

Validez y confiabilidad del instrumento para evaluar la producción de textos narrativos

Validity and Reliability of an Instrument for Assessing Narrative Text Production

Hilda Victoria Dávila Costa ¹ 

Nolberto Arnildo Leyva Aguilar ² 

Resumen

Introducción: La evaluación de la producción de textos escritos constituye un proceso fundamental en la educación primaria, debido a que permite valorar el desarrollo de competencias relacionadas con la planificación, organización y elaboración textual. Para garantizar resultados consistentes, es necesario disponer de instrumentos con evidencias adecuadas de validez y confiabilidad. **Objetivo:** Analizar la validez de constructo y la confiabilidad del instrumento AOU para evaluar la producción de textos escritos en estudiantes de cuarto grado de educación primaria de la institución educativa Abraham Valdelomar de Virú, Perú. **Metodología:** Se desarrolló una investigación cuantitativa de tipo instrumental. Se aplicó una lista de cotejo de 20 ítems, organizada en tres dimensiones, a una muestra de 30 estudiantes. La validez de contenido se determinó mediante el coeficiente V de Aiken, mientras que la validez de constructo se examinó mediante análisis factorial. La confiabilidad se estimó mediante el coeficiente alfa de Cronbach. **Resultados:** El instrumento presentó una validez de contenido alta, con un V de Aiken promedio de 0,82. El análisis factorial mostró condiciones adecuadas para su aplicación, con un índice KMO de 0,73 y una prueba de esfericidad de Bartlett estadísticamente significativa ($p < 0,001$). La confiabilidad global fue excelente ($\alpha = 0,96$). Por dimensiones, los coeficientes alcanzaron valores de 0,74, 0,90 y 0,89, evidenciando niveles aceptables y altos de consistencia interna. **Conclusiones:** El instrumento AOU presenta evidencias favorables de validez y confiabilidad para evaluar la producción de textos escritos en estudiantes de educación primaria. Sus propiedades psicométricas respaldan su utilización en contextos educativos similares, aunque se recomienda continuar su validación con muestras más amplias y diversas.

Palabras clave: Confiabilidad; Evaluación educativa; Instrumento de medición; Producción de textos; Validez

Abstract

Introduction: The assessment of written text production constitutes a fundamental process in primary education, as it enables the evaluation of competencies related to planning, organization, and text composition. To ensure consistent results, assessment instruments must provide adequate evidence of validity and reliability. **Objective:** To analyze the construct validity and reliability of the AOU instrument for assessing written text production among fourth-grade primary school students at Abraham Valdelomar Educational Institution in Virú, Peru. **Methodology:** A quantitative instrumental study was conducted. A 20-item checklist organized into three dimensions was administered to a sample of 30 students. Content validity was determined using Aiken's V coefficient, while construct validity was examined through factor analysis. Reliability was estimated using Cronbach's alpha coefficient. **Results:** The instrument demonstrated high content validity, with a mean Aiken's V of 0.82. Factor analysis showed adequate conditions for

^{1,2} Universidad César Vallejo. Trujillo, Perú

Artículo recibido 30 de julio 2026 | Aceptado 19 de agosto 2026 | Publicado 1 de octubre 2026

Autor de correspondencia: hdavilaco@ucvvirtual.edu.pe

Conflicto de interés: Los autores declaran no tener conflicto de intereses.

Cómo citar: Dávila Costa, H. V., & Leyva Aguilar, N. A. (2026). Validez y confiabilidad del instrumento para evaluar la producción de textos narrativos. *TRIBUNAL, Revista en Ciencias de la Educación y Ciencias Jurídicas*, 6(17), 1-16. <http://doi.org/10.59659/revistatribunal.v6i17.577>

its application, with a KMO index of 0.73 and a statistically significant Bartlett's test of sphericity ($p < 0.001$). Overall reliability was excellent ($\alpha = 0.96$). By dimension, the coefficients were 0.74, 0.90, and 0.89, indicating acceptable to high levels of internal consistency. **Conclusions:** The AOU instrument provides favorable evidence of validity and reliability for assessing written text production among primary school students. Its psychometric properties support its use in similar educational contexts, although further validation with larger and more diverse samples is recommended.

Keywords: Reliability; Educational assessment; Measurement instrument; Text production; Validity

Introducción

La producción de textos escritos ocupa un lugar central en el proceso educativo de los estudiantes de educación primaria, dado que la escritura constituye una herramienta esencial para la comunicación, el aprendizaje y el desarrollo del pensamiento crítico (Almukhanbet et al., 2025). Las competencias letradas se adquieren de forma progresiva durante los primeros años de escolaridad y que su dominio temprano guarda relación con el rendimiento académico posterior. En ese contexto, la evaluación de la escritura demanda instrumentos que reflejen con precisión las dimensiones del proceso compositivo, desde la adecuación al género discursivo hasta el uso de recursos ortográficos y reglas de puntuación. La ausencia de herramientas estandarizadas limita la posibilidad de formular diagnósticos certeros sobre el estado de las competencias escritas en la población escolar (Galegane y Ntereke, 2022).

En el panorama latinoamericano, la evaluación de la producción textual enfrenta desafíos particulares derivados de la diversidad lingüística, las disparidades socioeconómicas y la escasa disponibilidad de instrumentos validados para las poblaciones escolares de la región (Villa, 2026). Los docentes de educación primaria requieren herramientas que permitan identificar con rigor las fortalezas y debilidades de sus estudiantes en la composición escrita, pero la literatura revela que la mayor parte de los recursos evaluativos proviene de contextos anglosajones y no refleja las particularidades del español como lengua materna (Higgs y Kim, 2022). Esta situación genera una brecha entre las necesidades evaluativas del aula y los instrumentos disponibles, lo cual compromete la calidad de la retroalimentación pedagógica y las decisiones orientadas al mejoramiento de las competencias comunicativas en el nivel primario.

En esa línea, diversos autores han propuesto instrumentos para la evaluación de la escritura que abarcan desde rúbricas analíticas hasta escalas de valoración global (Heidari et al., 2022; Li, 2022). Los instrumentos de evaluación de textos producidos por estudiantes deben contemplar criterios que reflejen la complejidad del acto de escritura, como la adecuación al tipo textual, la coherencia interna, la cohesión discursiva y el dominio de las convenciones ortográficas (Siekmann et al., 2022). Las investigaciones coinciden en que un instrumento robusto requiere evidencias de validez de contenido y de constructo, así como indicadores de confiabilidad que garanticen la estabilidad de las mediciones (Anggara y Abdillah, 2023). La mayor parte de estos desarrollos se centran

en poblaciones de educación secundaria o superior, lo cual deja un vacío notable en el nivel de educación primaria.

Por esta razón, la necesidad de contar con instrumentos válidos y confiables para la evaluación de la escritura en el nivel primario adquiere relevancia particular (Shaheen et al., 2023). La evaluación de la escritura en los primeros grados escolares debe considerar las características cognitivas y lingüísticas de los niños en esta etapa, lo cual supone el diseño de herramientas específicas que no constituyan adaptaciones simplificadas de instrumentos para niveles educativos superiores (Sanmartín et al., 2023). La validación psicométrica de un instrumento de evaluación de la producción de textos escritos implica un proceso riguroso que incluye la demostración de su validez de contenido mediante juicio de expertos, la comprobación de su estructura interna mediante análisis factorial y la verificación de su consistencia interna mediante coeficientes de confiabilidad (Saw et al., 2025).

A partir de lo anterior, resulta pertinente señalar que en el Perú las políticas educativas vigentes exigen la evaluación periódica de las competencias comunicativas, pero los docentes cuentan con escasas herramientas validadas para tal fin, en particular en zonas rurales donde las condiciones socioeconómicas inciden de forma directa en el rendimiento escolar (Mu et al., 2022). La institución educativa Abraham Valdelomar de Virú atiende a una población con características socioculturales que demandan instrumentos contextualizados. Ante la ausencia de listas de cotejo validadas para evaluar la producción de textos, los docentes recurren a instrumentos de diseño propio sin respaldo psicométrico, lo cual compromete la validez de sus evaluaciones. Surge así la pregunta: ¿Es válido y confiable el instrumento AOU para medir la producción de textos escritos en estudiantes del nivel primario?

Frente a esta interrogante, el objetivo del presente estudio fue analizar la validez de constructo y la confiabilidad del instrumento AOU para medir la producción de textos escritos en estudiantes de cuarto grado de educación primaria de la institución educativa Abraham Valdelomar de Virú, Perú.

Los resultados contribuirán a fortalecer las prácticas evaluativas de los docentes de educación primaria al disponer de un instrumento con evidencias psicométricas sólidas, además de ofrecer un referente para futuras investigaciones sobre la evaluación de la escritura en contextos rurales de América Latina. La relevancia de este estudio radica en la combinación de procedimientos rigurosos de validación con una muestra proveniente de un contexto poco representado en la literatura especializada.

Método

La investigación adoptó un enfoque cuantitativo con un diseño de estudio instrumental, pues el propósito central consistió en analizar las propiedades psicométricas de un instrumento de medición. Este tipo de diseño se orientó a la construcción y validación de pruebas, escalas y listas de cotejo destinadas a la recopilación de datos en contextos

específicos de investigación. El estudio se circunscribió al paradigma positivista, para establecer relaciones objetivas entre las variables medidas a través de procedimientos estadísticos estandarizados. El nivel de investigación adoptado fue de carácter descriptivo, dado que el trabajo se propuso caracterizar las propiedades de validez y confiabilidad del instrumento AOU sin establecer relaciones causales entre las variables medidas.

La población estuvo conformada por los estudiantes matriculados en el cuarto grado de educación primaria de la institución educativa Abraham Valdelomar, ubicada en el distrito de Virú, provincia de Virú, Perú. La muestra incluyó a 30 estudiantes que cumplieron con los criterios de inclusión siguientes: estar matriculados de forma oficial en el cuarto grado durante el periodo lectivo 2024, tener entre 9 y 10 años de edad y contar con la autorización escrita de sus padres o apoderados para participar en el estudio. Se excluyó a los estudiantes con diagnósticos de trastornos del lenguaje o discapacidades que impidieran la realización de la tarea. Del total de participantes, 18 fueron varones (60 %) y 12 mujeres (40 %), todos provenientes de una zona rural del distrito de Virú.

El instrumento AOU consiste en una lista de cotejo elaborada para evaluar la producción de textos escritos en educación primaria. Este instrumento incluyó 20 ítems distribuidos en tres dimensiones: adecuación al tipo textual (4 ítems), coherencia y cohesión (10 ítems) y recursos ortográficos y puntuación (6 ítems). Cada ítem describió un aspecto de la producción textual y el evaluador marcó la presencia o ausencia de dicho aspecto en el texto producido por el estudiante. Las dimensiones reflejaron los componentes de la competencia escritora según los estándares nacionales del Perú sobre la evaluación de la composición. La puntuación total del instrumento oscila entre 0 y 20 puntos, donde valores más altos indican un mayor dominio en la producción textual.

Para establecer la validez de contenido, ocho expertos con formación doctoral y experiencia docente en lingüística o educación emitieron su juicio sobre los 20 ítems del instrumento. Los evaluadores valoraron cada ítem en cuatro criterios: suficiencia, claridad, coherencia y relevancia, mediante una escala de Likert de cuatro puntos. El coeficiente V de Aiken permitió cuantificar el grado de concordancia entre los expertos para cada criterio evaluado. Los valores del coeficiente se interpretaron según rangos, donde valores superiores a 0,80 indican una concordancia alta entre los jueces. Se realizaron ajustes pertinentes en los ítems cuyo valor resultó inferior al umbral establecido, los cuales fueron reformulados de acuerdo con las observaciones de los expertos hasta alcanzar una concordancia satisfactoria en los cuatro criterios evaluados.

La recolección de datos se ejecutó durante el periodo lectivo 2024 en la institución educativa seleccionada. Los estudiantes elaboraron un texto narrativo a partir de un estímulo visual, con un tiempo límite de 45 minutos. Dos evaluadores aplicaron la lista de cotejo AOU sobre los textos escritos producidos por los estudiantes. En cuanto a los aspectos éticos, el estudio respetó los principios de la Declaración de Helsinki. Los padres o apoderados de los participantes firmaron un consentimiento informado donde se detallaron los propósitos del estudio, la naturaleza voluntaria de la participación y la

garantía de confidencialidad en el tratamiento de los datos. El protocolo de investigación obtuvo la aprobación del comité de ética de la Universidad César Vallejo antes del inicio de la recolección de datos.

El análisis estadístico se ejecutó en el software SPSS versión 29. La prueba de normalidad de Shapiro-Wilk permitió verificar la distribución de los datos, requisito indispensable para la aplicación del análisis factorial. Se calculó la medida de adecuación muestral de Kaiser-Meyer-Olkin (KMO) y la prueba de esfericidad de Bartlett para determinar la factibilidad del análisis factorial exploratorio. Este procedimiento identificó la estructura latente del instrumento y confirmó la agrupación de los ítems en las tres dimensiones teóricas propuestas. La confiabilidad del instrumento se determinó mediante el coeficiente alfa de Cronbach, tanto para el instrumento completo como para cada dimensión de forma individual. Se adoptó un nivel de significancia de 0,05 para todas las pruebas estadísticas. Los resultados se interpretaron según los criterios de [Hair et al. \(2019\)](#).

Resultados

En esta sección se presentan los resultados de la validación del instrumento AOU, organizados en torno a los principales indicadores psicométricos para avalar su uso en contextos educativos. El análisis incluye los datos de la validez de contenido mediante el coeficiente V de Aiken, la suficiencia factorial determinada por la medida de Kaiser-Meyer-Olkin y la prueba de esfericidad de Bartlett, así como los indicadores de confiabilidad obtenidos a partir del alfa de Cronbach para el instrumento completo y para cada una de sus tres dimensiones. Estos resultados ofrecen una visión integral de las propiedades métricas del instrumento y constituyen la base para sustentar su empleo en la evaluación de la producción textual en educación primaria. Los datos se organizan en cuatro tablas que sintetizan los procedimientos estadísticos aplicados.

Cada evaluador analizó los ítems del instrumento en cuatro criterios: suficiencia, claridad, coherencia y relevancia para establecer la validez de contenido del instrumento AOU. La suficiencia alude a la capacidad del ítem para representar el aspecto evaluado; la claridad se refiere a la comprensibilidad de su redacción; la coherencia evalúa la correspondencia del ítem con la dimensión teórica asignada; y la relevancia pondera la importancia del ítem para el constructo global. Los expertos emitieron valoraciones en una escala de Likert de cuatro puntos y los resultados se procesaron mediante el coeficiente V de Aiken.

Los resultados de la Tabla 1 muestran que los valores del coeficiente V de Aiken oscilan entre 0,81 y 0,82, superiores al umbral de 0,80 establecido como referencia para una concordancia alta entre jueces. El promedio global del instrumento alcanzó un valor de 0,82, lo cual indica que los ocho expertos coincidieron de forma notable en la pertinencia de los ítems para medir la producción de textos escritos. La coherencia entre los criterios evaluados refleja una consistencia sólida en las valoraciones emitidas, lo que avala la representatividad del contenido del instrumento respecto al constructo teórico que se

pretende medir. El criterio de relevancia obtuvo el valor más bajo del conjunto, aunque su cifra de 0,81 se mantiene dentro del rango de concordancia alta según la literatura metodológica.

Tabla 1. Validez de contenido del instrumento AOU según coeficiente V de Aiken

Criterio	Valor V de Aiken
Suficiencia	0.82
Claridad	0.82
Coherencia	0.82
Relevancia	0.81
Promedio	0.82

Para verificar la factibilidad del análisis factorial, se calcularon la medida de adecuación muestral de Kaiser-Meyer-Olkin (KMO) y la prueba de esfericidad de Bartlett. Estos indicadores constituyen requisitos fundamentales para la aplicación del análisis factorial exploratorio, dado que evalúan la magnitud de las correlaciones entre las variables y la idoneidad de los datos para la extracción de componentes latentes. La medida de KMO valora el grado de solapamiento entre las variables, mientras que la prueba de Bartlett contrasta la hipótesis nula de que la matriz de correlaciones corresponde a una matriz de identidad. La prueba de normalidad de Shapiro-Wilk confirmó la distribución no normal de los datos, lo cual respalda la pertinencia del análisis factorial. La tabla 2 presenta los valores obtenidos en ambos indicadores muestrales.

Los resultados presentados en la Tabla 2 ofrecen evidencia inicial sólida sobre la validez de constructo del instrumento AOU, específicamente en lo referente a su estructura interna. El valor del índice de Kaiser-Meyer-Olkin (KMO) de 0,73 se sitúa en el rango considerado como aceptable o medio según la literatura psicométrica (valores entre 0,70 y 0,79). Este hallazgo es relevante porque confirma que las correlaciones entre los ítems o variables del instrumento no son demasiado bajas (lo que impediría la factorización) ni excesivamente altas (lo que indicaría redundancia). En consecuencia, el KMO de 0,73 sugiere que la matriz de correlaciones presenta un grado de estructura latente suficiente para proceder con la extracción de factores, un paso indispensable para demostrar que el instrumento mide efectivamente el constructo multidimensional de producción de textos escritos (que abarca desde la planificación y redacción hasta la revisión y edición).

Complementariamente, la prueba de esfericidad de Bartlett arroja un chi-cuadrado aproximado de 737,40 con 190 grados de libertad y una significancia de $p < 0,001$. Este resultado es estadísticamente robusto y permite rechazar la hipótesis nula de que la matriz de correlaciones es una matriz identidad (es decir, que las variables no están correlacionadas entre sí). La significancia alcanzada respalda que existen relaciones significativas entre los ítems, justificando la pertinencia del análisis factorial.

Asimismo, la mención de que la prueba de Shapiro-Wilk arrojó valores significativos (lo que indica que los datos no siguen una distribución normal) refuerza la decisión metodológica de utilizar un análisis factorial exploratorio, ya que esta técnica es apropiada cuando no se asume normalidad multivariante. En conjunto, estos tres indicadores —KMO aceptable, Bartlett significativo y no normalidad confirmada— demuestran la suficiencia factorial del instrumento AOU y la idoneidad de los datos para descomponer el constructo evaluado en dimensiones interpretables (por ejemplo, coherencia, cohesión, gramática, ortografía y edición textual). No obstante, para completar la validación, sería necesario reportar la varianza total explicada por los factores extraídos y los índices de confiabilidad, como el alfa de Cronbach o el omega de McDonald.

Tabla 2. *Medida de adecuación muestral de Kaiser-Meyer-Olkin y prueba de Bartlett*

Indicador	Valor
Medida de adecuación muestral de Kaiser-Meyer-Olkin	0.73
Prueba de esfericidad de Bartlett: Chi-cuadrado aproximado	737.40
gl	190
Sig.	<0.001

La confiabilidad del instrumento AOU en su conjunto se determinó mediante el coeficiente alfa de Cronbach, el cual constituye uno de los indicadores más utilizados en la investigación psicométrica para estimar la consistencia interna de una escala. Este coeficiente evalúa el grado en que los ítems del instrumento miden un mismo constructo latente y producen resultados coherentes entre sí. Valores superiores a 0,70 se consideran aceptables para fines de investigación, mientras que valores superiores a 0,90 indican una consistencia interna excelente. En el caso del instrumento AOU, el cálculo del alfa de Cronbach se realizó sobre la totalidad de los 20 ítems con el fin de obtener un indicador global de la homogeneidad de la escala. La tabla 3 resume el valor obtenido para el instrumento completo.

En la Tabla 3 se muestran los resultados que evidencian una confiabilidad excepcionalmente alta para el instrumento AOU destinado a evaluar la producción de textos escritos. El coeficiente alfa de Cronbach obtenido es de 0,96 sobre un total de 20 ítems, una cifra que supera ampliamente el mínimo recomendado de 0,70 para investigaciones en ciencias sociales y educativas, e incluso excede el umbral de 0,90 considerado como excelente por especialistas en psicometría. En términos concretos, este valor indica que el 96% de la varianza observada en las puntuaciones totales puede atribuirse a la varianza verdadera del constructo, mientras que únicamente un 4% correspondería a errores de medición, lo que denota una precisión notoria.

Desde la perspectiva de la medición de habilidades de escritura, un constructo complejo que involucra dimensiones como planificación, textualización, revisión y edición, un alfa de 0,96 sugiere que los 20 ítems del instrumento AOU operan de manera altamente

homogénea para medir un constructo latente común. Esta elevada consistencia interna tiene dos implicaciones metodológicas relevantes. La correlación media entre los ítems es muy robusta, lo que implica que todos contribuyen de forma equilibrada a la puntuación global y, por otra parte, resulta prácticamente improbable que la eliminación de algún ítem específico mejore significativamente la confiabilidad, lo cual es una señal técnica positiva porque indica que no existen ítems con comportamientos atípicos o que midan dimensiones completamente diferentes. Este elevado nivel de consistencia interna indica que los 20 ítems del instrumento miden un mismo constructo latente de forma homogénea y que las respuestas de los estudiantes mantienen una alta coherencia entre sí.

No obstante, vale señalar que coeficientes extremadamente altos (superiores a 0,95) podrían insinuar cierto grado de redundancia entre los ítems, por lo que se recomienda complementar este análisis con estudios de validez discriminante y un examen de la matriz de correlaciones inter-ítem para confirmar que no exista solapamiento excesivo que afecte la eficiencia del instrumento. En conjunto, se demuestra la solidez psicométrica del AOU desde la perspectiva de la confiabilidad por consistencia interna.

Tabla 3. *Confiabilidad global del instrumento AOU*

Indicador	Valor
Alfa de Cronbach	0.96
Número de ítems	20

Además del análisis de confiabilidad global, se calculó el coeficiente alfa de Cronbach para cada una de las tres dimensiones del instrumento. Este análisis desagregado permite verificar si cada subescala posee una consistencia interna adecuada de forma independiente, lo cual resulta esencial para garantizar la fiabilidad de las puntuaciones obtenidas en cada dimensión de forma aislada. La evaluación por dimensiones ofrece información complementaria al indicador global, pues permite identificar si alguna subescala presenta debilidades en su composición que requieran ajustes en futuras versiones del instrumento. La dimensión de coherencia y cohesión concentra el mayor número de ítems, mientras que la dimensión de adecuación al tipo textual comprende solo cuatro ítems, una diferencia que puede influir en los valores del coeficiente de confiabilidad. La tabla 4 presenta los resultados por dimensión.

Los resultados presentados en la Tabla 4 complementan y matizan el hallazgo de confiabilidad global excepcionalmente alta ($\alpha = 0,96$) al desagregar el comportamiento psicométrico de cada dimensión que compone el instrumento AOU para evaluar la producción de textos escritos. Las tres dimensiones evaluadas, adecuación al tipo textual (4 ítems, $\alpha = 0,74$), coherencia y cohesión (10 ítems, $\alpha = 0,90$), y recursos ortográficos y puntuación (6 ítems, $\alpha = 0,89$), superan todas el umbral mínimo de 0,70 recomendado por la literatura especializada, lo que confirma que cada subescala posee una consistencia interna aceptable por sí misma.

Desde una perspectiva interpretativa, la dimensión de coherencia y cohesión presenta el coeficiente más alto ($\alpha = 0,90$), lo cual resulta esperable dado que contiene el mayor número de ítems (10). Este valor, clasificado como excelente, indica que los ítems que evalúan la organización lógica de las ideas, el uso de conectores y la progresión temática mantienen una alta homogeneidad entre sí. La dimensión de recursos ortográficos y puntuación ($\alpha = 0,89$) también muestra un desempeño sobresaliente, ubicándose en el rango de bueno a excelente. Por su parte, la dimensión de adecuación al tipo textual ($\alpha = 0,74$), si bien es la más baja, se sitúa dentro del rango aceptable y su valor puede explicarse parcialmente por su reducido número de ítems (solo 4), ya que el coeficiente alfa de Cronbach es sensible a la longitud de la escala.

Es importante destacar que, aunque el alfa global de 0,96 es notablemente superior a cada coeficiente dimensional, esta diferencia no representa una contradicción metodológica. Por el contrario, confirma que el instrumento AOU mide un constructo general unitario (producción de textos escritos) al tiempo que posee la sensibilidad necesaria para evaluar de forma diferenciada sus tres componentes esenciales. La estructura trifactorial queda así respaldada empíricamente, demostrando que el instrumento es tanto homogéneo en su conjunto como discriminativo en sus partes.

Tabla 4. *Confiabilidad del instrumento AOU por dimensiones*

Dimensión	Número de ítems	Alfa de Cronbach
Adecuación al tipo textual	4	0.74
Coherencia y cohesión	10	0.90
Recursos ortográficos y puntuación	6	0.89

Los resultados obtenidos confirman que el instrumento AOU presenta propiedades psicométricas sólidas para evaluar la producción de textos escritos. La validez de contenido, sustentada en el juicio de expertos, arrojó niveles de concordancia altos en todos los criterios evaluados. El análisis factorial exploratorio demostró ser pertinente, pues los indicadores de adecuación muestral y esfericidad mostraron valores aceptables y estadísticamente significativos, lo que evidencia una estructura factorial interpretable. La confiabilidad global del instrumento resultó excelente, mientras que el análisis por dimensiones confirmó una consistencia interna adecuada en cada una de las tres subescalas que lo componen.

Discusión

Los hallazgos derivados del análisis psicométrico del instrumento AOU ofrecen evidencia empírica que respalda su uso para evaluar la producción de textos escritos en el contexto educativo peruano. En esta sección se discuten de manera integrada los resultados obtenidos en las fases de validez de contenido, validez de constructo y confiabilidad

global y dimensional. Se comparan estos hallazgos con los reportados por investigaciones previas en el ámbito de la evaluación de la escritura, se analizan las implicaciones metodológicas de los coeficientes alcanzados, especialmente la elevada consistencia interna, y se formulan recomendaciones para la aplicación del instrumento en futuros estudios o prácticas docentes.

Los resultados de validez de contenido muestran un promedio de V de Aiken de 0,82, cifra que se sitúa dentro del rango de concordancia alta según la literatura metodológica. Arauzo (2022) y Merino (2023), señalan que valores del coeficiente V de Aiken superiores a 0,80 reflejan un acuerdo sustancial entre los jueces y que su estabilidad depende de la claridad de las definiciones operacionales y de la experiencia de los expertos convocados. En el caso del instrumento AOU, la selección de ocho expertos con formación doctoral garantizó una evaluación cualificada de cada ítem. La coincidencia de los valores en los cuatro criterios evaluados sugiere que los ítems poseen una redacción comprensible y pertinente, aspecto que los expertos reconocieron de forma unánime en sus evaluaciones de validez de contenido del instrumento, lo que concuerda con los resultados de Roebianto et al. (2023).

En relación con el análisis factorial, el valor de KMO de 0,73 clasificado como aceptable confirma la pertinencia de esta técnica para examinar la estructura interna del instrumento, resultado que concuerda con los hallazgos de Ávila y Juárez (2023) y Abarzúa et al. (2024), en estudios de validación psicométrica de instrumentos educativos, quienes reportaron valores de KMO similares en la validación de un instrumento para evaluar secuencias didácticas en educación superior, lo cual indica que los valores entre 0,70 y 0,80 son frecuentes en investigaciones con muestras de tamaño moderado. La prueba de esfericidad de Bartlett con $\chi^2 = 737,40$ ($p < 0,001$) confirma la existencia de correlaciones significativas entre los ítems, un resultado esperado en un instrumento diseñado para medir un constructo unitario con múltiples dimensiones interrelacionadas entre sí en el marco de la producción textual, lo que concuerda con lo planteado por Ndruru et al. (2026).

Respecto a la confiabilidad, el alfa de Cronbach de 0,96 para el instrumento completo supera los valores reportados en estudios de validación de instrumentos de evaluación de escritura. Stuart y Barnett (2023), desarrollaron la Writing Quality Scale para identificar dificultades de escritura en estudiantes y obtuvieron valores de confiabilidad inferiores a 0,90, lo cual contrasta con el nivel de consistencia alcanzado por el instrumento AOU. Esta diferencia se explica por la naturaleza de la lista de cotejo, cuyo formato dicotómico reduce la variabilidad de las respuestas y favorece mayor homogeneidad en las puntuaciones, aspecto que refiere Elosua (2022). Las tres dimensiones del instrumento mostraron valores adecuados de alfa de Cronbach, con cifras que oscilan entre 0,74 y 0,90, lo cual confirma la contribución de cada dimensión a la medición global del constructo evaluado.

En el ámbito de los instrumentos validados para la evaluación de competencias en educación primaria, los resultados del presente estudio se alinean con las tendencias

reportadas por [Cervantes y Canto \(2025\)](#), quienes analizaron los instrumentos validados para evaluar la comprensión lectora en educación primaria entre 2015 y 2025 y concluyeron que la mayoría alcanza valores aceptables de validez de contenido y confiabilidad, aunque persiste una escasez de herramientas para la evaluación de la producción textual. El instrumento AOU contribuye a llenar este vacío al ofrecer evidencias psicométricas sólidas para la evaluación de la escritura en estudiantes de cuarto grado de primaria, concordando con [Rogers et al. \(2022\)](#), quienes consideran que este nivel educativo ha recibido menos atención en comparación con la educación secundaria y los niveles superiores del sistema educativo.

En cuanto a la validación de instrumentos de evaluación educativa, [Figueroa y Malet \(2024\)](#), construyeron y validaron rúbricas para valorar la idoneidad didáctica en educación superior mediante un proceso que incluyó juicio de expertos, análisis factorial y cálculo de confiabilidad, una secuencia metodológica similar a la empleada en el presente estudio. Estos autores encontraron que la combinación de múltiples evidencias de validez fortalece la argumentación a favor del uso del instrumento en contextos aplicados, una premisa que el instrumento AOU cumple al haber demostrado adecuación en sus tres dimensiones evaluadas. En esta línea, [Payan et al. \(2022\)](#) y [McGrath \(2023\)](#), consideran que la correspondencia entre los procedimientos de validación aplicados en distintos contextos educativos confirma la vigencia del enfoque clásico de validación propuesto por la teoría clásica de la medición en el campo de la evaluación educativa.

Sobre la motivación y el rendimiento en la escritura, [Camacho et al. \(2021\)](#) y [Graham et al. \(2025\)](#), realizaron un metaanálisis que reveló la importancia de contar con instrumentos fiables para evaluar las competencias de escritura en los primeros años de escolaridad. Estos autores subrayaron que la precisión de las mediciones influye de forma directa en la calidad de las intervenciones pedagógicas, ya que los diagnósticos basados en instrumentos poco confiables conducen a conclusiones equivocadas sobre el estado real de las competencias de los estudiantes. El alto valor de alfa de Cronbach obtenido por el instrumento AOU reduce este riesgo al proporcionar una medición estable y consistente de la producción de textos escritos, lo cual favorece la toma de decisiones pedagógicas fundamentadas en evidencias empíricas de carácter robusto y replicable en distintos contextos escolares.

En el terreno de la evaluación de la escritura en aulas de educación elemental, [Kennedy y Shiel \(2022\)](#), desarrollaron y validaron una escala de compromiso formativo con la escritura para estudiantes de los primeros grados, un esfuerzo que guarda similitudes con el presente estudio en cuanto al grupo etario de referencia y la intención de proveer herramientas evaluativas adaptadas a las características de los niños. Estos investigadores destacaron la necesidad de que los instrumentos contemplen tanto los aspectos textuales como los procesos cognitivos implicados en la composición. El instrumento AOU responde a esta demanda al incluir dimensiones que abarcan desde la adecuación textual hasta los recursos ortográficos, lo cual permite una evaluación integral de la producción de textos escritos en el contexto del aula de educación primaria.

Concerniente a la evaluación de la producción de textos académicos, [Chaverra et al. \(2022\)](#) y [Murrieta \(2024\)](#), analizaron las características de los textos producidos por estudiantes universitarios en un centro de escritura digital y encontraron que la coherencia y la cohesión constituyen las dimensiones con mayor peso en la evaluación de la calidad textual. Este hallazgo respalda la estructura del instrumento AOU, donde la dimensión de coherencia y cohesión concentra el mayor número de ítems (10 de 20), lo cual refleja la relevancia teórica y práctica de estos componentes en la producción de textos escritos. La coincidencia entre los hallazgos en educación superior y los del presente estudio en educación primaria sugiere que las dimensiones evaluadas por el instrumento AOU poseen un carácter transversal que trasciende los niveles educativos y los contextos geográficos.

No obstante, el presente estudio presenta limitaciones que deben considerarse al interpretar los resultados. La muestra de 30 estudiantes, si bien cumple con los criterios estadísticos para el análisis factorial, reduce la capacidad de generalización de los hallazgos a otras poblaciones y contextos educativos. [Matos et al. \(2025\)](#), señalaron que las investigaciones sobre evaluación formativa en educación primaria suelen trabajar con muestras pequeñas, una restricción compartida por la mayoría de los estudios de validación instrumental realizados en instituciones educativas únicas. Otra limitación radica en la procedencia rural de los participantes, lo cual podría influir en las características de los textos producidos y en las puntuaciones obtenidas. El diseño transversal del estudio tampoco permite evaluar la estabilidad temporal del instrumento en mediciones sucesivas.

A partir de las limitaciones identificadas, se recomienda que futuras investigaciones repliquen el proceso de validación del instrumento AOU con muestras de mayor tamaño y en contextos educativos urbanos de diferentes regiones de Perú y de otros países de América Latina. También resulta conveniente aplicar análisis factorial confirmatorio para contrastar la estructura trifactorial hallada en el presente estudio con modelos teóricos alternativos. Asimismo, la incorporación de la validez convergente y discriminante mediante la correlación del instrumento AOU con otras pruebas estandarizadas de evaluación de la escritura fortalecería el cuerpo de evidencias sobre sus propiedades psicométricas. La construcción de baremos normativos por grado escolar y por contexto geográfico constituye otra línea de investigación prometedora que ampliaría la utilidad práctica del instrumento en el sistema educativo.

Conclusiones

El instrumento AOU demostró poseer evidencias favorables de validez de contenido y de constructo, así como una confiabilidad elevada, lo cual permite afirmar que esta herramienta mide de forma fidedigna la producción de textos escritos en estudiantes de cuarto grado de educación primaria. La concordancia alta entre los expertos evaluadores, la suficiencia factorial confirmada por las pruebas de KMO y Bartlett y la consistencia

interna del instrumento en sus tres dimensiones configuran un perfil psicométrico robusto que respalda su empleo en contextos educativos similares al de este estudio. El proceso de validación seguido en esta investigación aborda una necesidad concreta del sistema educativo peruano: la disponibilidad de instrumentos evaluativos con respaldo científico para la evaluación de la competencia de producción de textos escritos en el nivel primario.

En consecuencia, los hallazgos de este estudio poseen implicaciones directas para la práctica pedagógica y la investigación educativa en el contexto latinoamericano. Provee a los docentes de educación primaria de una herramienta evaluativa con fundamentos psicométricos que les permite formular diagnósticos precisos sobre las competencias de escritura de sus estudiantes, identificar las áreas que requieren refuerzo pedagógico y monitorear el progreso de los aprendizajes a lo largo del periodo escolar. La estructura trifactorial del instrumento ofrece una visión integral de la producción textual al contemplar la adecuación al tipo discursivo, la coherencia y cohesión del texto y el uso de recursos ortográficos y puntuaciones, dimensiones que corresponden a los estándares curriculares vigentes en el sistema educativo peruano y en otros países de la región latinoamericana.

A modo de cierre, resulta preciso subrayar que la validación de instrumentos de evaluación educativa constituye un proceso continuo que trasciende un único estudio empírico. El instrumento AOU, con sus evidencias psicométricas favorables, representa un avance significativo en el campo de la evaluación de la escritura en educación primaria, pero su consolidación como herramienta de uso generalizado requiere replicaciones en diversos contextos y la acumulación progresiva de evidencias de validez. La contribución principal de este trabajo radica en ofrecer a la comunidad educativa un instrumento accesible, fácil de aplicar y con fundamentos técnicos que respaldan su uso en el aula. La mejora de las prácticas evaluativas de la escritura, sustentada en instrumentos rigurosos, constituye un pilar para la formación de estudiantes competentes en la producción textual.

Acerca de

Contribución de los autores: Los autores participaron en el diseño del estudio, el trabajo de campo, el análisis de la información y la redacción del manuscrito, cuyo contenido fue revisado y aprobado.

Financiamiento: Los autores declaran que no recibieron financiamiento para esta investigación.

Certificación ética: El estudio involucró a estudiantes de educación primaria, por lo que corresponde garantizar el consentimiento informado de sus representantes legales, el asentimiento de los participantes, la confidencialidad de los datos y el respeto a los principios éticos de la investigación.

Referencias

- Abarzúa, L., Ambrós, A., y Ruiz, A. (2024). Construcción y validación de un cuestionario sobre prácticas de lectura digital académica para estudiantado universitario de formación inicial de profesorado. *Revista de Investigación Educativa*, 42(1), 33-59. <https://doi.org/10.6018/rie.548111>
- Almukhanbet, S., Moldabek, K., Orynbassarova, G., Yesnazar, A., y Zhorabekova, A. (2025). Developing primary school students' language skills from a critical thinking perspective: A methodological system. *Frontiers in Education*, 10, 1565783. <https://doi.org/10.3389/feduc.2025.1565783>
- Anggara, D. S., y Abdillah, C. (2023). Content validity analysis of literacy assessment instruments. *Jurnal Cakrawala Pendidikan*, 42(2), 447-459. <https://doi.org/10.21831/cp.v42i2.55900>
- Arauzo, C. (2022). Instrumento de medida de la comprensión lectora. *Alborada de la Ciencia*, 2(2), 50-58. <https://doi.org/10.26490/uncp.alboradaciencia.2022.2.1116>
- Ávila, G., y Juárez, L. G. (2023). Propiedades psicométricas de un instrumento para evaluar la ejecución de una secuencia didáctica en educación superior. *Revista Electrónica Educare*, 27(3), 2. <https://dialnet.unirioja.es/servlet/articulo?codigo=9006362>
- Camacho, A., Alves, R. A., y Boscolo, P. (2021). Writing motivation in school: A systematic review of empirical research in the early twenty-first century. *Educational psychology review*, 33(1), 213-247. <https://doi.org/10.1007/s10648-020-09530-4>
- Cervantes, A. A., y Canto, P. J. (2025). Análisis de instrumentos validados para evaluar la comprensión lectora en educación primaria: Una revisión sistemática (2015 a 2025). *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 16(31). <https://doi.org/10.23913/ride.v16i31.2633>
- Chaverra, D. I., Yair, G., Hurtado, R. D., y Bolívar, W. A. (2022). A Review of Research on Academic Writing for the Creation of an Online Writing Center in Higher Education. *Íkala, Revista de Lenguaje y Cultura*, 27(1), 224-247. <https://doi.org/10.17533/udea.ikala.v27n1a11>
- Elosua, P. (2022). Validity evidences for scoring procedures of a writing assessment task. *A case study on consistency, reliability, unidimensionality and prediction accuracy. Assessing Writing*, 54, 100669. <https://doi.org/10.1016/j.asw.2022.100669>
- Figuerola, M. V., y Malet, O. (2024). Construcción y validación de rúbricas para valorar la idoneidad didáctica de clases de matemática. *Revista latinoamericana de investigación en matemática educativa*, 27(1), 43-72. <https://doi.org/10.12802/relime.24.2712>
- Galegane, G., y Ntereke, B. B. (2022). Examining the effectiveness of the process-genre writing instructional framework in EAP writing course. *Journal of Pedagogical Sociology and Psychology*, 4(2), 112-130. <https://doi.org/10.33902/JPSP.202217703>
- Graham, S., Ng, C., Hebert, M., Santangelo, T., Aitken, A. A., Camping, A., y Nusrat, A. (2025). Can teaching writing enhance students' writing self-efficacy: A meta-analysis. *Journal of Educational Psychology*. <https://doi.org/10.1037/edu0000923>

- Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24. <https://doi.org/10.1108/EBR-11-2018-0203>
- Heidari, N., Ghanbari, N., y Abbasi, A. (2022). Raters' perceptions of rating scales criteria and its effect on the process and outcome of their rating. *Language Testing in Asia*, 12(1), 20. <https://doi.org/10.1186/s40468-022-00168-3>
- Higgs, J. M., y Kim, G. M. (2022). Interpreting old texts with new tools: digital multimodal composition for a high school reading assignment. *English Teaching: Practice and Critique*, 21(2), 128-142. <https://doi.org/10.1108/ETPC-07-2020-0079>
- Kennedy, E., y Shiel, G. (2022). Writing assessment for communities of writers: Rubric validation to support formative assessment of writing in Pre-K to grade 2. *Assessment in Education: Principles, Policy and Practice*, 29(2), 127-149. <https://doi.org/10.1080/0969594X.2022.2047608>
- Li, W. (2022). Scoring rubric reliability and internal validity in rater-mediated EFL writing assessment: Insights from many-facet Rasch measurement. *Reading and Writing*, 35(10), 2409-2431. <https://doi.org/10.1007/s11145-022-10279-1>
- Matos, L., Duran, K. L., y Mucha, L. F. (2025). Evaluación formativa para el mejoramiento de las competencias comunicativas en estudiantes de primaria. *Episteme Koinonía. Revista Electrónica de Ciencias de la Educación, Humanidades, Artes y Bellas Artes*, 8(15), 138-156. <https://doi.org/10.35381/e.k.v7i14.4398>
- McGrath, R. E. (2023). A summary of construct validity evidence for two measures of character strengths. *Journal of Personality Assessment*, 105(3), 302-313. <https://doi.org/10.1080/00223891.2022.2120402>
- Merino, C. (2023). Aiken's V Coefficient: Differences in Content Validity Judgments. *MHSalud*, 20(1), 23-32. <https://doi.org/10.15359/mhs.20-1.3>
- Mu, E., Florek, A., y Pereyra, M. (2022). Development of a framework to assess challenges to virtual education in an emergency remote teaching environment: A developing country student perspective—The case of Peru. *Education sciences*, 12(10), 704. <https://doi.org/10.3390/educsci12100704>
- Murrieta, R. (2024). Escritura académica en educación superior. *Dificultades y desarrollo de saberes con apoyo de las TIC. RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 15(29). <https://doi.org/10.23913/ride.v15i29.2044>
- Ndruru, F., Suryati, N., Basthomi, Y., Rovikasari, M., y Laia, R. D. (2026). Developing and Validating a Multimodal Composing Scale in EFL Academic Writing: An Exploratory Factor Analysis. *Literacy Research and Instruction*, 65(2), 208-230. <https://doi.org/10.1080/19388071.2025.2592211>
- Payan, R., Sacau, A., Rebelo, H., Sebastião, L., y Pnevmatikos, D. (2022). Development and validation of a critical thinking assessment-scale short form. *Education sciences*, 12(12), 938. <https://doi.org/10.3390/educsci12120938>
- Roebianto, A., Savitri, S. I., Aulia, I., Suciyan, A., y Mubarakah, L. (2023). Content validity: Definition and procedure of content validation in psychological research.

Testing, Psychometrics, Methodology in Applied Psychology, 30(1), 5-18.
<https://doi.org/10.4473/TPM30.1.1>

Rogers, P. M., Marine, J. M., Ives, S. T., Parsons, S. A., Horton, A., y Young, C. (2022). Validity evidence for a formative writing engagement assessment in elementary grades. *Assessment in Education: Principles, Policy and Practice*, 29(2), 262-284.
<https://doi.org/10.1080/0969594X.2022.2054942>

Sanmartín, M. R., Curipoma, G. E., Bermeo, J. E., y Pullaguari, B. L. (2023). Principales problemáticas en el proceso del aprendizaje lecto escritor en los primeros años de escolaridad. *Ciencia Latina Revista Científica Multidisciplinar*, 7(2), 10910-10929.
https://doi.org/10.37811/cl_rcm.v7i2.6172

Saw, Z. K., Yuen, J. J. X., Ashari, A., Ibrahim Bahemia, F., Low, Y. X., Nik Mustapha, N. M., y Lau, M. N. (2025). Forward-backward translation, content validity, face validity, construct validity, criterion validity, test-retest reliability, and internal consistency of a questionnaire on patient acceptance of orthodontic retainer. *PloS one*, 20(1), e0314853.
<https://doi.org/10.1371/journal.pone.0314853>

Shaheen, N., Shaheen, A., Ramadan, A., Hefnawy, M. T., Ramadan, A., Ibrahim, I. A., Ashour, M., y Flouty, O. (2023). Appraising systematic reviews: a comprehensive guide to ensuring validity and reliability. *Frontiers in research metrics and analytics*, 8, 1268045. <https://doi.org/10.3389/frma.2023.1268045>

Siekmann, L., Parr, J. M., y Busse, V. (2022). Structure and coherence as challenges in composition: A study of assessing less proficient EFL writers' text quality. *Assessing Writing*, 54, 100672. <https://doi.org/10.1016/j.asw.2022.100672>

Stuart, N. J., y Barnett, A. L. (2023). The writing quality scale (WQS): A new tool to identify writing difficulties in students. *British Journal of Special Education*, 50(2), 258-267. <https://doi.org/10.1111/1467-8578.12464>

Villa, S. (2026). Towards inclusive publishing practices in language testing: Voices from the Global South. 43(1), 100-115. <https://doi.org/10.1177/02655322251383958>



Este artículo está bajo la licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0).