que la identificación de las modalidades lingüísticas.

Sobre la interacción alumno – objeto referente: El contacto del alumno con el objeto referente del discurso didáctico requiere que su desempeño cumpla necesariamente los criterios morfológicos y funcionales ante objetos y situaciones reales o sucedáneas. En estos episodios instruccionales, el alumno ya debió cumplir las competencias básicas del plan de estudios, y está desarrollando las competencias específicas para su dominio. Esta interacción didáctica la podemos corresponder con el ámbito didáctico “Práctica pertinente”.

Luego, Ibáñez presenta algunas extensiones tecnológicas para la planeación de la conducta de estudio, donde se destaca que, “se parte de la suposición de que los estudiantes pueden desarrollar la capacidad de auto administrarse criterios disciplinares y condiciones idóneas para generar su propio aprendizaje, como competencias de estudio efectivo” (p.132). Las recomendaciones metodológicas van desde un pequeño análisis conceptual sobre “estudiar”, definición y procedimientos para los objetivos instruccionales, ubicar el nivel de competencia profesional, hasta el cómo organizar las competencias didácticas para desarrollar las competencias. Finalmente, Ibáñez ha planteado la planeación de interacciones didácticas autoadministradas por el propio estudiante o conducta autodidacta del estudiante.

2.4.10. Niveles de análisis de las interacciones educativas

Todos los modelos de interacciones didácticas consideran dos niveles de análisis para las interacciones que suceden en el ámbito de la educación, el primero es el puramente conceptual y modela como se establecen las interacciones didácticas, bajo una lógica relacional o funcional de los factores del modelo. El segundo es el metodológico, el cual identifica factores y sus relaciones, mediante un lenguaje operacional para su estudio empírico según

enfoque (ideográfico o nomotético), su estrategia de investigación (manipulativa, asociativa y descriptiva) y su tipo (transversal y longitudinal). El lenguaje operacional a veces considera factores como entidades (como el docente o el objeto referente físico) o acontecimientos, situaciones o funciones (como el desempeño docente, la modalidad del objeto referente o el requerimiento de logro).

En cuanto al primer nivel de análisis, considerando la noción de campo psicológico como un sistema de elementos con interdependencia funcional, hemos seleccionado al segundo modelo de interacciones didácticas, del grupo de investigación de la FES Iztacala – UNAM, como la base del presente análisis. Esto no quiere decir que los otros modelos no sean pertinentes, porque se valoran y reconsideraran los estudios empíricos que han elaborados, integrándolos en el análisis.

La noción “episodio didáctico” propuesta por Morales et al. (2016) permite concretar el abordaje empírico de la interacción didáctica como noción conceptual, delimitando los segmentos en los cuales el estudiante desarrolla habilidades y competencias para los requerimientos de logro en diferentes ámbitos didácticos, como la exploración competencial, la explicación de criterios, la ilustración, etc.

2.5. Ámbitos didácticos del desempeños docente y estudiantil

Se presentan las definiciones de los ámbitos o dominios didácticos desde el punto de vista del docente o el estudiante. Es menester considerar en simultáneo las perspectivas de los actores educativos en el análisis de la interacción didáctica, ya que un análisis dividido o solitario de las actuaciones de cada uno no es suficiente (Morales et al., 2017).

En una versión temprana del desarrollo de la propuesta conceptual y metodológica de la presente tesis, se consideró nombrar los ámbitos didácticos del desempeño docente y desempeño estudiantil por separado, lo cual se expone en la tabla 4. En la segunda versión de la adaptación de los cuestionarios también se incluyó los nombres nuevos de los ámbitos

didácticos, esta vez presentados como dimensiones o subfactores de las variables “desempeño docente” y “desempeño estudiantil”.

Tabla 4

Propuesta de correspondencia entre ámbitos didácticos del desempeño docente y desempeño estudiantil

Ámbitos didácticos del desempeño del docente	Ámbitos del desempeño estudiantil que se ajustan al desempeño docente
Exploración competencial	Precurrentes y potenciales habilidades
Explicitación de Criterios	Identificación de criterios
Ilustración	Participación
Supervisión de la práctica	Práctica pertinente
Retroalimentación	Mejoramiento
Evaluación	Desempeño resultante

En una posterior reflexión, se entendió que es innecesario nombrar un ámbito según la conducta de quien analizar, porque un ámbito es la red de convenciones de las prácticas características según un dominio en particular, sea disciplinar, didáctico o cotidiano, por lo que ambos participantes (docente y estudiante) se ven involucrados en el mismo necesario de prácticas convencionales. Al enfatizar en el comportamiento de alguno de ellos, sea docente o estudiantes, solo se está desplazando el eje de análisis a su conducta, considerando la conducta del otro como parte de las circunstancias situacionales. Para las próximas investigaciones, no se tendrán en cuenta los nombres propuestos y se asumirán los nombres originales de los ámbitos didácticos.

A continuación, se presentan las definiciones de los ámbitos didácticos, según la perspectiva del docente y el estudiante. Para explicitar el centro de análisis como la conducta

del estudiante, se agregará el subtítulo “Para el estudiante”, seguido del nombre colocado en la primera versión del cuestionario.

2.5.1. Planeación didáctica

Son todas las acciones dirigidas al diseño de las situaciones, materiales, circunstancias, tareas y secuencias a implementar en las interacciones con los estudiantes con el objetivo de probabilizar el desarrollo de las habilidades y destrezas (Morales et al., 2021, p. 6). El docente debe ser competente para responder las preguntas ¿Qué debe aprender el estudiante?, ¿para qué debe aprender?, ¿en qué circunstancias o situaciones debe aprenderlo?, y ¿Cómo obtener evidencias de que ha aprendido?. A partir de la resolución de estas preguntas, se puede estructurar los planes de estudio, que agrupan los objetivos instruccionales (Silva et al., 2014).

Para planificar las interacciones didácticas se demanda satisfacer por lo menos tres condiciones fundamentales: 1) Quien enseña debe ser competente en el cumplimiento de los requerimientos de logro de lo que se enseñara, en los ámbitos disciplinar y didáctico; 2) La especificación del o los requerimientos de logro que los estudiantes deben satisfacer; y 3) Las situaciones-actividades mediante las cuales se enseña y se estudia de acuerdo con los requerimientos planificados en una unidad o programa instruccional, expresados generalmente en el currículo.

La planeación competencial ocurre en un tiempo previo al desarrollo de las clases, pero puede ser modificado en función del progreso de los estudiantes. No fue considerado en la adaptación de los cuestionarios para la presente investigación, pero las posibilidades de ideas de investigación en este ámbito son variadas y de importancia.

Finalmente, se especifica que el ámbito de la planeación didáctica, en el esquema de las interacciones didácticas, se ubica en la relación docente – tarea disciplinaria.

2.5.2. Exploración competencial

Es la identificación y evaluación de la conducta del estudiante en términos de las habilidades precurrentes necesarias para cursar la asignatura, con el objetivo de hacer ajustes al plan de enseñanza según sea necesario (Silva et al., 2014). Por “precurrente” nos referimos a una habilidad o destreza necesaria para aprender otra habilidad, destreza o competencia.

Para que el docente logre explorar los precurrentes de los estudiantes debe ...

diseñar condiciones de evaluación diagnóstica o inicial del desempeño estudiantil que servirá para establecer el punto de arranque de clase, curso o nivel, así como para implementar acciones correctivas que contribuyan al desarrollo de la clase en condiciones apropiadas para todos los estudiantes (Morales et al., 2021, p. 6).

2.5.2.1. Para el estudiante “Precurrentes y potenciales habilidades”. Desde el punto de vista del estudiante, ante las actividades evaluativas-diagnosticas del docente, el estudiante debe estar en la disposición de mostrar sus habilidades y destrezas precurrentes, respondiendo efectivamente a las actividades. En la definición original de Morales et al. (2017, p. 31) son “Precurrentes, que el estudiante identifica como condiciones necesarias para el aprendizaje de formas de actuación disciplinaria, determinando qué necesita saber y qué va a requerir para esa situación didáctica, curso, materia o semestre”. En la historia de la educación se ha descrito este desempeño como los “conocimientos previos”.

2.5.3. Explicitación de criterios

También denominado “Prescripción de criterios”. Desde el docente es...

...poner en contacto al aprendiz con los criterios disciplinarios y didácticos que debe satisfacer con su desempeño. Por criterios disciplinarios deben entenderse todos aquellos requerimientos conductuales que son exclusivos del ámbito de aquello que se está enseñando (e.g., lo que se debe decir y lo que no en la práctica). Por criterios didácticos deben entenderse aquellos requerimientos conductuales vinculados a la

organización de las actividades didácticas cotidianas (e.g., a qué hora llegar, cómo relacionarse con quien enseña, cómo relacionarse con otros compañeros) (Silva et al., 2014, p. 38).

2.5.3.1. Para el estudiante “Identificación de criterios”. Ante la prescripción de criterios o requerimientos de logro, sea de tipo didáctico o disciplinar, el estudiante debe identificarlos para saber qué y cómo decir o hacer, cuando y donde poner en práctica las habilidades o destrezas para satisfacer estos requerimientos. En un segundo momento, el estudiante debe estar en la capacidad de formular nuevos requerimientos de logro que le ayuden a lograr los requerimientos originalmente impuestos (Morales et al., 2021, p.10).

2.5.4. Ilustración

Para el docente:

... se describe al aprendiz las características de las habilidades de un experto en la disciplina para la solución de un problema específico. Es decir, quien enseña debe poner en contacto al que aprende con los elementos del comportamiento de otro, señalando en esta descripción: a) lo que hacía, b) la situación en la que lo hacía, y c) para qué lo hacía. Ilustrar consiste en relacionar lingüísticamente al que aprende con las habilidades de otro, incluso cuando ese otro es quien enseña (i.e. quien enseña se pone como ejemplo) ... (Silva et al., 2014, p. 38).

La ilustración o demostración del ejercicio ejemplar (de tipo disciplinario) requiere una alta confluencia con lo que el docente puede hacer y decir según los criterios de tipo didáctico. No cumplir con esto puede confundir o desinteresar a los estudiantes.

2.5.4.1. Para el estudiante “Participación”. Son los desempeños mostrados durante la clase, que pueden expresarse de diferentes formas, como por ejemplo “decir frente a otros lo que se opina”, “hacer frente a otros lo que se ha aprendido y lo que no”, etc. de acuerdo a los requerimientos de logro establecidos por el docente (Morales et al., 2017).

2.5.5. Supervisión de la práctica

Para el docente:

En este ámbito se debe involucrar a quien aprende en situaciones controladas de solución de problemas. El control de estas situaciones permite supervisar y corregir momento a momento el desempeño de quien aprende. Las habilidades y competencias a desarrollar se vinculan con la regulación de aquellos que aprenden en el transcurso de la situación didáctica, reduciendo las posibilidades de un desempeño inefectivo, y aprovechando al máximo la situación en beneficio de su desarrollo conductual. (Silva et al., 2014, pp 38-39).

2.5.5.1. Para el estudiante “Práctica pertinente”. Práctica desplegada por el estudiante que se encuentra enmarcada en el criterio de logro en correspondencia con el qué, cómo, cuándo y dónde dicha práctica puede ser supervisada momento a momento como forma de evaluación permanente o en aislado como situación de prueba para evidenciar una forma de comportarse ya aprendida. (Morales et al., 2017, p. 31).

2.5.6. Retroalimentación

El docente, en interacción sincrónica con el estudiante, logra que este evalúe y cambie su propio desempeño desplegado en la tarea disciplinaria, en función del requerimiento de logro y el desempeño ideal que se exige. Silva et al. (2014, p. 39) especifican que “Quien enseña describe al aprendiz lo que hizo, para qué lo hizo y en qué situación lo hizo. También pueden describirse las variantes del hacer (el desempeño) del aprendiz, que en ese momento eran posibles y que sin embargo no se concretaron. En este sentido, retroalimentar es algo más complejo que decir “correcto” o “incorrecto”, sino que implica habilitar al aprendiz para entrar en contacto con su propio comportamiento y sus variantes posibles...”.

2.5.6.1. Para el estudiante “Mejoramiento”. Además de que el estudiante logre establecer las correspondencias entre su desempeño desplegado y el desempeño ideal exigido,

debe lograr la independencia de supervisar y monitorear su propio desempeño, para identificar de manera más efectiva que tan cerca se encuentra de cumplir el requerimiento de logro o que ajustes debe implementar (Morales et al., 2017).

2.5.7. Evaluación

El docente debe diseñar situaciones evaluativas generales (no basta con un examen) donde pueda contrastar el desempeño del estudiante desplegado con el desempeño ideal propuesto en el plan o programa de estudio. Silva et al. (2014, p. 39) especifican que:

En este contraste es necesario contar con un sistema de valoración que permita determinar el grado de similitud entre el desempeño esperado y el real...la finalidad didáctica de este contraste no solo descansa en la posibilidad de extender una calificación de la conducta, sino sobre todo en crear la posibilidad de realizar ajustes estratégicos a interacciones didácticas posteriores...

En el caso de la evaluación general final que se suele hacer en las últimas fechas de la asignatura, las situaciones evaluativas planteadas deben ser lo suficientemente efectivas para determinar si los desempeños de los estudiantes se pueden igualar a las habilidades o competencias descritas en el plan o programa de estudio.

2.5.7.1. Para el estudiante “Desempeño resultante”. Además de que el estudiante logre las habilidades o competencias esperadas, el propio estudiante debe lograr un nivel de “autoevaluación” como comparación entre su desempeño desplegado y el desempeño ideal, con el objetivo de implementar mejoras para acercarse a lo exigido (Morales et al., 2017).

2.5.8. Autodirección del estudio

El comportamiento del estudiante que es parte de un programa educativo en general ha sido estudiado históricamente como “rendimiento académico”. Conforme fue progresando las investigaciones sobre el aprendizaje y el estudio, más variables fueron consideradas para

abarcar conceptualmente el comportamiento del estudiante. Actualmente se realizan estudios de las variables denominadas “desempeño académico”, “desempeño estudiantil”, “autoeficacia académica”, “percepción estudiantil”, “autodidactismo”, etc. Estas variables mantienen cierta similitud con la conceptualizada en el presente estudio, cuyo rotulo “autodirección del estudio” será explicada en los párrafos finales de esta sección.

Desde la investigación interconductual de los procesos educativos, Ibañez (1999) ejecuto el primer estudio empírico sobre la conducta de estudio. Con esa investigación y otros estudios posteriores, Ibañez (2007) concluyo que el estudiante puede monitorear su propio desempeño ante un caso-problema, solo cuando es capaz de establecer correspondencias entre las propiedades morfológicas de su desempeño y los criterios pertinentes provistos en el discurso didáctico, o autoimpuestos. En palabras del autor:

Un estudiante será capaz de monitorear su propio desempeño – y el de los demás – solo cuando ya sea competente ... cuando ya pertenezca a la comunidad epistémica para la cual se está preparando...Debe ser monitoreado en su proceso de observación por agentes expertos externos (Ibañez, 2007, p. 139).

Esta noción sobre el comportamiento del estudiante, en el marco de las interacciones didácticas, se ubicó en las relaciones de estudiar: los ámbitos didácticos para las interacciones docente-estudiante y la autoadministración o autodirección del estudio para la relación estudiante-tarea disciplinaria (también llamada estudiante-objeto referente). Sobre esta segunda relación, Morales et al. (2021) ampliaron su modelo de interacciones didácticas para describir y proponer ideas de investigación sobre lo que ellos denominaron “desempeño autodidacta”, como respuesta al problema educativo-social de la implementación urgente y masiva de la modalidad virtual en las universidades de todo el planeta, que trajo consigo nuevas exigencias a los estudiantes universitarios.

En el trabajo referenciado, inician enlistando las prácticas de enseñanza que limitan la “autonomía intelectual”; exponen el modelo y los ámbitos; delimitan la noción de “desempeño autodidacta”; actualizan los ámbitos del desempeño estudiantil; exploran la evolución del desempeño autodidacta y finalizan con algunas recomendaciones e ideas de procedimientos para su promoción. A continuación, se muestra los ámbitos del desempeño estudiantil en relación a la tarea disciplinar:

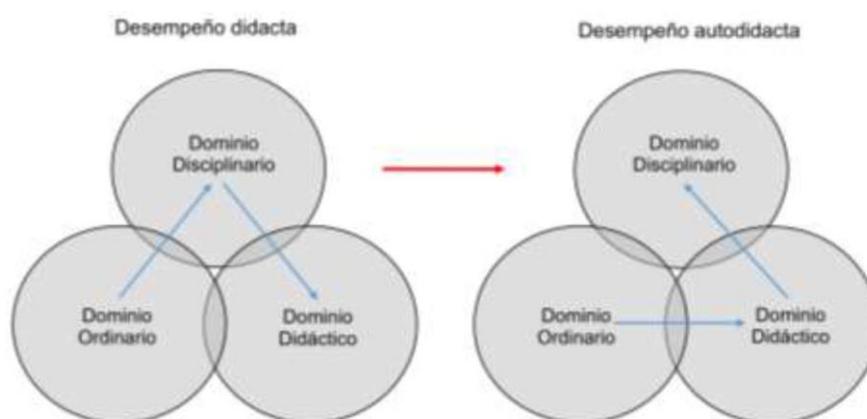
- a) Identificación de estados orgánicos y tendencias comportamentales que favorezcan o interfieran las actividades de estudio, con el objetivo de planificar situaciones pertinentes para el estudio. Por ejemplo: El caso de un estudiante que identifique su tendencia a demorar innecesariamente el cumplimiento de sus tareas, por lo que decide aplicar técnicas conductuales para cumplir a tiempo sus tareas.
- b) Estrategias y recursos elaborados por el estudiante para alcanzar los requerimientos de logro de la asignatura o que sean autoimpuestos. Las estrategias, de dominio didáctico, tienen la dirección de alcanzar las destrezas y habilidades del dominio disciplinar. Los recursos incluyen la administración, búsqueda y selección de materiales educativos. Por ejemplo: El caso de un estudiante de medicina que elabora detalladamente dibujos sobre partes del cuerpo humano, para alcanzar alguna destreza de identificación.
- c) Aplicación de “lo aprendido” en otras situaciones (variadas y/o novedosas). En esta transferencia se despliega la capacidad para imponer requerimientos de logro en nuevas situaciones, diferentes a las que aprobó en el curso académico. Por ejemplo: El caso de un estudiante que aprendió técnicas estadísticas con un determinado programa informático y con variables sanitarias, transfiriendo esas técnicas aprendidas en otro programa informático, con variables económicas y resolviendo nuevos problemas.

Estos ámbitos, explica Morales et al. (2021), se pueden incluir en una secuencia hipotética entre el desempeño docente didáctico hacia el desempeño autodidacta del estudiante,

que se inicia con el tránsito del docente entre el dominio cotidiano, luego el dominio disciplinar, para llegar al dominio didáctico. Para el caso del estudiante, el tránsito inicia en el dominio ordinario en dirección al dominio didáctico, luego avanzando al dominio de una disciplina. Estas secuencias se pueden graficar de la siguiente manera:

Figura 8

Ruta hipotética del desempeño autodidacta a partir del desempeño didáctico



Nota. Figura original de Morales, G., Tapia, F., & Escobedo, A. (2021). La emergencia del desempeño autodidacta como aspiración en educación superior. *Revista de Investigación Académica Sin Frontera: División de Ciencias Económicas y Sociales*, 14(36), 1-17.

Esta secuencia/tránsito de ámbitos abre posibilidades de investigación que puedan brindar luces sobre el estudio del comportamiento o desempeño estudiantil. Al respecto, Morales et al. (2021) proponen que el perfil ideal de un alto desempeño estudiantil implica necesariamente la “emergencia del desempeño autodidacta” (p. 7) que, más adelante explican, es posible caracterizarse como una noción (autodidactismo) donde el estudiante está en la capacidad de disponer condiciones que promueven el logro de competencias y habilidades disciplinarias (lo que estudia). Al respecto, consideramos adecuado clasificar al autodidactismo como una categoría de “noción” porque brinda un alcance conceptual de lo que se quiere estudiar, sin que sus usos conceptuales y/o procedimentales sean prematuramente rigurosos. Sin embargo, algunos teóricos conductistas han advertido de los riesgos del uso del prefijo

“auto” en los constructos psicológicos (por ejemplo: autoeficacia, autocontrol, etc.), para la explicación del comportamiento controlado por la propia persona. Por ejemplo, Catania (1975) analizó el término “autoreforzamiento”, encontrando: 1) contradicciones conceptuales internas, 2) lo que pretendía describir podía ser descrito bajo reglas de reforzamiento más simples y 3) que encubrió, de alguna manera, fenómenos que si eran interesantes, como lo era la autodiscriminación condicional, desde la perspectiva del autor (p. 198). A estas desventajas del uso de constructos con el prefijo auto, se debe agregar que son susceptibles a la tautología y al desplazamiento de la explicación a un hipotético “funcionamiento interno” del individuo. Antes estos riesgos, la noción que se propone hace referencia a las decisiones que toma un individuo respecto a las actividades “de estudio” que elige para lograr los objetivos de aprendizaje, tratando de describir a un individuo que, según la dirección que tomen en estas actividades, puede llegar a cumplir o no los objetivos de aprendizaje, por lo que la finalidad “didáctica” no estaría garantizada (Del Valle et al., 2015). Para que un estudiante sea capaz de elegir las actividades y recursos adecuados para un estudio efectivo, debe haber seguido un tránsito desde los dominios ordinarios, didáctico y disciplinario, por lo que las rutas o direcciones que tomen serán importantes. Para resumir esta descripción de la noción, se considera que “autodirección del estudio” puede ser una alternativa, ya que hace referencia a los comportamientos de la persona ante el estudio.

III. Método

3.1. Tipo de investigación

El enfoque es cuantitativo. La investigación reúne dos clases de investigación, de acuerdo con Ato et al. (2013):

a) instrumental, donde se comprueban las propiedades psicométricas de dos cuestionarios (objetivo específico 1); y

b) empírica, donde se estudian los fenómenos a partir de estrategias de investigación (objetivo principal y objetivos específicos 2, 3 y 4). La investigación actual siguió una estrategia asociativa, estudiando dos variables y sus respectivas dimensiones mediante la relación entre las variables (diseño correlacional simple) y la comparación según grupos (diseño de cohortes transversal) (Ato et al., 2013).

3.2. Ámbito temporal y espacial

La investigación tomó lugar en una facultad de psicología de una universidad pública ubicada en el distrito de Cercado de Lima, en la ciudad de Lima, Perú. El tiempo de recojo de datos para la investigación fue entre los meses de setiembre a diciembre del año 2019.

3.3. Variables

3.3.1. Variables de estudio

- Desempeño docente
- Desempeño estudiantil

Tabla 5

Operacionalización de la variable “Desempeño docente”

Variable	Dimensiones	Ítems	Puntuación por dimensión	Escala de respuesta	Instrumento y puntuación general
Desempeño docente percibido por el estudiante	Exploración competencial	1 al 5	0 a 15	Nunca = 0	Cuestionario sobre el desempeño docente
	Explicitación de criterios	6 al 11	0 al 18	Casi nunca = 1	
	Ilustración	12 al 17	0 al 18		
	Supervisión de la práctica	18 al 22	0 a 15	Casi siempre = 2	
	Retroalimentación	23 al 28	0 al 18		
	Evaluación	29 al 33	0 a 15	Siempre = 3	

Tabla 6

Operacionalización de la variable “Desempeño estudiantil”

Variable	Dimensiones	Ítems	Puntuación por dimensión	Escala de respuesta	Instrumento y puntuación general
Desempeño estudiantil percibido por el propio estudiante	Precurrentes y potenciales habilidades	1 al 5	0 a 15	Nunca = 0	Cuestionario sobre el desempeño estudiantil*
	Identificación de criterios	6 al 10	0 al 15	Casi nunca = 1	
	Participación	11 al 16	0 al 18		

Práctica pertinente	17 al 22	0 a 18	Casi siempre = 2
Mejoramiento	23 al 27	0 al 15	
Desempeño resultante	28 al 33	0 a 18	Siempre = 3

Nota. *Basado en la versión final del cuestionario (Anexo A)

3.3.2. Variables de control

A continuación, se explica cómo se conformaron los grupos para cada variable de agrupación:

- a. Sexo: Masculino y femenino.
- b. Edad: Se dividió en dos grupos, el primer grupo lo conformó los estudiantes cuya edad es igual y menor a 22 años, y el segundo grupo abarcó las edades igual y mayor a 23 años.
- c. Ciclo de estudios: Corresponde al ciclo que finalizó y del curso que evaluó en los cuestionarios.
- d. Turno de estudios: matutino y vespertino
- e. Especialidad de carrera: Los ciclos primero, tercero y quinto se clasificaron como “General” y los ciclos séptimo y noveno como “Especialidad”.
- f. Eje curricular: Se clasificó los cursos evaluados en tres ejes curriculares: “Investigación”, “básica” y “especialidad. El eje curricular de cada curso fue asignado en la tabla 7.

3.4. Población y muestra

3.4.1. Población

Alumnos matriculados en el año 2019, desde el primero hasta el décimo ciclo de la carrera de pregrado en psicología, impartida por una universidad pública de Lima, Perú.

3.4.2. Muestreo

El tipo de muestreo fue intencional – no probabilístico. Se solicitó la colaboración de los estudiantes, logrando abarcar un 40% de la población total.

3.4.3. Muestra

La muestra fue de 475 estudiantes. Los participantes provienen de 22 grupos o salones que cursaron el 1er, 3ro, 5to, 7mo y 9no semestre de estudio durante el periodo académico abril – julio, en el año 2019.

Tabla 7

Distribución de la muestra por salón y tipo de asignatura

Ciclo de estudios	Tipo de asignatura	Número de grupos o salones
1er ciclo	Investigación científica	2
	Fundamentos básicos de psicología	2
3er ciclo	Investigación científica	2
	Fundamentos básicos de psicología	2
5to ciclo	Investigación científica	2
	Fundamentos básicos de psicología	2
7mo ciclo	Especialización profesional	5
9no ciclo	Especialización profesional	5

Criterios de inclusión:

- a. Ser estudiante de la carrera de psicología de la universidad pública.

- b. Haber cursado los ciclos 1ro, 3ro, 5to, 7mo y 9no, del periodo académico abril – julio 2019.
- c. Haber cursado las asignaturas durante el periodo académico abril – julio 2019.

Criterios de exclusión:

- a. Estudiantes que cuenten con más del 25% de inasistencia en alguna de las asignaturas.
- b. Estudiantes que invaliden el instrumento.
- c. Estudiantes que no asistan a clases el día de aplicación del instrumento.
- d. Estudiantes que no acepten participar de forma voluntaria.

3.5. Instrumentos

Se emplearon tres instrumentos:

1. Ficha de datos sociodemográficos elaborado para el estudio (Anexo B)
2. Batería de Cuestionarios sobre desempeño docente y desempeño estudiantil, compuesto de:
 - a. Cuestionario de Desempeño Docente (CDD)
 - b. Cuestionario de Desempeño estudiantil (CDE)

Los cuestionarios fueron creados originalmente por Velarde (2020) para población mexicana, pero no fue posible acceder al reporte de sus propiedades psicométricas. Los cuestionarios originales cuenta con una versión abreviada elaborada por Bazán-Ramírez y Velarde (2021). La presente adaptación a población peruana se realizó a partir de la versión original (Anexo F) pero se describirán las propiedades psicométricas de la versión abreviada:

Bazán-Ramírez y Velarde (2021) adaptaron una versión del cuestionario de desempeño docente y desempeño estudiantil (abreviada) con las siguientes características: 1) La escala de respuesta fue de cuatro alternativas de frecuencia de conducta “nunca”, ”casi nunca”, ”casi siempre” y “siempre”; 2) Son dos cuestionarios, uno dirigido a evaluar la percepción del alumno sobre el desempeño de su profesor, y el segundo cuestionario el alumno autoevalúa su

comportamiento en función de la enseñanza; 3) ambos cuestionarios cuentan con 18 ítems; 4) los cuestionarios tienen cinco dimensiones compartidas.

La muestra de adaptación, seleccionada aleatoriamente, fueron 124 alumnos de la carrera de psicología de una universidad estatal de Morelos, México. Al aplicar el instrumento, el estudiante tuvo la potestad de elegir el curso a evaluar, con la condición de que sea un curso que estaba concluyendo en su ciclo de estudios actual.

Sobre el cuestionario de desempeño docente, superó las pruebas previas de Bartlett ($p < .05$) y $KMO = 0.89$. El AFE se hizo a partir de una matriz de correlaciones policóricas, pero los autores no especificaron el método de extracción y rotación, así como la técnica para identificar el número de factores. Los ítems obtuvieron pesos factoriales mayores a 0.3, pero solo el ítem 1 no superó este corte. Si bien el AFE identificó 5 factores, 1 factor agrupó la mayoría de ítems de tres factores predefinidos. Considerando este agrupamiento, el 77.7% de los ítems coincidió con su factor predefinido. En cuanto al AFC, se obtuvieron adecuados pesos factoriales (entre 0.58 a 0.90), adecuados índices de covarianza (entre 0.43 a 0.84) y aceptables índices de ajuste ($CFI = 0.95$; $RMSEA = 0.06$), confirmando al modelo teórico de 5 factores para desempeño docente.

Respecto al cuestionario de desempeño estudiantil, que también superó las pruebas previas (Test de Bartlett $p < .05$; $KMO = 0.713$), se obtuvo 5 factores por análisis factorial exploratorio, donde dos de estos factores conservaron sus ítems predefinidos. El resto de factores conservó al menos dos ítems predefinidos y todos los ítems superaron el peso factorial de 0.3. En el análisis factorial confirmatorio, del cual tampoco se especificó el método de estimación, se obtuvieron aceptables pesos factoriales (entre 0.26 a 0.94) aunque un poco menores al AFC de desempeño docente. Los índices de covarianza fueron más bajos (entre 0.06 a 0.74) y los índices de ajuste alcanzaron el límite de tolerancia de corte ($CFI = 0.90$; $RMSEA = 0.07$).

Sobre la fiabilidad de los cuestionarios, se aplicó el coeficiente alfa al conjunto completo de ítems, obteniendo un $\alpha = 0.93$ para desempeño docente y $\alpha = 0.78$ para desempeño estudiantil.

Los resultados psicométricos obtenidos por Bazán-Ramírez y Velarde (2021) respaldan el modelo de 5 dimensiones (ámbitos didácticos) para las variables desempeño docente y desempeño estudiantil, respaldando su uso y adaptación a nuevas muestras.

Para la adaptación a la muestra de estudiantes peruanos de la carrera de psicología, se modificó el contenido de los cuestionarios originales (Anexo F), con las siguientes características:

3.5.1. Cuestionario de desempeño docente - CDD:

- Muestra dirigida: Estudiantes de pregrado de universidades peruanas, de la carrera de psicología.
- Escala de respuesta: cuatro alternativas de frecuencia de conducta “Nunca”, “Casi Nunca”, “Casi siempre” y “Siempre”.
- Ítems o reactivos: 33.
- Instrucción al encuestado: Evaluar el desempeño del docente de una asignatura de estudios determinada, que haya finalizado o este por finalizar. Al leer el ítem, debe elegir la frecuencia con la que el docente ejecuta tal conducta.
- Dimensiones: 6 dimensiones, conformadas por entre 5 a 6 ítems.
- Método de calificación: Suma directa de 33 ítems, con un rango de puntaje de 0 a 99.

3.5.2. Cuestionario de desempeño estudiantil - CDE:

- Muestra dirigida: Estudiantes de pregrado de universidades peruanas, de la carrera de psicología.
- Escala de respuesta: cuatro alternativas de frecuencia de conducta “Nunca”, “Casi Nunca”, “Casi siempre” y “Siempre”.

- Ítems o reactivos: 34 conformaron la escala inicial, luego se retiró 1 ítem, quedando 33.
- Instrucción al encuestado: Evaluar su propio desempeño ante la enseñanza del docente, en una asignatura de estudios determinada, que haya finalizado o este por finalizar. Al leer el ítem, debe elegir la frecuencia con la que realiza tal conducta.
- Dimensiones: 6 dimensiones, conformadas por entre 5 a 6 ítems.
- Método de calificación: Suma directa de 33 ítems, con un rango de puntaje de 0 a 99. En la primera versión de la adaptación de 34 ítems existía un ítem inverso (ítem 24), pero fue retirado.

3.6. Procedimientos

La investigación se desarrolló en tres etapas: Evaluación de la muestra, recolección de los datos obtenidos y análisis de datos. La descripción de la tercera etapa se realizará en la sección 3.7.

3.6.1. Evaluación de la muestra

Se obtuvo la autorización escrita de las autoridades de la institución educativa, aplicando los instrumentos durante los horarios coordinados con los docentes a cargo y estudiantes informantes. Se tomó en cuenta la malla curricular vigente, la programación de clases y salones de los grupos, seleccionando los grupos de acuerdo con los criterios de inclusión y exclusión. Se solicitó a cada docente de los grupos elegidos, su participación voluntaria en el estudio, en caso aceptará se le pidió firmar un consentimiento informado de autorización, hasta completar la cuota por el tipo y nivel de las asignaturas.

Con el permiso del profesor de grupo, los aplicadores asistieron a los salones de clase para informar al grupo de estudiantes e invitarlos a participar de manera voluntaria. Se les entregó en forma impresa los cuestionarios. La instrucción fue:

Estimados participantes, lo que se le entregó es un cuestionario sobre la opinión o percepción que tienen ustedes sobre el desempeño de un docente y como ustedes

respondieron ante ello. Se les pide que recuerden la clase de la asignatura (se menciona la asignatura seleccionada al grupo) de principio a fin. Una vez que lo recuerden, por favor contesten la frecuencia con la cual su docente y ustedes realizaban los enunciados. Si tienen alguna consulta, pueden llamarme para resolverla.

La aplicación del cuestionario se realizó entre el 25 de setiembre al 7 de diciembre, durante el semestre académico agosto – diciembre 2019.

3.6.2. Recolección de datos

Durante la segunda etapa se digitó las respuestas de las pruebas en una base de datos creada en una hoja de cálculo del programa Microsoft Excel, del paquete Office 365. Se construyó la base según todas las variables del estudio y se depuró las respuestas nulas para que no existan problemas a la hora de ejecutar el procesamiento de datos.

3.7. Análisis de datos

Debido a que la cantidad de análisis estadísticos empleados requieren de una explicitación de su uso y delimitación de los criterios de interpretación, se elaboró una “Guía para los análisis de datos”, ubicado en el Anexo C, que servirá como orientación para la lectura de los resultados y contraste con la literatura científica en la presente tesis. A continuación, se presenta un resumen de los análisis empleados.

Se realizó un análisis preliminar para evaluar y/o comprobar el funcionamiento psicométricos de los instrumentos en las unidades muestrales consignadas (Muñiz, 2018). En primera instancia, se analizó las fuentes de evidencias de validez del instrumento: estadísticos descriptivos de los ítems, dimensiones y variables; adaptación del contenido del test; evidencia basada en la estructura interna a través del análisis factorial confirmatorio general; y, en instancias posteriores, multigrupo mediante invarianza factorial. En consecuencia, las evidencias de fiabilidad se analizaron por cada dimensión con el coeficiente alfa ordinal (α) y la fiabilidad del factor general mediante el coeficiente omega jerárquico (ω_H).

En la segunda instancia se obtuvo la información de los estadígrafos y las pruebas de hipótesis de las variables y sus dimensiones. En tal sentido, la prueba de bondad de ajuste a la distribución normal basada en el coeficiente Shapiro Wilk y situó sus datos en torno a la ausencia de distribución normal (Mohd y Bee, 2011), en consecuencia, se empleó pruebas no paramétricas para el análisis inferencial del contraste según variables de control de dos categorías como la prueba U de Mann-Whitney y más de dos categorías como la prueba H de Kruskal Wallis, y en el análisis correlacional a través de la prueba de correlación por rangos de Spearman (ρ) con una base interpretativa de los siguientes puntos de corte propuestos por Akoglu (2018): espúrico ($\rho = 0$); débil (± 0.1 a ± 0.3); moderado (± 0.4 a ± 0.6), fuerte (± 0.7 a ± 0.9), y perfecto ($\rho = \pm 1.00$). Adicionalmente, en el Anexo H se presenta un análisis comparativo alternativo, empleando los tamaños del efecto d de Cohen y Omega al cuadrado (ω^2).

En el contraste de medidas de tendencia central, se emplearon coeficientes con mayor robustez estadística para atenuar el efecto de los datos atípicos y control de variables de naturaleza ordinal como el Rango Promedio (Rp) también conocido como desviación absoluta media. El Rp es una medida de dispersión que evalúa cuánto varían los valores individuales de un conjunto de datos en relación con su media. Para calcularlo, se resta cada dato de la media para obtener desviaciones individuales, se toma el valor absoluto de cada una y luego se calcula el promedio de estas desviaciones absolutas.

Los análisis estadísticos se configuraron y ejecutaron mediante el lenguaje R v.4.1.3 (R Core Team, 2017) y la interfaz de usuario Posit (RStudio Team, 2020), empleando los siguientes paquetes de funciones:

1. tidyverse y dplyr: Estos paquetes forman parte de un conjunto de paquetes de R para manipulación y análisis de datos. Sirven para realizar tareas de limpieza, gestión, transformación y visualización de datos (Wickham et al., 2019).

2. `haven`: Se utiliza para importar y exportar datos en formatos de software estadístico, como SPSS, SAS y Stata (Wickham et al., 2023).
3. `foreign`: Se utiliza para importar y exportar datos desde y hacia otros programas estadísticos. Ofrece soporte para varios formatos de archivo (R Core Team, 2021).
4. `psych`: El paquete proporciona un gran conjunto de funciones para realizar análisis psicométricos como análisis factorial, análisis de consistencia interna y generación de estadísticas descriptivas y gráficos (Revelle, 2017).
5. `MVN`: Se utiliza para realizar pruebas de normalidad multivariada en datos. Ayuda a evaluar si las variables en un conjunto de datos siguen una distribución multivariada normal (Korkmaz et al., 2014).
6. `MBESS`: Se emplea para realizar cálculos de tamaños del efecto y potencia en estudios de investigación. Es especialmente útil en el contexto de análisis de medidas repetidas y análisis de varianza (ANOVA) (Kelley, 2007).
7. `lavaan`: Paquete especializado en análisis de modelado de ecuaciones estructurales (SEM) en R. Permite especificar y estimar modelos de SEM, incluyendo modelos de confirmación, de medición y estructurales (Rosseel, 2012).
8. `semPlot`: Se usa para visualizar modelos de ecuaciones estructurales. Se puede crear gráficos de sendero y diagramas de modelos de SEM para una representación visual de los resultados (Epskamp, 2015).
9. `semTools`: Es un paquete complementario para el análisis de modelado de ecuaciones estructurales (SEM) que incluye herramientas para calcular estadísticas específicas de SEM, como índices de ajuste, intervalos de confianza y otros (Jorgensen et al., 2022).
10. `BifactorIndicesCalculator`: Este paquete se utiliza para calcular índices de modelos bifactor, que son una extensión de los modelos de factores para evaluar estructuras latentes en datos psicométricos (Dueber, 2021).

3.7.1. Aspectos éticos

Se contó con el consentimiento informado y anónimo de todas las autoridades, docentes y estudiantes participantes de la muestra de la institución educativa de nivel superior universitario.

3.7.2. Limitaciones

La principal limitación del estudio es la lejanía entre el tiempo de aplicación de las encuestas y la finalización del curso evaluado en algunos grupos.

IV. Resultados

4.1. Análisis de ítems

Posteriormente a la depuración de la base de datos, se realizó los análisis descriptivos de los ítems del cuestionarios, clasificadas según sus respectivas dimensiones, consignando medidas de tendencia central, de forma y correlación ítem-subescala corregida. Para la asimetría y curtosis se consideró como límite adecuado cuando el valor se encuentra ± 2 , conforme a una escala de cuatro opciones de respuesta (Lloret-Segura et al., 2014). Para la correlación ítem-subescala (índice de homogeneidad corregida) se consideró como límite un valor > 0.3 .

En la tabla 8 se observa adecuadas medidas para los ítems de las dimensiones de desempeño docente.

Tabla 8

Estadísticos descriptivos de los ítems de la variable desempeño docente

Subescalas e ítems	<i>M</i>	<i>DE</i>	<i>g₁</i>	<i>g₂</i>	<i>r_{is}</i>
Exploración competencial					
1	1.47	1.01	-0.07	-1.12	0.620
2	1.69	0.86	-0.29	-0.53	0.763
3	1.69	0.82	-0.24	-0.44	0.758
4	2.01	0.76	-0.45	-0.09	0.658
5	1.92	0.78	-0.32	-0.36	0.658
Explicitación de criterios					
6	2.19	0.71	-0.56	0.07	0.703
7	2.16	0.72	-0.52	-0.06	0.706
8	2.02	0.83	-0.45	-0.50	0.692
9	1.99	0.81	-0.56	-0.06	0.732

10	2.02	0.83	-0.52	-0.34	0.703
11	2.08	0.79	-0.52	-0.24	0.719
Ilustración					
12	2.14	0.76	-0.64	0.09	0.680
13	1.99	0.72	-0.39	0.03	0.637
14	2.08	0.72	-0.39	-0.16	0.673
15	1.94	0.80	-0.40	-0.31	0.694
16	2.02	0.73	-0.38	-0.14	0.723
17	1.81	0.82	-0.23	-0.52	0.558
Supervisión de la práctica					
18	1.98	0.83	-0.46	-0.41	0.768
19	2.04	0.76	-0.49	-0.09	0.757
20	2.07	0.78	-0.58	0.02	0.732
21	1.98	0.79	-0.53	0.01	0.741
22	1.86	0.82	-0.24	-0.59	0.718
Retroalimentación					
23	2.05	0.81	-0.51	-0.33	0.734
24	2.09	0.81	-0.55	-0.31	0.748
25	1.91	0.81	-0.38	-0.37	0.752
26	1.92	0.85	-0.39	-0.52	0.747
27	1.96	0.79	-0.45	-0.2	0.771
28	1.96	0.81	-0.42	-0.36	0.766
Evaluación					
29	1.98	0.87	-0.46	-0.55	0.649
30	1.88	0.81	-0.33	-0.42	0.718

31	1.91	0.77	-0.43	-0.05	0.747
32	2.02	0.78	-0.43	-0.27	0.746
33	1.97	0.85	-0.53	-0.31	0.705

Nota. *M*: Media; *DE*= Desviación estándar; g_1 = Asimetría; g_2 = Curtosis; r_{is} = Correlación ítem-subescala.

En la tabla 9 se observa que el ítem 24 de la subescala “mejoramiento” tiene una baja correlación ítem-subescala (< 0.3) por lo que se decidió retirarlo del cuestionario. El resto de ítems muestran adecuadas medidas descriptivas.

Tabla 9

Estadísticos descriptivos de los ítems de la variable desempeño estudiantil

Subescalas e ítems	<i>M</i>	<i>DE</i>	g_1	g_2	r_{is}
Precurrentes y potenciales					
habilidades					
1	1.7	0.59	-0.33	0.09	0.479
2	1.57	0.66	0.3	-0.38	0.495
3	1.58	0.67	0.13	-0.3	0.480
4	1.91	0.7	-0.25	-0.07	0.442
5	1.91	0.66	-0.27	0.22	0.486
Identificación de criterios					
6	2.01	0.55	-0.29	1.33	0.565
7	2.08	0.6	-0.21	0.36	0.570
8	2.17	0.6	-0.31	0.67	0.679
9	2.16	0.63	-0.29	0.03	0.686
10	2.07	0.61	-0.26	0.54	0.558

Participación

11	2.07	0.64	-0.4	0.67	0.535
12	2.09	0.64	-0.32	0.31	0.586
13	2.1	0.59	-0.21	0.61	0.658
14	1.83	0.66	-0.34	0.32	0.545
15	2.01	0.62	-0.38	0.81	0.664
16	1.92	0.64	-0.33	0.48	0.578

Práctica pertinente

17	1.99	0.59	-0.37	1.15	0.564
18	2.05	0.68	-0.43	0.38	0.631
19	1.78	0.74	-0.13	-0.32	0.648
20	1.9	0.66	-0.15	-0.13	0.659
21	1.99	0.61	-0.22	0.43	0.587
22	1.9	0.73	-0.22	-0.32	0.532

Mejoramiento

23	2.01	0.64	-0.29	0.29	0.410
24	2.13	0.92	-0.69	-0.56	0.053
25	1.95	0.59	-0.31	0.85	0.395
26	1.96	0.59	-0.36	0.99	0.541
27	1.86	0.59	-0.34	0.69	0.385
28	2.08	0.58	-0.33	1.12	0.576

Desempeño resultante

29	2.02	0.64	-0.45	0.78	0.650
30	2.07	0.68	-0.37	0.07	0.583
31	2.04	0.52	-0.13	1.44	0.516
32	2	0.61	-0.34	0.8	0.645

33	2.13	0.64	-0.31	0.19	0.602
34	2.09	0.75	-0.54	-0.01	0.613

Nota. *M*: Media; *DE*= Desviación estándar; g_1 = Asimetría; g_2 = Curtosis; r_{is} = Correlación ítem-subescala.

4.2. Análisis factorial confirmatorio

Se seleccionó la técnica de análisis factorial confirmatorio (AFC) porque ya se cuenta con especificaciones de los modelos para las variables, provenientes de la teoría-matriz. El propósito del presente análisis es evaluar y comparar las estructuras factoriales: oblicua, jerárquico indirecto y jerárquico directo (bifactor), considerando a desempeño docente y desempeño estudiantil como factores generales (FG) y sus respectivas dimensiones como factores específicos (FE).

En primer lugar, se comprobaron los supuestos metodológicos del uso de esta técnica: 1) comprobación de la normalidad univariante y multivariada, 2) naturaleza de las variables (Lloret-Segura et al., 2014). Para las comprobaciones de normalidad en los ítems de desempeño docente, se evaluó la normalidad univariada mediante la prueba de Shapiro-Wilk ya que ha demostrado su robustez estadística en distintos tamaños de muestra (Mohd Razaki & Bee Wah, 2011), se identificó que la puntuaciones no se distribuyen de acuerdo a la curva de normalidad ($p < .001$) en todos los ítems. Para evaluar la normalidad multivariada se usó la prueba de Mardia a partir del coeficiente de asimetría de 12810.58 ($p < .001$) y coeficiente de curtosis de 60.19 ($p < .001$), comprobándose que no se ajusta a la distribución de normalidad multivariante. En el caso de la variable desempeño estudiantil, ningún ítem se ajustó a la normalidad univariada según la prueba S-W ($p < .001$). De igual forma para la normalidad multivariada, la prueba de Mardia obtuvo un coeficiente de asimetría de 12565.91 ($p < .001$) y un coeficiente de curtosis de 59.94 ($p < .001$).

En relación con el segundo supuesto, el estilo de respuesta de los reactivos de ambas variables se mide en escala ordinal. Una vez establecidas la comprobación de los supuestos de la técnica, se procedió a determinar un estimador que sea robusto frente a la ausencia de normalidad multivariada y examine los reactivos no lineales; por lo tanto se seleccionó el método de Mínimos cuadrados ponderados con media y varianza ajustada (DiStefano et al., 2018), abreviado como WLSMV (Brown, 2015).

En la comprobación de ajuste a los modelos propuestos, se analizaron los índices de ajuste absolutos tales como: Test de Chi-cuadrado (χ^2); Chi-cuadrado relativa (χ^2/df); Raíz del Error Cuadrático Medio (RMSEA); Media Cuadrática Residual Ponderada (WRMR); y relativos como el índice de Tucker-Lewis (TLI); Índice de Ajuste Comparativo (CFI); En cuanto a los puntos de corte límites de “ajuste óptimo”, se asumió los siguientes límites propuestos (Baumgartner & Homburg, 1996; DiStefano et al., 2018; Hu & Bentler, 1999): Chi-cuadrado relativa $> \chi^2/df > 1$; TLI $> .95$; CFI $> .95$; RMSEA $< .06$ y WRMR ≤ 1 .

En cuanto a los índices estadísticos para evaluar el modelo bifactor, mediante la comprobación de la presencia de un factor general (FG) se examinó la varianza común explicada (ECV) y el porcentaje de correlaciones no contaminadas (PUC), el índice de replicabilidad de constructo (H), el porcentaje de varianza confiable (PVR) y el omega jerárquico (ω_{H_S}). Se tomaron las consideraciones de valores ECV por encima de .60 (Reise et al., 2013), no obstante si en el análisis en conjunto con el coeficiente PUC (abreviatura en inglés) refleja valores de .80, se asumiría unidimensionalidad esencial (Ventura-León et al., 2021). En el análisis del omega jerárquico entre el FG (H_S), magnitudes por encima de $> .30$ en el evidenciaría fiabilidad sustancial; valores entre .20 a .30 implicaría fiabilidad moderada; y, por debajo de .20, fiabilidad baja (Smits et al., 2015; Ventura-León et al., 2021). Por otro lado, para el PVR se optó por coeficientes por encima de 50% (Ventura-León et al., 2021). Además, el índice de replicabilidad de constructo se consideró aceptables a partir de .60

(Dominguez-Lara & Rodriguez, 2017). Finalmente, se evaluaron las cargas factoriales del modelo jerárquico directo y se diagramó las variables en un “path diagram” por cada variable de estudio.

4.2.1. Análisis factorial confirmatorio de Desempeño docente

Se evaluó los índices de ajuste para los posibles modelos de desempeño docente: el modelo oblicuo, compuesto de únicamente de los FE (dimensiones) inter correlacionados, sin la presencia de un FG; el modelo jerárquico indirecto, que supone una carga factorial lineal desde el FG (orden superior) a los FE (orden inferior) y de los FE a los ítems; y el modelo bifactor, que asume cargas factoriales en simultáneo del FG y los FE dirigido a los ítems.

Tabla 10

Índices de bondad de ajuste para tres modelos de Desempeño docente (n=475)

Modelos	χ_2	gl	p	TLI	CFI	RMSEA	IC 95%		WRMR
							LI	LS	
Oblicuo	1236.496	480	.000	0.974	0.976	0.058	0.054	0.062	1.210
Jerárquico indirecto	1262.630	489	.000	0.974	0.976	0.058	0.054	0.062	1.255
Bifactor	616.176	441	.000	0.993	0.994	0.029	0.023	0.034	0.991

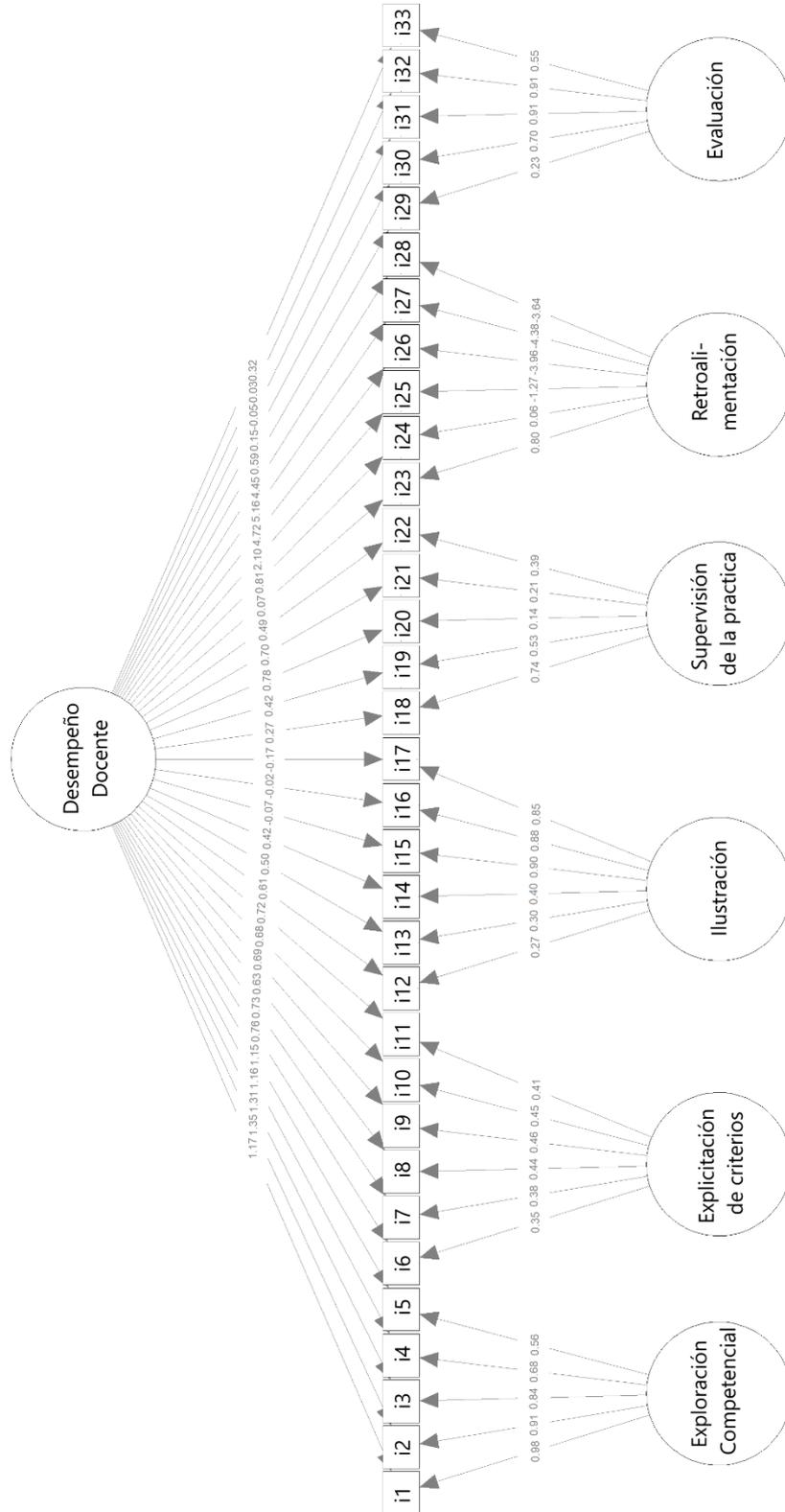
Nota. χ_2 : Prueba chi cuadrado robusta; gl: grados de libertad; p: nivel de significancia estadística; TLI: índice de Tucker-Lewis; CFI: índice de ajuste comparativo; RMSEA: raíz de la media cuadrática del error de aproximación; WRMR: Media Cuadrática Residual Ponderada.

De acuerdo con los criterios de índices de ajuste establecidos, todos los modelos obtienen óptimo ajuste en sus índices relativos TLI y CFI (>.95). En el caso de RMSEA, los modelos oblicuo y jerárquico indirecto obtienen buenas medidas (RMSEA<.08) y el modelo

bifactor obtiene una excelente medida ($<.05$). En el caso de WRMR, el modelo bifactor obtiene una excelente medida (≤ 1) mientras que los restantes modelos se encontraron por encima del criterio de corte. Se concluye que el modelo bifactor obtiene mejores índices de ajuste en comparación de los modelos oblicuo y jerárquico directo. Este modelo bifactor se someterá a un análisis de sus parámetros locales y a pruebas para confirmar la pertinencia de un factor general dado que los modelos de segundo orden suelen sobrestimar sus índices de ajuste (Dominguez-Lara & Rodriguez, 2017). En la figura 14, se muestran las cargas factoriales al factor general (FG) y a los factores específicos (FE) del modelo bifactor. La mayoría de ítems muestran adecuadas altas cargas factoriales ($>.30$), aunque se observan algunas cargas bajas (<0.2) y cargas anómalas (negativas o muy altas).

Figura 9

Diagrama de senderos del modelo bifactor de desempeño docente



En la tabla 11 se muestran los índices estadísticos para evaluar el modelo bifactor, obteniéndose medidas adecuadas para PUC ($>.60$), coeficiente H ($>.70$) y un alto omega jerárquico del FG ($\geq .70$) y superior a los omega jerárquico de los FE. La ECV estuvo a punto de llegar a $.60$ pero por su cercanía y por la evaluación global se podría asumir que el comportamiento de las dimensiones de las variables se adapta a la unidimensionalidad esencial en simultaneidad con los factores específicos.

Tabla 11

Índices del modelo Bifactor del cuestionario de Desempeño docente

	ECV	PUC	ω_{HS}	H	PVR
Factor General	.589	0.857	.926	1.143	95.168
Factores Específicos			ω_{hs}		
Exploración competencial			.624	.965	62.469
Explicitación de criterios			.222	.558	25.054
Ilustración			.667	.916	84.495
Supervisión de la práctica			.312	.650	40.011
Retroalimentación			.275	1.321	29.875
Evaluación			.709	.913	89.400

Nota. ECV: Varianza común explicada del factor general; PUC: Porcentaje de correlaciones no contaminadas; ω_{HS} : Coeficiente Omega jerárquico; H: coeficiente H.; PVR: Porcentaje de varianza confiable.

A partir de los favorables índices de ajuste del AFC del modelo de desempeño docente y sus respectivos factores/dimensiones, y las pruebas estadísticas para modelos bifactor, se concluye que existen evidencias de validez de la estructura interna para el cuestionario de desempeño docente.

4.2.2. Análisis factorial confirmatorio de Desempeño estudiantil

Se especificaron tres modelos de desempeño estudiantil, similares a los especificados para desempeño docente. Los modelos fueron: modelo oblicuo; modelo jerárquico indirecto y modelo bifactor. Los índices de ajuste para los tres modelos fueron los siguientes:

Tabla 12

Índices de bondad de ajuste para tres modelos de Desempeño estudiantil (n=475)

Modelos	χ_2	gl	p	TLI	CFI	RMSEA	IC 95%		WRMR
							LI	LS	
Oblicuo	1205.555	480	.000	0.951	0.955	0.056	0.053	0.060	1.289
Jerárquico indirecto	1266.256	489	.000	0.948	0.952	0.058	0.054	0.062	1.349
Bifactor	621.579	441	.000	0.987	0.989	0.029	0.024	0.035	0.996

Nota. χ_2 : Prueba chi cuadrado robusta; gl: grados de libertad; p: nivel de significancia estadística; TLI: índice de Tucker-Lewis; CFI: índice de ajuste comparativo; RMSEA: raíz de la media cuadrática del error de aproximación; WRMR: Media Cuadrática Residual Ponderada.

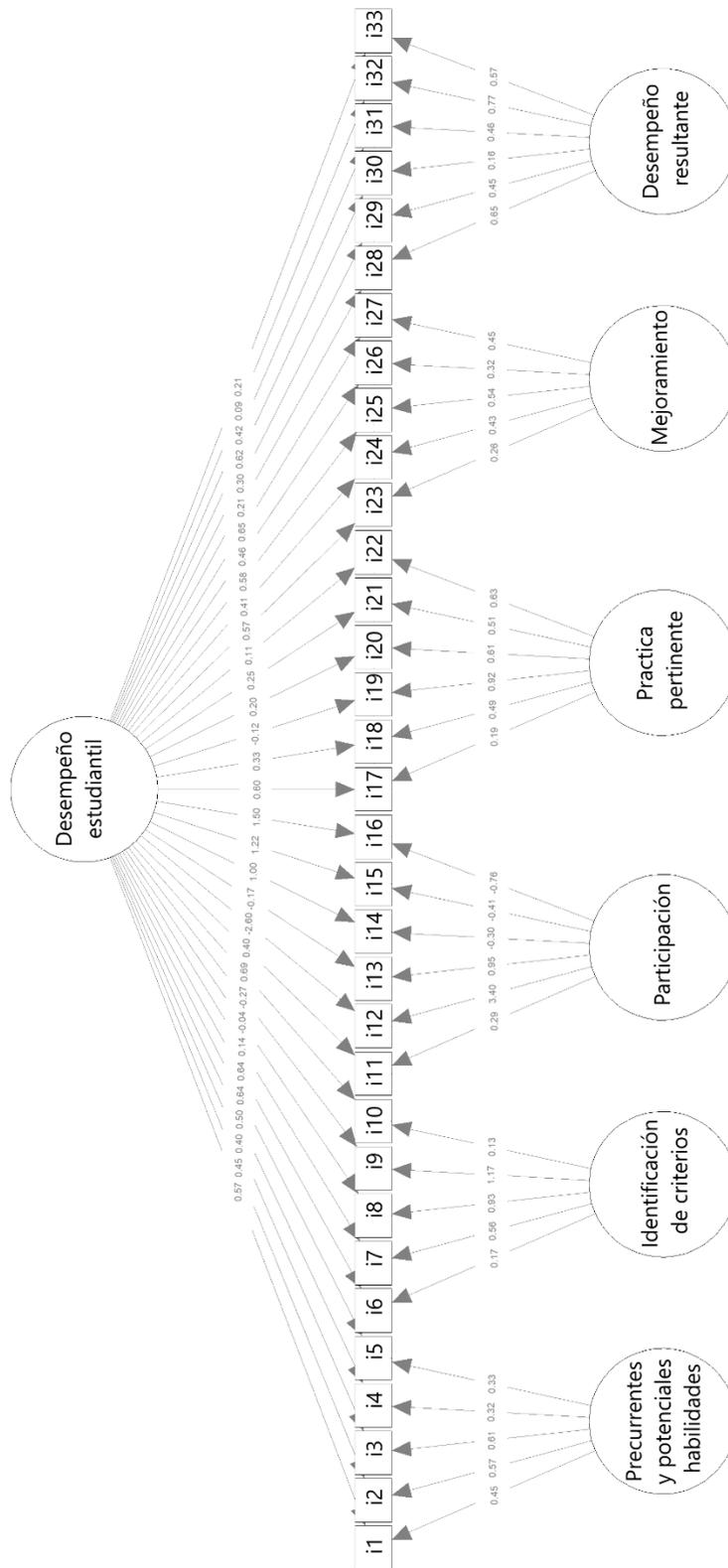
Los resultados hallados fueron muy similares a los índices de la variable desempeño docente. Quien obtuvo los menos adecuados índices de ajuste fue el modelo jerárquico indirecto, cuyo TLI fue menor a 0.95 y WRMR fue muy alto. En la comparación entre el modelo oblicuo y el modelo bifactor, ambos obtienen óptimos ajustes en los índices relativos TLI y CFI ($>.95$). Para la evaluación del RMSEA, se destaca la medida del modelo bifactor ($<.05$). También se destacó en el índice WRMR (≤ 1), mientras que el modelo oblicuo mostró una medida elevada. Se concluye que el modelo bifactor de desempeño estudiantil obtiene mejores índices de ajuste en comparación de los modelos oblicuo y jerárquico directo. Este

modelo bifactor también se someterá a un análisis de sus parámetros locales y pruebas confirmatorias del efecto de un factor general.

En la figura 15, se muestran las cargas factoriales al factor general (FG) y a los factores específicos (FE) del modelo bifactor para desempeño estudiantil. Al igual que el modelo bifactor de la otra variable de estudio, la mayoría de ítems tienen buenas cargas factoriales ($>.30$) y se observa la presencia de ítems con cargas bajas o anómalas.

Figura 10

Diagrama de senderos del modelo bifactor de desempeño estudiantil



En la tabla 15 se muestran los índices estadísticos para evaluar el modelo bifactor de desempeño estudiantil, obteniéndose medidas adecuadas para coeficiente H ($>.70$) y PVR ($>.50$) y. El omega jerárquico del FG superó a casi todos los omega jerárquico de los FE, a excepción de “Identificación de criterios”. La ECV no alcanzó el corte de .60 pero el PUC supero el .80.

Con base en las evaluaciones del PUC, H y PVR, se optó por asumir de manera conservadora la unidimensionalidad de desempeño estudiantil. Se abordarán adicionales observaciones en la sección de discusión.

Tabla 13

Índices del modelo Bifactor del cuestionario de Desempeño estudiantil

	ECV	PUC	ω_{Hs}	H	PVR
Factor General	.434	0.857	.665	.998	75.201
Factores Específicos			ω_{hs}		
Precurrentes y potenciales habilidades			.333	.610	44.274
Identificación de criterios			.703	.731	89.933
Participación			.044	.904	5.019
Práctica pertinente			.635	.876	86.599
Mejoramiento			.270	.520	36.619
Desempeño resultante			.512	.765	71.656

Nota. ECV: Varianza común explicada del factor general; PUC: Porcentaje de correlaciones no contaminadas; ω_{Hs} : Coeficiente Omega jerárquico; H: coeficiente H.; PVR: Porcentaje de varianza confiable.

Con base en los resultados de los índices de ajuste del AFC del modelo de desempeño estudiantil y sus respectivos factores/dimensiones, y las pruebas estadísticas para modelos

bifactor, se concluye que existen evidencias de validez de la estructura interna para el cuestionario de desempeño estudiantil.

4.3. Invarianza de la medición

En tercer lugar, se analizó los modelos de invarianza de la medición para las puntuaciones del cuestionario de desempeño docente según los grupos, siguiendo el procedimiento de Jöreskog (2002) y Wu & Estabrook (2016) restringiendo los componentes de un análisis factorial multigrupo, por lo mismo se adoptará un enfoque híbrido (Little, 1997) alternando entre los coeficiente χ_2 e índices de ajuste con diferenciación comparativa de ajustes en RMSEA y CFI (Chen, 2007) como reglas de decisión para transitar entre niveles de IM, a partir de ello, se utilizará los criterios de Hu & Bentler (1999), DiStefano et al. (2018) y Svetina et al. (2020) para definir lo que es un “óptimo” ajuste en los índices de ajuste (TLI, CFI y RMSEA), Media cuadrática residual ponderada (WRMR) y diferencia estandarizada de los modelos (Δ CFI, Δ RMSEA), respectivamente, todo esto para modelos de IM con desenlaces ordinales. Al tratarse de un estudio piloto orientado a verificar, detectar, evitar y corregir errores de los instrumentos (Barbero et al., 2015; Muñiz, 2018), se prescindirá de análisis parcializados que podrían segmentar los ítems para cada agrupación.

Tabla 14

Modelos de invarianza de las puntuaciones de Desempeño Docente (n=475) según agrupaciones sociodemográficas en ...

IM	χ_2	gl	p	TLI	CFI	Δ CFI	RMSEA	Δ RMSEA	WRMR
Sexo									
IM-1	1168.66	960	.000	0.955	0.959		0.030		0.764
IM-2	1119.08	987	.000	0.972	0.974	.015	0.024	-.007	0.923
IM-3	1148.82	1014	.000	0.972	0.973	-.001	0.024	.000	0.937
IM-4	1176.61	1047	.000	0.974	0.974	.001	0.023	-.001	0.952

Edad									
IM-1	1165.05	960	.000	0.955	0.959		0.030		0.777
IM-2	1100.56	987	.000	0.976	0.977	.018	0.022	-.008	0.893
IM-3	1128.55	1014	.000	0.976	0.977	.000	0.022	.000	0.905
IM-4	<i>1169.88</i>	<i>1047</i>	<i>.000</i>	<i>0.975</i>	<i>0.975</i>	<i>-.002</i>	<i>0.022</i>	<i>.000</i>	<i>0.934</i>

Nota. En *cursiva* modelos que no alcanzaron un determinado nivel de invarianza ($p < .05$); IM: Modelos de invarianza; IM-1: invarianza configural, IM-2: invarianza métrica; IM-3: invarianza escalar; IM-4: invarianza estricta; χ_2 : Prueba chi cuadrado robusta; gl: grados de libertad; p: nivel de significancia; TLI: índice de Tucker-Lewis; CFI: índice de ajuste comparativo; Δ CFI: diferencia de CFI con el IM anterior; RMSEA: raíz de la media cuadrática del error de aproximación; Δ RMSEA: diferencia de RMSEA con el IM anterior; WRMR: Media Cuadrática Residual Ponderada. La presente leyenda se aplica a todas las tablas de modelos de invarianza.

En la tabla 14, se visualizó que el modelo hexafactorial del desempeño docente alcanzó equivalencia en su configuración, carga factorial, interceptos y residuos según sexo ($p < .05$) con óptimos índices de ajuste en CFI ($> .95$), TLI ($> .95$), RMSEA ($< .06$) y WRMR (< 1), así como en sus diferencias (Δ RMSEA $\leq .01$; Δ CFI $\geq -.002$). Por otro lado, según edad, no se identificó equivalencia en los residuos ($p > .05$) con leves mejoras en los índices de ajuste a través de los niveles de invarianza.

Tabla 15

Modelos de invarianza de las puntuaciones de Desempeño Docente (n=475) según agrupaciones socioeducativas en ...

IM	χ_2	gl	p	TLI	CFI	Δ CFI	RMSEA	Δ RMSEA	WRMR
Ciclo de estudios									

IM-1	2590.11	2400	0.004	0.958	0.962		0.029		1.235
IM-2	2633.75	2508	0.040	0.973	0.975	.013	0.023	-.006	1.566
IM-3	2772.00	2616	0.017	0.968	0.968	.000	0.025	.000	1.625
IM-4	2911.92	2748	0.015	0.968	0.967	.001	0.025	-.001	1.679
Turno									
IM-1	1195.27	960	.000	0.952	0.956		0.032		0.838
IM-2	1208.82	987	.000	0.956	0.959	.002	0.031	-.001	1.034
IM-3	1236.87	1014	.000	0.957	0.959	.000	0.030	.000	1.045
IM-4	1236.87	1014	.000	0.957	0.959	.000	0.030	.000	1.045
Especialidad									
IM-1	1210.82	960	.000	0.944	0.949		0.033		0.802
IM-2	1162.21	987	.000	0.962	0.965	.015	0.027	-.006	0.937
IM-3	1210.58	1014	.000	0.959	0.960	.000	0.029	.000	0.967
IM-4	1245.50	1047	.000	0.959	0.960	.000	0.028	.000	0.989

En la tabla 15, se observó los niveles de invarianza obtenidos por la composición de desempeño docente según ciclo de estudios, turno y especialidad, los cuales alcanzan la equidad entre cargas factoriales mas no en interceptos y residuos. Intergrupalmente, en el grupo ciclo y turno, se identifica óptimos índices de ajuste (CFI: > .95; TLI: > .95; RMSEA: < .06) y sus diferencias ($\Delta RMSEA \leq .01$; $\Delta CFI \geq .002$), no obstante, en WRMR obtuvieron un ajuste pobre (>1). En la última agrupación, se obtuvo índices de ajuste, de diferencias y WRMR óptimos en todos los niveles de invarianza, a pesar de estos, apenas alcanzó la equivalencia en sus cargas factoriales.

Tabla 16

Modelos de invarianza de las puntuaciones de Desempeño Docente (n=475) según agrupaciones de encuesta educativa en ...

IM	χ_2	gl	p	TLI	CFI	Δ CFI	RMSEA	Δ RMSEA	WRMR
Cursos para evaluar									
IM-1	1628.20	1440	.000	0.956	0.960		0.029		0.980
IM-2	1612.11	1494	.017	0.973	0.975	.015	0.022	-.006	1.181
IM-3	1691.62	1548	.006	0.969	0.970	.000	0.024	.000	1.228
IM-4	1771.79	1614	.003	0.967	0.967	.001	0.025	-.001	1.273

En la tabla 16, se identifica que las puntuaciones de desempeño docente se ubican en el nivel métrico de IM, con índices de ajuste óptimos (CFI: > .95; TLI: > .95; RMSEA: < .06) y sus diferencias (Δ RMSEA \leq .01; Δ CFI \geq -.002), por otro lado, su WRMR empeora en tanto transita en la jerarquía de invarianza (>1).

Tabla 17

Modelos de invarianza de las puntuaciones de Desempeño Estudiantil (n=475) según agrupaciones sociodemográficas en ...

IM	χ_2	gl	p	TLI	CFI	Δ CFI	RMSEA	Δ RMSEA	WRMR
Sexo									
IM-1	1144.00	960	.000	0.938	0.944		0.028		0.966
IM-2	1116.46	987	.000	0.958	0.961	.015	0.024	-.007	1.084
IM-3	1150.65	1014	.000	0.957	0.958	-.002	0.024	.000	1.105
IM-4	1188.49	1047	.000	0.957	0.957	.000	0.024	-.001	1.132
Edad									
IM-1	1162.70	960	.000	0.929	0.936		0.030		0.967

IM-2	1120.94	987	.000	0.955	0.958	.022	0.024	-.006	1.075
IM-3	1147.28	1014	.000	0.956	0.958	-.002	0.024	.000	1.087
IM-4	1200.43	1047	.000	0.951	0.951	-.006	0.025	.001	1.129

En la tabla 17, se observó que el modelo de desempeño docente demuestra su equivalencia en su estructura, cargas factoriales e interceptos. En el análisis de sus índices de ajuste se ubicaron como “óptimos” (CFI: > .95; TLI: > .95; RMSEA: < .06) y sus diferencias (Δ RMSEA \leq .01; Δ CFI \geq -.002) para sus agrupaciones sexo y edad. Por otro lado, en referencia a los WRMR tienden a empeorar (>1) a medida que se restringen componentes.

Tabla 18

Modelos de invarianza de las puntuaciones de Desempeño Estudiantil (n=475) según agrupaciones socioeducativas en ...

IM	χ_2	gl	p	TLI	CFI	Δ CFI	RMSEA	Δ RMSEA	WRMR
Ciclo de estudios									
IM-1	2583.03	2400	0.005	0.932	0.938		0.028		1.412
IM-2	2625.01	2508	0.051	0.958	0.961	.022	0.022	-.006	1.742
IM-3	2741.25	2616	0.043	0.957	0.958	-.003	0.023	.000	1.782
IM-4	2882.35	2748	0.037	0.956	0.955	-.003	0.023	.000	1.840
Turno									
IM-1	1144.00	960	.000	0.938	0.944		0.028		0.966
IM-2	1116.46	987	.000	0.958	0.961	.021	0.024	-.006	1.084
IM-3	1150.65	1014	.000	0.957	0.958	-.003	0.024	.000	1.105
IM-4	1188.49	1047	.000	0.957	0.957	.001	0.024	-.001	1.132
Especialidad									
IM-1	1182.14	960	.000	0.920	0.928		0.031		0.975

IM-2	1148.63	987	.000	0.944	0.947	.020	0.026	-.005	1.112
IM-3	1178.00	1014	.000	0.944	0.947	-.001	0.026	.000	1.126
IM-4	1219.86	1047	.000	0.943	0.944	-.003	0.026	.000	1.157

En la tabla 18, el modelo de desempeño estudiantil según ciclo de estudios alcanza la invarianza estricta ($p > .05$) con óptimos índices de ajuste CFI: $> .95$; TLI: $> .95$; RMSEA: $< .06$) y sus diferencias estandarizadas ($\Delta RMSEA \leq .01$; $\Delta CFI \geq -.002$), sin embargo, en sus coeficientes WRMR tiende a sobrecargarse a medida que se restringen valores. En el análisis de la variable según especialidad, no se obtuvo equivalencia entre los residuos, en tanto se repite el panorama anterior: índices de ajuste y diferencia de estos óptimos, con coeficientes WRMR cada vez más restringidos. En la misma tabla, se aprecia la comparación de la variable de estudio según turno con un nivel de invarianza métrica alcanzado, resaltando sus índices de ajuste óptimo y la diferencia de los mismos acordes a los criterios de Svetina et al. (2020), y coeficientes WRMR aumentando su carga a medida que se restringen valores.

Tabla 19

Modelos de invarianza de las puntuaciones de Desempeño Estudiantil (n=475) según agrupaciones de encuesta educativa en ...

IM	χ_2	gl	p	TLI	CFI	ΔCFI	RMSEA	$\Delta RMSEA$	WRMR
Cursos para evaluar									
IM-1	1595.83	1440	0.002	0.946	0.951		0.026		1.162
IM-2	1619.58	1494	0.012	0.958	0.960	.010	0.023	-.003	1.356
IM-3	1688.88	1548	0.007	0.954	0.955	-.005	0.024	.001	1.395
IM-4	1763.70	1614	0.005	0.953	0.952	-.003	0.024	.000	1.436

En la tabla 19, se observó la comparación del desempeño estudiantil según el curso a evaluar se comprobó la equivalencia de las cargas factoriales ($p > .05$), mas no de los interceptos

y residuos ($p < .05$). Además, se reportan índices de ajuste óptimos para los criterios previamente establecidos (CFI: $> .95$; TLI: $> .95$; RMSEA: $< .06$) y sus diferencias estandarizadas ($\Delta\text{RMSEA} \leq .01$; $\Delta\text{CFI} \geq -.002$), mientras que en todos los niveles de invarianza se obtuvo valores por encima de lo esperado para la naturaleza del estudio (< 1).

4.4. Fiabilidad

Se evaluó como consistencia interna de las dimensiones de las variables y fue medido con los siguientes coeficientes: alfa (α) y alfa ordinal (α_{ordinal}). También se reportó los intervalos de confianza para el coeficiente alfa (95% IC). Para una evaluación general de la fiabilidad de las variables, se optó por el coeficiente omega jerárquico (ω_{HS}).

En la tabla 20 se muestra excelentes coeficientes alfa ordinal para las dimensiones de desempeño docente.

Tabla 20

Coefficientes alfa y alfa ordinal de las dimensiones de desempeño docente

Dimensiones	α	95% IC de α	α_{ordinal}
Exploración competencial	0.863	[.846; .887]	0.904
Explicitación de criterios	0.889	[.872; .908]	0.925
Ilustración	0.864	[.844; .886]	0.903
Supervisión de la práctica	0.895	[.880; .911]	0.930
Retroalimentación	0.911	[.898; .925]	0.940
Evaluación	0.878	[.857; .899]	0.915

La fiabilidad de desempeño docente fue $\omega_{\text{HS}}=0.926$.

En la tabla 21 se muestra adecuados coeficientes alfa ordinal para las dimensiones de desempeño estudiantil. La única dimensión que mostró una regular medida de alfa ordinal fue “precurrentes y potenciales habilidades”.

Tabla 21*Coefficientes alfa y alfa ordinal de las dimensiones de desempeño estudiantil*

Dimensiones	α	95% IC de α	α_{ordinal}
Precurrentes y potenciales habilidades	0.717	[.672; .762]	0.785
Identificación de criterios	0.819	[.792; .850]	0.884
Participación	0.824	[.792; .859]	0.877
Práctica pertinente	0.829	[.805; .859]	0.879
Mejoramiento	0.745	[.705; .793]	0.821
Desempeño resultante	0.828	[.805; .859]	0.884

La fiabilidad para desempeño estudiantil fue $\omega_{H_S}=0.665$

4.5. Resultados descriptivos

El procesamiento de datos permitió la sistematización de las variables de estudio y sociodemográfica para sus análisis estadísticos.

4.5.1. Descripción de la población

En la tabla 22 se observa la distribución de los participantes de acuerdo con las características sociodemográficas. Se identifica una mayor frecuencia en las participantes de sexo femenino (72.6%) en comparación al sexo masculino (27.4%). Para las frecuencias por especialidad, se describe específicamente una mayor frecuencia a 275 estudiantes de cursos generales (57.9%), 88 de Psicología Clínica (18.5%), 63 participantes de Psicología Organizacional (13.3%), 35 de Psicología Educativa (7.4%) y 14 de Psicología Social (2.9%).

Tabla 22

Características sociodemográficas de los participantes (N= 475).

Característica	n	%
Sexo		
Masculino	130	27.4
Femenino	345	72.6
Edad		
22 años a menos	245	51.6
Mayor a 23 años	230	48.4
Lugar de Procedencia		
Lima Este	127	26.7
Lima Norte	129	27.2
Lima Sur	72	15.2
Lima Centro	84	17.7
Callao	48	10.1

Otros	2	0.4
NS/NC	13	2.7
Especialidad		
Cursos Generales	275	57.9
Especialidad Prof.	200	42.1
Psicología Clínica	88	18.5
Psicología Organizacional	63	13.3
Psicología Educativa	35	7.4
Psicología Social	14	2.9
Ciclo de estudios		
Primer Ciclo	83	17.5
Tercer Ciclo	106	22.3
Quinto Ciclo	87	18.3
Séptimo Ciclo	116	24.4
Noveno Ciclo	83	17.5
Turno de estudios		
Mañana	205	43.2
Tarde	270	56.8
Tipo de curso		
Básica	112	23.6
Investigación	167	35.2
Especialidad	196	41.3

4.5.2. Descripción de las variables de estudio

Se procedió a realizar un análisis exploratorio de las variables desempeño docente y desempeño estudiantil percibidos, así como sus dimensiones a partir de estadígrafos de resumen. En la tabla 23 se observa la dimensión de “retroalimentación” es la que presenta mayor desviación estándar (4.07).

Tabla 23

Estadígrafos para las dimensiones de la variable desempeño docente

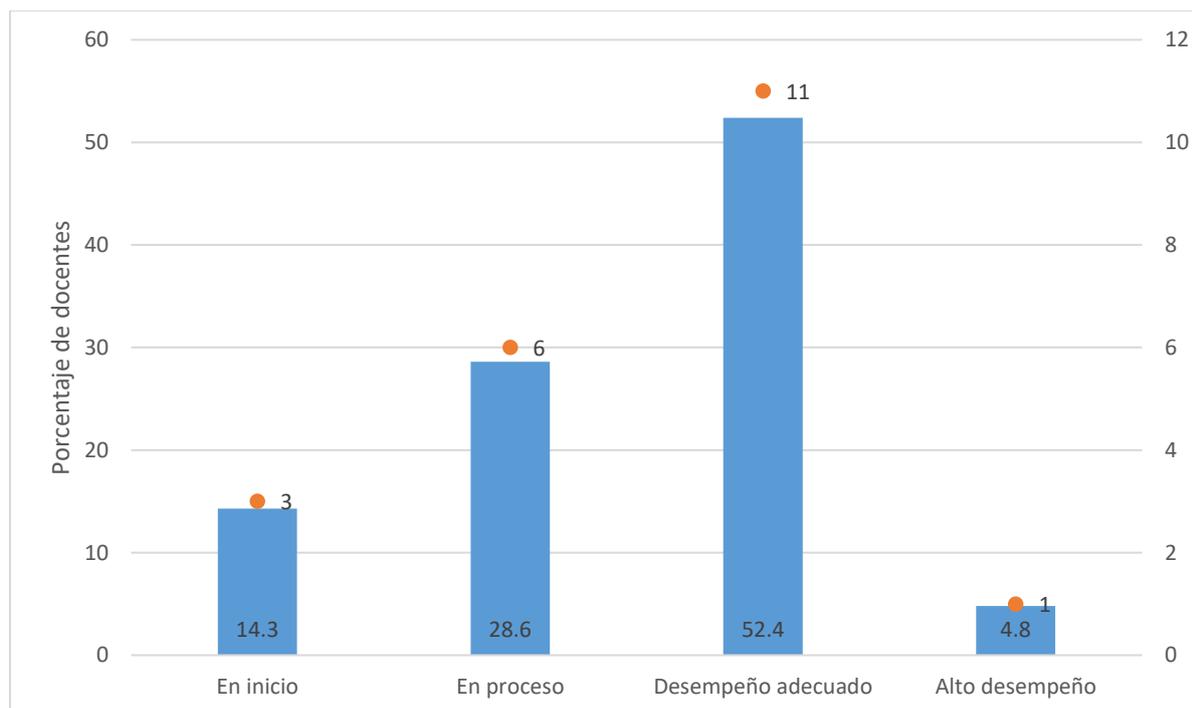
Dimensiones y total	M	Mediana	DE	V	Mínimo	Máximo
Exploración Competencial	8.78	9	3.42	11.7	0	15
Explicitación de criterios	12.45	12	3.77	14.2	0	18
Ilustración	11.99	12	3.51	12.3	0	18
Supervisión de la práctica	9.93	10	3.35	11.2	0	15
Retroalimentación	11.89	12	4.07	16.6	1	18
Evaluación	9.76	10	3.35	11.2	0	15
Total	64.79	66	19.02	361.6	4	99

Nota: M = media; DE = desviación estándar; V = varianza.

Para calificar el desempeño del docente se usaron cuatro niveles (Anexo J), los cuales son indicadores de las habilidades de enseñanza evaluadas por los estudiantes. Se evaluaron a los 21 docentes participantes del estudio. En la figura 11 se muestran las frecuencias de docentes (en naranja) y sus porcentajes (en azul), de acuerdo con el nivel de desempeño de enseñanza alcanzado. Se destaca que la mayoría de docentes (52.4%) alcanzó un nivel “desempeño adecuado”.

Figura 11

Frecuencia y porcentaje de docentes según nivel de desempeño alcanzado



En la tabla 24 se estableció el resumen de cada dimensión, a partir de este análisis se observa que “Precurrentes y potenciales habilidades” presenta menor desviación de sus datos (DE=2.24) respecto a su media (8.68) en comparación a las demás subvariables, no obstante “Práctica Pertinente” obtiene mayor dispersión de sus datos (DE= 2.95) respecto a su tendencia central (11.60). En la puntuación general se destaca una media de 65.01 y una desviación de 12.61.

Tabla 24

Estadígrafos para las dimensiones de la variable desempeño estudiantil

Dimensiones y total	M	Mediana	DE	V	Mínimo	Máximo
Precurrentes y potenciales habilidades	8.68	9	2.24	5.01	1	15
Identificación de criterios	10.5	10	2.28	5.22	1	15
Participación	12.02	12	2.76	7.61	0	18
Práctica pertinente	11.6	12	2.96	8.73	1	18

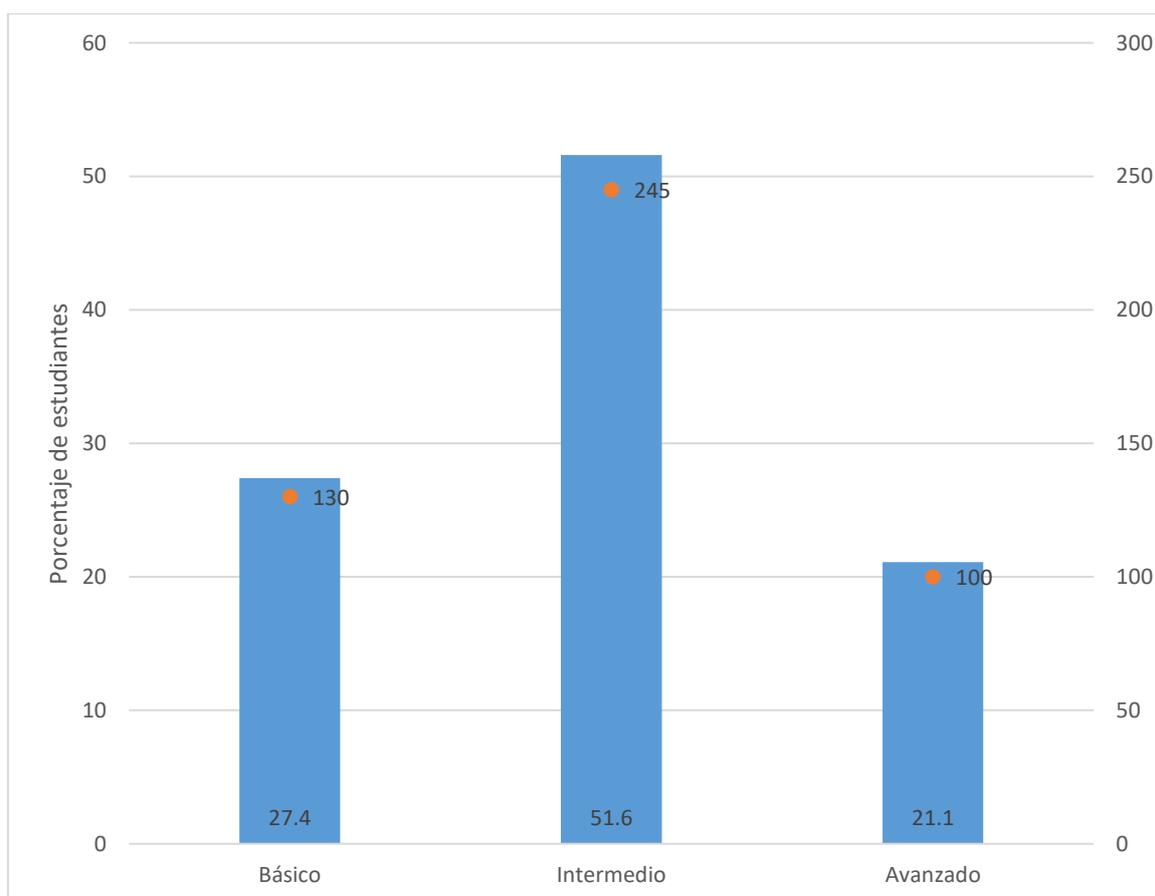
Mejoramiento	9.85	10	2.11	4.45	2	15
Desempeño resultante	12.36	12	2.84	8.04	2	18
Total	65.01	64	12.61	158.91	17	99

Nota: M = media; DE = desviación estándar; V = varianza.

Para calificar el desempeño de los alumnos se emplearon tres niveles (Anexo J), donde los 475 alumnos autocalificaron sus conductas de estudio. En la figura 12 se muestra la frecuencia de alumnos (en naranja) y sus porcentajes (en azul), de acuerdo con el nivel de desempeño estudiantil alcanzado. La mayoría de alumnos (51.6%) se ubica en un nivel “intermedio”.

Figura 12

Frecuencia y porcentaje de alumnos según nivel de desempeño estudiantil



En el establecimiento de pruebas acordes a la naturaleza y comportamiento de los datos, se empleó la prueba de bondad de ajuste a la curva normal de Kolgomorov Smirnov con la corrección de Lillieford (KSL en adelante) para las variables de estudio. Se obtuvieron valores p estadísticamente significativos ($p < .05$) evidenciando la necesidad de usar pruebas no paramétricas para su análisis inferencial: Rho de Spearman para las correlaciones; U Mann Whitney para analizar la diferencia entre dos grupos; y H Kruskall Wallis para identificar la diferencia entre más de dos grupos.

4.6. Correlaciones entre variables

Se establece una matriz de correlaciones acorde a la naturaleza de los datos y la distribución no paramétrica, seleccionando la prueba Rho de Spearman. En la tabla 25 se muestra que las correlaciones son estadísticamente significativas ($p < .05$). Se confirma una correlación directa y grande entre desempeño docente y desempeño estudiantil (Rho = 0.596). También se evaluó un modelo de regresión lineal, mediante el coeficiente de determinación, obteniendo un $R^2 = 0.367$, cifra que indica la variabilidad explicada del desempeño docente sobre el desempeño estudiantil.

En cuanto a las correlaciones entre dimensiones y variables, se encuentran correlaciones “grandes” entre desempeño docente total y la dimensión “desempeño resultante” de desempeño estudiantil (Rho = 0.637), seguido de una correlación grande entre las dimensiones “evaluación” y “desempeño resultante” (Rho = 0.627), siendo esta última dimensión quien obtiene mayor intensidad con cada una de las dimensiones en las correlaciones existentes.

Tabla 25

Matriz de correlación entre las variables y dimensiones de desempeño docente y desempeño estudiantil percibido

	Precurrentes y potenciales	Identificaci ón de criterios	Particip ación	Práctica pertinente	Mejora miento	Desempe ño resultante	Total DE
Exploración Competencial	.317*	.357*	.368*	.364*	.334*	.462*	.446*
Explicitación de criterios	.268*	.458*	.448*	.444*	.348*	.565*	.518*
Ilustración	.290*	.428*	.482*	.452*	.372*	.520*	.519*
Supervisión de la práctica	.294*	.444*	.465*	.484*	.396*	.574*	.545*
Retroalimentac ión	.321*	.437*	.447*	.475*	.417*	.603*	.553*
Evaluación	.304*	.454*	.484*	.471*	.409*	.627*	.568*
Total DD	.333*	.487*	.509*	.507*	.431*	.637*	.596*

Nota. * $p < .001$

4.7. Resultados comparativos

En las tablas 26 y 27 se observa la comparación de las variables desempeño docente, desempeño estudiantil y sus respectivas dimensiones según la variable sexo, ubicándose valores p por encima de 0.05 evidenciando diferencias estadísticamente no significativas, no obstante, en la dimensión “Precurrentes y potenciales habilidades” se identifica una mayor puntuación en los hombres ($R_p=260.64$) en comparación a las mujeres ($R_p=229.47$) siendo esta diferencia estadísticamente significativa pero sin presentar alguna magnitud del efecto ($P_s=0.326$).

Tabla 26

Prueba de U de Mann Whitney para el análisis de la variable desempeño docente percibido según sexo

Desempeño Docente	Sexo		U	p	Ps
	Masculino (n=130)	Femenino (n=345)			
Exploración Competencial	248.98	233.86	20998	0.283	0.361
Explicitación de criterios	241.95	236.51	21912	0.699	0.383
Ilustración	236.50	238.56	22230.5	0.884	0.394
Supervisión de la práctica	239.51	237.43	22228.5	0.882	0.391
Retroalimentación	244.82	235.43	21538.5	0.505	0.374
Evaluación	235.00	239.13	22034.5	0.768	0.392
Total	242.55	236.29	21833.5	0.657	0.381

Tabla 27

Prueba de U de Mann Whitney para el análisis de la variable desempeño estudiantil percibido según sexo

Desempeño Estudiantil	Sexo		U	p	Ps
	Masculino (n=130)	Femenino (n=345)			
Precurrentes y potenciales habilidades	260.64	229.47	19482	0.026	0.326
Identificación de criterios	233.83	239.57	21883	0.678	0.391
Participación	242.43	236.33	21848.5	0.662	0.381
Práctica pertinente	243.27	236.02	21740.5	0.605	0.379
Mejoramiento	245.12	235.32	21499.5	0.479	0.373

Desempeño resultante	239.68	237.37	22206	0.868	0.390
Total	241.77	236.58	21935.5	0.713	0.384

En la tabla 28 y 29 se observan las comparaciones según la variable edad, dividida en dos grupos, los estudiantes de 22 años a menos y los estudiantes que tienen 23 años a más. Solo se hallaron diferencias estadísticamente significativas en la variable desempeño estudiantil ($p=0.035$) y su dimensión “Precurrentes y potenciales habilidades” ($p=0.034$) a favor de los estudiantes mayores (23 años a más).

Tabla 28

Prueba de U de Mann Whitney para el análisis de la variable desempeño docente percibido según edad

Desempeño Docente	Edad (en años)		U	p	Ps
	22 años a menos (n=245)	Mayor a 23 años (n=230)			
Exploración Competencial	233.70	242.58	27121	0.479	0.478
Explicitación de criterios	232.40	243.97	26803	0.357	0.473
Ilustración	230.66	245.82	26375.5	0.227	0.465
Supervisión de la práctica	230.84	245.63	26420	0.238	0.466
Retroalimentación	229.17	247.40	26012	0.147	0.459
Evaluación	231.15	245.29	26497.5	0.259	0.467
Total	229.72	246.82	26146.5	0.175	0.461

Tabla 29

Prueba de U de Mann Whitney para el análisis de la variable desempeño estudiantil percibido según edad

Desempeño Estudiantil	Edad (en años)		<i>U</i>	<i>p</i>	<i>Ps</i>
	22 años a menos (n=245)	Mayor a 23 años (n=230)			
Precurrentes y potenciales habilidades	225.18	251.66	25034	0.034	0.442
Identificación de criterios	229.51	247.04	26095	0.156	0.460
Participación	226.26	250.51	25298.5	0.052	0.446
Práctica pertinente	229.43	247.13	26075	0.157	0.460
Mejoramiento	230.27	246.24	26280.5	0.196	0.463
Desempeño resultante	230.36	246.14	26303.5	0.204	0.464
Total	225.14	251.70	25023.5	0.035	0.442

En la tabla 30 y 31 se observan las comparaciones según la variable nivel de especialidad, donde se divide a los estudiantes que cursan estudios generales y los estudiantes que están cursando alguna de las cuatro especialidades que ofrece la facultad. Se visualiza valores *p* por encima de 0.05, a excepción de tres dimensiones: Precurrentes y Potenciales Habilidades ($p=0.023$), Práctica Pertinente ($p=0.032$) y Desempeño Estudiantil Total ($p=0.018$).

Tabla 30

Prueba de U de Mann Whitney para el análisis de la variable desempeño docente percibido según nivel de especialidad

Desempeño Docente	Nivel de Especialidad		<i>U</i>	<i>p</i>	<i>Ps</i>
	General (n=275)	Especialidad Prof. (n=200)			
Exploración Competencial	228.31	251.33	24834.5	0.070	0.433

Explicitación de criterios	233.02	244.85	26130	0.351	0.458
Ilustración	234.81	242.39	26623	0.551	0.468
Supervisión de la práctica	230.15	248.79	25341.5	0.142	0.443
Retroalimentación	229.20	250.11	25079	0.100	0.438
Evaluación	230.76	247.95	25509.5	0.175	0.446
Total	230.57	248.22	25455.5	0.166	0.445

Tabla 31

Prueba de U de Mann Whitney para el análisis de la variable desempeño estudiantil percibido según nivel de especialidad

Desempeño Estudiantil	Nivel de Especialidad		U	p	Ps
	General (n=275)	Especialidad Prof. (n=200)			
Precurrentes y potenciales habilidades	225.93	254.59	24182	0.023	0.420
Identificación de criterios	232.92	244.98	26103.5	0.334	0.457
Participación	228.29	251.35	24830.5	0.067	0.433
Práctica pertinente	226.58	253.71	24358.5	0.032	0.424
Mejoramiento	230.10	248.86	25328.5	0.134	0.442
Desempeño resultante	232.30	245.84	25932	0.282	0.454
Total	225.33	255.42	24017	0.018	0.417

En la tabla 32 y 33 se observan las comparaciones según la variable turno de estudios, donde se observa que, en todas las dimensiones y puntaje total de Desempeño docente, existen diferencias estadísticamente significativas, en favor de los docentes del turno tarde. Caso

contrario en las dimensiones y total de desempeño estudiantil, donde no se observa ningún resultado estadísticamente significativo.

Tabla 32

Prueba de U de Mann Whitney para el análisis de la variable desempeño docente percibido según turno de estudios.

Desempeño Docente	Turno de estudios		U	p	Ps
	Mañana (n=205)	Tarde (n=270)			
Exploración Competencial	197.77	268.55	19427.5	0.000	0.351
Explicitación de criterios	216.11	254.62	23188.5	0.002	0.419
Ilustración	212.81	257.12	22512	0.000	0.407
Supervisión de la práctica	214.69	255.7	22895.5	0.001	0.414
Retroalimentación	212.87	257.08	22522.5	0.000	0.407
Evaluación	215.68	254.95	23098.5	0.002	0.417
Total	209.2	259.87	21771	0.000	0.393

Tabla 33

Prueba de U de Mann Whitney para el análisis de la variable desempeño estudiantil percibido según turno de estudios.

Desempeño Estudiantil	Turno de estudios		U	p	Ps
	Mañana (n=205)	Tarde (n=270)			
Precurrentes y potenciales habilidades	241.07	235.67	27046.5	0.668	0.489
Identificación de criterios	235.36	240	28215.5	0.71	0.510
Participación	236.98	238.77	27883.5	0.887	0.504

Práctica pertinente	236.69	239	27944.0	0.855	0.505
Mejoramiento	237.45	238.41	27787.0	0.939	0.502
Desempeño resultante	224.75	248.06	30392.0	0.063	0.549
Total	235.66	239.78	27195.5	0.746	0.491

En la tabla 34 y 36 se observan las comparaciones según la variable eje curricular, donde se clasificó a los docentes y cursos evaluados en los tres ejes curriculares. Para la variable desempeño docente y sus dimensiones se obtuvo diferencias estadísticamente significativas. A partir de este hallazgo es factible establecer comparaciones por parejas mediante un análisis post-hoc, donde se observa que entre los ejes curriculares “investigación” y “especialidad” no existen diferencias estadísticamente significativas, pero si entre estos dos ejes con el eje “básica” en todas las dimensiones y total de desempeño docente.

Tabla 34

Prueba H de Kruskal Wallis para el análisis de la variable desempeño docente según eje curricular

Desempeño Docente	Eje curricular			H	p	ϵ^2
	Investigación (n=167)	Básica (n=112)	Especialidad (n=196)			
Exploración Competencial	238.76	214.08	251.02	5.212	0.074	0.011
Explicitación de criterios	267.58	184.10	243.59	25.585	0.000	0.054
Ilustración	257.49	201.09	242.48	11.771	0.003	0.025
Supervisión de la práctica	256.23	192.03	248.73	16.894	0.000	0.036

Retroalimentación	262.69	179.85	250.19	27.240	0.000	0.057
Evaluación	258.53	192.75	246.36	16.820	0.000	0.035
Total	261.10	185.73	248.19	22.059	0.000	0.047

Tabla 35

Análisis post-hoc mediante la prueba Bonferroni para los grupos de eje curricular de la variable desempeño docente

Desempeño Docente	Eje curricular	BSD	<i>p</i>
Explicitación de criterios	básica*-especialidad	-59.492	0.000
	básica*-investigación	83.478	0.000
Ilustración	básica*-especialidad	-41.386	0.011
	básica*-investigación	56.400	0.001
Supervisión de la práctica	básica*-especialidad	-56.703	0.000
	básica*-investigación	64.199	0.000
Retroalimentación	básica*-especialidad	-70.343	0.000
	básica*-investigación	82.843	0.000
Evaluación	básica*-especialidad	-53.615	0.001
	básica*-investigación	65.780	0.000
Total	básica*-especialidad	-62.461	0.000
	básica*-investigación	75.371	0.000

Nota: Los símbolos denotados con “*” indican el grupo de eje curricular con menor puntuación en la comparación.

Para el caso de desempeño estudiantil, no se hallaron diferencias estadísticamente significativas en “Identificación de criterios” y “Participación”. En las restantes dimensiones y total se ejecutó el análisis post-hoc, destacándose un resultado similar al anterior, donde los

puntajes de desempeño estudiantil del eje curricular “básico” fueron bajos en comparación con los otros dos ejes.

Tabla 36

Prueba H de Kruskal Wallis para el análisis de la variable desempeño estudiantil percibido según eje curricular

Desempeño Estudiantil	Eje curricular			H	p	ε^2
	Investigación (n=167)	Básica (n=112)	Especialidad (n=196)			
Precurrentes y potenciales habilidades	231.16	215.46	256.71	7.224	0.027	0.015
Identificación de criterios	243.09	214.89	246.87	4.402	0.111	0.009
Participación	235.13	218.02	251.86	4.552	0.103	0.010
Práctica pertinente	242.92	198.89	256.16	12.945	0.002	0.027
Mejoramiento	243.32	207.91	250.66	7.594	0.022	0.016
Desempeño resultante	253.83	202.76	244.65	10.352	0.006	0.022
Total	241.00	199.37	257.52	12.928	0.002	0.027

Tabla 37

Análisis post-hoc mediante la prueba Bonferroni para los grupos de eje curricular de la variable desempeño estudiantil

Desempeño Estudiantil	Eje curricular	BSD	p
Precurrentes y potenciales habilidades	básica*-especialidad	-41.245	0.010

Explicitación de criterios	296.51	208.44	204.13	232.21	260.84	28.052	0.000	0.059
Ilustración	284.64	215.34	213.34	237.19	247.26	15.791	0.003	0.033
Supervisión de la práctica	305.73	199.12	199.41	241.17	255.94	37.484	0.000	0.079
Retroalimentación	316.55	189.95	195.72	234.40	270.16	53.432	0.000	0.113
Evaluación	287.43	211.42	205.58	234.95	260.76	22.176	0.000	0.047
Total	308.06	198.84	198.07	236.63	261.71	40.112	0.000	0.085

Tabla 39

Análisis post-hoc mediante la prueba Bonferroni para los grupos de ciclos de estudios de la variable desempeño docente.

Desempeño Docente	Ciclo de estudios	BSD	<i>p</i>
Exploración Competencial	Tercero*-Séptimo	-54.701	0.029
	Tercero*-Noveno	-56.537	0.048
	Tercero*-Primero	102.308	0.000
	Quinto*-Primero	83.361	0.001
Explicitación de criterios	Quinto*-Primero	92.374	0.000
	Tercero*-Primero	88.063	0.000
	Séptimo*-Primero	64.299	0.011
Ilustración	Quinto*-Primero	71.3	0.007
	Tercero*-Primero	69.3	0.005
Supervisión de la práctica	Tercero*-Noveno	-56.817	0.045
	Tercero*-Primero	106.606	0.000

	Quinto*-Primero	106.315	0.000
	Séptimo*-Primero	64.561	0.01
	Tercero*-Noveno	-80.21	0.001
	Tercero*-Primero	126.601	0.000
Retroalimentación	Quinto*-Noveno	-74.444	0.004
	Quinto*-Primero	120.836	0.000
	Séptimo*-Primero	82.158	0.000
	Quinto*-Primero	81.853	0.001
Evaluación	Tercero*-Primero	76-014	0.001
	Quinto*-Séptimo	-63.642	0.025
	Quinto*-Primero	109.991	0.000
Total	Tercero*-Noveno	-62.867	0.018
	Tercero*-Primero	109.216	0.000
	Séptimo*-Primero	71.427	0.003

Nota: Los símbolos denotados con “*” indican la especialidad con menor puntuación en la comparación.

Encontramos un resultado diferente en el desempeño estudiantil según ciclo de estudios, donde solo se hallaron diferencias estadísticamente significativas en las dimensiones “práctica pertinente”, “mejoramiento” y el puntaje total. En las comparaciones por pareja, el desempeño estudiantil del tercer ciclo es el que menos puntaje tiene en comparación con los otros ciclos.

Tabla 40

Prueba H de Kruskal Wallis para el análisis de la variable desempeño estudiantil percibido según ciclo de estudios.

Desempeño Estudiantil	Ciclo de estudios					H	p	ϵ^2
	Primero (n=83)	Tercero (n=106)	Quinto (n=87)	Séptimo (n=116)	Noveno (n=83)			
Precurrentes y potenciales habilidades	227.49	211.91	242.53	246.50	265.19	8.283	0.082	0.017
Identificación de criterios	249.96	219.87	233.10	247.23	241.44	3.301	0.509	0.007
Participación	230.48	224.49	233.00	246.81	255.72	3.331	0.504	0.007
Práctica pertinente	256.22	202.12	226.25	262.91	243.10	13.492	0.009	0.028
Mejoramiento	256.22	202.12	226.25	262.91	243.10	10.858	0.028	0.023
Desempeño resultante	269.02	217.53	218.99	249.31	237.24	9.318	0.054	0.020
Total	248.80	207.86	224.47	254.19	257.26	9.726	0.045	0.021

Tabla 41

Análisis post-hoc mediante la prueba Bonferroni para los grupos de ciclos de estudios de la variable desempeño estudiantil.

Desempeño Estudiantil	Ciclo de estudios	BSD	p
	Tercero*-Noveno	-40.974	0.040
Práctica pertinente	Tercero*-Primero	54.094	0.007
	Tercero*-Séptimo	-60.791	0.001
Mejoramiento	Tercero*-Noveno	-54.737	0.006

	Tercero*-Primero	51.948	0.008
	Tercero*-Primero	40.937	0.042
Total	Tercero*-Séptimo	-46.327	0.012
	Tercero*-Noveno	-49.401	0.014

Nota: Los símbolos denotados con “*” indican la especialidad con menor puntuación en la comparación.

V. Discusión de resultados

En la primera parte se discutió los resultados comparativos según las variables sociodemográficas y educativas. En la segunda parte se analizó la relación entre el desempeño docente y el desempeño estudiantil, a la luz de los resultados actuales y el marco de investigación. A partir de las evidencias de invarianza de la medición obtenidas para las variables y dimensiones, se hace posible las comparaciones de los puntajes por grupos. Se discuten cada una de acuerdo a los agrupamientos:

Género

Para desempeño docente y todas sus dimensiones no se hallaron diferencias entre el género masculino y femenino (D. docente: $p = .657$; $P_s = 0.381$; $d = 0.005$). De acuerdo con los resultados hallados en la muestra de estudiantes de psicología, se interpreta que, en la percepción del desempeño docente, el género no es un factor que diferencie los puntajes de la evaluación. En otros estudios se halló un resultado similar: En la muestra de estudiantes mexicanos de psicología (Bazán-Ramírez & Velarde, 2021) no se identificó grandes diferencias entre las medias de varones y mujeres al calificar a sus docentes. En la muestra de estudiantes peruanos provenientes de dos universidades públicas (Bazán-Ramírez et al., 2021), tampoco se hallaron diferencias estadísticamente significativas. Como lo explican estos últimos autores, el género, como variable de agrupación, no afecta diferencialmente la evaluación de la enseñanza de los docentes de psicología.

Para desempeño estudiantil, únicamente se hallaron diferencias estadísticamente significativas para la dimensión “Precurrentes y potenciales habilidades”, en favor de los varones ($p = .026$; $P_s = 0.326$; $d = 0.250$). Esto se puede interpretar en que los varones ($M = 9.08$; $DE = 2.37$) perciben estar mejor preparados que las mujeres ($M = 8.53$; $DE = 2.17$), en la etapa inicial de los cursos. Un resultado similar se halló con una muestra mexicana (Bazán-Ramírez & Velarde, 2021) y en una muestra peruana con la variable autoeficacia académica

(Dominguez-Lara & Fernández-Arata, 2019). La similaridad de estos resultados se puede ser causada por las características de las muestras: peruanos, universitarios de la licenciatura de psicología, siendo la mayoría mujeres.

Respecto al resto de dimensiones y la variable general desempeño estudiantil, no se hallaron diferencias (D. estudiantil: $p = .713$; $Ps = 0.384$; $d = 0.072$), resultado parecido al reportado por Bazán-Ramírez et al. (2022b), a excepción de la dimensión “desempeño resultante”, donde los autores si hallaron diferencias estadísticamente significativas. Como síntesis, se puede afirmar que, en la autoevaluación transversal del comportamiento de estudio, se obtienen resultados muy similares entre ambos géneros. Bazán-Ramírez et al. (2022b) explica que este resultado es consistente con la evidencia anterior y que otras metodologías más ideográficas podrían develar nueva información.

Edad

Se clasificó a los estudiantes en dos grupos, los estudiantes que tenían 22 años a menos y los estudiantes que tenían 23 años a más. En cuanto al desempeño docente, no se hallaron diferencias en ninguna de las dimensiones y variable general (D. docente: $p = .175$; $Ps = 0.461$; $d = 0.114$). Este resultado es diferente al antecedente de Bazán-Ramírez et al. (2021), donde si se hallaron diferencias estadísticamente significativas entre dos grupos de edad, para los puntajes de cinco dimensiones de desempeño docente. Una causa de esta diferencia en los resultados es que, en el antecedente mencionado, las edades de clasificación eran 21 y 22 años, a diferencia del presente estudio, que estableció el corte entre 22 y 23 años.

Bazán-Ramírez et al. (2021) explica que la variable edad parece un indicador relevante en la forma de evaluar el desempeño docente, así como estar relacionada con otras variables de evolución educativa. Las diferencias de percepción de la enseñanza según la edad aumentan conforme el rango de edad es mayor en la muestra (Abellán-Roselló et al., 2023), pero cuando

el rango de edad es pequeño, la influencia de la edad parece disminuir (Acevedo & Mairena, 2006).

Sobre el desempeño estudiantil, se hallaron diferencias estadísticamente significativas en la dimensión “Precurrentes y potenciales habilidades” y en la variable total (D. estudiantil: $p = .035$; $P_s = 0.442$; $d = 0.189$), en favor de los estudiantes mayores de 23 años. Como explicación a este resultado, se puede interpretar que los estudiantes de mayor edad, al contar con la experiencia de cursar las asignaturas universitarias, perciben mejor su desempeño estudiantil en general, así como se sienten mejor preparados para iniciar los cursos, en comparación con los estudiantes menores a 22 años. En el resto de dimensiones no se hallaron diferencias, resultado similar al hallado en una muestra de estudiantes argentinos de psicología que autoevaluaron sus estrategias de estudio (Hendrie & Bastacini, 2019). La causa de esta coincidencia en los resultados podría deberse a la similaridad de la carrera y rango de edad de las muestras.

Especialidad de carrera

En el desempeño docente, no se observaron diferencias entre los tres grupos de nivel de especialidad del curso evaluado (D. docente: $p = .166$; $P_s = 0.445$; $d = 0.147$). Bazán-Ramírez et al. (2021) realizó una clasificación similar con dos grupos, uno de “etapa disciplinaria” (4^{to} y 6^{to} semestres) y “etapa final” (8^{vo} y 10^{mo} semestre). Al comparar estos grupos según los puntajes de la variable desempeño docente, obtuvo diferencias estadísticamente significativas y un tamaño del efecto moderado ($\eta^2 = 0.102$). Si bien estos resultados pueden parecer contradictorios, como explicación de estas diferencias se debe precisar que la agrupación de “estudios generales” en la presente investigación incluyó a los estudiantes que finalizaron el 1^{er}, 3^{er} y 5^{to} ciclo de estudios, donde los estudiantes de primer año calificaron a sus profesores con puntuaciones altas y pudieron influir en la varianza del grupo “estudios generales”, haciéndolo similar al grupo “especialidad”. Cuando se observan

las comparaciones por eje curricular (básico vs especialidad) y ciclo de estudios (3^{er} y 5^{to} vs 7^{mo} y 9^{no}) en los resultados del estudio actual, si se observan diferencias en la calificación del desempeño docente, a favor de los ciclos superiores (y de “especialidad”), siendo consistente este resultado con el antecedente presentado. Como explicación a este resultado, Bazán-Ramírez et al. (2021) refiere que la etapa académica de la formación profesional puede influir en la percepción que tiene el estudiante sobre la enseñanza del docente, ya sea por el interés que tiene el estudiante en un tipo de curso básico o de especialidad, en las habilidades ya adquiridas por el estudiante en su formación, y por las competencias didácticas del docente para enseñar en los ámbitos científico y tecnológico de la psicología (Silva et al., 2016).

En cuanto al desempeño estudiantil, se hallaron diferencias estadísticamente significativas en las dimensiones “Precurrentes y potenciales habilidades”, “Práctica pertinente” y en la variable total (D. estudiantil: $p = .018$; $P_s = 0.417$; $d = 0.148$), en favor de los estudiantes que están cursando sus especialidades. La interpretación de este resultado es que los estudiantes de especialidad perciben un mejor desempeño suyo a la hora de mostrar sus habilidades previas en el inicio de los cursos y en adaptarse a la retroalimentación brindada por sus docentes, en comparación de los estudiantes que siguen cursos generales. La razón de esta diferencia puede ser que los estudiantes de especialidad, al recibir entrenamiento en análisis e intervención sobre el comportamiento en sus cursos de especialidad, podrían extrapolar estas destrezas para mejorar su desempeño durante los cursos.

Turno de estudios

Se agruparon a los alumnos según su horario de estudio. En la comparación de puntajes en desempeño docente, en todas las dimensiones y variable general se hallaron diferencias estadísticamente significativas, en favor del turno tarde (D. docente: $p < .000$; $P_s = 0.393$; $d = 0.406$). Es decir, los estudiantes perciben un mejor desempeño de enseñanza de los profesores que dirigen cursos en el turno tarde o noche, a diferencia de los que dictan en el turno de

mañana. Una posible interpretación de este resultado es que el turno tarde agrupo a los docentes que enseñan los cursos de especialidad profesional, que también han sido evaluados con mayor puntaje, en comparación con los docentes de los cursos básicos, que dictan en el turno mañana en su mayoría. Como antecedente, en Morales y Chávez (2019) investigaron la satisfacción académica en universitarios mexicanos de la carrera de psicología, hallando también diferencias estadísticamente significativas ($t = 2.10$; $p = .036$) entre los turnos matutino y vespertino, aunque en favor del turno matutino. La diferencia de este resultado podría ser causada en como las instituciones universitarias gestionan el dictado de los cursos según la disponibilidad de los docentes y estudiantes, así como el grado de especialidad del curso.

Respecto al desempeño estudiantil, no se hallaron diferencias estadísticamente significativas según el turno de estudios (D. estudiantil: $p = .746$; $P_s = 0.491$; $d = 0.058$), por lo que se puede interpretar que las habilidades de estudio permanecen similares entre ambos turnos. En el antecedente de Huaringa y Aliaga (2012), el turno de estudios influyó en el rendimiento académico de los alumnos en sus primeros años, pero dejó de influir cuando los estudiantes cursaban los últimos años de estudio, lo cual es parcialmente similar al resultado actual, y su causa puede estar en la similaridad de la muestra, ya que el antecedente evaluó los efectos de variables socioeconómicas y educativas en el rendimiento académico de alumnos de una universidad nacional en Perú.

Respecto a estos resultados, generalmente la variable turno de estudios no es considerada en análisis de las interacciones docente-estudiante universitario, pero dependiendo de las características de una institución educativa, se podrían hallar condiciones favorables o desfavorables para la enseñanza y el estudio según sus horarios. Estas condiciones educativas (infraestructura, planes de estudio, cantidad y horarios de servicios, personal docente, etc.) han sido estudiadas como satisfacción académica (Montes de Oca et al., 2023; Morales & Chávez, 2019), y para el análisis de los estudiantes, se consideran sus características socioeconómicas

y educativas (Huaranga & Aliaga, 2012; Maturana, 2020). Estas perspectivas de análisis explicarían la diferencia entre los antecedentes y los resultados de la investigación actual.

Eje curricular

En la comparación del desempeño docente, se hallaron diferencias estadísticamente en cinco de las seis dimensiones, así como en la variable total (D. docente: $p < .000$; $\varepsilon^2 = 0.047$; $\omega^2 = 0.027$), como se observa en la tabla 36. Las diferencias más fuertes las podemos detectar con el tamaño del efecto épsilon al cuadrado, en las dimensiones Explicitación de criterios ($\varepsilon^2=0.054$); Retroalimentación ($\varepsilon^2=0.057$); y la variable total ($\varepsilon^2=0.047$). Los puntajes de los cursos de “Especialidad” e “Investigación” resultaron similares, por lo que las diferencias son con los cursos “Básicos”, que tienen menor puntaje. En el análisis post-hoc mediante la prueba de Bonferroni (tabla 37), se confirmó que las diferencias son con el grupo de cursos “Básica”, obteniendo diferencias estadísticamente significativas contra los otros ejes curriculares.

Un resultado similar se halló en Bazán-Ramírez y Velarde (2021), que clasificaron a los cursos de psicología en “General”, “Disciplinar”, y “Profesional”; y compararon los puntajes de las dimensiones de desempeño docente, obteniendo puntajes promedio más bajos en los docentes de la etapa de estudios “General”. Respecto a estos hallazgos, se observa que los docentes que enseñan cursos básicos o generales reciben puntajes más bajos en comparación de los docentes que dictan cursos de investigación o especialidad. Las hipótesis de explicación de esta similitud en los resultados y antecedentes se pueden ubicar en algunas características del docente, como su edad, capacitaciones recibidas, certificaciones profesionales, tiempo de enseñanza, condiciones socioeconómicas, etc. Sin embargo, no se descarta también irregularidades y defectos en los planes de enseñanza de los cursos básicos, los cuales no contengan situaciones de enseñanza donde se ejerciten las habilidades (Carpio & Irigoyen, 2005).

En cuanto al desempeño estudiantil según el eje curricular, se obtuvieron diferencias estadísticamente significativas en las dimensiones “Precurrentes y potenciales habilidades”, “Práctica pertinente”, “Mejoramiento”, “Desempeño resultante” y en la variable total (D. estudiantil: $p = .002$; $\varepsilon^2 = 0.027$; $\omega^2 = 0.009$). En el análisis post-hoc (tabla 39), se vuelve a confirmar que los puntajes bajos provienen del eje curricular “básica” aunque la interpretación en este caso es que los estudiantes que llevan cursos “básicos” de la carrera perciben su desempeño estudiantil con un puntaje menor, en comparación con los estudiantes que cursan los otros ejes curriculares.

En comparación al estudio de Bazán-Ramírez y Velarde (2021) con muestra mexicana de estudiantes de psicología, los puntajes promedio fueron más cercanos entre etapas de estudio, pero la etapa “disciplinar” obtuvo promedios mayores al resto en las primeras tres dimensiones de desempeño estudiantil. La etapa “general” obtuvo los promedios más bajos en la primera y tercera dimensión, por lo que los resultados de este antecedente coinciden parcialmente con lo hallado en la presente investigación. Al buscar explicaciones a estas coincidencias entre los hallazgos, los ámbitos didácticos nos brindan la información precisa de las áreas peor valoradas: la exploración que hace el docente a las habilidades previas y precurrentes de los estudiantes; la supervisión de los ejercicios prácticos; la retroalimentación brindada y cuanto mejoran los estudiantes; y las evaluaciones finales que hace el docente y los desempeños que alcanzaron los alumnos. Las bajas valoraciones podrían deberse al bajo desempeño docente, a la hora de formular situaciones de enseñanza pertinentes a estos ámbitos, en comparación a los docentes de especialidad profesional e investigación, donde puede que sean más competentes o creativos en diseñar y aplicar estas situaciones de enseñanza.

Ciclo de estudios

Desde primero hasta noveno, se compararon cinco grupos de estudiantes según su ciclo de estudios. Respecto al desempeño docente, se hallaron diferencias estadísticamente

significativas en todas sus dimensiones y total (D. docente: $p < .000$; $\varepsilon^2 = 0.085$; $\omega^2 = 0.074$), destacando Exploración Competencial ($\varepsilon^2=0.064$); Supervisión de la práctica ($\varepsilon^2=0.079$); Retroalimentación ($\varepsilon^2=0.113$) y el puntaje total ($\varepsilon^2=0.085$). Se destaca estas diferencias porque, de acuerdo a la percepción de los alumnos, en ciertos ciclos el desempeño de los docentes es percibido como superior, sobre todo en los ámbitos que involucra el intercambio constante, evaluación y mejora de las habilidades del alumno con la mediación del docente. Se procedió a analizar por par de ciclo de estudios mediante la prueba de Bonferroni (tabla 41), hallando que los puntajes del tercer y quinto ciclo resultaron los más bajos, obteniendo diferencias estadísticamente significativas con el resto de ciclos comparados. Este resultado es consistente con lo hallado en una muestra mexicana de estudiantes de psicología (Morales & Chávez, 2019) donde se compararon los niveles de satisfacción académica en los ciclos segundo, cuarto, sexto y octavo, hallando diferencias estadísticamente significativas, y donde coincidentemente los estudiantes del segundo y tercer año reportaron los menores puntajes. En cuanto a la explicación de esta coincidencia con el antecedente y nuestro resultado, las razones pueden ser las mismas que fueron expuestas para los análisis comparativos según especialidad de carrera, turno de estudios y eje curricular. Estas razones son características y competencias de los docente, la pertinencia del plan de estudios, así como el desarrollo de habilidades en evaluación de los estudiantes.

En cuanto al desempeño estudiantil, solo se hallaron diferencias estadísticamente significativas en “práctica pertinente”, “mejoramiento” y el puntaje total (D. estudiantil: $p = .045$; $\varepsilon^2 = 0.021$; $\omega^2 = 0.008$). Los tamaños del efecto coinciden con este resultado. En el análisis post hoc (tabla 43), los estudiantes de tercer y quinto ciclo volvieron a tener los puntajes más bajos en comparación a los otros ciclos, lo cual confirma que los estudiantes de tercer y quinto ciclo perciben que su desempeño no se está ajustando pertinentemente a la retroalimentación y supervisión del docente, a diferencia de los estudiantes del séptimo y noveno ciclo, donde estos

ámbitos didácticos están más presentes en sus cursos, por lo que las explicaciones pueden ser similares al análisis del desempeño estudiantil según eje curricular.

En el resto de dimensiones no se obtuvieron diferencias en los puntajes, lo cual coincide parcialmente con la investigación de Dominguez-Lara y Fernández-Arata (2019), donde compararon los puntajes de autoeficacia académica entre cinco grupos de años de estudio (desde primero a quinto año), y no hallaron diferencias estadísticamente significativas.

Respecto a la relación entre las variables desempeño docente y desempeño estudiantil, así como sus respectivas dimensiones, para comprobarlas, se correlacionaron estas variables, obteniendo en su mayoría correlaciones moderadas y altas, las cuales se discuten a continuación. Se halló la existencia de una correlación directa y grande entre las variables generales “desempeño docente” y “desempeño estudiantil” ($Rho = 0.596$), por lo cual se confirmó la hipótesis principal del estudio, la cual afirmó la existencia de una correlación entre los puntajes de las percepciones del desempeño docente con la autopercepción del desempeño estudiantil, en estudiantes de psicología. La hipótesis tuvo como base la conocida relación entre el desempeño docente y el desempeño académico de los estudiantes (Akram, 2019), mayormente estudiada en alumnos de educación básica regular, a nivel internacional (Akram, 2019; Vattøy & Smith, 2019) y nacional (Acuña, 2023; Pérez, 2018). Esta relación, así como su análisis en dimensiones (ámbitos didácticos), también ha sido comprobada en el presente estudio y coincide con los antecedentes con muestra de universitarios (Bazán-Ramírez et al., 2022a; González-Fernández, 2018; Salazar & Rosas, 2021).

Por ejemplo, a nivel nacional, Leon (2020) relacionó los puntajes de desempeño docente y aprendizaje autónomo en 650 estudiantes de un centro de idiomas de una universidad de Lima, obteniendo una correlación directa y moderada ($Rho = 0.389$; $p < .05$). A nivel internacional, Moreno (2023) relacionó los puntajes de práctica docente y aprendizaje autónomo en 70 estudiantes de enfermería de una universidad de Los Ríos, Ecuador, hallando

una correlación directa y grande ($Rho = 0.507$; $p < .05$). Las causas de la similaridad de los antecedentes con nuestro resultado principal son las características similares que comparten la muestras (estudiantes universitarios de pregrado); el parecido entre las variables estudiadas, así como los instrumentos que las evaluaron.

Como sustento teórico sobre este resultado general, Ibáñez y Ribes (2001) precisaron el supuesto de sincronidad recíproca entre la conducta del docente y el estudiante, como un constante intercambio funcional entre ambos actores, circunscritos en ámbitos de desempeño. Esto quiere decir que un análisis solitario del comportamiento del docente o del estudiante resultaría incompleto. Para analizar este fenómeno, se ha propuesto el concepto de “interacción didáctica”, como segmento analítico para estudiar las interacciones donde participan el docente, el estudiante y la tarea disciplinaria (Morales et al., 2013). Sus relaciones o interacciones son determinadas por el requerimiento de logro, que es una especificación para evaluar el éxito o error de un comportamiento (De la Sancha & Pérez-Almonacid, 2017). Asimismo, se ha definido el “ámbito funcional de desempeño” como el conjunto de prácticas (decires y haceres) que pueden ser ejercitadas, aprendidas, transmitidas y generadas por personas en circunstancias estructuradas por criterios y valoraciones (Acuña et al., 2013). Morales et al. (2017) explicitó que los ámbitos didácticos circunscriben a docentes y estudiantes, por lo que se puede obtener una perspectiva de cada actor educativo en los seis ámbitos didácticos, los cuales fueron plasmados en los cuestionarios CDD y CDE, mediante doce dimensiones. Los resultado hallados en las relaciones entre dimensiones confirman estas correspondencias.

VI. Conclusiones

- a. Se halló una correlación directa y grande ($r_s = 0.596$; $p < .001$) entre los puntajes de desempeño docente y desempeño estudiantil percibidos por estudiantes de psicología de una universidad pública de Lima – 2019.
- b. Se obtuvieron evidencias de validez y fiabilidad para los cuestionarios de desempeño docente – CDD y de desempeño estudiantil – CDE, por lo que se pueden emplear para propósitos de investigación y tecnología aplicada.
- c. Se hallaron correlaciones grandes, moderadas y pequeñas entre todos los cruces de las dimensiones de desempeño docente y desempeño estudiantil, destacándose las correlaciones entre la dimensión “desempeño resultante” con las dimensiones: Explicitación de criterios, Ilustración, Supervisión de la práctica, Retroalimentación y Evaluación.
- d. Se hallaron diferencias estadísticamente significativas entre los puntajes de desempeño docente en las variables de agrupación: turno de estudios ($p < .001$; $d = 0.406$), eje curricular ($p < .001$; $\omega^2 = 0.027$) y ciclo de estudios ($p < .001$; $\omega^2 = 0.074$).
- e. No se hallaron diferencias estadísticamente significativas entre los puntajes de desempeño docente en las variables de agrupación: género ($p = .657$; $d = 0.005$), edad ($p = .175$; $d = 0.114$) y especialidad de carrera ($p = .166$; $d = 0.147$).
- f. Se hallaron diferencias estadísticamente significativas entre los puntajes de desempeño estudiantil en las variables de agrupación: edad ($p = .035$; $d = 0.189$), especialidad de carrera ($p = .018$; $d = 0.148$), eje curricular ($p = .002$; $\omega^2 = 0.009$) y ciclo de estudios ($p = .045$; $\omega^2 = 0.008$).
- g. No se hallaron diferencias estadísticamente significativas entre los puntajes de desempeño estudiantil en las variables de agrupación: género ($p = .713$; $d = 0.072$) y turno de estudios ($p = .746$; $d = 0.058$).

VII. Recomendaciones

- a. Investigar la relación entre las dimensiones del desempeño docente y el desempeño estudiantil empleando diseños, estrategias y métodos de análisis de datos más completos y sofisticados.
- b. Considerar el rediseño del CDE, reformulando sus ítems y agregando nuevas dimensiones, con base en los ámbitos didácticos de la “Autodirección del estudio”.
- c. Continuar estudiando las razones de las diferencias y no diferencias halladas en las variables de agrupación. Así también considerar nuevas variables, como la realización de prácticas preprofesionales.
- d. Aportar evidencias de validez a los cuestionarios CDD y CDE, específicamente evidencias basadas en las consecuencias de la aplicación de los cuestionarios.
- e. Agregar más variables educativas al análisis, que puedan influir en el desempeño docente y desempeño estudiantil, bajo un diseño de aplicación de cuestionarios, pruebas de rendimiento y análisis documental.
- f. Ampliar las muestras de participantes a más carreras universitarias, para someter a prueba la validez del modelo de ámbitos didácticos para describir las prácticas de docentes y estudiantes de otras carreras tecnológicas y científicas.
- g. Investigar el punto de vista del profesor, mediante la creación de un cuestionario donde evalué su propio desempeño y otro cuestionario donde evalué el comportamiento de estudio de sus alumnos.

VIII. Referencias

- Abellán-Roselló, L., Fernández-Rodicio, C. I., & Reyes-Suarez, D. C. (2023). Diferencias en las percepciones del alumnado universitario sobre apoyos docentes según metodología, grado de estudios y edad. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 26(2), 193-205. <https://doi.org/10.6018/reifop.555341>
- Acevedo, R., & Mairena, N. (2006). Factores de sesgo asociados a la validez de la evaluación docente universitaria: un modelo jerárquico lineal. *Archivos Analíticos de Políticas Educativas*, 14(34), 1-22. <https://www.redalyc.org/pdf/2750/275020543034.pdf>
- Acuña, D. (2023). *Desempeño docente en el aprendizaje autónomo de estudiantes del VII ciclo de una institución educativa de Lima, 2022* [Tesis de Maestría, Universidad César Vallejo]. <https://repositorio.ucv.edu.pe/handle/20.500.12692/106413>
- Acuña, K., Irigoyen, J., & Jiménez, M. (2013). *La comprensión de contenidos científicos en estudiantes universitarios*. Qartuppi. <https://doi.org/10.29410/QTP.13.01>
- AERA, APA, & NCME. (2018). *Estándares para Pruebas Educativas y Psicológicas*. American Educational Research Association. <https://doi.org/10.2307/j.ctvr43hg2>
- Akram, M. (2019). Relationship between Students' Perceptions of Teacher Effectiveness and Student Achievement at Secondary School Level. *Bulletin of Education and Research*, 41(2), 93-108. <https://eric.ed.gov/?id=EJ1229453>
- American Psychological Association. (2021). *Manual de Publicaciones de la American Psychological Association* (Cuarta ed.). Manual Moderno.
- Ato, M., López, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038-1059. <https://doi.org/10.6018/analesps.29.3.178511>
- Barbero, M., Vila, E., & Holgado, F. (2015). *Psicometría (Teoría y Formulario)*. Editorial UNED.

- Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing*, 13(2), 139-161. [https://doi.org/10.1016/0167-8116\(95\)00038-0](https://doi.org/10.1016/0167-8116(95)00038-0)
- Bazán-Ramírez, A., Capa-Luque, W., Bello-Vidal, C., & Quispe-Morales, R. (2022). Influence of Teaching and the Teacher's Feedback Perceived on the Didactic Performance of Peruvian Postgraduate Students Attending Virtual Classes During the COVID-19 Pandemic. *Frontiers in Education*, 7. <https://doi.org/10.3389/feduc.2022.818209>
- Bazán-Ramírez, A., Pérez-Morán, J. C., & Bernal-Baldenebro, B. (2021). Criteria for Teaching Performance in Psychology: Invariance According to Age, Sex, and Academic Stage of Peruvian Students. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.764081>
- Bazán-Ramírez, A., Quispe-Morales, R., De La Cruz, C., & Henostroza-Mota, C. (2022). Teacher-Student Performance Criteria During Online Classes due to COVID-19: Self-Report by Postgraduate Students in Education. *European Journal of Educational Research*, 11(4), 2101-2114. <https://doi.org/10.12973/eu-jer.11.4.2101>
- Bazán-Ramírez, A., & Velarde, N. (2021). Auto-reporte del estudiantado en criterios de desempeño didáctico en clases de Psicología. *Journal of Behavior, Health & Social Issues*, 13(1), 22-35. <http://revistas.unam.mx/index.php/jbhsi/article/view/78071>
- Beltrán, R., Amésquita, J., & Turpo-Gebera, O. (2022). Desempeño docente en las competencias investigativas de estudiantes de maestría. *Comuni@cción: Revista de Investigación en Comunicación y Desarrollo*, 13(4), 262-271. <https://doi.org/10.33595/2226-1478.13.4.747>
- Brown, T. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.
- Camacho, I., Arroyo, R., & Serrano, M. (2012). Psicología y lógica: Una relación transdisciplinaria. *Interdisciplinaria*, 28(2), 221-230.

<https://www.redalyc.org/pdf/180/18022339004.pdf>

Carpio, C., & Irigoyen, J. J. (Eds.). (2005). *Psicología y educación. Aportaciones desde la teoría de la conducta*. UNAM-FESI.

Carpio, C., Pacheco, V., Canales, C., & Flores, C. (1998). Comportamiento inteligente y juegos de lenguaje en la enseñanza de la psicología. *Acta Comportamentalia: Revista Latina de Análisis de Comportamiento*, 6(1), 47-60.
<http://revistas.unam.mx/index.php/acom/article/view/18244>

Catania, C. (1975). The myth of self-reinforcement. *Behaviorism*, 3(2), 192-199.
<https://www.jstor.org/stable/27758845>

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464-504.
<https://doi.org/10.1080/10705510701301834>

De la Sancha, E. (2016). *Análisis experimental del desarrollo de competencias intrasituacionales: un ejemplo con el teorema de Pitágoras* [Tesis de Doctorado, Universidad Veracruzana].
https://www.academia.edu/33834802/Tesis_doctoral_Análisis_experimental_del_desarrollo_de_competencias_intrasituacionales_un_ejemplo_con_el_teorema_de_Pitágoras

De la Sancha, E., & Pérez-Almonacid, R. (2017). Competencias y habilidades: implicaciones de su naturaleza disposicional. En J. J. Irigoyen, K. Acuña, & M. Jiménez (Eds.), *Aportes conceptuales y derivaciones tecnológicas en psicología y educación* (pp. 43-56). Qartuppi. <http://www.qartuppi.com/2017/APORTES.pdf>

Del Valle, M., García, M., Serralunga, G., Yañez, L., & García, L. (2015). Autodirección del aprendizaje percibida en estudiantes de medicina de la Universidad Nacional del Sur. *Revista Argentina de Educación Médica*, 6(3), 87-94.

- DiStefano, C., Liu, J., Jiang, N., & Shi, D. (2018). Examination of the Weighted Root Mean Square Residual: Evidence for Trustworthiness? *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 453-466. <https://doi.org/10.1080/10705511.2017.1390394>
- Dominguez-Lara, S., & Fernández-Arata, M. (2019). Autoeficacia académica en estudiantes de Psicología de una universidad de Lima. *Revista Electrónica de Investigación Educativa*, 21(e32), 1-13. <https://doi.org/10.24320/redie.2019.21.e32.2014>
- Dominguez-Lara, S., & Rodriguez, A. (2017). Índices estadísticos de modelos bifactor. *Interacciones: Revista de Avances en Psicología*, 3(2), 59-65. <https://doi.org/10.24016/2017.v3n2.51>
- Dueber, D. (2021). *BifactorIndicesCalculator: Bifactor Indices Calculator* (0.2.2). <https://cran.r-project.org/package=BifactorIndicesCalculator>
- Epskamp, S. (2015). semPlot: Unified Visualizations of Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(3), 474-483. <https://doi.org/10.1080/10705511.2014.937847>
- Espinoza, C. (2014). *Desempeño docente y calidad educativa en las facultades de Ingeniería del Perú* [Tesis de Doctorado, Universidad Nacional del Centro del Perú]. <http://hdl.handle.net/20.500.12894/1145>
- González-Fernández, M. O. (2018). Percepción del desempeño docente-estudiante en la modalidad mixta desde una mirada ecosistémica. *RIDE - Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 8(16), 370-393. <https://doi.org/10.23913/ride.v8i16.346>
- Gutiérrez, G. (Ed.). (2018). *Teorías en psicología. Integración y el futuro de la disciplina*. El Manual Moderno.
- Hendrie, K., & Bastacini, M. D. C. (2019). Autorregulación en estudiantes universitarios:

- Estrategias de aprendizaje, motivación y emociones. *Revista Educación*, 44(1), 190–207. <https://doi.org/10.15517/revedu.v44i1.37713>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Huaranga, M., & Aliaga, M. (2012). Efecto de factores económico - sociales y modalidad de ingreso en el rendimiento académico: Facultad de Economía 2005-2010. *Prospectiva Universitaria*, 9(1), 96-101. <http://200.60.61.131/index.php/prospectiva/article/view/299>
- Ibáñez, C. (1994). Pedagogía y psicología interconductual. *Revista Mexicana de Análisis de la Conducta*, 20(1), 99-113. <http://revistas.unam.mx/index.php/rmac/article/view/23451>
- Ibáñez, C. (1999). Conducta de estudio: el papel de identificar criterios en el discurso didáctico. *Acta Comportamentalia: Revista Latina de Análisis de Comportamiento*, 7(1), 47-66. <http://revistas.unam.mx/index.php/acom/article/view/18241>
- Ibáñez, C. (2007). *Metodología para la Planeación de la Educación Superior: Una aproximación desde la psicología interconductual*. Universidad de Sonora.
- Ibáñez, C. (2011). La noción de discurso didáctico en el análisis psicológico de los procesos educativos. *Acta Comportamentalia: Revista Latina de Análisis de Comportamiento*, 19(1), 125-134. <http://www.revistas.unam.mx/index.php/acom/article/view/28021>
- Ibáñez, C., & Ribes, E. (2001). Un análisis interconductual de los procesos educativos. *Revista Mexicana de Psicología*, 18(3), 359-371. https://www.academia.edu/7266287/Un_análisis_interconductual_de_los_procesos_educativos_2001_
- Irigoyen, J., & Jiménez, M. (2004). *Análisis Funcional del Comportamiento y Educación*. Universidad de Sonora.

https://www.academia.edu/4223580/Irigoyen_J_J_y_Jiménez_M_2004_Análisis_Funcional_del_Comportamiento_y_Educación_Hermosillo_Universidad_de_Sonora

Irigoyen, J., Jiménez, M., & Acuña, K. (2007). Aproximación a la pedagogía de la ciencia. En Juan Jose Irigoyen, M. Jiménez, & K. Acuña (Eds.), *Enseñanza, aprendizaje y evaluación. Una aproximación a la pedagogía de las ciencias* (pp. 13-44). Universidad de Sonora. <https://www.uv.mx/rmipe/files/2019/07/Ensenanza-aprendizaje-y-evaluacion.-Una-aproximacion-a-la-Pedagogia-de-las-Ciencias.pdf>

Irigoyen, J., Jiménez, M., & Acuña, K. (2016). Discurso didáctico e interacciones sustitutivas en la enseñanza de las ciencias. *Enseñanza e Investigación en Psicología*, 21(1), 68-77. <https://www.redalyc.org/pdf/292/29248180008.pdf>

Jöreskog, K. (2002). *Structural equation modeling with ordinal variables using LISREL*. <http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf>

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). *semTools: Useful tools for structural equation modeling. R package version 0.5-6*. <https://cran.r-project.org/package=semTools>

Kantor, J. R. (1978). *Psicología interconductual: un ejemplo de construcción científica sistemática*. Trillas.

Kantor, J. R., & Smith, N. (2015). *La ciencia de la psicología: Un estudio interconductual*. Universidad de Guadalajara.

Kelley, K. (2007). Methods for the Behavioral, Educational, and Social Sciences: An R package. *Behavior Research Methods*, 39(4), 979-984. <https://doi.org/10.3758/BF03192993>

Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R Package for Assessing Multivariate Normality. *The R Journal*, 6(2), 151. <https://doi.org/10.32614/RJ-2014-031>

- León, A., Morales, G., Silva, H., & Carpio, C. (2011). Análisis y evaluación del comportamiento docente en el nivel educativo superior. En V. Pacheco & C. Carpio (Eds.), *Observación y métricas en el análisis del comportamiento*. (pp. 84-104). FES-Iztacala UNAM.
- Leon, H. (2020). *Desempeño Docente Universitario y Aprendizaje Autónomo de los Estudiantes del Centro de Idiomas de la Universidad Nacional de Ingeniería, 2018* [Tesis de Maestría, Universidad Nacional de Educación Enrique Guzmán y Valle]. <https://repositorio.une.edu.pe/handle/20.500.14039/5964>
- Little, T. D. (1997). Mean and Covariance Structures (MACS) Analyses of Cross-Cultural Data: Practical and Theoretical Issues. *Multivariate Behavioral Research*, 32(1), 53-76. https://doi.org/10.1207/s15327906mbr3201_3
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3). <https://doi.org/10.6018/analesps.30.3.199361>
- López-Valadez, F. (2015). *Teoría de la conducta, modos de conocimiento y modos de conocer* [Tesis de Doctorado, Universidad Veracruzana]. <https://cdigital.uv.mx/handle/123456789/41149>
- Maturana, F. (2020). *Aprendizaje autorregulado y procrastinación académica en estudiantes de nivel secundario del distrito de Chorrillos* [Tesis de Licenciatura, Universidad Peruana de Ciencias Aplicadas]. <https://repositorioacademico.upc.edu.pe/handle/10757/652410>
- Mohd Razaki, N., & Bee Wah, Y. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, 2(1), 21-33.
- Montes de Oca, J. H., Bazán-Ramírez, A., & Tirado, J. (2023). Estrategias de aprendizaje y

- desempeño docente en la satisfacción académica en universitarios de Lima. *Propósitos Y Representaciones*, 11(2), e1792. <https://doi.org/10.20511/pyr2023.v11n2.1792>
- Morales, G., Alemán, M., Canales, C., Arroyo, R., & Carpio, C. (2013). Las modalidades de las interacciones didácticas: entre los disensos esperados y las precisiones necesarias. *Conductual*, 1(2), 73-89. [http://www.conductual.com/articulos/Las modalidades de las interacciones didacticas.pdf](http://www.conductual.com/articulos/Las%20modalidades%20de%20las%20interacciones%20didacticas.pdf)
- Morales, G., Chávez, E., Peña, B., & Carpio, C. (2016). Estudiar y aprender: precisiones sobre su naturaleza analítica e investigación empírica. *Revista de educación y desarrollo*, 37(2), 5-14. https://www.cucs.udg.mx/revistas/edu_desarrollo/anteriores/37/37_Morales.pdf
- Morales, G., Pacheco, V., Canales, C., Silva, H., Arroyo, R., & Carpio, C. (2010). Apuntes para la transformación de la enseñanza de la ciencia desde la perspectiva interconductual. En C. Carpio (Ed.), *Comportamiento creativo en estudiantes universitarios* (pp. 79-109). Universidad Nacional Autónoma de México.
- Morales, G., Peña, B., Hernández, A., & Carpio, C. (2017). Competencias didácticas y competencias de estudio: su integración funcional en el aprendizaje de una disciplina. *Alternativas en Psicología*, 37(1), 24-35. <https://www.alternativas.me/26-numero-37-febrero-julio-2017/143-competencias-didacticas-y-competencias-de-estudio-su-integracion-funcional-en-el-aprendizaje-de-una-disciplina>
- Morales, G., Pichardo, A., Arroyo, R., Canales, C., Silva, H., & Carpio, C. (2005). Enseñanza de la Psicología a través de la lectura: un ejemplo del abordaje experimental en la comprensión de textos. En C. Carpio & J. J. Irigoyen (Eds.), *Psicología y Educación. Aportaciones desde la Teoría de la Conducta* (pp. 127-173). Universidad Nacional Autónoma de México.
- Morales, G., Tapia, F., & Escobedo, A. (2021). La emergencia del desempeño autodidacta

- como aspiración en educación superior. *Revista de Investigación Académica Sin Frontera: División de Ciencias Económicas y Sociales*, 14(36), 1-17.
<https://doi.org/10.46589/rdiasf.vi36.418>
- Morales, M., & Chávez, J. (2019). Satisfacción académica en estudiantes de nivel superior: variables escolares y personales asociadas. *Revista Electrónica sobre Cuerpos Académicos y Grupos de Investigación*, 6(12).
<https://www.cagi.org.mx/index.php/CAGI/article/view/198>
- Moreno, A. (2023). *Práctica docente y aprendizaje autónomo en estudiantes de enfermería de una universidad pública de los Ríos, 2023* [Tesis de Maestría, Universidad César Vallejo]. <https://repositorio.ucv.edu.pe/handle/20.500.12692/120722>
- Muñiz, J. (2018). *Introducción a la Psicometría. Teoría clásica y TRI*. Ediciones Pirámide.
- Padilla, M. A. (2008). ¿Pueden entrenarse competencias de investigación en Psicología al margen de las teorías psicológicas? *Revista de Educación y Desarrollo*, 9(3), 45-53.
https://www.cucs.udg.mx/revistas/edu_desarrollo/anteriores/9/009_Padilla.pdf
- Padilla, M. A., Buenrostro, J., & Loera, V. (2009). *Análisis del entrenamiento de un nuevo científico. Implicaciones para la pedagogía de la ciencia*. Universidad de Guadalajara.
https://www.academia.edu/11322447/Análisis_del_entrenamiento_de_un_nuevo_científico_Implicaciones_para_la_pedagogía_de_la_ciencia
- Peña-Correal, T., Ordóñez, S., Fonseca, J., & Fonseca, L. C. (2012). La investigación empírica de la función sustitutiva referencial. En M. A. Padilla & R. Pérez-Almonacid (Eds.), *La función sustitutiva referencial: análisis histórico-crítico* (pp. 35-100). University Press of the South.
https://www.uv.mx/rmipe/files/2014/01/la_funcion_sustitutiva_referencial._analisis_historicocritico._avances_y_perspectivas.pdf
- Pérez-Almonacid, R. (2018). Límites de la integración teórica en psicología. En G. Gutiérrez

(Ed.), *Teorías en Psicología* (pp. 58-101). El Manual Moderno.

Pérez-Almonacid, R., & Gómez, D. (Eds.). (2014). *Emilio Ribes Iñesta. Una historia de proyectos institucionales, de identidad disciplinar e innovación educativa*. Universidad Veracruzana.

https://www.academia.edu/36141745/Una_historia_de_proyectos_institucionales_de_identidad_disciplinar_e_innovación_educativa_2014_

Pérez, I. (2018). *Percepción del desempeño docente en relación con el aprendizaje de los estudiantes de la Institución Educativa José Carlos Mariátegui del Centro Poblado Naranjillo, Nueva Cajamarca, San Martín* [Tesis de Maestría, Universidad Nacional de Educación Enrique Guzmán y Valle].

<https://repositorio.une.edu.pe/handle/20.500.14039/3909>

R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.r-project.org>

R Core Team. (2021). *foreign: Read Data Stored by «Minitab», «S», «SAS», «SPSS», «Stata», «Systat», «Weka», «dBase», R package version 0.8-81*. <https://cran.r-project.org/package=foreign>

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and Structural Coefficient Bias in Structural Equation Modeling. *Educational and Psychological Measurement*, 73(1), 5-26. <https://doi.org/10.1177/0013164412449831>

Revelle, W. (2017). *An Overview of the Psych Package*. University of California. <http://personality-project.org/r/overview.pdf>

Reyes, M., Ibáñez, C., & Mendoza, G. (2009). *Una propuesta psicopedagógica para el aprendizaje de competencias básicas en educación superior*. Universidad Autónoma de Chihuahua.

https://www.academia.edu/38097676/Una_propuesta_Psicopedagógica_para_el_Apre

ndizaje_de_Competicencias_Básicas_en_Educación_Superior_pdf

- Ribes, E. (1981). Reflexiones sobre el concepto de inteligencia y su desarrollo. *Revista Mexicana de Análisis de la Conducta*, 7(2), 107-116.
<http://www.revistas.unam.mx/index.php/rmac/article/view/25707>
- Ribes, E. (1982). *El conductismo: Reflexiones Críticas*. Fontanella.
- Ribes, E. (1989). La inteligencia como comportamiento: un análisis conceptual. *Revista Mexicana de Análisis de la Conducta*, 15, 51-67.
<http://www.revistas.unam.mx/index.php/rmac/article/view/23485>
- Ribes, E. (1990). *Psicología general*. Trillas.
- Ribes, E. (1993). La práctica de la investigación científica y la noción de juego del lenguaje. *Acta Comportamental: Revista Latina de Análisis de Comportamiento*, 1(1), 63-82.
<http://revistas.unam.mx/index.php/acom/article/view/18207>
- Ribes, E. (Ed.). (2002). *Psicología del aprendizaje*. Manual Moderno.
- Ribes, E. (2008). Educación básica, desarrollo psicológico y planeación de competencias. *Revista Mexicana de Psicología*, 25(2), 193-207.
<https://www.redalyc.org/pdf/2430/243016308001.pdf>
- Ribes, E. (2009). Reflexiones sobre la aplicación del conocimiento psicológico: ¿qué aplicar o cómo aplicar? *Revista Mexicana de Análisis de la Conducta*, 35(1), 3-17.
<https://doi.org/10.5514/rmac.v35.i1.395>
- Ribes, E. (2010). *Teoría de la conducta 2*. Trillas.
- Ribes, E. (2018). *El estudio científico de la conducta individual: Una introducción a la teoría de la psicología*. Manual Moderno.
- Ribes, E., Cortés, A., & Romero, P. (1992). Quizá el lenguaje no es un proceso o tipo especial de comportamiento: algunas reflexiones basadas en Wittgenstein. *Revista Latina de Pensamiento y Lenguaje*, 1(1), 58-73.

https://www.academia.edu/7174322/Quizá_el_lenguaje_no_es_un_proceso_o_tipo_especial_de_comportamiento_algunas_reflexiones_basadas_en_Wittgenstein_1992_

Ribes, E., & Lopez, F. (1985). *Teoría de la conducta: un análisis de campo y paramétrico*. Trillas.

Ribes, E., Moreno, R., & Padilla, M. A. (1996). Un análisis funcional de la práctica científica: extensiones de un modelo psicológico. *Acta Comportamentalia: Revista Latina de Análisis de Comportamiento*, 4(2), 205-235.
<http://revistas.unam.mx/index.php/acom/article/view/18284>

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2). <https://doi.org/10.18637/jss.v048.i02>

RStudio Team. (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC.
<http://www.rstudio.com>

Ryle, G. (2005). *El concepto de lo mental*. Paidós.

Salazar, R., & Rosas, F. (2021). Modelo de ecuación estructural explicativo del desempeño estudiantil en la asignatura de Estadística general en la UNALM. *Anales Científicos*, 82(1), 83-91. <https://doi.org/10.21704/ac.v82i1.1744>

Silva, H., Morales, G., Pacheco, V., Camacho, G., Garduño, H., & Carpio, C. (2014). Didáctica como conducta: Una propuesta para la descripción de las habilidades de enseñanza. *Revista Mexicana de Análisis de la Conducta*, 40(3), 32-46.
<https://doi.org/10.5514/rmac.v40.i3.63679>

Silva, H., Ruiz, D., Aguilar, F., Canales, C., & Guerrero, J. (2016). Enseñanza de la ciencia, la tecnología y la técnica interconductual: hacia una distinción entre prácticas psicológicas. *Revista Electrónica de Psicología Iztacala*, 19(1), 220-247.
www.revistas.unam.mx/index.php/rep/rep/article/view/54901/0

Smits, I. A. M., Timmerman, M. E., Barelds, D. P. H., & Meijer, R. R. (2015). The Dutch

- Symptom Checklist-90-Revised. *European Journal of Psychological Assessment*, 31(4), 263-271. <https://doi.org/10.1027/1015-5759/a000233>
- Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-Group Invariance with Categorical Outcomes Using Updated Guidelines: An Illustration Using M plus and the lavaan/semTools Packages. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(1), 111-130. <https://doi.org/10.1080/10705511.2019.1602776>
- Tamayo, J. (2014). La investigación en modos lingüísticos: Una breve mirada retrospectiva y un ejercicio de clasificación. En G. Mendoza, M. A. Reyes, & P. Barrera (Eds.), *Algunas aportaciones psicológicas y sociológicas a la educación* (pp. 81-107). Universidad Autónoma de Chihuahua.
- Tupayachi-Díaz, C., & Lopez-Vega, M. (2018). *Implementación de Nuevas Políticas para SUNEDU en el Proceso de Licenciamiento de Universidades Interculturales del Perú* [Tesis de Maestría, Universidad César Vallejo]. <https://repositorio.ucv.edu.pe/handle/20.500.12692/31748>
- Vattøy, K.-D., & Smith, K. (2019). Students' perceptions of teachers' feedback practice in teaching English as a foreign language. *Teaching and Teacher Education*, 85, 260-268. <https://doi.org/10.1016/j.tate.2019.06.024>
- Velarde, N. (2020). *Análisis de los criterios de desempeño didáctico en docentes y estudiantes de educación media superior como componentes de la interacción didáctica* [Tesis de doctorado no publicada]. Universidad Autónoma del Estado de Morelos
- Ventura-León, J. L., Quiroz-Burga, L., Caycho-Rodríguez, T., & Valencia, P. (2021). BifactorCalc: An Online Calculator for Ancillary Measures of Bifactor Models. *Revista Evaluar*, 21(3), 01-14. <https://doi.org/10.35670/1667-4545.v21.n3.36272>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller,

- K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H., Miller, E., & Smith, D. (2023). *haven: Import and Export «SPSS», «Stata» and «SAS» Files (2.5.3)*. <https://haven.tidyverse.org>
- Wittgenstein, L. (1999). *Investigaciones filosóficas*. Altaya.
- Wu, H., & Estabrook, R. (2016). Identification of Confirmatory Factor Analysis Models of Different Levels of Invariance for Ordered Categorical Outcomes. *Psychometrika*, 81(4), 1014-1045. <https://doi.org/10.1007/s11336-016-9506-0>
- Yamada, G., Castro, J., & Rivera, M. (2012). *Educación Superior en el Perú: retos para el Aseguramiento de la Calidad*. SINEACE. <https://repositorio.minedu.gob.pe/handle/20.500.12799/937>

IX. Anexos

Índice

- Anexo A. Cuestionario sobre desempeño docente - CDD (Versión final) y Cuestionario sobre desempeño estudiantil – CDE (Versión final)
- Anexo B. Ficha Sociodemográfica
- Anexo C. Guía para los análisis de datos
- Anexo D. Formato de consentimiento informado dirigido al docente
- Anexo E. Formato de consentimiento informado dirigido al estudiante
- Anexo F. Cuestionario de Interacciones Didácticas – CID prueba original de 2017
- Anexo G. Discusión de los resultados psicométricos
- Anexo H. Análisis comparativo alternativo
- Anexo I. Matriz de consistencia de la investigación
- Anexo J. Baremos del CDD y CDE

Anexo A. Batería: Cuestionario sobre desempeño docente – CDD y Cuestionario sobre desempeño estudiantil – CDE. Versiones finales

BATERIA DE CUESTIONARIOS SOBRE DESEMPEÑO DOCENTE Y DESEMPEÑO ESTUDIANTIL

Nombre de la asignatura a valorar en este cuestionario:

Ciclo en el cual curso la asignatura a valorar:

Código asignado:

1- CUESTIONARIO SOBRE DESEMPEÑO DOCENTE - CDD

Instrucciones: De acuerdo con tu experiencia en clase durante el semestre académico, Marca con una **X** la opción que mejor describa la frecuencia de las **conductas de tu profesor** durante las clases de esta asignatura. Debes marcar una de las cuatro opciones en cada afirmación o pregunta de acuerdo con la siguiente escala: *Nunca, Casi nunca, Casi siempre, Siempre*. Te pedimos que seas lo más sincero (a) posible.

N°	Indicadores	Nunca	Casi Nunca	Casi Siempre	Siempre
	Exploración competencial				
1	Al inicio del curso, el docente evaluó nuestros conocimientos previos, de manera escrita u oral				
2	Al inicio de clase, el docente exploró nuestras habilidades previas				
3	El docente explora mis conocimientos del curso al inicio de cada clase				
4	El docente pregunta sobre conceptos relacionados al tema antes de explicarlo				
5	El docente ajustó los contenidos del curso de acuerdo con nuestros conocimientos previos				
N°	Indicadores	Nunca	Casi Nunca	Casi Siempre	Siempre
	Explicitación de criterios				

6	El docente explica los criterios necesarios para realizar algún trabajo en clase				
7	El docente explica los criterios y requerimientos necesarios para realizar una práctica del curso				
8	El docente menciona los objetivos de aprendizaje que debemos lograr en el curso				
9	En cada clase, el docente menciona qué es lo que debemos aprender				
10	Con relación al curso, el docente menciona los criterios que un profesional de mi carrera debe cumplir.				
11	El docente menciona las habilidades necesarias que debemos desarrollar en el curso				
N°	Indicadores Ilustración	Nun ca	Casi Nun ca	Casi Siem pre	Siem pre
12	El docente explica con claridad el tema de la clase.				
13	El docente muestra ejemplos de cómo desarrollar una tarea o práctica.				
14	El docente resuelve problemas frente a los estudiantes, a partir del tema planteado				
15	El docente describe cómo un experto resolvería un problema relevante del curso				
16	Sobre una solución, el docente describe en qué consiste la solución, en que situaciones aplicarla y por qué aplicar esa solución.				
17	El docente describe las soluciones incorrectas a un problema y luego muestra la solución correcta				
N°	Indicadores Supervisión de la práctica	Nun ca	Casi Nun ca	Casi Siem pre	Siem pre
18	Durante las prácticas del curso, el docente supervisa nuestro desempeño				

19	En las actividades del curso, el docente corrige nuestro desempeño				
20	El docente colabora con nosotros en resolver un problema del curso				
21	El docente controla las reglas y condiciones de las actividades prácticas, para garantizar que aprendamos efectivamente				
22	Todos mis compañeros reciben la supervisión del docente durante las actividades del curso				
N°	Indicadores Retroalimentación	Nunca	Casi Nunca	Casi Siempre	Siempre
23	El docente corrige nuestro desempeño en las actividades de clase, señalando los aciertos y errores				
24	El docente retroalimenta lo que hice bien o mal en una actividad y me enseña las formas en las que lo puedo resolver				
25	El docente enseña distintas formas en las que se puede satisfacer los criterios de logro de las actividades de su clase				
26	Después de la explicación de la clase y la práctica, el docente nos brinda sugerencias para mejorar nuestro desempeño				
27	Sobre una solución, el docente describe el procedimiento que seguimos y muestra cómo podemos mejorar.				
28	El docente nos brinda sugerencias sobre cómo mejorar nuestras habilidades, en relación con el curso.				
N°	Indicadores Evaluación	Nunca	Casi Nunca	Casi Siempre	Siempre
29	El docente revisa los ejercicios y tareas, señalando lo que se debe corregir o mejorar en mi trabajo				

30	El docente evalúa constantemente mis conocimientos y habilidades con relación al curso				
31	El docente realiza evaluaciones de carácter aplicado y la solución de problemas prácticos, derivados de la asignatura				
32	El docente evalúa a los estudiantes según los objetivos de aprendizaje que mostro al inicio del curso, o que se encuentra en el silabo				
33	El sistema de evaluación del docente es pertinente para medir nuestros conocimientos y habilidades logrados en el curso				

2- CUESTIONARIO SOBRE DESEMPEÑO ESTUDIANTIL - CDE

Instrucciones: De acuerdo con tu experiencia en clase durante el semestre académico, Marca con una **X** la opción que mejor describa la frecuencia de las **conductas que tú y tus compañeros realizaron** durante las clases de esta asignatura. Debes marcar una de las cuatro opciones en cada afirmación o pregunta de acuerdo con la siguiente escala: *Nunca, Casi nunca, Casi siempre, Siempre*. Te pedimos que seas lo más sincero (a) posible.

N°	Indicadores	Nunca	Casi Nunca	Casi Siempre	Siempre
	Precurrentes y potenciales habilidades				
1	Respondo las preguntas que el docente hace acerca de un tema nuevo.				
2	Formulo preguntas sobre un nuevo tema				
3	Menciono mis conocimientos previos sobre algún nuevo tema de clase				
4	Participó de las evaluaciones previas que realiza el docente				
5	Demuestro en la práctica mis competencias y habilidades previas, sobre algún nuevo tema				
N°	Indicadores	Nunca	Casi Nunca	Casi Siempre	Siempre
	Identificación de criterios				
6	Alcanzo satisfactoriamente los criterios que el docente indica para realizar algún trabajo en clase				
7	Con relación al curso, sé cuáles son los criterios que un profesional de mi carrera debe alcanzar				
8	Conozco qué es lo que debo aprender en este curso.				
9	Conozco las habilidades que debo desarrollar en el curso.				
10	Cumplo con los criterios para desarrollar correctamente una tarea o práctica.				

N°	Indicadores Participación	Nunca	Casi Nunca	Casi Siempre	Siempre
11	Comprendo con claridad el tema que el docente explica en la clase				
12	Desarrollo una tarea o práctica según los ejemplos que mostro el docente				
13	Resuelvo los problemas que el docente plantea para la clase				
14	A la hora de resolver un problema del curso, lo hago lo más parecido a como lo hace un experto				
15	Identifico en qué consiste la solución que explicó el docente				
16	Identifico las soluciones incorrectas a un problema				
N°	Indicadores Práctica pertinente	Nunca	Casi Nunca	Casi Siempre	Siempre
17	Identifico en que situaciones aplicar una solución y porque seleccionarla, de acuerdo a las clases del docente				
18	Recibo y aplico las correcciones que el docente me brinda durante las actividades del curso				
19	Participo, en conjunto con el docente, en la resolución de algún problema del curso				
20	Participo de manera efectiva en las actividades que son similares a los problemas que debo resolver como profesional				
21	Participo en las actividades de aprendizaje controladas por el docente				
22	Busco la supervisión del docente respecto a las actividades del curso				

N°	Indicadores Mejoramiento	Nunca	Casi Nunca	Casi Siempre	Siempre
23	Identifico en mi desempeño durante la clase, si he realizado correcta o incorrectamente las prácticas				
24	Cambio mis acciones para cumplir con los criterios de la clase o tarea, según las sugerencias del docente				
25	Recibo y aplico la retroalimentación que el docente me brinda, después de la práctica o actividad de clase				
26	Aplico una forma diferente de solución al problema o actividad de clase, de acuerdo con las observaciones del docente.				
27	Aplico las sugerencias del docente para mejorar mis habilidades en relación con el curso				
N°	Indicadores Desempeño resultante	Nunca	Casi Nunca	Casi Siempre	Siempre
28	Conozco las observaciones que me brinda el docente sobre mis ejercicios y tareas				
29	Mis conocimientos y habilidades son evaluados por el docente				
30	Resuelvo satisfactoriamente los problemas y/o actividades que forman parte de la evaluación				
31	He desarrollado los conocimientos esperados del curso				
32	Conozco los objetivos de aprendizaje del curso				
33	Entiendo el sistema de calificación con el cual el docente nos evalúa				

Anexo B. Ficha Sociodemográfica



FICHA DE DATOS SOCIODEMOGRÁFICOS

Estimado(a) estudiante, por favor llenar los siguientes datos previos a esta investigación:

Datos personales

Sexo: M / F	Edad:
Distrito de residencia:	

Datos de carrera

Ciclo de estudios actual:	Salón:
Turno: mañana / tarde	Especialidad:
Año de ingreso:	Alumno Regular: Si / No

Marcar con una X en una de las respuestas según corresponda: **¿Por qué estudias la carrera?**

<input type="checkbox"/>	Ayudar a los demás	<input type="checkbox"/>	Facilidad de la carrera
<input type="checkbox"/>	Comprender el comportamiento humano	<input type="checkbox"/>	Ayudarse a sí mismo
<input type="checkbox"/>	Otros _____		

Valoración del curso

Curso por evaluar:

Del 1 al 10, siendo 10 la máxima puntuación y 1 la mínima puntuación ¿cuánto crees que sabe tu docente sobre el curso a evaluar?

1	2	3	4	5	6	7	8	9	10

Del 1 al 10, siendo 10 la máxima puntuación y 1 la mínima puntuación ¿cuánta experiencia tiene tu docente sobre el curso en cuestión?

1	2	3	4	5	6	7	8	9	10

Datos laborales Vigentes

Estoy en un Voluntariado: (Si) / (No)	Realizo practicas preprofesionales: (Si) / (No)	Estoy en un trabajo no relacionado a la carrera: (Si) / (No)
--	--	---

Anexo C. Guía para los análisis de datos

En este apartado se brindarán precisiones y actualizaciones sobre los conceptos y procedimientos de análisis psicométrico y metodológicos que se desarrolló en la presente tesis, asumiendo los estándares para pruebas educativas y psicológicas (AERA et al., 2018), los lineamientos del manual de la APA (American Psychological Association, 2021) y los actuales lineamientos de publicación de artículos en revistas indexadas. Agradezco a Pedro Arroyo por la revisión de este apartado.

1 Fiabilidad

Representa una propiedad psicométrica descrita en la precisión, estabilidad o consistencia interna de una o más mediciones de una variable. Los métodos para la recolección de evidencia en la fiabilidad se sitúan en los siguientes procesos: formas paralelas, test-retest, dos mitades y consistencia interna, siendo este último el más usado y se define como el grado de relación entre los ítems y la puntuación general de la prueba (Muñiz, 2018), por lo tanto, el coeficiente alfa (α) ha sido el indicador más usado para evaluar la consistencia interna en ciencias sociales. En tal sentido, el coeficiente α representa tres supuestos: (a) tau equivalencia, (b) naturaleza continua de las variables, y (c) unidimensionalidad de los ítems; sin embargo, Oliden & Zumbo (2008) han alertado la impertinencia de la utilización del α para estimar la consistencia interna en cuestionarios con escalamiento ordinal, es decir, menores a cinco categorías de respuestas (Brown, 2015), produciendo una sobreestimación de los valores.

Frente a las circunstancias metodológicas de la utilización del coeficiente α , se enfatiza la naturaleza de medición de la variable como mecanismo para seleccionar el coeficiente “alfa ordinal” ya que los estudios de comprobación y simulación han consolidado su eficiencia frente a las deficiencias aplicativas del α en el análisis de los instrumentos de medición (Oliden & Zumbo, 2008; Zumbo et al., 2007), de esta manera, se justificó su empleo en la escala de cuatro

categorías de respuesta para evaluar la consistencia interna de las seis dimensiones de cada variable general.

En tal sentido, la evaluación de la fiabilidad de las variables generales que contienen dimensiones denominadas “estructuras multidimensionales” constituye una tendencia errada en la comunidad de psicólogos investigadores de aplicar el coeficiente alfa al test completo, sin embargo, la formulación del α se basa en el supuesto de que la escala o test tiene una estructura unidimensional (Muñiz, 2018; Rodríguez et al., 2016). Por ende, la utilización del alfa en una escala multidimensional representa una estimación sesgada de la fiabilidad en dos casos: 1) Sobreestimar la fiabilidad cuando los ítems están fuertemente correlacionados entre sí pero miden diferentes constructos; y, 2) subestimar la fiabilidad cuando los ítems están débilmente correlacionados pero miden el mismo constructo (Abad et al., 2011; Zinbarg et al., 2007). Por consiguiente, se han propuesto diferentes coeficientes alternativos de consistencia interna como el coeficiente omega total de McDonald; el coeficiente de glb de Sijtsma; el coeficiente H de Bentler, etc. (Cho, 2016).

Por lo tanto, en el presente estudio, se decide emplear el coeficiente omega jerárquico (ω_H) cuya estimación se instaura en la formulación del análisis factorial y cuando la proporción de varianza de los reactivos no se atribuyen a un factor general. De este modo, el ω_H no asume la equiparabilidad de las cargas factoriales de los ítems y redistribuye las cargas con base a la presencia de dimensiones (Kalkbrenner, 2023; Rodríguez et al., 2016; Viladrich et al., 2017).

El criterio interpretativo del coeficiente ω_H se ubica entre un rango de cociente de proporción de 0.00 a 1.00, delimitando que mayores cociente de probabilidad es el indicador de mayor consistencia interna o fiabilidad. Al tratarse de probabilidad, su interpretación puede situarse denotando la proporción de varianza total de los ítems atribuida a un factor general. Por ejemplo, un cociente de 0.80 representa que los reactivos explican el 80% de la variabilidad atribuida al factor general, y 20% explica las variaciones de las puntuaciones de los factores o

componentes específicos y el error de medición (Abad et al., 2011; Rodríguez et al., 2016). Por dicha razón, se propone un punto de corte a partir de 0.80 como un coeficiente “aceptable” (Dueber, 2017).

2 Validez y sus fuentes de evidencia

La validez representa el grado en que una teoría se corresponde con los datos disponibles para brindar soporte interpretativo a las puntuaciones de un cuestionario en una muestra específica (AERA et al., 2018). A pesar de la presencia de un consenso disciplinar en la evaluación de las fuentes de evidencia son difundidas en los *Estándares para pruebas educativas y psicológicas*, persiste una resistencia en su adherencia en el contexto nacional cimentada en la configuración de la validez como propiedades de la puntuación y no del constructo, conduciendo a diversas interpretaciones que serán analizadas en un proceso de recopilación de evidencias para respaldar la interpretación de las puntuaciones (Abad et al., 2011). Por lo tanto, no existe un tipo de validez, si no conjunto de evidencias de validez circunscritas en cinco clases:

- Evidencias basadas en el contenido del test: Se obtiene del análisis lógico y empírico de la relación entre el contenido de una prueba y el constructo que se intenta medir. Los criterios que componen el contenido de la prueba o cuestionario que son susceptibles de evaluación son: la temática a evaluar; la redacción; el formato de los ítems, tareas o preguntas; la administración o instrucciones; y el sistema de puntaje (Abad et al., 2011). Para el presente trabajo se aplicó el siguiente procedimiento: primero se analizó la representatividad de los constructos (desempeños y ámbitos didácticos) en los ítems de la prueba original y todos sus aspectos (escala de respuesta, administración, etc.); segundo, se agregó, modificó y retiró los ítems de acuerdo a su representatividad lógica de los constructos; tercero, seis expertos en metodología de la investigación y psicología educativa evaluaron la primera versión de la adaptación peruana, recibiendo sus sugerencias para agregar, modificar o

retirar ítems; finalmente, se construyó la segunda versión de la adaptación del cuestionario, la que fue aplicada.

- Evidencias basadas en la estructura interna: Es el grado en que las relaciones entre los ítems y los componentes del test se ajustan al constructo en el que se basan las interpretaciones propuestas de las puntuaciones del test (AERA et al., 2018). Las evidencias de este tipo se pueden obtener en tres formas: dimensionalidad, invarianza de la medición (funcionamiento diferencial de los ítems) y fiabilidad (AERA et al., 2018; Rios & Wells, 2014). En tal sentido, la comprobación de la dimensionalidad representa el contraste de las relaciones entre los ítems con la estructura hipotética proporcionada por un marco referencial en una propuesta teórica o fundamentación estadística (Abad et al., 2011), a modo de ejemplo, las cargas factoriales de un cuestionario con estructura multidimensional deben ser más altas en su dimensión y más bajas en otras dimensiones. De esta manera, el método estadístico predilecto para indagar la estructura interna de un cuestionario es el análisis factorial basado en matrices de correlaciones o covarianzas entre los ítems, el cual tiene dos perspectivas (Ferrando et al., 2022): 1) El análisis factorial exploratorio o no restringido se aplica cuando no hay una estructura hipotética o preestablecida de las dimensiones de los ítems, por lo que se pretende indagar sobre esta estructura a partir de los datos empíricos. 2) Cuando ya se cuenta con una estructura de dimensiones preestablecida, se aplica el análisis factorial confirmatorio o de restricción progresiva. En la presente tesis, se aplicó el análisis factorial confirmatorio, tanto a la muestra total y a muestras subdivididas, para verificar la invarianza del modelo del medición. Ambas etapas de análisis son explicadas a continuación:

3 *Análisis factorial confirmatorio*

Propósito de la técnica

Con esta técnica se pretende estudiar las dimensiones de un grupo de ítems, considerando previamente modelos de medición predefinidos por los autores de la prueba y sustentados en evidencias psicométricas previas. De acuerdo con Abad et al. (2011), lo que se plantea de manera predeterminada sobre el modelo de medición es: 1) El número de dimensiones; 2) la existencia de correlación entre las dimensiones; 3) las cargas factoriales de los ítems en las dimensiones, 4) los residuos y sus posibles correlaciones.

Características de la técnica

A continuación, se detallan algunas características resaltantes de esta técnica:

- El Análisis factorial confirmatorio (AFC) se basa en el modelado de ecuaciones estructurales.
- El AFC es una matriz de correlaciones o covarianzas entre todos los ítems del cuestionario, definida como la matriz de datos del modelo real (de la muestra).
- De la matriz de datos reales se descomponen la matriz del modelo teórico y los residuos, donde el mejor modelo de medición será el que tenga mínimos residuos y una matriz del modelo teórico más similar a la matriz de datos reales.
- El modelo de medición es propuesto por el marco de referencia desarrollado por la comunidad científica y contrastada directamente con los datos.
- Las diferencias con el Análisis factorial exploratorio (AFE) es: 1) No se calcula un peso factorial del ítem en todos los factores, sino en los que fue prefijado; 2) los errores se pueden correlacionar; y 3) se pueden aplicar restricciones en los pesos factoriales.
- El AFC representa un proceso iterativo de ida y vuelta (Furr & Bacharach, 2014), porque se elabora y evalúa un modelo de medición de la variable, pero a menudo el

primer modelo evaluado no se acepta, sino que le revisa y se aplican modificaciones conformando un “modelo revisado” al cual se le reevalúa.

- Los procedimientos para realizar el AFC se pueden agrupar en seis fases: especificación del modelo, identificación, estimación de parámetros, ajustes del modelo, interpretación y reespecificación.
- Previamente a iniciar los pasos para el AFC. se requiere hacer una evaluación de los ítems y comprobación de supuestos estadísticos, los cuales predeterminaran la configuración del AFC.

Evaluación de los ítems

Es necesario verificar los estadísticos descriptivos univariados y multivariados de los ítems, para comprobar si son adecuados para tratamiento de los datos y si cumplen supuestos estadísticos. De acuerdo con esta evaluación, se predeterminará una configuración para aplicar el AFC:

Estadísticos descriptivos de los ítems

Es la evaluación de la media, desviación estándar, curtosis y asimetría de los ítems. Con la media se busca identificar ítems con “efecto suelo” o “efecto techo”. Con la desviación estándar se evalúa la variabilidad de los ítems. En cuanto a la asimetría y curtosis, generalmente se sugiere el punto de corte de 1 o -1 como límite a no sobrepasar (Ferrando et al., 2022). Si el ítem no obtiene buenas propiedades en estas medidas de tendencia central y de dispersión, se sugiere eliminarlo.

Comprobación de supuestos

Es importante aclarar que el no cumplimiento de estos supuestos no implica la imposibilidad de aplicar el AFC sino que se debe tomar en cuenta otras configuraciones, de acuerdo con las características de los datos (Ferrando et al., 2022).

Como primera comprobación, se debe evaluar si el modelo de análisis factorial es de relación lineal entre los factores y los ítems. De acuerdo con el modelo elegido, Ferrando et al. (2022) detallan que para el modelo lineal: las variables son continuas; las regresiones entre los ítems y el factor son lineales; y en la matriz de correlaciones se utiliza el coeficiente de Pearson. En contraste, en el modelo no lineal, las variables son discretas, las regresiones ítem-factor no son lineales y la matriz de correlaciones sigue una matriz policórica o tetracórica si la escala de respuesta es dicotómica.

Además, se debe comprobar la normalidad univariada y multivariada. En la normalidad univariada de los ítems se comprueba su ajuste estadístico a la distribución normal estándar mediante la prueba de contraste de hipótesis Shapiro Wilk debido a su robustez estadística a pesar de las variaciones en el tamaño de muestra (Mohd Razaki & Bee Wah, 2011). De esta manera, si el p valor de la prueba Shapiro-Wilk es mayor a .05 indica que es probable la existencia de una distribución normal univariada, por el contrario, si el p valor es menor a .05, indica que probablemente no existe evidencia de una distribución normal en las puntuaciones.

Por otro lado, la estimación de la normalidad multivariada se sitúa en torno al coeficiente de Mardia que describe el patrón de asimetría y curtosis multivariada de las puntuaciones de una muestra. En ese sentido, la asimetría multivariada describe la ausencia de simetría de la distribución de los datos en todas las dimensiones en términos de la proporción de datos tendiente a lo positivo o negativo, mientras que la curtosis multivariada evalúa la forma de la distribución según la anchura y altura del histograma (Gana & Broc, 2019). La prueba arroja dos p valores, de la asimetría y la curtosis multivariada, donde si el p valor es menor a .05 se asumirá que probablemente no existe una distribución multivariada. Además, Si el p valor es superior a .05 en la asimetría y curtosis, se indica que hay evidencia probable a favor de la normalidad multivariada.

Finalmente, Brown (2015) propone que la verificación de la escala de respuesta del ítem, donde se recomienda una escala igual a seis o más categorías de respuesta para considerarlo como una variable continua, por el contrario, si la escala tiene 5 o menos categorías de respuesta, se le considera ordinal. El número de respuestas es importante porque predeterminara las configuraciones a realizar en el AFC y considerando que en las investigaciones en ciencias sociales y la salud se suelen utilizar pocas categorías de respuesta. A modo de ejemplo, Abad et al. (2011) sugiere la siguiente configuración en el caso de una encuesta que tiene ítems con 5 categorías o más, tamaño de muestra grande y cumple con el supuesto de normalidad multivariada, se puede elegir rápidamente el estimador Máxima verosimilitud (ML).

Fases del análisis factorial confirmatorio

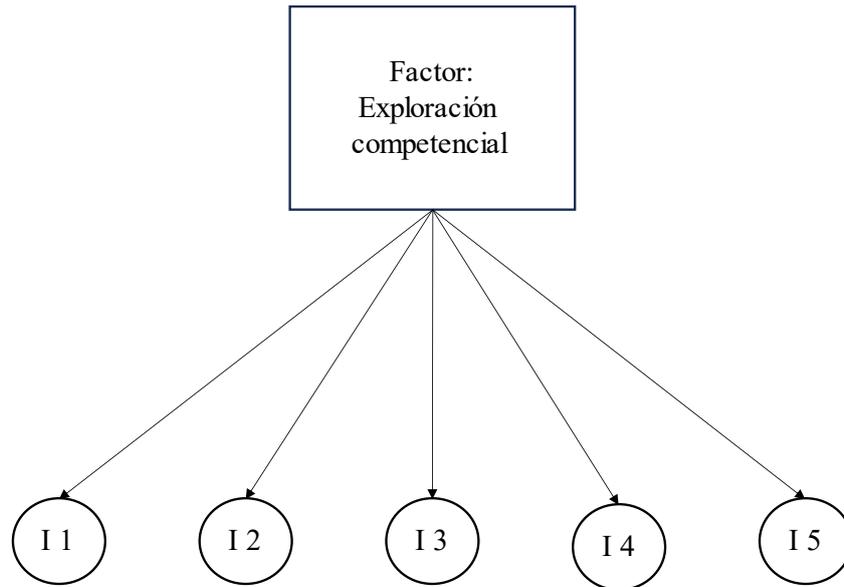
Generalmente se puede considerar cinco grandes pasos en la ejecución de un análisis factorial confirmatorio, los cuales se explican a continuación, enfatizando en los conceptos relevantes para entender los resultados de la presente tesis.

1) Especificación

El modelo de medición hipotetizado se representa gráficamente con un diagrama de senderos, donde los indicadores son representados con cuadrados y las variables latentes con círculos. Las relaciones o vinculaciones se representan con líneas unidireccionales y las covariaciones con curvas bidireccionales. La ausencia de curvas bidireccionales o unidireccionales entre los factores implica que son factores independientes (Abad et al., 2011).

Figura C.1

Ejemplo de diagrama de senderos con un factor



El diagrama de senderos del modelo de medición hipotetizado es implementado en el programa informático de análisis de datos. Las opciones de configuración varían de acuerdo con el programa, pero generalmente abarcan los siguientes pasos:

- 1- Especificar el número de dimensiones o factores.
- 2- Seleccionar los ítems que conformaran las dimensiones o factores.
- 3- Si se cuenta con dos o más dimensiones, se debe configurar las posibles asociaciones entre factores.

Adicionalmente, se pueden restringir más configuraciones, como igualar parámetros o establecer valores exactos, por ejemplo, a las cargas factoriales (Furr & Bacharach, 2014).

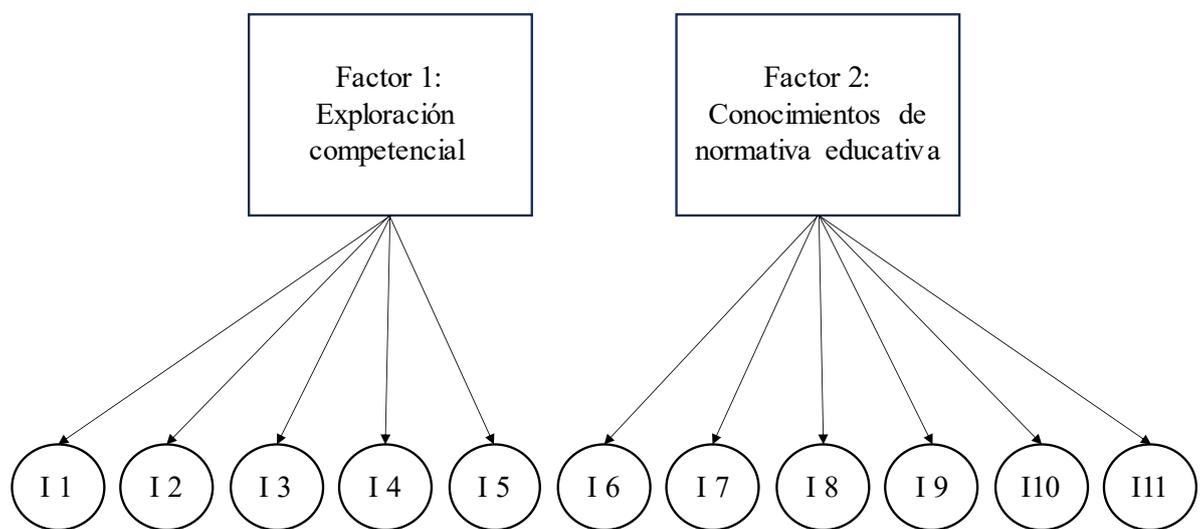
En el primer paso, se debe evaluar el número de factores, en consecuencia, si existe solo un factor o dimensión a evaluar, el modelo de medición se denomina “unidimensional”. Por otro lado, si cuenta con dos dimensiones a más, que pueden estar organizados en diferentes jerarquías, se le denomina “multidimensional” como modelo de medición. A continuación, se exponen los tipos de modelos multidimensionales, de acuerdo con Canivez (2016):

1.1) Modelo de factores no correlacionados

Se denominan “modelo de factores independientes” se basan en la premisa de que los ítems agrupados en dimensiones no presentan correlación con otras dimensiones, siendo un punto de corte de 0.33 para establecer la presencia o ausencia de relación (Tabachnick & Fidell, 2007). Por consiguiente, la ausencia de relación refleja cierto grado de independencia entre las dimensiones (Figura C.2).

Figura C.2

Ejemplo de diagrama de senderos con dos factores no correlacionados

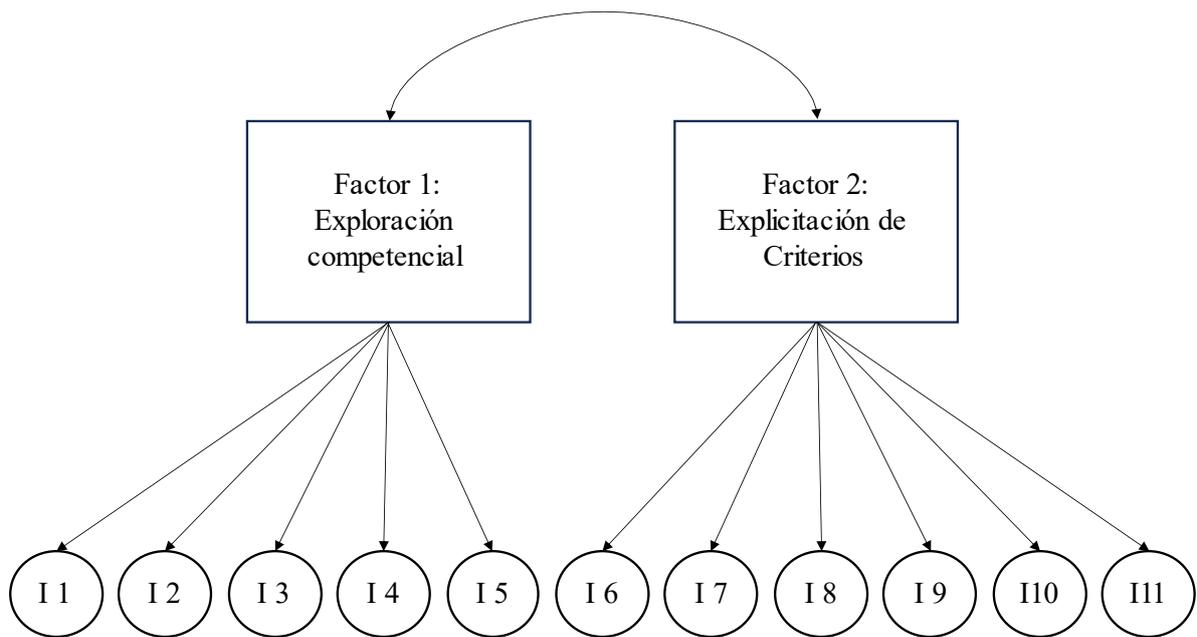


1.2) Modelo de factores correlacionados

Se denominan “modelos oblicuos” ya que contienen factores correlacionados y no existe una variable latente jerárquica que los agrupe. De este modo, en la figura C.3, se describe un diagrama de senderos basado en la covariación de los factores concatenados en curvas bidireccionales. Esencialmente la principal ventaja del modelo subyace a que correlaciones altas dificultan la capacidad discriminativa entre los factores, por lo cual se podría asumir la presencia de un factor superior no especificado en el análisis (Canivez, 2016).

Figura C.3

Ejemplo de diagrama de senderos con dos factores correlacionados

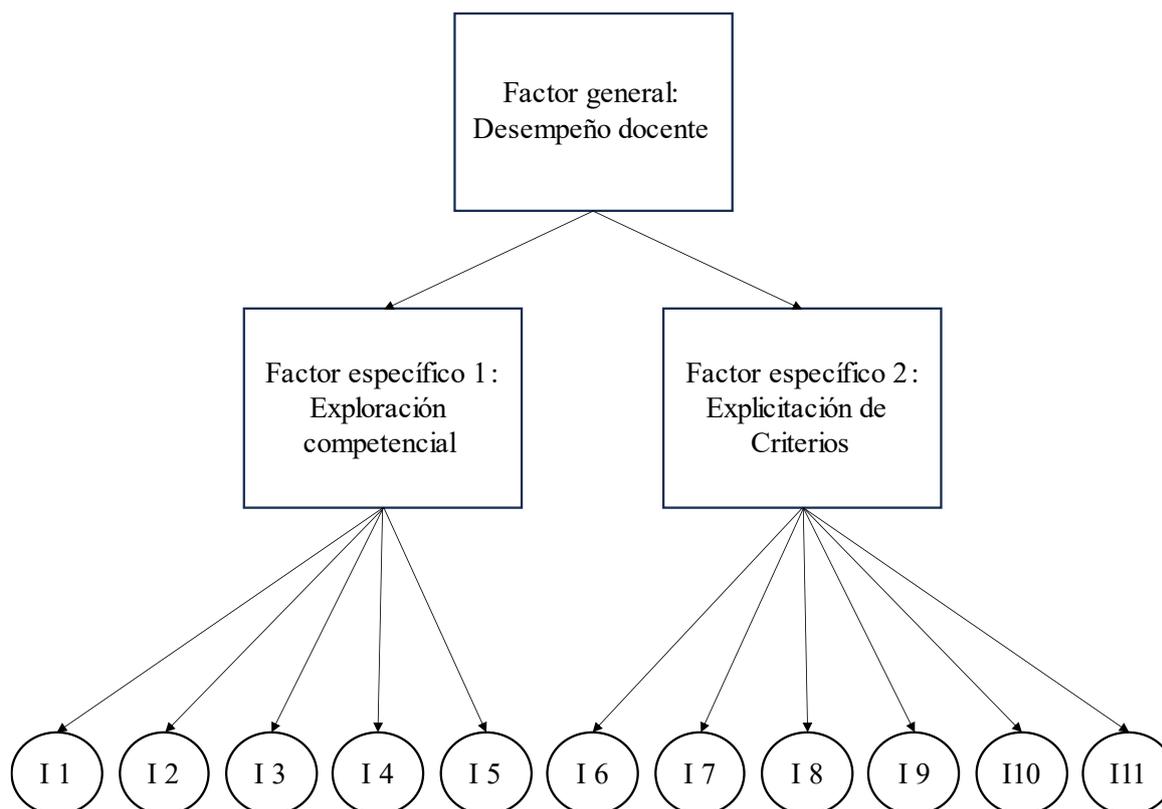


1.3) Modelo Jerárquico Indirecto

Secuencialmente, la presencia de factores implicados en otros factores converge en la premisa de una variable latente superior, es decir, un factor general de segundo orden que dirige el patrón de dispersión de los factores de primer orden denominados específicos que agrupan los indicadores o ítems (Canivez, 2016), como se puede ilustrar en la figura C.4.

Figura C.4

Ejemplo de diagrama de senderos con un factor general y dos factores específicos, con jerarquía indirecta



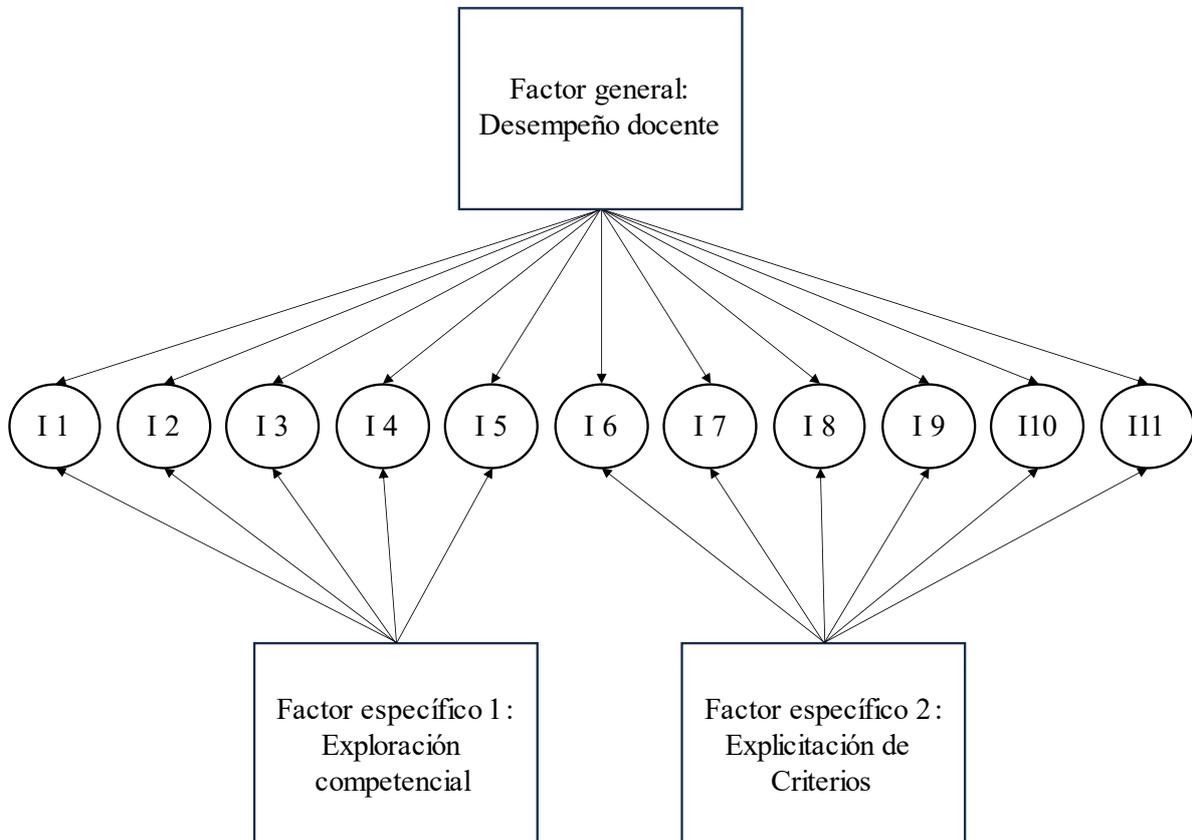
En el modelo jerárquico indirecto, el factor de segundo orden se relaciona de manera indirecta mediado por la presencia de factores específicos, quienes regresionan con los ítems o reactivos. Por esta razón, McDonald (1999) la nombro como “modelo jerárquico indirecto”. Siendo su principal desventaja la dificultad para determinar la influencia o carga factorial del factor de segundo orden en los ítems o indicadores observables, y si estas deben estar mediadas siempre por el factor de primer orden (Canivez, 2016).

1.4) Modelo Jerárquico directo

Denominado como modelo “bifactor” ya que el factor de segundo orden se regresiona con las variables observables y sus cargas estandarizadas no son mediadas por factores específicos (Canivez, 2016). Por lo tanto, en la figura 13, los reactivos o variables observables se establecen directamente en los factores específicos y factor general, denotando la influencia simultánea y representando menor varianza de error en contraste con modelos jerárquicos indirectos.

Figura C.5

Ejemplo de diagrama de senderos con un factor general y dos factores específicos, con jerarquía directa



Las ventajas del modelo bifactor son, de acuerdo con Canivez (2016):

- Su evaluación simultánea de cargas factoriales de los reactivos y factores específicos.
- La estimación de los índices de ajustes relativos y absolutos pueden discriminar la presencia de factores específicos y generales.
- Los criterios prácticos para la evaluación mediante instrumentos permiten obtener dimensiones de análisis de las unidades de comportamiento humano, y obtener un patrón generalizado de las puntuaciones.

En el AFC, el tratamiento de instrumentos multidimensionales suele basarse en un la comparación de distintas soluciones factoriales para analizar la eficiencia de los modelos

teóricos con base al contraste de modelos empíricos según sus índices de ajuste relativos y absolutos. No obstante, la evaluación de modelos bifactoriales implica un espectro de evaluaciones adicionales con la premisa de comprobar los datos en modelos jerárquicos directos.

2) Identificación

En el proceso de identificación, se establece que identificar el modelo AFC implica la estimación de parámetros para detectar el modelo propuesto a través de una solución factorial (Abad et al., 2011). Los aspectos por considerar para la identificación del modelo de AFC son: 1) Fijar la métrica de las variables; y 2) Comprobar si el número de grados de libertad es positivo.

3) Estimación

En ese sentido, la estimación de parámetros del modelo posibilita la reducción de la diferencia entre la matriz de covarianzas observada y la matriz de covarianza reproductiva en el modelo factorial propuesto (Abad et al., 2011). Por otro lado, se han establecido métodos matemáticos para estimar parámetros de acuerdo a determinadas circunstancias del contexto, por lo tanto, se han disgregado en una serie de estimadores.

Máxima Verosimilitud (ML: Maximum Likelihood)

Representa el método de estimación más usual en los paquetes informáticos de modelamiento de ecuaciones estructurales, se cimenta en la detección de factores mediadores y parámetros entre los errores para garantizar la máxima verosimilitud entre las respuestas hipotetizadas y empíricas de la muestra analizada (Brown, 2015). Se presentan una serie de condiciones para su aplicación como la normalidad univariada y multivariada de las puntuaciones, relación lineal entre las variables de estudio segmentadas en factores específicos, y tamaño muestral grande ($n > 300$).

Aunque exista omisiones en el cumplimiento de los supuestos previamente descritos, puede emplearse la Máxima Verosimilitud Robusto (MLR: Maximum Likelihood with Robust Standard Errors), en consecuencia, se analiza los errores estandarizados robustos para su aproximación a los errores de los parámetros en consideración de la heterogeneidad de las varianzas de los reactivos (Kline, 2016).

Mínimos Cuadrados no Ponderados (ULS: Unweighted Least Squares)

El método de mínimos cuadrados no ponderados se basa en la similitud estadística de las respuestas observadas para reducir las discrepancia cuadradas entre las covarianzas reales y covarianzas estimadas del modelo factorial (Brown, 2015).

Entre sus condiciones de uso, ULS es adecuado en muestras de tamaño moderado a grande ($n > 200$ sujetos), tolera las violaciones al supuesto de normalidad derivados de valores atípicos o faltantes. Asimismo, tiende a producir menores índices de ajuste en matrices de varianzas-covarianzas (Abad et al., 2011). Por lo tanto, se recomienda la estandarización de las variables antes de aplicar ULS.

Por otro lado, la obtención de estimadores no lineales se han utilizado para reducir las discrepancias entre los modelos teóricos y empíricos según la prueba, variables latentes, muestra y aspectos relacionados. Por esa razón, los más usados se han establecido en torno a: Mínimos cuadrados generalizados (GLS); y Mínimos cuadrados ponderados (WLS). En los últimos años ha destacado el estimador WLSMV adquirió mayor versatilidad de uso frente a estimadores complementarios. En consecuencia, se establecerá las ventajas y desventajas concernientes al análisis psicométrico en el presente estudio.

Mínimos Cuadrados Ponderados con media y varianza ajustada (WLSMV: Weighted Least Squares Mean and Variance adjusted)

Brown (2015) recomienda el uso del estimador WLSMV cuando los ítems son categóricos, es decir, que tienen una escala de respuesta, dicotómica, politómica u ordinal.

Aunque existen otros estimadores también recomendados para datos categóricos, tienen algunas desventajas que, según lo investigado actualmente, WLSMV puede equiparar y obtener mayor eficiencia. El WLSMV o WLS robusto, se basa en la matriz de covarianza policórica, donde se estima una matriz de covarianza ajustada que tiene en cuenta la estructura de relación ordinal y las relaciones de varianza entre las categorías de las variables observadas. WLSMV minimiza las diferencias entre la matriz de covarianza ajustada y la matriz de covarianza observada, ponderando las discrepancias en función de la información de orden y varianza. Además, el WLSMV también ajusta las medias de las variables observadas (Brown, 2015).

En estudios psicométricos, se halló que una muestra entre 150 y 200 puede ser lo suficientemente útil para modelos factoriales medianos. Asimismo, los estadísticos de resultado han sido precisos en varias situaciones y sus configuraciones del A.F.C. (Flora & Curran, 2004).

4) Evaluación

La evaluación de las propiedades de un modelo se realiza mediante la contrastación del ajuste del modelo a los datos, donde se quiere que los residuos sean lo mínimo posibles (Abad et al., 2011). Para lograr este objetivo, se han diseñado formulas estadísticas para medir estas diferencias, denominadas “Índices de ajuste”, los cuales, de acuerdo a su objetivo de medición, se pueden clasificar en dos:

1. Los índices de ajuste absolutos: Establece el grado en que el modelo factorial preestablecido reproduce los datos.
2. Los índices de ajuste comparativo: Evalúa el grado en que el modelo factorial es mejor que otros modelos factoriales suplentes.

Según esta clasificación, se pasará a explicar los índices de ajuste usados en esta investigación:

Índices de ajuste inferenciales

Asumen una distribución estadística y evalúan con base a inferencias de lo que ocurre en la población, con el presupuesto de que los residuos son cero en la población (Abad et al., 2011). A continuación, se exponen los índices inferenciales de ajuste absoluto y ajuste comparativo:

- 1) Estadístico X^2 : Se obtiene del estimador ML. Se busca contrastar la hipótesis nula de que la discrepancia o diferencia entre el modelo reproducido y modelo real es cero en la población. Su principal desventaja es que, al usar el test de significancia estadística, es muy sensible al tamaño de muestra, produciendo errores de sobreestimaciones y subestimaciones.
- 2) Estadístico χ^2 en modelos anidados: Se usa para comparar modelos que pueden contenerse dentro del modelo propuesto, es decir, modelos anidados al modelo confirmatorio principal. La diferencia de χ^2 se distribuye con los grados de libertad, donde se puede obtener una diferencia a favor del mejor modelo.

Índices de ajuste descriptivos

No asumen una distribución estadística conocida y solo se hacen evaluaciones de lo que ocurre en la muestra. Existe una gran cantidad de índices de ajuste descriptivos, pero que también se han desestimado su uso debido a los defectos de medición que pueden producir (Abad et al., 2011). Para la presente investigación se han seleccionado 4 índices de ajuste idóneos a las características de la investigación y que tienen menos defectos de medición en comparación al resto:

- 1) X^2/gl : Es la división del valor de chi cuadrado del modelo factorial por sus grados de libertad. El criterio de corte es que $X^2/gl < 3$.

- 2) *RMR*, *SRMR* y *WRMR*: La estrategia de estos índices es calcular los residuos de la matriz de varianzas-covarianzas. En primer lugar, surgió el índice “Raíz del promedio de los residuos al cuadrado” (*RMR*), pero al no poder interpretarse directamente, Hu y Bentler (1999) propusieron una versión estandarizada llamada *SRMR*, donde un valor menor a 0.8 indicaría buen ajuste a los datos (Abad et al., 2011). Sin embargo, *SRMR* se creó para muestras con residuos que siguen una distribución de normal y variables continuas, por lo que es un índice muy exigente para variables ordinales y con residuos que no tienen una distribución normal.

Para datos categóricos basados en el estimador *WLSMV*, se recomienda el uso del índice de ajuste “Residuo cuadrático medio ponderado” (*Weighted Root Mean Square Residual: WRMR*) que es una versión ponderada de *RMR*, que tiene en cuenta la varianza y covarianza de los residuos. De este modo, más reducidos se encuentren los residuos, mejor funcionará el modelo para reproducir la matriz de covarianza propuesto. De acuerdo a la investigación de diseño simulado de DiStefano et al. (2018), se ha demostrado eficiencia del coeficiente *WRMR* para modelos con datos categóricos y residuos no ajustados a la normalidad. El punto de corte recomendado es una cifra menor o igual a uno ($WRMR \leq 1$).

- 3) *RMSEA*: Es la raíz del error cuadrático medio de aproximación, con la que se trata de evaluar el error de aproximación del modelo, como la diferencia entre el valor de la discrepancia de la muestra y el valor esperado de tal discrepancia si el modelo fuera verdadero. Los puntos de corte para este índice son: buen ajuste para valores menores a 0.05 y ajuste aceptable para valores menores a 0.08 (Schumacker & Lomax, 2015).

- 4) CFI: Este índice compara el error de aproximación del modelo propuesto y un modelo de independencia (modelo donde no hay factores comunes). El rango de valores oscila entre 0 y 1, siendo el punto de corte recomendado un valor superior a 0.95. Se recomienda para muestras pequeñas (Abad et al., 2011).
- 5) TLI o NNFI: El índice Tucker – Lewis es una mejora del índice NFI, que tenía el defecto de no penalizar modelos complejos. Con la corrección aplicada por estos autores, se puede evaluar mejor la discrepancia entre el modelo propuesto y el modelo de independencia (Abad et al., 2011). El punto de corte recomendado es un valor mayor a 0.95.

Los índices de ajuste descritos deben interpretarse en conjunto con la evaluación global de los modelos factoriales propuesto con base a los puntos de corte recomendados, posteriormente, se prioriza una evaluación local a través de las cargas y residuos factoriales (Furr & Bacharach, 2014).

Evaluación para modelos bifactor

Generalmente los modelos bifactor resultan ganadores de la comparación con otros tipos de modelos especificados para las pruebas, esto se debe a que, por sus características, son mejor evaluados por los índices de ajuste tradicionales (Gignac, 2016). Por esta razón, los modelos bifactor requieren de pruebas adicionales (Canivez, 2016; Dominguez-Lara & Rodriguez, 2017; Rodriguez et al., 2016; Ventura-León et al., 2021), los cuales se expondrán a continuación:

- 1) Coeficiente Omega jerárquico (ω_{H_S}): Se emplea para valorar la varianza total que puede ser atribuida al factor general, se obtiene de los factores específicos, que son interpretadas como la varianza confiable de puntuaciones por sobre el factor general. De acuerdo a estas mediciones, la interpretación para un instrumento multidimensional bifactor es que tiene que tener un coeficiente omega muy alto para el factor general y coeficiente omegas bajos para los factores específicos. En consecuencia, resultados inversos indicarían una mayor importancia de los factores específicos (Canivez, 2016). Los puntos de corte recomendados son $\omega_{H_S} \geq 0.70$ de coeficiente omega jerárquico del factor general, y $\omega_{h_s} \geq 0.30$ para los factores específicos (Dominguez-Lara & Rodriguez, 2017).
- 2) Índice de replicabilidad de constructo – Coeficiente H_h : Evalúa a un conjunto de ítems si tiene la calidad suficiente de representatividad de un factor. El Coeficiente H puede variar de 0 a 1, donde un valor más cercano a 1 indica una mayor replicabilidad del constructo. En este caso este estadístico se aplica al factor general y se espera un valor superior a 0.70 (Dominguez-Lara & Rodriguez, 2017).
- 3) Varianza común explicada – ECV (Explained Common Variance): Se aplica al Factor general, para evaluar la varianza común que aporta. Se espera una cantidad mayor a 0.60, que se interpreta como poca varianza común entre los factores específicos, más allá que al del factor general (Dominguez-Lara & Rodriguez, 2017; Reise et al., 2013).

- 4) Porcentaje de correlaciones no contaminadas – PUC (Percentage of Uncontaminated Correlations): Aplicado al caso de los modelos bifactor, indica que porcentaje del total de correlaciones ítems – factores no están contaminadas por la multidimensionalidad (Rodríguez et al., 2016). Su interpretación se instaura en conjunto con el ECV. Rodríguez et al. (2016) brindaron los siguientes criterios: Si en una evaluación de un bifactor el $ECV > 0.70$ y $PUC > 0.70$, esto indica evidencia a favor de la unidimensionalidad. Otro ejemplo es el de Reise et al. (2013), donde si el PUC es alto (> 0.80), la medida de ECV ya no resultaría relevante.
- 5) Porcentaje de varianza confiable – PRV (Percentage of Reliable Variance): Es el porcentaje de fiabilidad total de los reactivos que puede atribuirse a la fiabilidad del factor general. Su cálculo se realiza como el cociente entre el coeficiente omega jerárquico y el coeficiente omega, ambos del factor general (Ventura-León et al., 2021). El punto de corte sugerido por Ventura-León et al. (2021) es $PRV > 50$.

5) Re-especificación

Es la fase que se aplica cuando el modelo factorial especificado no supera la evaluación de los índices de ajuste, por lo que habría volver a especificar el modelo, con la esperanza matemática de obtener mejores resultados. Las razones del rechazo del modelo pueden ser:

- 1) El modelo factorial obtiene buenos ajustes, pero no los pesos factoriales, lo cual puede indicar que las relaciones factor-ítem especificadas no sean las correctas.
- 2) Se configuraron parámetros fijados a cero, pero en realidad deben ser estimados.
- 3) Los errores de determinados ítems deben correlacionarse debido a que comparten una misma fuente de varianza sistemática.
- 4) Se deben fijar las correlaciones entre determinados ítems con sus interceptos.

En el segundo, tercer y cuarto caso se deben emplear los Índices de modificación (IM), que indican cual sería la mejora en el índice de ajuste si al parámetro se le libera para obtener una estimación con base a la reducción de la implicancia del coeficiente χ^2 . Los parámetros que obtienen grandes índices de modificación son los propuestos a reevaluarse (Abad et al., 2011). En la mayoría de programas informáticos se pueden obtener los IM y aplicarlos, pero se sugiere reespecificar con medida, porque existe la posibilidad de que las modificaciones sugeridas no tengan sentido desde el punto de vista teórico del cuestionario.

4 Problema del NHST y tamaños del efecto

El test de significancia de la hipótesis nula, en adelante “NHST” (Null Hypothesis Significance Testing), ha sido el método paradigmático para el análisis de datos cuantitativos inferenciales en las disciplinas de las psicologías y las ciencias sociales, tanto a nivel internacional como nacional (Castillo-Blanco & Alegre-Bravo, 2015). Sin embargo, este método tiene críticas a su aplicación científica. A continuación, se brinda una lista no exhaustiva de estas críticas:

- Interpretaciones erróneas de los resultados, como por ejemplo, calificar las diferencias estadísticamente significativas como una sentencia certera de las diferencias en las muestras, cuando en realidad este estadístico solo informa un aspecto reducido de este resultado (Castillo-Blanco & Alegre-Bravo, 2015; Cumming, 2014).
- La arbitrariedad de demarcar la cifra de 0.05 como estadísticamente significativa o no, asumiendo un enfoque binario de conclusión de los resultados (Wasserstein & Lazar, 2016).
- La interpretación probabilística del NHST de hallar verdadera a la hipótesis nula en una de veinte posibilidades (5%) de replicación del estudio es errada (Kline, 2004; Normile et al., 2019).

- Las dificultades de replicar y reproducir los estudios que usaron únicamente el NHST (Ioannidis, 2005).
- La excesiva dependencia del tamaño de muestra, siendo fácil de obtener un resultado estadísticamente significativo en muestras grandes, originando falsos positivos (Simmons et al., 2011).
- El error de confundir la significancia estadística con la “significancia práctica” (Goodman, 2008).
- La mayoría de investigadores cree en el mito de que el NHST les brindara la información necesaria para las conclusiones de sus análisis, cuando en realidad no aporta información relevante (Cohen, 1994).
- La aplicación y enseñanza del NHST es poco intuitiva y contraria al sentido común (Cumming, 2014), obstaculizando la formación educativa de investigadores (Normile et al., 2019).

Con base en estas críticas, se concluye que es urgente el uso de métodos alternativos para el análisis de datos cuantitativos en las disciplinas de las psicologías y las ciencias sociales, considerando las deficiencias del NHST.

Como alternativa de solución a este problema, se han propuesto el “tamaño del efecto” como métodos de análisis de datos cuantitativos más pertinentes, de acuerdo a los objetivos de investigación. Diversos investigadores han impulsado su uso desde los años sesenta (Normile et al., 2019) y el Manual de Publicaciones de la APA ha recomendado explícitamente reportar los tamaños del efecto e intervalos de confianza en las investigaciones cuantitativas en general (American Psychological Association, 2021, p. 87).

Se ha definido al tamaño del efecto como “el grado en que el fenómeno se presenta en la población” (Cohen, 1988) o como “la medida del grado en que el fenómeno de interés está presente en la muestra” (Castillo-Blanco & Alegre-Bravo, 2015). Se debe aclarar que el tamaño

del efecto no es una sola formula, sino un conjunto de fórmulas que se aplican según los objetivos y diseño de investigación cuantitativa. A continuación, se expondrán los tamaños del efecto que se emplean para las comparaciones de grupos independientes:

Tamaños del efecto para la comparación de dos grupos independientes

Para cuantificar la magnitud de la diferencia observada entre dos grupos se ha usado la d de Cohen (Castillo-Blanco, 2014), cuyo calculo se expresa de la siguiente manera:

$$d_{Cohen} = \frac{M_1 - M_2}{S_{agrupado}}$$

Donde: M_1 = Media del grupo 1

M_2 = Media del grupo 2

$S_{agrupado}$ = Estimación muestral de la desviación estándar poblacional

El último elemento descrito también se puede entender como una medida de la variabilidad total de las dos muestras combinadas y se calcula promediando las desviaciones estándar de cada grupo, ponderadas por el tamaño de muestra de cada grupo.

La d de Cohen ha sido la medida de tamaño del efecto más empleada en la investigación contemporánea (Castillo-Blanco & Alegre-Bravo, 2015), pero hay algunos casos en las que no se recomienda usarla:

- Cuando las escalas de medida de los cuestionarios son diferentes.
- Cuando las muestras son muy pequeñas, en las cuales la d de Cohen puede no proporcionar estimaciones confiables.
- Cuando la distribución no tiene normalidad estadística: Si las distribuciones de las variables son muy asimétricas o no se ajustan bien a la distribución normal, la d de Cohen puede verse afectado y no reflejar adecuadamente la magnitud de la diferencia entre los grupos.

Como alternativa que pueda proporcionar una medición más robusta, se ha recomendado el estadístico Probabilidad de Superioridad – PS (Grissom, 1994; Ventura-León, 2016), que se define como “la probabilidad de que una puntuación muestreada aleatoriamente de una población sea mayor que la de otra puntuación...” (Castillo-Blanco, 2014). Su fórmula es:

$$PS = \frac{U}{mn}$$

Donde: U = Valor de la U de Mann Whitney

m = cantidad de participantes del grupo 1

n = cantidad de participantes del grupo 2

Sobre las cifras de corte para definir como pequeño o grande una magnitud del efecto, los autores de cada una de estas fórmulas han ofrecido cifras tentativas para calificar un resultado. Sin embargo, los investigadores recomiendan colocar como punto de referencia en primer lugar las cifras de corte surgidas a partir de la línea de investigación respecto a algún fenómeno determinado. Por ejemplo, las cifras de corte en la medición del tamaño del efecto de un programa de enseñanza probablemente sean diferentes a la medición de un tratamiento médico, a pesar de que obtengan la misma cifra. En función a esta advertencia, se exponen (tabla C.1) las cifras de corte de los tamaños del efecto para dos grupos independientes, considerándolas como referenciales, esperando que, conforme surjan más investigaciones sobre los fenómenos estudiados, se establezcan cifras de corte propios de esta línea de investigación.

Tabla C.1

Cifras de corte referenciales en tamaños del efecto para la comparación de dos grupos independientes

Tamaño del efecto	Nivel	Corte	Referencia académica
-------------------	-------	-------	----------------------

d de Cohen	Grande	0.80	Cohen, J. (1988). <i>Statistical Power Analysis for the Behavioral Sciences</i> (2nd ed.). Lawrence Erlbaum Associates.
	Moderado	0.50	
	Pequeño	0.20	
Probabilidad de Superioridad - PS	Grande	0.71	Grissom, R. J. (1994). Probability of the superior outcome of one treatment over another. <i>Journal of Applied Psychology</i> , 79(2), 314-316.
	Mediano	0.64	
	Pequeño	0.56	
	No efecto	0.00	

Tamaños del efecto para la comparación de más de dos grupos independientes

En cuanto a la comparación de tres grupos a más, el estadístico más empleado ha sido Eta cuadrado (η^2) que mide la proporción de la varianza total de una variable dependiente que es explicada por las diferencias entre los grupos. Sin embargo, este estimador tiene un sesgo positivo porque el numerador de su fórmula es susceptible de aumentar por el error de variabilidad (Castillo-Blanco, 2014). Debido a esta sobreestimación, se ha sugerido como alternativa el estimador Omega cuadrado (ω^2) que ajusta el sesgo de estimación que se encuentra en Eta cuadrado, brindando una estimación de la proporción de variabilidad explicada por las diferencias entre los grupos (Dominguez-Lara, 2017; Fritz et al., 2012). Los valores de ω^2 oscilan entre 0 y 1, donde 0 indica que no hay diferencias entre los grupos y 1 indica que la totalidad de la variabilidad se explica por las diferencias entre los grupos. A continuación, se expone la fórmula:

$$\omega^2 = \frac{SS_b - (k - 1)MS_w}{SS_{total} + MS_w}$$

Donde: SS_b = Suma de desviaciones al cuadrado entre grupos

SS_{total} = Suma total de las desviaciones al cuadrado

k = cantidad de grupos

MS_w = Suma de varianzas ponderadas

Otra alternativa de estimador de tamaño del efecto es Épsilon al cuadrado (ε^2) que ha sido recomendado como reemplazo a la prueba NHST no paramétrica de Kruskal-Wallis (Ventura-León, 2017). Para calcular este tamaño del efecto, se necesita del coeficiente H de Kruskal-Wallis (Tomczak & Tomczak, 2014), como se expone en la formula:

$$\varepsilon^2 = \frac{H}{(n^2 - 1)(n + 1)}$$

Donde: H = Coeficiente H de Kruskal-Wallis

n = Número de observaciones o de muestra

La lectura de la cifra de ε^2 es similar a ω^2 , porque los valores tienen el rango de 0 a 1, donde entre más cercano a uno mayor será la variabilidad explicada por las diferencias entre los grupos.

Para la interpretación de estos tamaños del efecto, se usarán las mismas cifras de corte referenciales (Dominguez-Lara, 2017), teniendo en cuenta la advertencia expuesta en el anterior apartado.

Tabla C.2

Cifras de corte referenciales en tamaños del efecto para la comparación de más de dos grupos independientes

Tamaño del efecto	Nivel	Corte	Referencia académica
Omega cuadrado (ω^2)	Fuerte	0.64	Dominguez-Lara, S. (2017).
	Moderada	0.25	Magnitud del efecto en comparaciones entre 2 o más
	Mínima	0.04	

Fuentes de información del Anexo C

- Abad, F., Olea, J., Ponsoda, V., & García, C. (2011). *Medición en ciencias sociales y de la salud*. Editorial Síntesis.
- AERA, APA, & NCME. (2018). *Estándares para Pruebas Educativas y Psicológicas*. American Educational Research Association. <https://doi.org/10.2307/j.ctvr43hg2>
- American Psychological Association. (2021). *Manual de Publicaciones de la American Psychological Association* (Cuarta ed.). Manual Moderno.
- Brown, T. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.
- Canivez, G. L. (2016). Bifactor Modeling in Construct Validation of Multifaceted Tests: Implications for Understanding Multidimensional Constructs and Test Interpretation. En K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements* (pp. 247-271). Hogrefe.
- Castillo-Blanco, R. W. (2014). *Reporte del tamaño del efecto en los artículos de tres revistas de psicología peruanas en los años 2008 al 2012* [Tesis de Licenciatura, Universidad Nacional Mayor de San Marcos]. <https://cybertesis.unmsm.edu.pe/handle/20.500.12672/3850>
- Castillo-Blanco, R. W., & Alegre-Bravo, A. (2015). Importancia del tamaño del efecto en el análisis de datos de investigación en psicología. *Persona*, 18, 137-148. <https://doi.org/10.26439/persona2015.n018.503>
- Cho, E. (2016). Making Reliability Reliable. *Organizational Research Methods*, 19(4), 651-682. <https://doi.org/10.1177/1094428116656239>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence

Erlbaum Associates.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49(12), 997-1003.

<https://doi.org/10.1037/0003-066X.49.12.997>

Cumming, G. (2014). The New Statistics. *Psychological Science*, 25(1), 7-29.

<https://doi.org/10.1177/0956797613504966>

DiStefano, C., Liu, J., Jiang, N., & Shi, D. (2018). Examination of the Weighted Root Mean

Square Residual: Evidence for Trustworthiness? *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 453-466.

<https://doi.org/10.1080/10705511.2017.1390394>

Dominguez-Lara, S. (2017). Magnitud del efecto en comparaciones entre 2 o más grupos.

Revista de Calidad Asistencial, 32(2), 121-122.

<https://doi.org/10.1016/j.cali.2016.04.002>

Dominguez-Lara, S., & Rodriguez, A. (2017). Índices estadísticos de modelos bifactor.

Interacciones: Revista de Avances en Psicología, 3(2), 59-65.

<https://doi.org/10.24016/2017.v3n2.51>

Ferrando, P. J., Lorenzo-Seva, U., Hernández-Dorado, A., & Muñiz, J. (2022). Decálogo para

el Análisis Factorial de los Ítems de un Test. *Psicothema*, 34(1), 7-17.

<https://doi.org/10.7334/psicothema2021.456>

Flora, D. B., & Curran, P. J. (2004). An Empirical Evaluation of Alternative Methods of

Estimation for Confirmatory Factor Analysis With Ordinal Data. *Psychological Methods*, 9(4), 466-491. <https://doi.org/10.1037/1082-989X.9.4.466>

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. <https://doi.org/10.1037/a0024338>

Furr, M., & Bacharach, V. (2014). *Psychometrics. An Introduction* (Second Edi). SAGE

Publications.

- Gana, K., & Broc, G. (2019). *Structural Equation Modeling with lavaan*. ISTE Ltd and John Wiley & Sons.
- Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence*, *55*, 57-68. <https://doi.org/10.1016/j.intell.2016.01.006>
- Goodman, S. (2008). A Dirty Dozen: Twelve P-Value Misconceptions. *Seminars in Hematology*, *45*(3), 135-140. <https://doi.org/10.1053/j.seminhematol.2008.04.003>
- Grissom, R. J. (1994). Probability of the superior outcome of one treatment over another. *Journal of Applied Psychology*, *79*(2), 314-316. <https://doi.org/10.1037/0021-9010.79.2.314>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, *2*(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Kalkbrenner, M. T. (2023). Alpha, Omega, and H Internal Consistency Reliability Estimates: Reviewing These Options and When to Use Them. *Counseling Outcome Research and Evaluation*, *14*(1), 77-88. <https://doi.org/10.1080/21501378.2021.1940118>
- Kline, R. (2004). What's Wrong With Statistical Tests--And Where We Go From Here. En *Beyond significance testing: Reforming data analysis methods in behavioral research*. (pp. 61-91). American Psychological Association. <https://doi.org/10.1037/10693-003>
- Kline, R. (2016). *Principles and practice of structural equation modeling* (4th Ed.). The Guilford Press.
- McDonald, R. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates.

- Mohd Razaki, N., & Bee Wah, Y. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, 2(1), 21-33.
- Muñiz, J. (2018). *Introducción a la Psicometría. Teoría clásica y TRI*. Ediciones Pirámide.
- Normile, C. J., Bloesch, E. K., Davoli, C. C., & Scherr, K. C. (2019). Introducing the new statistics in the classroom. *Scholarship of Teaching and Learning in Psychology*, 5(2), 162-168. <https://doi.org/10.1037/stl0000141>
- Oliden, P. E., & Zumbo, B. D. (2008). Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema*, 20(4), 896-901. <http://www.psicothema.com/pdf/3572.pdf>
- Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and Structural Coefficient Bias in Structural Equation Modeling. *Educational and Psychological Measurement*, 73(1), 5-26. <https://doi.org/10.1177/0013164412449831>
- Rios, J., & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema*, 26(1), 108-116. <https://doi.org/10.7334/psicothema2013.260>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137-150. <https://doi.org/10.1037/met0000045>
- Schumacker, R. E., & Lomax, R. G. (2015). *A Beginner's Guide to Structural Equation Modeling*. Routledge. <https://doi.org/10.4324/9781315749105>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology. *Psychological Science*, 22(11), 1359-1366. <https://doi.org/10.1177/0956797611417632>
- Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics* (5th ed.). Pearson Education.
- Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An

- overview of some recommended measures of effect size. *Trends in Sport Sciences*, 1(21), 19-25.
- Ventura-León, J. L. (2016). Tamaño del efecto para la U de Mann-Whitney: aportes al artículo de Valdivia-Peralta et al. *Revista chilena de neuro-psiquiatría*, 54(4), 353-354. <https://doi.org/10.4067/S0717-92272016000400010>
- Ventura-León, J. L. (2017). Tamaño del efecto para Kruskal-Wallis: aportes al artículo de Domínguez-González et al. *Investigación en Educación Médica*, 8(30), 135-136. <https://doi.org/10.1016/j.riem.2017.07.002>
- Ventura-León, J. L., Quiroz-Burga, L., Caycho-Rodríguez, T., & Valencia, P. (2021). BifactorCalc: An Online Calculator for Ancillary Measures of Bifactor Models. *Revista Evaluar*, 21(3), 01-14. <https://doi.org/10.35670/1667-4545.v21.n3.36272>
- Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). Un viaje alrededor de alfa y omega para estimar la fiabilidad de consistencia interna. *Anales de Psicología*, 33(3), 755. <https://doi.org/10.6018/analesps.33.3.268401>
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p -Values: Context, Process, and Purpose. *The American Statistician*, 70(2), 129-133. <https://doi.org/10.1080/00031305.2016.1154108>
- Zinbarg, R. E., Revelle, W., & Yovel, I. (2007). Estimating ω^2 for Structures Containing Two Group Factors: Perils and Prospects. *Applied Psychological Measurement*, 31(2), 135-157. <https://doi.org/10.1177/0146621606291558>
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for likert rating scales. *Journal of Modern Applied Statistical Methods*, 6(1), 21-29. <https://doi.org/10.22237/jmasm/1177992180>

Anexo D. Formato de consentimiento informado dirigido al docente

FORMATO DE CONSENTIMIENTO INFORMADO PARA DOCENTES

Universidad Nacional Federico Villarreal
Escuela de Formación profesional de psicología

Fecha:/...../.....

Yo.....identificado con DNI (carné de extranjería o pasaporte para extranjeros) N° he sido informado por el Señor..... acerca del trabajo de investigación que van a realizar en un salón donde dicte clases durante el ciclo académico de abril a julio. Me ha informado de los procedimientos, tiempo de duración y posibles ventajas / desventajas de aceptar que se realice en mi salón.

He realizado las preguntas que consideré oportunas, todas las cuales han sido absueltas y con repuestas que considero suficientes y aceptables.

Por lo tanto, en forma consciente y voluntaria doy mi consentimiento para que se realice la evaluación correspondiente mediante los formatos que se crean convenientes. Teniendo pleno conocimiento de los procedimientos, tiempo de duración y posibles ventajas/desventajas que podrían desprenderse de dicho acto.

<p>Firma del docente de aula Nombre: DNI:</p>	<p>Firma del evaluador Nombre: DNI:</p>

Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:

Fin de la hoja de registro de firmantes.

FORMATO DE CONSENTIMIENTO INFORMADO PARA ALUMNOS

Universidad Nacional Federico Villarreal
Escuela de Formación profesional de psicología

Fecha:/...../.....

Doy fe que he sido informado/a por el señor..... acerca del trabajo de investigación que van a realizar en mi salón. Me ha informado de los fines, procedimientos y tratamiento de la información.

He realizado las preguntas que consideré oportunas, las cuales han sido absueltas y con repuestas que considero suficientes y aceptables.

Por lo tanto, en forma conciente y voluntaria doy mi consentimiento para formar parte de la evaluación correspondiente mediante los formatos que se crean convenientes, teniendo pleno conocimiento de los fines, procedimientos y tratamiento de la información que podrían desprenderse de dicho acto.

Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:
Firma del alumno(a) Nombre y apellido: DNI:	Firma del alumno(a) Nombre y apellido: DNI:

No.	Preguntas	Nunca	Casi Nunca	Casi Siempre	Siempre
10	Cumplo con los criterios que el profesor indica para realizar algún trabajo en clase.				
11	Realizo una práctica cumpliendo con los criterios que el docente explicó.				
12	Cumplo con los criterios que el profesor menciona al realizar algún ejercicio en clase.				
13	¿El profesor explica con claridad el tema de la clase?				
14	¿El profesor ejemplifica la tarea que asignará al estudiante?				
15	¿El profesor resuelve problemas frente a los estudiantes, a partir del tema planteado?				
16	Comprendo sin ningún problema el tema que el profesor explica en la clase.				
17	Realizo ejercicios en clase contemplando el modelo que el profesor realizó.				
18	Resuelvo los problemas que el profesor plantea para la clase.				
19	¿El profesor me corrige cuando no realizo bien alguna actividad de la clase?				
20	¿El profesor me pone en contacto con lo que hice mal y me enseña las formas en las que lo puedo resolver?				
21	¿El profesor me enseña distintas formas en las que puedo satisfacer los criterios de las actividades de su clase?				
22	¿El profesor menciona las reglas de la disciplina en clase?				
23	¿El profesor pide que guarde silencio en clase?				
24	¿El profesor pone en evidencia a algún alumno que no esté atento a la clase?				
25	Hago caso a las reglas de la clase.				
26	Guardo silencio durante la clase.				
27	Interrumpo la clase por platicar con mis compañeros.				
28	Utilizo mi teléfono en clase (para leer o enviar mensajes, para responder o hacer llamadas).				
29	¿El profesor habla de temas no relacionados con la clase?				
30	¿El profesor utiliza el tiempo de la clase en actividades no académicas?				
31	¿El profesor pierde tiempo de la clase en preparar los materiales de la clase?				
32	¿El profesor es puntual?				
33	Hablo en clase, de temas que no corresponden a la clase.				
34	Utilizo tiempo de las clases en actividades no académicas.				
35	Pierdo tiempo en preparar los materiales para hacer las actividades de la clase.				
36	Soy puntual.				
37	Estoy atento al tema de la clase.				
38	Repito y completo las frases que el profesor me indica.				

2



CUESTIONARIO SOBRE EL DESEMPEÑO DOCENTE E INTERACCIONES DIDÁCTICAS
Bazán & Velarde
Universidad Autónoma del Estado de Morelos
Centro de Investigación Transdisciplinar en Psicología

Folio: _____ **Fecha:** _____

Nombre: _____

Nombre de la asignatura a valorar en este cuestionario: _____

Nombre de la escuela: _____

Semestre que cursa: _____ **Grupo:** _____

Turno: _____

Edad: _____ **Sexo:**

H	M
---	---

Instrucciones: De acuerdo con tu experiencia en clase durante este último semestre o parcial, Marca con una X la opción que mejor describa la frecuencia de las acciones de tu profesor y las tuyas según sea el caso durante las clases de esta asignatura. Debes marcar una de las cuatro opciones en cada afirmación o pregunta de acuerdo con la siguiente escala: *Nunca, Casi nunca, Casi siempre, Siempre*. Te pedimos que seas lo más sincero (a) posible.

No.	Indicadores	Nunca	Casi Nunca	Casi Siempre	Siempre
1	¿El docente explora mis conocimientos de la materia al inicio de cada clase?				
2	¿El docente plantea problemas para que los resuelva antes de iniciar un tema?				
3	¿El docente pregunta conceptos referentes al tema antes de explicarlo?				
4	Respondo o comento cuando el docente me pregunta o pregunta a todos, antes de iniciar un tema.				
5	Resuelvo adecuadamente los problemas que el docente plantea antes de iniciar un tema.				
6	Respondo las preguntas que el profesor hace acerca de un tema nuevo.				
7	¿El docente explica los requerimientos necesarios para realizar algún trabajo en clase?				
8	¿El docente explica los criterios que se requieren para realizar una práctica?				
9	¿El docente explica que necesito hacer para completar un ejercicio en clase?				

1

No.	Preguntas	Nunca	Casi Nunca	Casi Siempre	Siempre
39	Tomo apuntes de los temas de la clase.				
40	Intervengo en clase para completar el tema que se está desarrollando.				
41	Emito mi opinión acerca del tema de la clase.				
42	Realizo aportaciones a los comentarios de mis compañeros.				
43	Realizo preguntas de acuerdo al tema que se está desarrollando.				
44	Realizo preguntas que están fuera del tema abordado.				
45	Cumplo con las tareas que se me piden en la clase.				
46	Realizo los ejercicios referentes al tema desarrollado.				
47	Identifico en mi actuar referente a la clase, si he realizado bien las prácticas.				
48	Identifico si algún compañero se ha equivocado y puedo decirle en que lo ha hecho.				
49	Cambio mis acciones para cumplir con los criterios de la clase o tarea.				
50	Cumplo con los criterios de las actividades de la clase.				
51	Tengo ciertas limitaciones para realizar las tareas de la clase.				
52	Son claros los criterios de la clase para poder cumplirlos.				
53	¿El profesor revisa los ejercicios y tareas que deja, y proporciona calificaciones?				
54	¿El profesor revisa los ejercicios y tareas que deja, y hace señalamientos para corregir, o para mejorar mi trabajo?				
55	¿El profesor realiza evaluaciones sobre mi conocimiento teórico y de conceptos o fundamentos básicos de la asignatura?				
56	¿El profesor realiza evaluaciones de carácter aplicado y la solución de problemas prácticos, derivados de la asignatura?				
57	¿El profesor evalúa mi capacidad de integrar conocimientos de otras asignaturas, con la asignatura que curso con él?				

Anexo G. Discusión de los resultados psicométricos de los cuestionarios CDD y CDE

Cuestionario de Desempeño docente - CDD

La presente versión aplicada cuenta con una estructura de 6 dimensiones y 33 ítems. Para el análisis factorial confirmatorio, se empleó el estimador WLSMV, que ya fue usado en otras versiones del cuestionario (Bazán-Ramírez, et al., 2022a) y en un instrumento similar (Dominguez-Lara et al., 2019). De la comparación de modelos, el modelo bifactor fue el que obtuvo los mejores índices de ajuste (modelo bifactor de desempeño docente: CFI = 0.994; TLI = 0.993; RMSEA = 0.029; RMSEA-IC = 0.023-0.034; WRMR = 0.991) y resultados en las pruebas adicionales para modelos bifactor (ECV = 0.589; PUC = 0.857; H = 1.143; PVR = 95.168).

En comparación con la evidencia previa, se obtuvo un resultado adecuado y similar con otras versiones del cuestionario CDD: Bazán-Ramírez y Velarde (2021) aplicaron una primera versión abreviada del CDD en una muestra mexicana, obteniendo aceptables índices de ajuste para un modelo factorial oblicuo de cinco dimensiones de desempeño docente (CFI = 0.95; RMSEA = 0.06); Bazán-Ramírez et al. (2021) volvieron a aplicar la versión abreviada del CDD, esta vez a una muestra peruana muy similar a la del presente estudio, obteniendo adecuados índices de ajuste (CFI = 0,95; NNFI = 0,94; SRMR = 0,050; RMSEA = 0,052); Bazán-Ramírez et al. (2022b) reelaboraron el CDD con 24 ítems, seis dimensiones de desempeño docente y la aplicaron a una muestra peruana de estudiantes de posgrado, donde los resultados de índices de ajuste fueron favorables (GFI = 0,98; CFI = 1,11; RMSEA = 0,04; RMR = 0,043; AGFI = 0,98; NFI = 0,98; NNFI = 1,11) para el modelo oblicuo de seis factores.

Los resultados psicométricos hallados en los antecedentes se asemejan a los resultados hallados en la presente investigación. Estas semejanzas se pueden deber a: 1- La muestra similar de universitarios, estudiantes de psicología; 2- Una estructura de dimensiones similar entre las versiones del cuestionario; 3- La redacción de los ítems, y 4- El comportamiento

evaluado de docentes peruanos de universidades públicas. Una diferencia entre los estudios psicométricos es que los antecedentes no reportan la prueba de modelos bifactor, pero si reportan índices de coeficiente alfa y omega del conjunto de ítems completos, por encima del 0.7, añadiendo evidencias en favor de un factor general.

Comparando nuestro cuestionario con la Escala de Autoeficacia Docente – EAD (Dominguez-Lara et al., 2019), donde se empleó el estimador WLSMV y las pruebas de idoneidad bifactor, se obtuvieron resultados similares en los índices de ajuste y las pruebas bifactor, lo cual sugiere que se puede evaluar de manera global el desempeño docente, a través de un factor general. Estos resultados respaldan el uso del CDD en estudiantes peruanos de la carrera de psicología, para evaluar el desempeño docente mediante un puntaje general, así como evaluar el despliegue de competencias en los seis ámbitos didácticos.

Una vez confirmada la estructura factorial para el CDD, esta se puso a prueba mediante el análisis factorial confirmatorio multigrupo (AFC-MG), donde se generan modelos factoriales para cada grupo segmentado y se comparan sus propiedades, en la búsqueda de mediciones equivalentes. Esta evaluación adopto un enfoque híbrido (Little, 1997), donde se consideró los coeficientes χ^2 y la comparación de índices de ajuste RMSEA y CFI (Chen, 2007) como reglas de decisión para transitar entre niveles de invarianza de la medición: IM-1: invarianza configural, IM-2: invarianza métrica; IM-3: invarianza escalar; y IM-4: invarianza estricta (residual). A continuación, se discutirá los resultados obtenidos del modelo hexafactorial de desempeño docente, según cada variable de agrupamiento.

La variable género alcanzo la invarianza estricta (IM-4), lo cual indica que la medición es equivalente entre varones y mujeres, cumpliendo los criterios de invarianza configural, métrica, escalar y residual. Un resultado similar se halló en Bazán-Ramírez et al. (2021) con estudiantes de pregrado de psicología y en Bazán-Ramírez et al. (2022a) con estudiantes de

posgrado de ciencia de la educación. En ambas investigaciones se empleó versiones similares del CDD al del presente estudio.

La variable edad alcanzó la invarianza escalar (IM-3), lo cual indica que la equivalencia entre los dos grupos de edad alcanzó los niveles configural (estructura de factores), métrica (cargas factoriales) y escalar (interceptos) de invarianza de la medición, pero no el nivel estricto (residual). Es necesario investigar porque existen diferencias residuales en las mediciones, porque en una investigación similar la invarianza si alcanzó el nivel estricto (IM-4), pero con la diferencia metodológica en que usaron un corte de edad diferente al presente estudio (Bazán-Ramírez et al., 2021).

Sobre las variables de agrupación ciclo de estudios, turno de estudios, nivel de especialidad y eje curricular, los grupos alcanzaron la invarianza métrica (IM-2), lo que indica que la estructura de los factores y las cargas factoriales son equivalentes en las mediciones de estas variables. Es necesario investigar porque no se alcanzó la equivalencia de interceptos de medición, porque podría indicar que todavía existen diferencias sistemáticas en cómo se responden individualmente los ítems en cada grupo, lo cual puede deberse a factores específicos relacionados con las características de los grupos, que no fueron considerados en este estudio.

En el caso específico de la variable nivel de especialidad, en un antecedente similar también se halló una invarianza parcial (Bazán-Ramírez et al., 2021), porque se alcanzó el nivel escalar (IM-3) en las mediciones de los grupos “etapa académica” (4to y 6to semestre) y “etapa final” (8vo y 10mo semestre) en estudiantes peruanos de psicología. Este resultado puede sugerir que clasificar el desempeño evaluado según especialidad o estudios generales puede provocar fuentes de invalidez, lo cual es un desafío metodológico por resolverse. También existe la posibilidad de que los docentes, de acuerdo al curso que enseñan, prioricen algunos ámbitos didácticos, y ello conlleve a una evaluación inequivalente de los estudiantes. Por

ejemplo, un docente de primer ciclo puede enfocarse más en la “Ilustración” y no en la “retroalimentación”, en contraposición de un docente de noveno ciclo cuyo énfasis pedagógico sea inverso al ejemplo.

Los resultados obtenidos apoyan el uso del CDD en el establecimiento de diferencias grupales de los puntajes de la percepción del desempeño docente. La importancia de las variables de agrupación sociodemográficas y educativas se ha demostrado en la investigación de las interacciones didácticas (Acevedo & Mairena, 2006; Bazán-Ramírez et al., 2021; Bazán-Ramírez et al., 2022), por lo que se debe contar con instrumentos que evalúen estas variables, cumpliendo los estándares psicométricos para su aplicación.

Cuestionario de Desempeño estudiantil - CDE

La versión aplicada del CDE cuenta con 6 dimensiones y 33 ítems. En la terna original de ítems, el ítem 24 obtuvo una nula correlación ítem-test corregida y carga factorial, por lo que se decidió retirarlo del cuestionario. En el AFC, donde también se empleó el estimador WLSMV, el modelo que obtuvo los mejores índices de ajuste volvió a ser el modelo bifactor (modelo bifactor de desempeño estudiantil: CFI = 0.989; TLI = 0.989; RMSEA = 0.029; RMSEA-IC = 0.024-0.035; WRMR = 0.996). En cuanto a las mediciones adicionales para confirmar el modelo bifactor, el coeficiente omega jerárquico del factor general obtuvo un valor cercano al punto de corte sugerido ($\omega_{HS} = 0.665$) y cuatro dimensiones fueron mayores al límite recomendado 0.3 (Canivez, 2016). Estos resultados de varianza pueden indicar una varianza específica para las dimensiones “identificación de criterios”, “práctica pertinente” y “desempeño resultante”, que podrían usarse de manera individual. En cuanto al resto de mediciones (ECV = 0.434; PUC = 0.857; H = 0.995; PVR = 75.201), la varianza común explicada – ECV resultó bajo, pero de acuerdo con Reise et al. (2013) si el porcentaje de correlaciones no contaminadas – PUC supera el 0.80, la ECV ya no sería relevante. Esto se complementa con el resto de mediciones que superaron los cortes recomendados. Como

conclusión, si existen mediciones a favor de una estructura bifactor para el cuestionario de desempeño estudiantil, aunque valdría la pena estudiar sus dimensiones individuales como sub-cuestionarios.

Los antecedentes del cuestionario de desempeño estudiantil – CDE son los mismos de la discusión del CDD, por lo que las comparaciones serán más abreviadas. En el estudio de Bazán-Ramírez y Velarde (2021) donde aplicaron una versión abreviada del CDE en estudiantes mexicanos de Psicología, se obtuvo índices de ajuste cercanos a los puntos de corte limite (CFI = 0.90; RMSEA = 0.07) para un modelo factorial oblicuo de cinco dimensiones de desempeño estudiantil. Bazán-Ramírez et al. (2022b) aplicaron una versión reelaborada del CDE con 24 ítems y seis dimensiones de desempeño estudiantil, en una muestra peruana de estudiantes de posgrado en ciencias de la educación, donde se obtuvieron índices de ajuste aceptables (GFI = 0,98; CFI = 1,00; RMSEA = 0,04; RMR = 0,03; AGFI = 0,97; NFI = 0,98; NNFI = 0,99) para un modelo oblicuo de seis factores, que exactamente fue un segundo modelo corregido con el retiro de dos ítems, resultando en 22 ítems para el CDE. En una segunda etapa del estudio descrito anteriormente, Bazán-Ramírez et al. (2022a) sometieron a un análisis factorial confirmatorio una versión del CDE compuesta de cuatro dimensiones (Participación, práctica pertinente, retroalimentación-mejora, y evaluación-aplicación) y 18 ítems. Los resultados del AFC para el modelo multidimensional de desempeño estudiantil se encontraron dentro de los límites recomendados para los puntos de corte (CFI = 0.964, TLI = 0.955, RMSEA = 0.069, RMSEA IC = 0.057 - 0.081, SRMR = 0.075), pero adicionalmente los autores reportaron el coeficiente H (0.92) añadiendo evidencia sobre un factor general.

No se hallaron estudios con instrumentos similares en muestra peruana que tengan una estructura bifactor a comparar. Únicamente se halló el estudio instrumental de Dominguez-Lara et al. (2014) quienes aplicaron el AFC (con el estimador ML) a la Escala de Autoeficacia Percibida Específica de Situaciones Académicas (EAPESA) en una muestra de 448 estudiantes

universitarios de Perú, obteniendo buenos índices de ajuste (CFI = 0.978; GFI = 0.969; AGFI = 0.949; RMR = 0.029; RMSEA = 0.056) para un modelo unidimensional de 9 ítems.

En general, se observan resultados de AFC similares a los hallados en el presente estudio. Respecto a las otras versiones del CDE, pese a que las muestras sean levemente diferentes y existan variaciones en la estructura factorial de los ítems, todas han alcanzado adecuados índices de ajuste y cargas factoriales a sus respectivas dimensiones. Esto puede deberse a la idoneidad de los ítems y dimensiones en representar la frecuencia de conductas de estudio que aplican los estudiantes en sus clases y tareas. Respecto a una evaluación global del desempeño estudiantil, la investigación actual provee de indicios de evidencia de un factor general y existen otros cuestionarios similares que trabajan con un puntaje general (Dominguez-Lara et al., 2014; Dominguez-Lara & Fernández-Arata, 2019).

Estos resultados añaden soporte al uso del CDE, ya sea empleando un puntaje general para “desempeño estudiantil”, como también para obtener puntajes por dimensión (ámbito didáctico).

Para la comparación de modelos factoriales multigrupo al evaluar el desempeño estudiantil, se usaron los mismos criterios expuestos en la AFC-MG de desempeño docente. Se discute a continuación los resultados obtenidos del modelo hexafactorial de desempeño estudiantil, según cada variable de agrupamiento.

Las variables género, edad y nivel de especialidad alcanzaron la invarianza escalar (IM-3), lo cual indica que la medición de las autopercepciones de los grupos especificados tiene una equivalencia en la estructura de sus factores (modelo hexafactorial de desempeño estudiantil), en las cargas factoriales y en los interceptos. Es necesario indagar en la manera de alcanzar la invarianza estricta (de residuos), porque en un antecedente similar (Bazán-Ramírez et al., 2022a) si se alcanzó la invarianza de residuos, aunque con una muestra fue diferente. En el antecedente mencionado se empleó una versión del CDE compuesta de cuatro dimensiones,

aplicada en una muestra de estudiantes de posgrado y comprobaron la equivalencia de las mediciones entre los grupos de género, edad (entre 23 a 71 años) y nivel de posgrado (maestría y doctorado). Las diferencias expuestas podrían influir en el resultado de invarianza alcanzado.

Se destaca que la variable ciclo de estudios alcanzó la invarianza estricta (IM-4), lo cual indica que existe una medición equivalente entre los estudiantes de los ciclos primero, tercero, quinto, séptimo y noveno, cumpliendo con los criterios de invarianza configural (IM-1), métrica (IM-2), escalar (IM-3) y de invarianza de residuos (IM-4). Este resultado es muy útil en la medición de la autopercepción del desempeño estudiantil, porque a pesar de las diferencias entre los estudiantes de diferentes ciclos, los ítems de las seis dimensiones son interpretados de manera equivalente.

Finalmente, las variables turno de estudios y eje curricular alcanzaron la invarianza métrica (IM-2), lo cual indica que las mediciones son equivalentes a nivel de estructura de dimensiones y pesos factoriales, pero no a nivel de interceptos y residuos. Se debe investigar las fuentes de esta equivalencia débil entre los grupos del turno matutino y vespertino, así como los grupos de cursos básicos, de investigación y especialidad.

Como balance general, el Cuestionario de Desempeño Estudiantil – CDE obtuvo adecuados índices de ajuste en su estructura bifactorial, así como niveles de invarianza fuerte y estricta en 4 grupos, por lo que su uso interdisciplinar es válido y fiable.

Fuentes de información del Anexo G

- Acevedo, R., & Mairena, N. (2006). Factores de sesgo asociados a la validez de la evaluación docente universitaria: un modelo jerárquico lineal. *Archivos Analíticos de Políticas Educativas*, 14(34), 1-22. <https://www.redalyc.org/pdf/2750/275020543034.pdf>
- Bazán-Ramírez, A., Capa-Luque, W., Bello-Vidal, C., & Quispe-Morales, R. (2022). Influence of Teaching and the Teacher's Feedback Perceived on the Didactic Performance of Peruvian Postgraduate Students Attending Virtual Classes During the COVID-19

- Pandemic. *Frontiers in Education*, 7. <https://doi.org/10.3389/feduc.2022.818209>
- Bazán-Ramírez, A., Pérez-Morán, J. C., & Bernal-Baldenebro, B. (2021). Criteria for Teaching Performance in Psychology: Invariance According to Age, Sex, and Academic Stage of Peruvian Students. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.764081>
- Bazán-Ramírez, A., Quispe-Morales, R., De La Cruz, C., & Henostroza-Mota, C. (2022). Teacher-Student Performance Criteria During Online Classes due to COVID-19: Self-Report by Postgraduate Students in Education. *European Journal of Educational Research*, 11(4), 2101-2114. <https://doi.org/10.12973/eu-jer.11.4.2101>
- Bazán-Ramírez, A., & Velarde, N. (2021). Auto-reporte del estudiantado en criterios de desempeño didáctico en clases de Psicología. *Journal of Behavior, Health & Social Issues*, 13(1), 22-35. <http://revistas.unam.mx/index.php/jbhsi/article/view/78071>
- Canivez, G. L. (2016). Bifactor Modeling in Construct Validation of Multifactor Tests: Implications for Understanding Multidimensional Constructs and Test Interpretation. En K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements* (pp. 247-271). Hogrefe.
- Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464-504. <https://doi.org/10.1080/10705510701301834>
- Dominguez-Lara, S., & Fernández-Arata, M. (2019). Autoeficacia académica en estudiantes de Psicología de una universidad de Lima. *Revista Electrónica de Investigación Educativa*, 21(e32), 1-13. <https://doi.org/10.24320/redie.2019.21.e32.2014>
- Dominguez-Lara, S., Fernández-Arata, M., Merino-Soto, C., Navarro-Loli, J., & Calderón-De la Cruz, G. (2019). Escala de Autoeficacia Docente: análisis estructural e invarianza de medición en docentes peruanos de escuelas públicas. *Revista Argentina de Ciencias del*

Comportamiento, 11(3), 61-72. <https://doi.org/10.32348/1852.4206.v11.n3.24624>

Dominguez-Lara, S., Yauri, C., Mattos, E., & Ramírez, F. (2014). Autoeficacia para situaciones académicas en estudiantes universitarios peruanos: Un enfoque de ecuaciones estructurales. *Revista De Psicología*, 4, 43-54. <https://revistas.ucsp.edu.pe/index.php/psicologia/article/view/20>

Little, T. D. (1997). Mean and Covariance Structures (MACS) Analyses of Cross-Cultural Data: Practical and Theoretical Issues. *Multivariate Behavioral Research*, 32(1), 53-76. https://doi.org/10.1207/s15327906mbr3201_3

Anexo H. Análisis comparativo alternativo

Se realizó un análisis alternativo de las comparaciones de los puntajes de acuerdo a las seis agrupaciones, empleando los siguientes análisis: t de Student y d de Cohen para agrupaciones de dos; ANOVA de un factor y Omega al cuadrado para agrupaciones mayores a dos. El criterio para elegir estas pruebas es el teorema del límite central, una vez verificado las distribuciones descriptivas de los puntajes de las variables.

Tabla H.1

Pruebas comparativas para el análisis de las variables de desempeño docente y desempeño estudiantil percibido según sexo

Variables	Masculino (n=130)	Femenino (n=345)	t	p	d de Cohen
Exploración Competencial	9.00	8.70	0.847	0.397	0.087
Explicitación de criterios	12.48	12.44	0.101	0.920	0.010
Ilustración	11.82	12.05	-0.618	0.537	-0.064
Supervisión de la práctica	9.80	9.98	-0.513	0.608	-0.053
Retroalimentación	12.02	11.83	0.449	0.653	0.046
Evaluación	9.59	9.82	-0.653	0.514	-0.067
Total Desempeño docente	64.72	64.81	-0.051	0.960	-0.005
Precurrentes y potenciales habilidades	9.08	8.53	2.431	0.015	0.250
Identificación de criterios	10.45	10.51	-0.252	0.802	-0.026
Participación	12.12	11.98	0.505	0.614	0.052
Práctica pertinente	11.70	11.57	0.443	0.658	0.046
Mejoramiento	9.92	9.83	0.398	0.691	0.041
Desempeño resultante	12.39	12.34	0.162	0.871	0.017
Total Desempeño estudiantil	65.67	64.76	0.701	0.484	0.072

Tabla H.2

Pruebas comparativas para el análisis de las variables de desempeño docente y desempeño estudiantil percibido según rango de edad

Variables	22 años a menos (n=245)	Mayor a 23 años (n=230)	<i>t</i>	<i>p</i>	<i>d</i> de Cohen
Exploración Competencial	8.70	8.87	-0.559	0.576	-0.051
Explicitación de criterios	12.32	12.59	-0.776	0.438	-0.071
Ilustración	11.78	12.20	-1.293	0.197	-0.119
Supervisión de la práctica	9.74	10.13	-1.248	0.213	-0.115
Retroalimentación	11.59	12.20	-1.631	0.104	-0.150
Evaluación	9.60	9.92	-1.020	0.308	-0.094
Total Desempeño docente	63.74	65.90	-1.241	0.215	-0.114
Precurrentes y potenciales habilidades	8.44	8.93	-2.416	0.016	-0.222
Identificación de criterios	10.36	10.64	-1.356	0.176	-0.125
Participación	11.75	12.31	-2.227	0.026	-0.204
Práctica pertinente	11.40	11.82	-1.572	0.117	-0.144
Mejoramiento	9.74	9.97	-1.171	0.242	-0.108
Desempeño resultante	12.17	12.56	-1.481	0.139	-0.136
Total Desempeño estudiantil	63.86	66.23	-2.061	0.040	-0.189

Tabla H.3

Pruebas comparativas para el análisis de las variables de desempeño docente y desempeño estudiantil percibido según nivel de especialidad de carrera

Variables	General (n=275)	Especialidad Prof. (n=200)	<i>t</i>	<i>p</i>	<i>d</i> de Cohen
Exploración Competencial	8.55	9.10	-1.724	0.085	-0.160
Explicitación de criterios	12.29	12.67	-1.093	0.275	-0.102
Ilustración	11.88	12.13	-0.740	0.460	-0.069
Supervisión de la práctica	9.73	10.20	-1.483	0.139	-0.138
Retroalimentación	11.59	12.30	-1.895	0.059	-0.176
Evaluación	9.57	10.02	-1.441	0.150	-0.134
Total Desempeño docente	63.61	66.41	-1.584	0.114	-0.147
Precurrentes y potenciales habilidades	8.51	8.92	-1.999	0.046	-0.186
Identificación de criterios	10.43	10.59	-0.717	0.474	-0.067
Participación	11.86	12.23	-1.457	0.146	-0.135
Práctica pertinente	11.37	11.91	-1.974	0.049	-0.183
Mejoramiento	9.79	9.94	-0.726	0.468	-0.067
Desempeño resultante	12.26	12.49	-0.898	0.369	-0.084
Total Desempeño estudiantil	64.23	66.08	-1.590	0.113	-0.148

Tabla H.4

Pruebas comparativas para el análisis de las variables de desempeño docente y desempeño estudiantil percibido según turno de estudios

Variables	Mañana (n=205)	Tarde (n=270)	<i>t</i>	<i>p</i>	<i>d</i> de Cohen
Exploración Competencial	7.79	9.54	-5.716	<.001	-0.530
Explicitación de criterios	11.79	12.95	-3.352	<.001	-0.311
Ilustración	11.26	12.53	-3.965	<.001	-0.367

Supervisión de la práctica	9.33	10.38	-3.425	<.001	-0.317
Retroalimentación	11.09	12.49	-3.754	<.001	-0.348
Evaluación	9.21	10.17	-3.098	0.002	-0.287
Total Desempeño docente	60.48	68.06	-4.386	<.001	-0.406
Precurrentes y potenciales habilidades	8.68	8.68	-0.017	0.987	-0.002
Identificación de criterios	10.42	10.55	-0.602	0.548	-0.056
Participación	11.97	12.06	-0.365	0.715	-0.034
Práctica pertinente	11.55	11.64	-0.358	0.720	-0.033
Mejoramiento	9.86	9.85	0.053	0.958	0.005
Desempeño resultante	12.12	12.54	-1.615	0.107	-0.150
Total Desempeño estudiantil	64.59	65.33	-0.630	0.529	-0.058

Tabla H.5

Pruebas comparativas para el análisis de las variables de desempeño docente y desempeño estudiantil percibido según eje curricular

Variables	Investigación (n=167)	Básica (n=112)	Especialidad (n=196)	<i>F</i>	<i>p</i>	ω^2
Exploración	8.75	8.29	9.10	2.022	0.133	0.004
Competencial						
Explicitación de criterios	13.13	11.12	12.63	10.357	0.000	0.038
Ilustración	12.35	11.21	12.12	3.772	0.024	0.012
Supervisión de la práctica	10.28	8.96	10.19	6.368	0.002	0.022
Retroalimentación	12.44	10.37	12.29	10.710	0.000	0.039

Evaluación	10.16	8.81	9.95	6.113	0.002	0.021
Total Desempeño docente	67.10	58.75	66.27	7.683	0.001	0.027
Precurrentes y potenciales habilidades	8.53	8.43	8.95	2.583	0.077	0.007
Identificación de criterios	10.57	10.19	10.61	1.361	0.257	0.002
Participación	11.95	11.72	12.25	1.391	0.250	0.002
Práctica pertinente	11.65	10.89	11.97	4.835	0.008	0.016
Mejoramiento	9.94	9.49	9.99	2.185	0.114	0.005
Desempeño resultante	12.66	11.74	12.45	3.743	0.024	0.011
Total Desempeño estudiantil	65.29	62.46	66.22	3.265	0.039	0.009

Tabla H.6

Pruebas comparativas para el análisis de las variables de desempeño docente y desempeño estudiantil percibido según ciclo de estudios

Variables	Primero (n=83)	Tercero (n=106)	Quinto (n=87)	Séptimo (n=116)	Noveno (n=83)	<i>F</i>	<i>p</i>	ω^2
Exploración Competencial	10.23	7.73	8.08	9.03	9.07	7.876	0.000	0.055
Explicitación de criterios	14.07	11.71	11.36	12.31	13.11	7.800	0.000	0.054
Ilustración	13.23	11.43	11.23	11.95	12.30	4.594	0.001	0.029

Supervisión de la práctica	11.55	8.99	8.99	9.97	10.43	9.858	0.000	0.069
Retroalimentación	14.12	10.57	10.49	11.83	12.88	14.267	0.000	0.100
Evaluación	11.00	9.12	8.92	9.67	10.31	6.011	0.000	0.040
Total								
Desempeño docente	74.21	59.54	59.07	64.76	68.11	10.487	0.000	0.074
Precurrentes y potenciales habilidades	8.47	8.28	8.82	8.76	9.15	2.044	0.087	0.009
Identificación de criterios	10.68	10.14	10.58	10.62	10.52	0.878	0.477	0.000
Participación	11.94	11.69	12.01	12.17	12.31	0.721	0.578	0.000
Práctica pertinente	12.00	10.79	11.43	12.14	11.68	3.480	0.008	0.020
Mejoramiento	10.21	9.42	9.77	9.82	10.19	2.326	0.056	0.011
Desempeño resultante	12.98	11.89	12.14	12.52	12.35	1.954	0.100	0.008
Total								
Desempeño estudiantil	66.27	62.21	64.74	66.03	66.19	1.911	0.107	0.008

TITULO	PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES	METODOLOGÍA
Desempeño docente y desempeño estudiantil percibidos por estudiantes de psicología de una universidad nacional de Lima, 2019	Problema general ¿Cuál es la relación entre el desempeño docente y desempeño estudiantil percibidos por estudiantes de psicología de una universidad pública de Lima – 2019?	Objetivo general. Determinar la relación entre el desempeño docente y desempeño estudiantil percibidos por estudiantes de psicología de una universidad estatal de Lima – 2019. Objetivos específicos. 1) Identificar las fuentes de evidencias de fiabilidad, contenido del test, estructura interna global y por grupos, de los Cuestionarios sobre desempeño docente (CDD) y desempeño estudiantil (CDE) en estudiantes de la carrera de psicología de una universidad pública de Lima. 2) Identificar la relación entre las dimensiones del desempeño docente y desempeño estudiantil percibidos en estudiantes de la carrera de psicología de una universidad pública de Lima. 3) Identificar las diferencias grupales en la variable desempeño docente, según las variables de control, en estudiantes de la carrera de psicología de una universidad pública de Lima. 4) Identificar las diferencias grupales en la variable desempeño estudiantil, según las variables de control, en estudiantes de la carrera de psicología de una universidad pública de Lima.	Hipótesis general. Existe una correlación positiva entre el desempeño docente y desempeño estudiantil percibidos por estudiantes de psicología de una universidad pública de Lima, 2019 Hipótesis específicas. 1) Existen relaciones entre las dimensiones del desempeño docente y desempeño estudiantil percibidos en estudiantes de la carrera de psicología de una universidad pública de Lima. 2) Existen diferencias grupales en la variable desempeño docente, según las variables de control, en estudiantes de la carrera de psicología de una universidad pública de Lima. 3) Existen diferencias grupales en la variable desempeño estudiantil, según las variables de control, en estudiantes de la carrera de psicología de una universidad pública de Lima.	Desempeño docente percibido por el estudiante Dimensiones: a. Exploración competencial b. Explicitación de criterios c. Ilustración d. Supervisión de la práctica e. Retroalimentación f. Evaluación Desempeño estudiantil percibido por el propio estudiante a. Precurrentes y potenciales habilidades b. Identificación de criterios c. Participación d. Practica pertinente e. Mejoramiento f. Desempeño resultante	Enfoque: cuantitativo Clase: investigación empírica e investigación instrumental Estrategia: asociativa comparativa y predictiva Diseño: diseño correlacional simple diseño de cohortes transversal (Ato et al., 2013).

Anexo J. Baremos del CDD y CDE

Calificación del Cuestionario de Desempeño docente - CDD

- 1- El puntaje general se obtiene mediante suma directa de todos los ítems.
- 2- Se agrupan todas las puntuaciones por docente evaluado y se obtiene la mediana
- 3- La cifra de la mediana se ubica en la tabla de baremos, para calificar el nivel de desempeño docente alcanzado, de acuerdo a la percepción de un salón de estudiantes:

Nivel de desempeño	Puntajes de corte
Alto desempeño	>82.5
Desempeño adecuado	82.5-66
En proceso	65-50
En inicio	< 50

Calificación del Cuestionario de Desempeño estudiantil - CDE

- 1- El puntaje general se obtiene mediante suma directa de todos los ítems.
- 2- Cada estudiante tiene su propio puntaje, el cual puede ser promediado según la agrupación de interés (género, salón, especialidad o alguna otra característica). El puntaje por grupo obtenido debe ser la mediana
- 3- La cifra de la mediana se ubica en la tabla de baremos, para calificar el nivel de desempeño estudiantil alcanzado en el grupo segmentado. También se puede calificar a cada estudiante individualmente.

Nivel de desempeño	Puntajes de corte
Avanzado	>75
Intermedio	75-59
Básico	< 59