

# INFORME TÉCNICO

---

PROPIEDADES PSICOMÉTRICAS DE LA  
EVALUACIÓN PARA LA OBTENCIÓN DEL  
CERTIFICADO DE ESPAÑOL NIVEL AVANZADO



**FILO:UBA**  
Facultad de Filosofía y Letras



AGOSTO DE 2024

---

Laboratorio de Idiomas, Facultad de Filosofía y Letras –  
Universidad de Buenos Aires

Autor: Dr. Juan Pablo Barreyro





**FILO:UBA**  
Facultad de Filosofía y Letras



**LABORATORIO DE  
IDIOMAS**  
UNIVERSIDAD DE BUENOS AIRES

El presente informe técnico tiene como objetivo principal el análisis exhaustivo de las propiedades psicométricas de la evaluación escrita del Certificado de Español Avanzado (CEA) del Laboratorio de Idiomas de la Facultad de Filosofía y Letras de la Universidad de Buenos Aires. Este trabajo se enmarca en el servicio tecnológico (STAN) denominado "Consultoría para la adaptación de instrumentos psicométricos" (ST3868), del Centro Interdisciplinario de Investigaciones en Psicología Matemática y Experimental (CIIPME), contratado a través de los servicios tecnológicos de CONICET, a cargo del Dr. Juan Pablo Barreyro, Investigador Independiente del Consejo de Investigaciones Científicas y Técnicas.



<b>ÍNDICE</b>	<b>PÁG.</b>
<b>INTRODUCCIÓN</b>	4
<b>1. ANÁLISIS DE LA EVALUACIÓN INTER-JUECES</b>	6
<b>1.1. Objetivos</b>	6
<b>1.2. Metodología</b>	6
1.2.1. Descripción de los jueces	7
1.2.2. Procedimiento	7
<b>1.3. Resultados</b>	8
1.3.1. Análisis Cuantitativo de la Confiabilidad Inter-jueces	8
1.3.2. Análisis Cualitativo de la Confiabilidad Inter-jueces	10
<b>1.4. Conclusiones</b>	11
<b>2. ANÁLISIS PSICOMÉTRICO Y ESTADÍSTICO DE LA EVALUACIÓN ESCRITA PARA OBTENER EL CEA</b>	13
<b>2.1. Objetivos</b>	13
<b>2.2. Aspectos Metodológicos</b>	13
2.2.1. Muestra	13
2.2.2. Análisis de datos	14
<b>2.3. Análisis Descriptivos de la Evaluación</b>	15
2.3.1. Estadísticos Descriptivos Globales de la Evaluación	15
2.3.2. Estructura Percentilar de la Evaluación	16
2.3.3. Estadísticos Descriptivos de las Diferentes Partes de la Evaluación	17
<b>2.4. Análisis de la Confiabilidad de la Evaluación</b>	19
2.4.1. Análisis de la Confiabilidad Global	19
2.4.2. Análisis de la Consistencia Interna de los Ítems	20
<b>2.5. Análisis de la Validez de la Evaluación</b>	22
2.5.1. Análisis de Correlación entre las diferentes partes de la prueba	23
2.5.2. Análisis Factorial Confirmatorio de la Evaluación	24
<b>2.6. Análisis de Relaciones con Nacionalidad, Lengua Materna, Edad y Estudios Previos</b>	26
<b>2.7. Análisis de la Probabilidad de Obtención del CEA en Función del Desempeño en la Prueba Escrita</b>	28
<b>2.8. Conclusiones</b>	31

## INTRODUCCIÓN

---

El presente informe técnico **aborda** el análisis correspondiente con los objetivos **de análisis de las propiedades psicométricas** de la evaluación escrita del Certificado de Español Avanzado, del Laboratorio de Idiomas de la Facultad de Filosofía y Letras de la Universidad de Buenos Aires.

El **Certificado de Español Avanzado (CEA)** es un examen de la **Facultad de Filosofía y Letras de la Universidad de Buenos Aires**, diseñado para evaluar la capacidad de los estudiantes para hablar, entender, leer y escribir en un entorno universitario. El examen consta de dos partes: una escrita y una oral. La parte escrita incluye una actividad de comprensión oral, dos de comprensión lectora y dos de producción escrita, en las cuales los aspirantes deben leer textos académicos, escribir textos breves, y escuchar fragmentos de clases o conferencias. Tras aprobar la parte escrita, los candidatos deben realizar una entrevista oral con un profesor y una breve exposición sobre un tema específico. El perfil del alumno que alcanza el nivel avanzado refleja su capacidad para desempeñarse con fluidez en una amplia gama de situaciones en lengua segunda o extranjera, tanto en el ámbito laboral como en el académico. El alumno **es capaz** de escribir con claridad y complejidad, producir textos de opinión, ensayos académicos, y narraciones, y expresarse oralmente con fluidez, espontaneidad y un repertorio amplio de vocabulario. Además, puede comprender textos extensos y detallados, incluso aquellos que no están relacionados con su área de especialidad, y mantener diálogos sobre temas abstractos y complejos, reconociendo expresiones idiomáticas y diferencias en registros formales e informales.

En primer lugar, **y con el propósito** asegurar la calidad, validez y fiabilidad del instrumento de evaluación escrita **se llevó a cabo** un proceso de evaluación Inter-jueces. En esta fase, un panel de jueces expertos **revisó y evaluó** detalladamente la prueba, con el fin de identificar fortalezas y

debilidades en los ítems, asegurando que la prueba cumpla con los estándares necesarios en aspectos como claridad, pertinencia y equidad.

Posteriormente, se procedió con el análisis psicométrico de la evaluación de la prueba en su fase final. El objetivo de este análisis consistió en obtener indicadores de la confiabilidad de la prueba en su conjunto, y asimismo analizar la validez del instrumento. Se analizarán los ítems de la prueba y también el desempeño de la prueba en su conjunto.

Ambos objetivos serán presentados con un enfoque detallado y riguroso, en dos apartados de este informe. El primer apartado se centrará en el análisis Inter-jueces y el segundo se enfocará en el análisis de la evaluación escrita para la obtención del certificado de español avanzado.

## 1. ANÁLISIS DE LA EVALUACIÓN INTER-JUECES

---

### 1.1. *Objetivos*

El propósito del análisis inter-jueces consiste en garantizar la confiabilidad y validez de la prueba diseñada para la obtención del Certificado de Español Avanzado, a partir de la revisión del examen por parte de jueces expertos. Para ello cuatro jueces expertos analizaron diversos aspectos de la prueba, como la claridad, relevancia, cobertura de contenidos, equidad, variedad de ítems, dificultad global, tiempos de realización, distribución de puntajes, y las claves de corrección y escalas utilizadas. El análisis de esta revisión por parte de los jueces se muestra en este apartado.

### 1.2. *Metodología*

Se realizó un análisis de confiabilidad Inter-jueces empleando los métodos **Kappa ponderado de Cohen** y **Kappa de Fleiss**. El Kappa ponderado de Cohen es una medida estadística que evalúa el grado de acuerdo entre dos evaluadores (jueces) teniendo en cuenta la magnitud de las diferencias en sus observaciones, el análisis ponderado estima un promedio de todas las evaluaciones por pares entre los evaluadores. Por otro lado, el Kappa de Fleiss es una extensión del Kappa de Cohen, diseñada para medir el acuerdo entre más de dos jueces, proporcionando una medida de confiabilidad global. En cuanto a los indicadores de referencia, un valor de Kappa superior a .75 indican un buen grado de acuerdo, valores entre .40 y .75 sugieren un acuerdo moderado, y valores inferiores a .40 indican un acuerdo bajo. En este estudio, cuatro jueces expertos evaluaron la prueba diseñada por el equipo de profesionales del Laboratorio de Idiomas de la Facultad de Filosofía y Letras de la Universidad de Buenos Aires, para la obtención del Certificado de Español Avanzado por parte de la Universidad.

Los jueces, con amplia experiencia en el campo de la enseñanza y evaluación del español como Lengua Segunda y Extranjera (ELSE), fueron seleccionados por su expertise en áreas

específicas como intervención en el lenguaje, lingüística aplicada, teoría literaria, y tecnología educativa en ELSE.

#### 1.2.1. Descripción de los Jueces

Juez 1: Experto en intervención en el lenguaje, con un doctorado en esta área y un magíster en Psicología Cognitiva y Aprendizaje (Facultad Latinoamericana de Ciencias Sociales, sede argentina). Su experiencia incluye enseñanza en universidades y asesoría en el Ministerio de Educación de la Nación, especialmente en la enseñanza de aspectos gramaticales y léxicos a estudiantes sordos.

Juez 2: Doctor en Lingüística aplicada a ELSE, con un enfoque de investigación en la interlengua fónica y su impacto en la didáctica de la pronunciación. Ha coordinado consorcios de ELSE y dirige proyectos de investigación sobre enseñanza y evaluación en ELSE.

Juez 3: Especialista en teoría literaria y enseñanza de español como segunda lengua. Posee un magíster en ELSE y ha coordinado el Consorcio Interuniversitario ELSE. Su investigación se centra en narrativa argentina contemporánea y la evaluación en ELSE.

Juez 4: Experta externa en interacción multimodal y tecnologías educativas en ELSE. Su investigación se enfoca en el diseño de materiales didácticos y enfoques interculturales y pragmáticos en la enseñanza de ELE.

#### 1.2.2. Procedimiento

Para garantizar la confiabilidad y validez de la evaluación de la prueba para obtener el Certificado de Español Avanzado, se llevó a cabo un proceso de revisión por los cuatro jueces expertos en el área. Se proporcionó a cada juez una copia completa de la prueba, que incluye ejercicios de comprensión auditiva, comprensión lectora y producción escrita. Junto con la prueba, se entregó una planilla de evaluación que contenía una escala Likert para puntuar diferentes aspectos de la prueba.

La planilla de evaluación utilizó una escala Likert de 5 puntos (que iba de 1 punto esto es, Muy en desacuerdo, a 5 puntos, esto es, Muy de acuerdo) para puntuar en cada uno de los ítems de la prueba en los aspectos de claridad, que se refiere a si las instrucciones o las preguntas son comprensibles para los estudiantes, y relevancia, que se refiere a si el ítem que se está evaluando es adecuado para el nivel de competencia avanzado. Por otro lado, también se solicitó a los jueces que evaluarán la prueba en su conjunto, tanto a nivel cuantitativo (empleando la misma escala Likert) como cualitativo. Para este punto se solicitó a los jueces evaluar: (a) la cobertura de contenidos (Evaluación de la exhaustividad de las competencias lingüísticas necesarias para el nivel avanzado), (b) la equidad (Evaluación de la justicia de la prueba y la ausencia de sesgo a favor de algún tipo de estudiantes en particular), (c) la variedad de Ítems (Evaluación de la variedad y complejidad de los ítems incluidos en la prueba), (d) la dificultad global (Evaluación del nivel de dificultad de las preguntas, adecuado para el nivel avanzado), (e) El tiempo de realización (Evaluación de la adecuación del tiempo estipulado para completar la prueba), (f) la distribución de puntajes (Evaluación de la adecuación de la distribución de puntajes por tareas) y (g) las claves de corrección y escalas (Evaluación de la claridad y completitud de las claves de corrección y escalas utilizadas para evaluar).

Los jueces completaron la planilla de evaluación, proporcionando tanto puntuaciones numéricas como comentarios cualitativos en cada una de las categorías mencionadas.

### **1.3. Resultados**

#### 1.3.1. Análisis Cuantitativo de la Confiabilidad Inter-jueces

El análisis de **Kappa ponderado de Cohen** mostró un valor de **.89**, indicando una muy buena confiabilidad entre los evaluadores. Los valores de confiabilidad por pares entre los evaluadores oscilaron entre 0.85 y 0.98. El valor de **Kappa de Fleiss** fue de **.52**, con un error estándar (EE) de **.05** y un intervalo de confianza (IC) del 95% entre .43 y .61 ( $z = 11.2, p < .001$ ), lo que indica una confiabilidad moderada y significativa.

La **proporción total de acuerdo** entre los cuatro observadores fue del **.85 (85.3%)**. El valor mínimo de acuerdo es 0 (sin acuerdo) y el valor máximo de acuerdo es 1 (acuerdo completo entre los cuatro jueces). En términos de **claridad**, el acuerdo fue del **.80**, mientras que en **relevancia** fue del **.92**. En el **análisis global de la evaluación**, el acuerdo fue del **.82**. En la Figura 1 pueden observarse los valores obtenidos promedio.

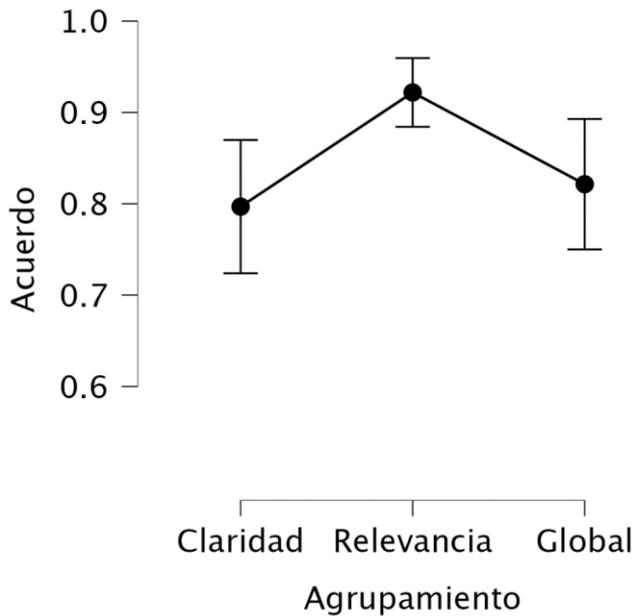


Fig. 1. Valores promedio del acuerdo de claridad, relevancia y análisis global.

Las observadoras evaluaron la claridad de las preguntas con un valor promedio de 4.84 ( $DE = 0.20$ ), la relevancia con un valor promedio de 4.92 ( $DE = 0.15$ ) y el análisis global de la prueba con un valor promedio de 4.82 ( $DE = 0.19$ ), en una escala de 1 (muy en desacuerdo) a 5 (muy de acuerdo). En la siguiente figura, pueden observarse los valores promedios para cada categoría analizada.

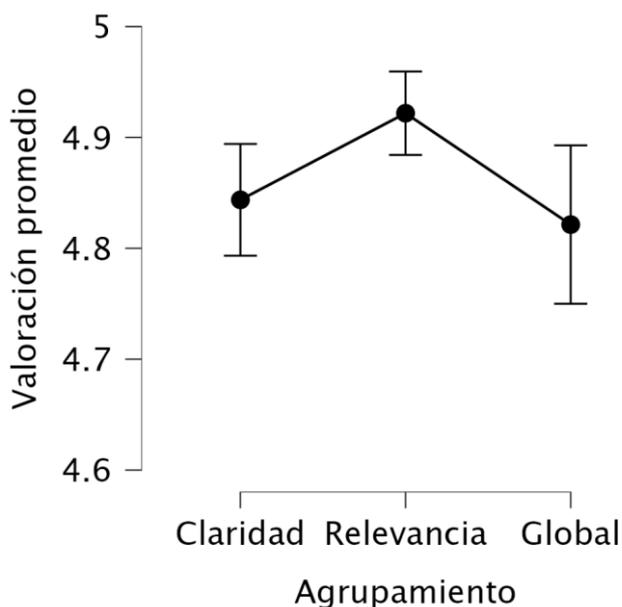


Fig. 2. Valores promedio de la valoración analizada de la claridad, la relevancia y del análisis global.

### 1.3.2. Análisis Cualitativo de la Confiabilidad Inter-jueces

Al revisar las observaciones realizadas por los jurados, se pudo delinear teniendo en cuenta convergencias y divergencias entre las observaciones los diferentes aspectos según cada uno de los criterios analizados. A continuación, se presentan el resultado del análisis según cada uno de los criterios.

**Cobertura de Contenidos:** Los jueces coincidieron en que la prueba abarca de manera exhaustiva las competencias lingüísticas necesarias para la obtención del certificado de nivel avanzado de español, específicamente en las áreas de comprensión auditiva, comprensión lectora y producción escrita.

**Equidad:** Todos los jueces estuvieron de acuerdo en que la prueba es justa y no favorece a ningún tipo de estudiante en particular. La evaluación fue diseñada para reflejar los géneros discursivos y contextos relevantes para los destinatarios previstos, incluyendo tanto contextos académicos como laborales.

**Variación de Ítems:** La prueba incluye una variedad de ítems que permiten una evaluación completa de las competencias que se requieren para obtener el nivel avanzado de español. Se destacó la inclusión de preguntas de comprensión auditiva y lectora, así como de producción escrita, utilizando tanto respuestas cerradas (verdadero o falso) como abiertas (desarrollo de respuestas y escritura de textos).

**Dificultad Global:** Los jueces coincidieron en que el nivel de las preguntas es adecuado para el nivel avanzado. La evaluación fue considerada consistente con las expectativas de competencia para el nivel evaluado.

**Tiempos de Realización de la Prueba:** En general, los jueces estuvieron de acuerdo en que el tiempo asignado para la realización de la prueba es adecuado. Sin embargo, un juez sugirió que el tiempo podría ser un poco ajustado y recomendó la inclusión de intervalos para completar distintas partes de la evaluación.

**Distribución de Puntajes:** La distribución de puntajes por tarea fue considerada adecuada. Se destacó que la ponderación de los puntajes refleja apropiadamente las competencias evaluadas, con una asignación significativa a la producción escrita. Sin embargo, un juez mencionó la necesidad de clarificar el puntaje total requerido para aprobar la prueba y obtener el nivel avanzado.

**Claves de Corrección y Escalas:** Las claves de corrección y las escalas utilizadas fueron consideradas completas y claras, adecuadas para evaluar el nivel avanzado. Un juez sugirió que se podría mejorar la guía para corregir casos de interferencia lingüística de lenguas cercanas, como el portugués o el italiano, que pueden afectar la interlengua de los estudiantes sin impedir la comprensión.

#### **1.4. Conclusiones**

El análisis de confiabilidad inter-jueces realizado en este indica un alto grado de consistencia y precisión en la **evaluación del Certificado de Español Avanzado**, especialmente en las áreas de

comprensión auditiva, comprensión lectora y producción escrita. Los resultados cuantitativos indican un acuerdo muy bueno entre los evaluadores, con un valor de Kappa ponderado de Cohen alto y una moderada pero significativa confiabilidad global reflejada en el valor de Kappa de Fleiss. La alta proporción de acuerdo total (85.3%) y los valores elevados en las categorías de claridad y relevancia, junto con las valoraciones cualitativas, sugieren que la prueba para la evaluación escrita del examen para obtener el certificado de español avanzado es exhaustiva, justa, y adecuada para medir las habilidades y competencias lingüísticas necesarias para el nivel avanzado de español, necesario para sobrellevar actividades académicas y profesionales.

## 2. ANÁLISIS PSICOMÉTRICO Y ESTADÍSTICO DE LA EVALUACIÓN ESCRITA PARA OBTENER EL CEA

---

### 2.1. *Objetivos*

El propósito del análisis psicométrico y estadístico de la evaluación escrita para la obtención del **Certificado de Español Avanzado (CEA)** fue recopilar datos, a partir de su administración en un amplio grupo de estudiantes, que proporcionaran información crucial sobre la fiabilidad y validez de la evaluación. Estos indicadores resultan esenciales para validar la prueba y garantizar su calidad antes de su implementación a gran escala. Los datos obtenidos se han analizado permitiendo obtener evidencia cuantitativa sobre la consistencia interna, la fiabilidad de los ítems y la estructura factorial de la prueba, facilitando así su optimización y validación final. Asimismo, se analizó el nivel de dificultad de cada uno de los ítems de la prueba, la relación de la prueba escrita con la prueba oral, el nivel de asociación entre la prueba, la nacionalidad de los estudiantes, su lengua materna y el haber realizado los cursos de español dentro del Laboratorio de Idiomas. Finalmente, se analizó la probabilidad de obtener el certificado de español avanzado, de acuerdo con el puntaje obtenido en cada una de las partes del examen escrito (comprensión lectora, comprensión oral y producción escrita).

### 2.2. *Aspectos Metodológicos*

#### 2.2.1. Muestra

La muestra estuvo conformada por 152 estudiantes que realizaron la evaluación escrita para la obtención del Certificado de Español Avanzado, de los cuales 67.10% (102) fueron mujeres y 31.89%, hombres. Con respecto a la edad de los estudiantes, el promedio de edad fue de 24.10 años ( $DE = 6.18$ ) con un rango de entre 18 a 56 años. Con respecto a la nacionalidad de los participantes 143 eran brasileños (94.07%), cuatro eran de origen ruso (2.63%), dos de nacionalidad argentina (1.31%), y los tres restantes eran, uno de la India, otro Italia y el último proveniente de Canadá. En relación con la primera lengua, 144 tenían como primera lengua el portugués (94.73%),

cuatro el ruso (2.63%), y los restantes cuatro estudiantes tenían como lengua materna, uno el hindi, otro el italiano, el siguiente el inglés y finalmente el último, el chino.

### 2.2.2. Análisis de datos

Los datos recolectados de la evaluación fueron analizados con el fin de obtener información acerca de las características psicométricas de la prueba y además obtener información estadística.

En primer lugar, se calcularon los estadísticos descriptivos de la evaluación, obteniendo así información sobre la tendencia central de la prueba a partir de la *media* y la *mediana*, y sobre la variabilidad y dispersión de las puntuaciones mediante los puntajes *mínimos*, *máximos* y *el desvío estándar*. Posteriormente, se analizaron los aspectos relacionados con la forma de la distribución. Para ello, se calcularon los indicadores de *asimetría* y *curtosis*, que proporcionaron información sobre la simetría y la forma de la distribución. Además, se presentó un histograma de los puntajes para visualizar gráficamente la distribución de los datos.

En segundo lugar, se obtuvieron los *percentiles* de la prueba desde el percentil 10 al 90. Se describieron los resultados de las medidas de las diferentes partes de la evaluación. La prueba escrita consta de tres partes, una que evalúa la comprensión lectora, otra la comprensión oral y la última que evalúa la producción escrita.

Se analizó la confiabilidad global de la prueba a partir de los coeficientes de *Alpha de Cronbach* y *Omega de McDonald*. También se analizaron los ítems, de ellos se obtuvo la información de la *media*, su *desvío estándar*, el *nivel de dificultad*, el valor de *Alpha de Cronbach* u *Omega de McDonald si el ítem es eliminado* y el coeficiente de *correlación del ítem con la prueba total*.

Con el fin de obtener indicadores, acerca de la validez de la prueba, se llevaron a cabo correlaciones entre las tres partes de la prueba escrita (comprensión lectora, comprensión oral y producción escrita), y también entre estas y el valor obtenido en la evaluación oral final (este análisis se realizó con aquellos que lograron obtener el puntaje mínimo necesario para pasar al examen oral). Se llevó a cabo un análisis factorial confirmatorio, con el propósito de testear la organización de los ítems en las diferentes partes de la prueba, y con el propósito de poner a prueba la hipótesis

de un único factor latente (de segundo orden) correspondiente con las habilidades y destrezas necesarias para alcanzar las aptitudes escritas necesarias para encaminar al examen oral y así obtener el certificado de español avanzado.

Se llevó a cabo un análisis de relaciones entre la puntuación global de la prueba escrita, la nacionalidad, la primera lengua, la edad y estudios previos dentro del Laboratorio de Idiomas de la Facultad de Filosofía y Letras de la Universidad. Finalmente, a partir de análisis de regresión logística binaria, se modeló a partir de los puntajes obtenidos la probabilidad de obtener el Certificado Avanzado de Español en función del desempeño en cada una de las partes de la prueba escrita y del desempeño global

### 2.3. Análisis Descriptivos de la Evaluación

#### 2.3.1. Estadísticos Descriptivos Globales de la Evaluación

En la siguiente tabla, se muestran los estadísticos descriptivos obtenidos de la evaluación escrita del examen para obtener el Certificado de Español Avanzado (ver tabla 1).

Tabla 1. Estadísticos Descriptivos de la Evaluación Escrita

Estadístico	Evaluación Escrita	
Tendencia Central	<i>Media</i>	68.26
	<i>Mediana</i>	71.25
	<i>Mínimo</i>	7.50
Dispersión	<i>Máximo</i>	100
	<i>Desvío Estándar</i>	16.59
	<i>Asimetría</i>	-0.65
Distribución	<i>Curtosis</i>	0.56

Al analizar la distribución de las puntuaciones obtenidas en la prueba, se observa que la distribución no es simétrica [ $A = -0.65$ ,  $z = -3.31$ ,  $p < .001$ ]. El valor negativo y significativo de la asimetría sugiere una inclinación hacia la izquierda, lo que indica que la distribución presenta una cola más larga en ese extremo. Es decir, hay una mayor cantidad de puntuaciones bajas (por debajo de la media) que se extienden a lo largo de un rango más amplio de valores (de 0 a 68). En contraste,

las puntuaciones por encima de la media se agrupan en un rango más estrecho (de 69 a 95), concentrándose la mayoría de los datos hacia el extremo derecho de la distribución. En cuanto a la curtosis, que analiza el aglutinamiento de datos en torno a los valores de tendencia central, se aprecia un valor normal o mesocúrtico ( $C = 0.56$ ,  $z = 1.43$ ,  $p = .16$ ). La curtosis observada sugiere una concentración de datos dentro de los valores de la tendencia central similar a una distribución normal. En la siguiente figura (ver Fig. 3) puede apreciarse el histograma de la distribución de los puntajes de la evaluación.

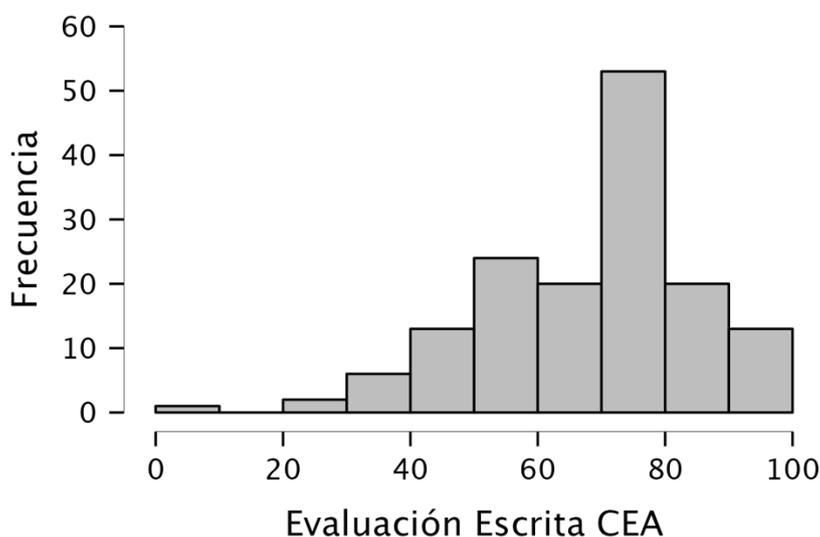


Fig. 3. Histograma de la Puntuación Obtenida en la Evaluación Escrita

### 2.3.2. Estructura Percentilar de la Evaluación

Los percentiles son medidas estadísticas que dividen un conjunto de datos en 100 partes iguales, cada una representando un porcentaje específico de los datos. En estadística descriptiva, los percentiles se utilizan para comprender la distribución de los datos y, en particular, para situar una puntuación o un valor en relación con el resto del conjunto de datos. Por ejemplo, si un estudiante obtiene un puntaje que se ubica en el percentil 20, esto significa que el 20% de los estudiantes obtuvieron una puntuación igual o menor, mientras que el 80% restante obtuvo una

puntuación mayor. De esta forma, los percentiles permiten comparar el rendimiento de un individuo con el de un grupo más amplio.

En la siguiente tabla (ver Tabla 3), se muestran los percentiles de la evaluación escrita. Observando la tabla, por ejemplo, un puntaje de 68.5 se encuentra en el percentil 40, lo que significa que el 40% de los estudiantes obtuvieron una puntuación igual o inferior a 68.5. De manera similar, un puntaje de 88.45 se encuentra en el percentil 90, indicando que solo el 10% de los estudiantes obtuvieron una puntuación superior.

Tabla 2. Percentiles de la Evaluación Escrita

Percentil	Puntaje de la Evaluación Escrita
10	47.05
20	54.00
30	60.15
40	68.5
50	71.25
60	73.50
70	77.35
80	81.90
90	88.45

### 2.3.3. Estadísticos Descriptivos de las Diferentes Partes de la Evaluación

A continuación, se presentan los estadísticos descriptivos de las diferentes partes que conforman el examen escrito para la obtención del Certificado de Español Avanzado, estas son, evaluación de la comprensión auditiva, evaluación de la comprensión lectora y producción escrita (ver Tabla 3).

Tabla 3. Estadísticos Descriptivos de las Diferentes Partes de la Evaluación Escrita

Estadístico		Comprensión	Comprensión	Producción
		Auditiva	Lectora	Escrita
Tendencia	<i>Media</i>	14.39	13.34	40.93
Central	<i>Mediana</i>	15.00	14.00	40.00
	<i>Mínimo</i>	2.50	4.00	0.00
	<i>Máximo</i>	20.00	20.00	60.00
Dispersión	<i>Desvío</i>	4.61	3.63	11.63
	<i>Estándar</i>			
	<i>Asimetría</i>	-0.56	-0.44	-0.58
Distribución	<i>Curtosis</i>	-0.64	-0.42	-0.47

Los valores de distribución sugieren en los tres casos una Asimetría negativa, esto es presencia de una cola más larga o un sesgo hacia la izquierda en las distribuciones. En cuanto a la Curtosis, las tres medidas, muestran valores mesocúrticos, con una concentración de datos en torno a los valores centrales similares a los de una distribución normal. Las siguientes Figuras (Fig. 4, 5, y 6) se corresponde con los histogramas de las puntuaciones obtenidas por los estudiantes en las tres partes de la evaluación escrita.

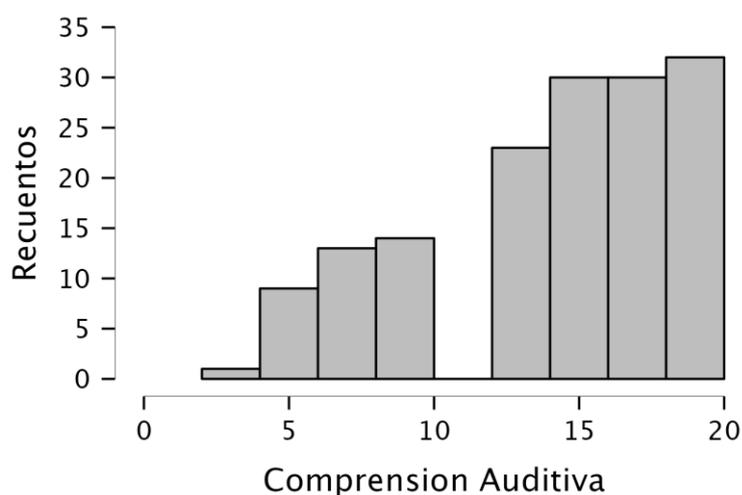


Fig. 4. Histograma de la Puntuación Obtenida en Comprensión Auditiva

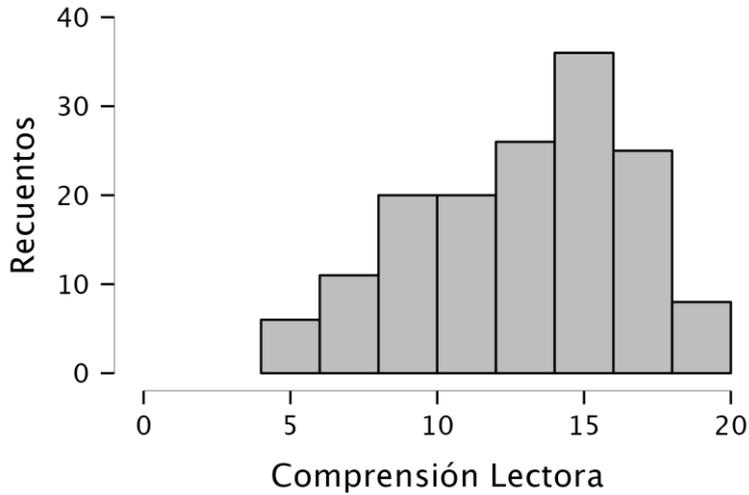


Fig. 5. Histograma de la Puntuación Obtenida en Comprensión Lectora

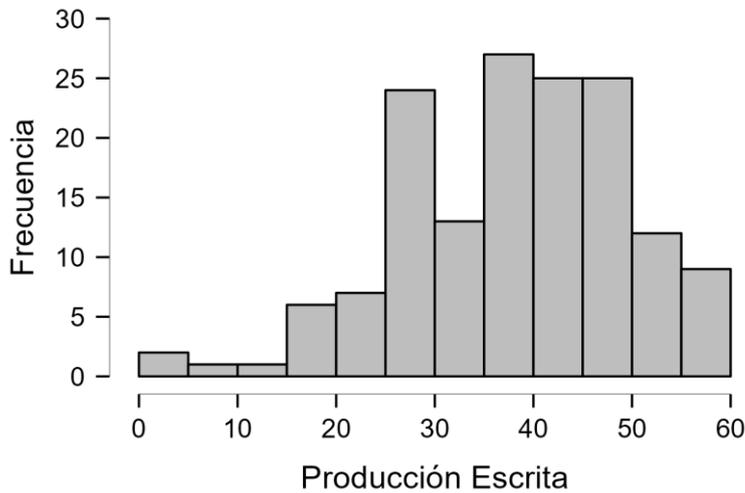


Fig. 6. Histograma de la Puntuación Obtenida en Producción Escrita

## 2.4. Análisis de la Confiabilidad de la Evaluación

### 2.4.1. Análisis de la Confiabilidad Global

La confiabilidad en una prueba se refiere a la consistencia y a la estabilidad de los resultados que produce. Esto significa que, si la prueba se aplica en diferentes momentos o a diferentes grupos similares, debería producir resultados semejantes. Las pruebas que presentan adecuados índices de confiabilidad cuentan con menores errores de medición y garantizan que las puntuaciones reflejen con precisión lo que se pretende medir. Para analizar la confiabilidad de la evaluación escrita para obtener el certificado de español avanzado se emplearon los índices de *Alpha de Cronbach* y *Omega*

de McDonald. El índice *Alpha de Cronbach* es uno de los más empleados en psicometría para evaluar la consistencia interna de un conjunto de ítems, y se obtiene a partir de correlaciones entre los ítems de la prueba. *Omega de McDonald* analiza las correlaciones entre ítems y también la contribución de cada ítem a la medición de lo que se propone evaluar. En ambos casos, valores superiores a .70 son considerados aceptables y adecuados de confiabilidad.

La confiabilidad de la evaluación escrita global muestra valores aceptables y muy buenos de confiabilidad, *Alpha de Cronbach* = .78 ( $IC_{95\%} = .72/.83$ ) y *Omega de McDonald* = .84 ( $IC_{95\%} = .75/.89$ ). Estos valores sugieren que la evaluación escrita para obtener el certificado de español avanzado muestra una buena consistencia interna, con valores de confiabilidad robustos, que sugieren que los ítems de la prueba están correlacionados de manera adecuada entre sí. Asimismo, los intervalos de confianza al 95% ( $IC_{95\%}$ ) respaldan la afirmación, mostrando que la confiabilidad estimada es bastante precisa y consistente, con valores que no son menores a .70 (límite para una confiabilidad adecuada).

#### 2.4.2. Análisis de la Consistencia Interna de los Ítems

A continuación, se analizan los ítems de la evaluación, la evaluación se encuentra conformada por 16 ítems, los ítems 1.1 al 1.4 evalúan la comprensión auditiva (4 ítems), los ítems 2.1 al 3.5 evalúan la comprensión lectora (10 ítems) y los ítems 4 y 5 producción escrita (2 ítems). En la siguiente tabla (ver Tabla 4) se muestra los valores de cada uno de los ítems referidos a su *media (M)*, *desvío estándar (DE)*, el *nivel de dificultad*, valor de *Alpha de Cronbach* u *Omega de McDonald* si el ítem es eliminado y el coeficiente de correlación del ítem con la prueba total. El índice de dificultad es un porcentaje que se interpreta como un valor que cuanto más cercano a 0 se encuentra mayor es la dificultad del ítem, y cuanto más cercano a 100 sea su valor menor será su dificultad. En cuanto al índice de correlación, los valores más cercanos a 0 sugieren que el ítem y la evaluación son independientes, pero valores superiores a .20 indican una asociación baja, valores superiores .40 indican una asociación moderada y los superiores a .60 una asociación alta.

Tabla 4. Estadísticos Descriptivos y de Confiabilidad de los Ítems de la Evaluación

Ítem	si se elimina el ítem		Correlación ítem-test	M	DE	Nivel de Dificultad
	Omega de McDonald	Alpha de Cronbach				
Ítem 1.1	.82	.76	.50	3.34	1.75	66.78
Ítem 1.2	.83	.76	.45	3.29	1.36	65.79
Ítem 1.3	.83	.76	.55	4.14	1.38	82.89
Ítem 1.4	.84	.77	.34	3.63	1.91	72.70
Ítem 2.1	.84	.78	.25	1.04	0.84	51.97
Ítem 2.2	.83	.77	.37	1.61	0.75	80.26
Ítem 2.3	.84	.77	.28	1.26	0.83	63.16
Ítem 2.4	.84	.78	.21	1.28	0.86	64.14
Ítem 2.5	.84	.79	.21	1.72	0.59	86.18
Ítem 3.1	.84	.77	.28	1.03	0.86	51.64
Ítem 3.2	.84	.77	.29	1.14	0.77	56.91
Ítem 3.3	.83	.77	.29	1.55	0.74	77.63
Ítem 3.4	.83	.76	.38	1.49	0.67	74.67
Ítem 3.5	.84	.78	.22	1.19	0.57	59.54
Ítem 4	.69	.75	.69	20.82	6.32	69.41
Ítem 5	.73	.76	.64	19.70	6.52	65.68

Los resultados del análisis de los ítems indica que la eliminación de la mayoría de ellos no altera los índices de confiabilidad general. A excepción de los ítems de producción escrita (ítem 4 y 5), estos ítems son sensibles y cruciales para la evaluación, puesto que no solo la confiabilidad global de la evaluación se reduce al excluirlos, sino también muestran valores elevados de asociación ítem-test con un índice de dificultad cercanos al promedio de la evaluación general.

Se observa que la mayoría de los ítems muestra valores adecuados de correlación ítem-test con un buen nivel de dificultad. Se presta atención, por otro lado, el ítem 2.5, vinculado a la evaluación de la comprensión auditiva que si bien muestra una ajustada correlación ítem-test,

presenta un valor de confiabilidad que al excluirlo mejora el índice *Alpha de Cronbach*, aunque no impacta su exclusión en el *Omega de McDonald*, en cuanto a su índice de dificultad se encuentra en torno al 86.18%, sugiriendo baja dificultad general.

A nivel global, la eliminación de la mayoría de los ítems no afecta significativamente los índices de *Alpha de Cronbach* ni de *Omega de McDonald*, lo que sugiere que los ítems en su conjunto están bien alineados con lo que se pretende evaluar. En cuanto a los índices de correlación, la mayoría de los ítems presentan valores adecuados de correlación ítem-test, indicando que cada ítem está bien asociado con la evaluación general y contribuyen en la medición. En relación con los índices de dificultad, se observa que los mismos varían de acuerdo con los ítems, pero en general, los ítems presentan niveles de dificultad que permiten una evaluación equilibrada.

### **2.5. *Análisis de la Validez de la Evaluación***

En el contexto de la evaluación psicométrica, la validez de una evaluación se refiere al grado en que esta mide efectivamente lo que se propone medir, en este caso y para esta evaluación, las aptitudes escritas necesarias para alcanzar el nivel avanzado de español. El análisis acerca de la validez es importante para asegurar que las inferencias y decisiones basadas en los resultados de la prueba sean precisas y adecuadas. El enfoque empleado para analizar la validez consiste en verificar la estructura teórica de la prueba, basada en un modelo de tres componentes. La prueba analiza las aptitudes necesarias para obtener el nivel avanzado de español a partir del desempeño de los estudiantes en las medidas de comprensión auditiva, comprensión lectora y producción escrita. En este caso, la validez de la prueba fue analizada mediante la correlación entre las diferentes medidas internas (comprensión auditiva, comprensión lectora, producción escrita, el puntaje total de la prueba y el puntaje obtenido en la evaluación oral). Estas correlaciones proporcionan evidencia sobre cómo se relacionan entre sí las distintas partes de la prueba, lo cual es fundamental para confirmar que todas las secciones están alineadas con el objetivo general de la evaluación. Como se ha señalado anteriormente los valores de correlaciones cercanos a 0 sugieren que las medidas

son independientes entre sí, valores superiores a .20 indican una relación baja, valores superiores a .40 indican una relación de moderada intensidad y los valores superiores a .60 da indicio de una asociación alta.

Se realizó un análisis factorial confirmatorio de segundo orden con el objetivo de examinar la estructura subyacente de la prueba. En este análisis, se identificaron tres factores de primer orden: comprensión auditiva (ítems 1.1 a 1.4), comprensión lectora (ítems 2.1 a 3.5) y producción escrita (ítems 4 y 5). Estos factores están agrupados bajo un factor de segundo orden, que representa las aptitudes escritas necesarias para obtener el Certificado de Español Avanzado (CEA). Este análisis permite evaluar si la prueba está conformada por componentes coherentes que, en conjunto, reflejan de manera precisa las competencias que se busca medir. Se examina también el grado de ajuste del modelo a los datos, evaluando la bondad de ajuste, el nivel de error asociado, y la comparación con un modelo nulo, además de analizar los valores internos del modelo propuesto.

#### 2.5.1. Análisis de Correlación entre las diferentes partes de la prueba

A continuación, se presenta la tabla de correlaciones entre las medidas de las puntuaciones de comprensión lectora, comprensión auditiva, producción escrita, evaluación escrita y evaluación oral (ver Tabla 5). Para este análisis se empleó el coeficiente de correlación por rangos de *Spearman* (*Rho*).

Tabla 5. Correlaciones entre las medidas de la Evaluación

	<b>Compr. Auditiva</b>	<b>Compr. Lectora</b>	<b>Prod. Escrita</b>	<b>Evaluación Escrita</b>
Compr. Auditiva	—			
Compr. Lectora	.41 ***	—		
Prod. Escrita	.43 ***	.42 ***	—	
Evaluación Escrita	.68 ***	.64 ***	.92 ***	—
Evaluación Oral	.30 **	.13	.24 *	.36 ***

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

Los resultados indican que las tres medidas que conforman la puntuación total de la evaluación escrita se encuentra asociadas con moderada intensidad. A su vez, todas ellas se relacionan con alta intensidad con la puntuación total de la evaluación escrita. Con respecto a las relaciones con la evaluación oral, tanto las medidas de comprensión auditiva, producción escrita y puntaje global de la evaluación escrita mostraron relaciones de baja intensidad. Estos resultados permiten sostener que la evaluación escrita mide de manera coherente las habilidades de comprensión auditiva, lectora y de producción escrita, las cuales contribuyen de forma significativa al puntaje total de la prueba escrita. A su vez las correlaciones de la evaluación escrita con la evaluación oral indican sugieren que ambas medidas están en sintonía en la evaluación general, aunque se enfoquen en aspectos diferentes.

#### 2.5.2. Análisis Factorial Confirmatorio de la Evaluación

A continuación, se presentan los resultados del análisis factorial confirmatorio realizado. Para este análisis se seleccionaron los siguientes indicadores: Para analizar el ajuste del modelo se empleó la prueba de *Chi cuadrada* como prueba de bondad de ajuste y el índice *GFI (Goodness of Fit Index)*; para analizar el nivel de error en el ajuste del modelo se seleccionó el índice *RMSEA (Root Mean Square Error of Approximation)* y como medida de comparación entre el modelo propuesto y un modelo nulo se emplearon los índices *CFI (Comparative Fit Index)* y *TLI (Tucker-Lewis Index)*. El análisis factorial confirmatorio analizó el ajuste de un modelo de segundo orden, que incluyó tres factores de primer orden, cada uno de los cuales fue definido como de comprensión lectora, comprensión auditiva y producción escrita, subordinados a un factor de segundo orden, que representa las aptitudes escritas necesarias de español avanzado.

Para interpretar la calidad del ajuste del modelo en el análisis factorial confirmatorio, se utilizaron los siguientes puntos de corte: un *p*-valor superior a .05 de la *Chi cuadrada* prueba indica un buen ajuste de los datos al modelo propuesto. El *Goodness of Fit Index (GFI)* se considera adecuado con valores superiores a .90. El *Root Mean Square Error of Approximation (RMSEA)* es aceptable con valores inferiores .08, y excelente con valores inferiores .05. El *Comparative Fit Index*

(CFI) y el *Tucker-Lewis Index* (TLI) son indicativos de un ajuste adecuado con valores superiores .90, siendo preferibles valores superiores .95, que reflejan un ajuste excelente del modelo propuesto.

Para realizar el Análisis Factorial Confirmatorio se utilizó el estimador de mínimos cuadrados ponderados con medias y varianzas ajustadas (*DWLS*, por sus siglas en inglés), el cual es adecuado para datos ordinales minimizando sesgos resultantes con estimadores menos robustos. Se empleó, también, un procedimiento robusto para calcular los *p*-valores resultantes, para llevar a cabo estimación más precisas y confiables de los parámetros debido al incumpliendo del supuesto de normalidad multivariada.

Los indicadores del modelo, una vez realizado el análisis factorial confirmatorio, mostraron un ajuste excelente con un nivel de error despreciable:  $\chi^2_{(101)} = 69.54, p = .99$ ; *GFI* = .99; *RMSEA* < .01, *CFI* > .99 y *TLI* > .99. que permite sugerir que el modelo es altamente confiable y adecuado para representar las relaciones entre las medidas y los factores analizados. Las cargas factoriales del modelo propuesto y testeado puede apreciarse en la siguiente figura (ver Figura 7).

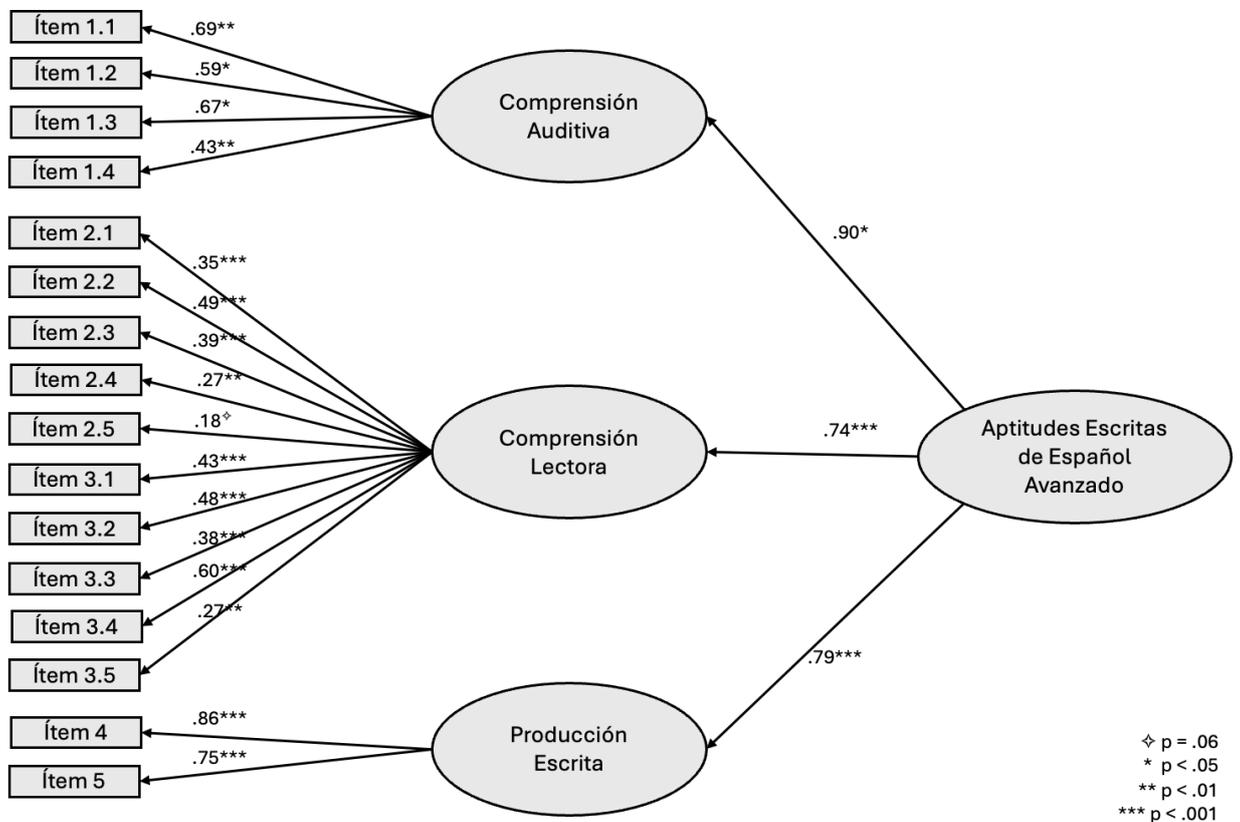


Fig. 7. Cargas factoriales del modelo propuesto

Al analizar las cargas factoriales, se puede apreciar que los factores de primer orden están bien representados en los ítems que los evalúan. En el caso de los ítems referidos a comprensión auditiva las cargas toman valores entre .43 y .69 siendo todas ellas de mediana y alta intensidad y significativas. En cuanto a comprensión lectora se aprecian cargas factoriales significativas sobre los ítems con valores de .18 a .60, de baja a alta carga factorial. Con respecto al factor de producción escrita, los ítems muestran alto nivel de saturación del factor con valores altos superiores a .76 y significativos. Respecto del factor de segundo orden definido como de las aptitudes escritas necesarias para alcanzar el nivel avanzado de español, también se aprecian cargas factoriales altas y significativas superiores a .74 sobre los factores latentes de primer orden.

En conclusión, a partir del análisis factorial confirmatorio se puede concluir que la validez de la prueba está bien sustentada, los ítems muestran cargas factoriales significativas y de mediana a alta intensidad en los factores de primer orden a los que están asociados (comprensión auditiva, comprensión lectora y producción escrita). Además, las cargas factoriales altas y significativas en el factor de segundo orden refuerzan la validez del modelo subyacente, que postula que estas habilidades están interrelacionadas y son esenciales para alcanzar un nivel avanzado de español.

## ***2.6. Análisis de Relaciones con Nacionalidad, Lengua Materna, Edad y Estudios Previos***

Se analizó la relación entre la puntuación obtenida en la evaluación escrita del CEA y variables relevantes, obtenidas de la evaluación, como, la nacionalidad de los estudiantes, su lengua materna, si realizaron estudios previos en el laboratorio de idiomas durante el cuatrimestre anterior al examen y su edad.

Para explorar la relación entre la edad y la puntuación obtenida, se utilizó el coeficiente de correlación por rangos de *Spearman* (*Rho*). El análisis no reveló una asociación entre estas variables ( $\rho = .05, p = .58$ ), lo que indica que la edad no tiene un impacto en las puntuaciones obtenidas.

Por otro lado, para examinar la relación entre la nacionalidad y la lengua materna con las

puntuaciones obtenidas, se realizó un análisis de comparación utilizando la prueba de *Kruskal-Wallis*. Mientras que, se aplicó la prueba *U* de *Mann-Whitney* para determinar si haber cursado estudios previos en el laboratorio de idiomas de español se asocia con un mejor desempeño en el examen escrito. Los resultados descriptivos de la evaluación escrita, en función de la nacionalidad, lengua materna y estudios previos, se presentan en la Tabla 6.

Tabla 6. Estadísticos descriptivos de la Evaluación Escrita por Nacionalidad, Lengua Materna, Estudio en el Laboratorio

		Resultado de la Evaluación Escrita		
		<i>M</i>	<i>DE</i>	<i>IC95%</i>
Nacionalidad	Brasileña	67.45	16.51	64.72 / 70.18
	Rusa	85.25	13.79	63.31 / 107.19
	Otras	76.70	12.52	61.15 / 92.25
Lengua Materna	Portugués	67.48	16.46	64.77 / 70.19
	Ruso	85.25	13.79	63.31 / 107.19
	Otras	77.88	14.14	55.38 / 100.37
Estudió en el Laboratorio	No	67.45	16.60	64.50 / 70.40
	Sí	71.63	16.36	65.28 / 77.97

Los análisis realizados no mostraron relaciones significativas entre la nacionalidad de los estudiantes y las puntuaciones obtenidas en la evaluación escrita ( $X^2_{(2)} = 5.70, p = .06$ ), ni entre la lengua materna y las puntuaciones ( $X^2_{(2)} = 5.89, p = .06$ ). Estos resultados sugieren que el examen no favorece ni perjudica a los estudiantes según su nacionalidad o lengua materna de origen.

Asimismo, no se encontró asociación entre haber realizado estudios previos en el laboratorio de idiomas durante el último cuatrimestre y las puntuaciones obtenidas en la evaluación escrita ( $U = 1470, p = .21$ ). Esto indica que estudiar en el laboratorio de idiomas de la Facultad de Filosofía y Letras de la UBA no influye en el desempeño en el examen.

Los análisis realizados muestran que ni la edad, ni la nacionalidad, ni la lengua materna, ni los estudios previos en el laboratorio de idiomas están relacionados significativamente con las puntuaciones obtenidas en la evaluación escrita para el Certificado de Español Avanzado (CEA).

Esto robustece la imparcialidad del examen, característica propia de un examen de estas características, indicando que la evaluación mide las competencias de manera objetiva, sin estar sesgada por las variables analizadas.

### ***2.7. Análisis de la Probabilidad de Obtención del CEA en Función del Desempeño en la Prueba Escrita***

Se realizó un análisis de regresión logística binaria para modelar la probabilidad de obtener el Certificado de Español Avanzado (CEA) en función del desempeño en las distintas partes de la prueba escrita, así como del desempeño global. Se llevaron a cabo cuatro regresiones logísticas, considerando como variable dependiente la obtención del nivel necesario para recibir el CEA tras la evaluación oral, y como variables independientes las medidas obtenidas de comprensión auditiva, comprensión lectora, producción escrita y la puntuación total de la prueba escrita para obtener CEA.

En relación con la comprensión auditiva, el modelo resultó significativo ( $X^2_{(150)} = 43.91, p < .001, R^2 = .34$ ), indicando que el desempeño en comprensión auditiva es un predictor del logro del CEA ( $\beta = .30, z = 5.40, p < .001$ ). El gráfico de estimaciones (ver Figura 8) muestra que una puntuación de 15 en comprensión auditiva se asocia con una probabilidad de menos del 50% de obtener el CEA, mientras que una puntuación cercana a 20 eleva esta probabilidad a más del 75%.

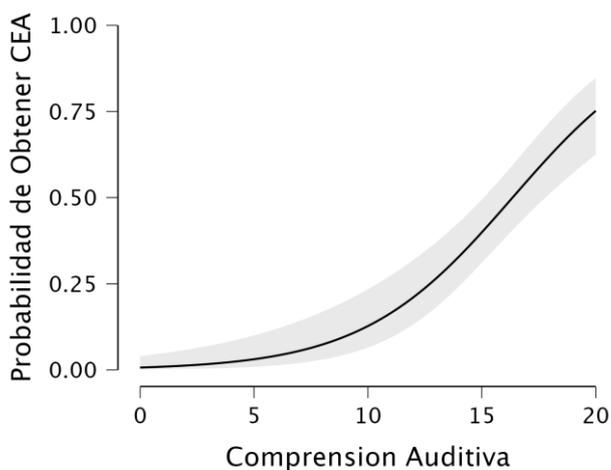


Fig. 8. Probabilidad de Obtener el CEA de acuerdo con el puntaje obtenido en Comprensión Auditiva

Con respecto a la puntuación en comprensión lectora, también se encontró que su desempeño predijo la obtención del CEA ( $X^2_{(150)} = 20.41, p < .001, R^2 = .17$ ). Según el gráfico de estimaciones (ver Figura 9), una puntuación de 15 en comprensión lectora se vincula con una probabilidad de algo más del 50% de obtener el CEA, y una puntuación cercana a 20 incrementa esta probabilidad a aproximadamente un 75%, siendo esta estimación significativa ( $\beta = .23, z = 4.11, p < .001$ ).

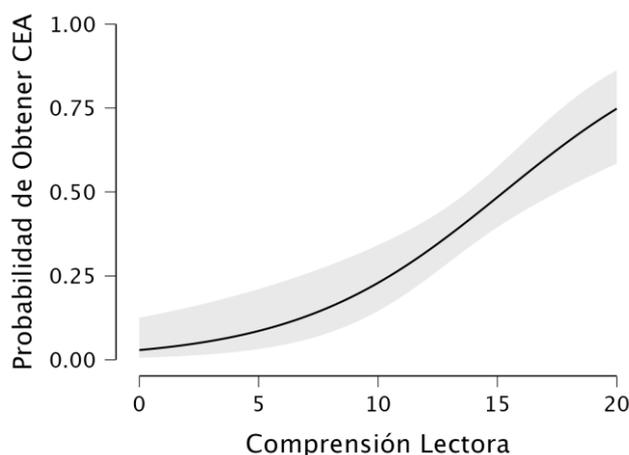


Fig. 9. Probabilidad de Obtener el CEA de acuerdo con el puntaje obtenido en Comprensión Lectora

Al analizar las puntuaciones en producción escrita, el modelo también resultó altamente significativo ( $X^2_{(150)} = 65.54, p < .001, R^2 = .47$ ). El gráfico de estimaciones (ver Figura 10) muestra que una puntuación de 40 en producción escrita se asocia con una probabilidad cercana al 50% de obtener el CEA, mientras que una puntuación de 50 eleva esta probabilidad al 75%, y una puntuación de 60 la acrecienta a casi un 99%, resultando esta estimación, también, significativa ( $\beta = .17, z = 5.97, p < .001$ ).

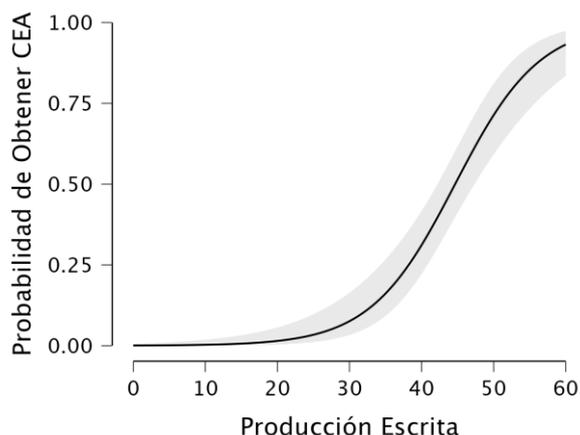
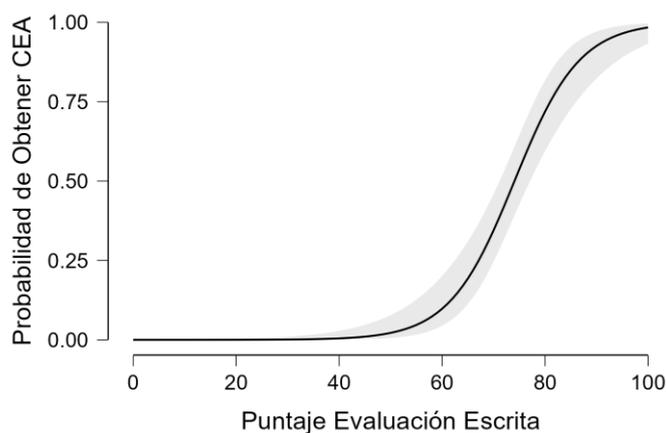


Fig. 10. Probabilidad de Obtener el CEA de acuerdo con el puntaje obtenido en Producción Escrita

Al analizar la puntuación total de la prueba escrita, el modelo también resultó significativo ( $X^2_{(150)} = 81.99, p < .001, R^2 = .56$ ), sugiriendo que el desempeño global en la prueba escrita es un fuerte predictor de la obtención del CEA significativa ( $\beta = .16, z = 5.82, p < .001$ ). Según el gráfico de estimaciones (ver Figura 11), una puntuación cercana a 70 en la prueba se asocia con una probabilidad del 25% de obtener el CEA, una puntuación cercana a 80 incrementa la probabilidad al 75%, y puntuaciones superiores a 90 se asocian con una probabilidad superior al 95% de obtener el CEA.



Los análisis de regresión logística binaria indican que las puntuaciones en comprensión auditiva, comprensión lectora, producción escrita y la puntuación total de la prueba escrita son predictores significativos de la probabilidad de obtener el Certificado de Español Avanzado (CEA).

Entre estos, el desempeño en la prueba escrita general y en producción escrita son los predictores más fuertes, seguidos por la comprensión auditiva y la comprensión lectora. Esto permite subrayar la importancia de un desempeño sólido en todas las áreas evaluadas para asegurar la obtención del CEA, con una especial relevancia en las competencias involucradas en la producción escrita.

## **2.8. Conclusiones**

El análisis psicométrico y estadístico de la evaluación escrita del Certificado de Español Avanzado (CEA) proporciona evidencia sólida sobre la fiabilidad y validez del instrumento.

Los resultados descriptivos y el análisis de las diferentes partes de la prueba revelan que, si bien las puntuaciones están distribuidas con una ligera asimetría negativa, la concentración de los puntajes se mantiene dentro de valores esperados para una evaluación de estas características. Esto sugiere que la prueba es capaz de diferenciar entre los distintos niveles de habilidad de los estudiantes.

Los análisis de confiabilidad, representados por los coeficientes Alpha de Cronbach y Omega de McDonald, indican que la prueba posee una consistencia interna adecuada, lo que sugiere que los ítems están alineados y miden constructos coherentes dentro de la evaluación. Además, la estructura factorial confirmatoria respalda la hipótesis de un factor latente de segundo orden, que engloba las habilidades necesarias para superar la evaluación escrita y avanzar al examen oral.

La ausencia de una asociación significativa entre la nacionalidad, la primera lengua y los estudios previos en el Laboratorio de Idiomas con el desempeño en la evaluación escrita sugiere que la prueba es justa y equitativa, evaluando las habilidades sin sesgos relacionados con estos factores.

Los análisis de regresión logística realizados modelaron la probabilidad de éxito en la obtención del CEA según el desempeño en las distintas partes de la prueba. Este análisis indicó la

importancia de un rendimiento equilibrado en todas las áreas evaluadas para obtener el certificado de español avanzado.

Teniendo en cuenta todos los análisis realizados, y en este contexto, los resultados obtenidos confirman que la evaluación escrita para la obtención del CEA es un instrumento confiable y válido para medir las competencias lingüísticas requeridas para el nivel avanzado de español, siendo adecuado para su uso en contextos educativos, laborales y certificativos.