



**DESAFÍOS DE LA EVALUACIÓN DE LA ESCRITURA EN LENGUA EXTRANJERA EN
EL NIVEL SUPERIOR: PERCEPCIONES DE DOCENTES SOBRE LA APLICACIÓN DE
UNA ESCALA ANALÍTICA**

**CHALLENGES OF FOREIGN LANGUAGE WRITING ASSESSMENT IN HIGHER
EDUCATION: TEACHERS' PERCEPTIONS OF THE APPLICATION OF AN ANALYTIC
SCORING SCALE**

María Elisa Romano¹

maria.elisa.romano@unc.edu.ar

Julia Inés Martínez²

juliamartinez@unc.edu.ar

Natalia Dalla Costa³

natalia.dalla.costa@unc.edu.ar

Facultad de Lenguas, Universidad Nacional de Córdoba

Argentina

Resumen

La evaluación de la escritura en lengua extranjera es un campo de acción complejo, que suele generar ansiedad y aprehensión entre docentes debido a sus efectos e implicancias, tanto a nivel pedagógico como ético y político (Ruecker y Crusan, 2018) y que continúa siendo un área de permanente investigación, especialmente debido a la variedad y complejidad de contextos en los que se lleva a cabo. La evaluación de ensayos en lengua extranjera (inglés) por parte de docentes en las carreras de Profesorado, Traductorado y Licenciatura en Inglés en la Universidad Nacional de Córdoba (UNC) ha sido nuestro objeto de investigación en los últimos seis años. Se aplicaron distintos instrumentos de evaluación y una escala analítica de evaluación demostró ser la más adecuada para este contexto académico específico: la cátedra de Lengua Inglesa II. Con el fin de determinar la confiabilidad de dicha escala analítica, se llevó a cabo un proceso de validación de los descriptores que conforman cada una de las categorías de la escala. Se conformó un corpus

de ejemplos textuales para ilustrar los descriptores y así lograr mayores niveles de confiabilidad, y se implementó un programa de entrenamiento de evaluadores en línea previo a la aplicación de la escala. El objetivo de este trabajo, que forma parte de un proyecto de investigación mayor cuyo objetivo es la validación del instrumento diseñado de evaluación de la escritura en lengua extranjera, es presentar las percepciones de los docentes evaluadores obtenidas a través de entrevistas.

Palabras clave: Escritura en lengua extranjera - Escritura académica - Escala analítica de evaluación - Entrevistas

Abstract

The assessment of writing in a foreign language is a complex field, which tends to generate anxiety and apprehension on the part of teachers given its effects and implications, both at a pedagogic and at an ethical and political level (Ruecker y Crusan, 2018) and is an area of continuous research, mainly due to the variety and complexity of contexts where it is carried out. The assessment of essays in a foreign language (English) on the part of teachers in charge of the English Language Teaching, Translation and Licentiate in English Language and Literature degree programmes at Universidad Nacional de Córdoba (UNC) has been our object of study for the last six years. Different assessment instruments were designed and an analytic scoring scale proved to be the most suitable one for this specific academic context: the Chair of English Language II. In order to determine the reliability of such analytic scale, a process of validation of the descriptors that conform each of the categories of the scale was carried out. A corpus of sample texts was formed to illustrate the descriptors and thus achieve greater levels of reliability, and an online rater training programme was implemented before the application of the scale. The purpose of this paper, which is part of a larger research project whose aim is the validation of the foreign language writing assessment instrument designed, is to present the perceptions of the raters obtained by means of interviews.

Keywords: Writing in a foreign language - Academic writing - Analytic scoring scale - Interviews

Recepción: 15-12-2022

Aceptación: 23-06-2023

INTRODUCCIÓN

Uno de los desafíos que plantean los procesos de enseñanza y de aprendizaje de la escritura en lengua extranjera está relacionado con la evaluación. Tanto los instrumentos como las prácticas de evaluación de la escritura suelen tener un impacto significativo en el desarrollo de la habilidad de escritura. Este impacto es aún mayor en entornos educativos en los que la evaluación resulta determinante a la hora de acreditar saberes.

Nuestro contexto de trabajo es una asignatura troncal del segundo año de las carreras de grado de inglés de una universidad argentina en la que la evaluación de la escritura cumple un rol muy importante, ya que en evaluaciones parciales y en el examen final los estudiantes deben aprobar una tarea de escritura de ensayo académico. Por lo tanto, durante varios años nos dedicamos a analizar, diseñar y aplicar distintos instrumentos de evaluación y la escala analítica de evaluación de textos académicos resultó ser la más apropiada para nuestro contexto educativo específico.

El objetivo de esta comunicación es presentar las percepciones de docentes evaluadores de escritura sobre el uso de una escala analítica para evaluar ensayos académicos escritos en inglés como lengua extranjera en el nivel superior. Dichas percepciones se obtuvieron a partir de entrevistas realizadas a los docentes como parte de la etapa final de un Programa de entrenamiento de evaluadores, en el marco de un proyecto de investigación enfocado en la validación de este instrumento de evaluación de la escritura.

1. Marco teórico y antecedentes

La evaluación de la escritura, que constituye nuestro objeto de estudio (evaluación de ensayos en exámenes), se puede considerar, según Goodman y Swann (2003) y Hyland (2019), como sumativa, ya que evalúa las producciones de los alumnos de manera formal y se lleva a cabo al final del proceso de enseñanza con el propósito de determinar si los objetivos del curso han sido alcanzados. Este tipo de evaluación también se identifica como evaluación directa de la escritura debido a que dicha habilidad es evaluada por medio de la producción individual de un texto en vez de hacerlo a través de ejercicios relacionados a la escritura (por ejemplo, ejercicios de vocabulario y gramática), lo que se define como evaluación indirecta de la escritura (Coombe, 2018; Purpura, 2016). La evaluación es, por lo tanto, una instancia de la escritura en la que existe una interacción fundamental entre los estudiantes y los evaluadores; es decir, es un tipo de actividad en el que el componente humano juega un rol central (Cushing Weigle, 2016; Ferris y Hedgcock, 2023; Ruecker y Crusan, 2018).

Dos conceptos básicos en el área de la evaluación que resultan relevantes para el presente estudio son la fiabilidad y la validez. El primero se refiere al grado en el que cierta respuesta (un texto, en este caso) es siempre evaluada de la misma manera con

los mismos resultados (Ferris y Hedgcock, 2023). Esto es muy importante debido a que implica que, dado el mismo párrafo o ensayo, diferentes evaluadores deberían asignar la misma nota. Una tarea de escritura puede ser considerada fiable si mide lo que intenta medir de manera consistente tanto en relación a los estudiantes (el mismo estudiante en diferentes ocasiones) como en relación a los evaluadores (la misma tarea realizada por diferentes personas) (Hyland, 2019). En cuanto al concepto de validez, este implica, en términos generales, que la que la prueba o examen focalice en las habilidades o contenidos que se intenta evaluar y no en otros en vez de otras habilidades o contenidos (Coombe, 2018; Hyland, 2019). A pesar de esta distinción conceptual, varios expertos en el campo han concluido que la fiabilidad y la validez están íntimamente relacionadas. Si bien se trata de términos diferentes, en la práctica, están conectadas y es muy difícil que funcionen de manera independiente (Alderson et al., 2017; Brookhart y Chen, 2015; Crusan, 2010; Ferris y Hedgcock, 2023).

En lo que respecta a los enfoques para la evaluación de la escritura, una distinción básica, en la que se enmarca el diseño de los instrumentos de evaluación para este estudio, es la que existe entre el enfoque holístico y el enfoque analítico (Huang y Han, 2013). En el presente estudio, el instrumento que intentamos validar se basa en el enfoque analítico. A diferencia del enfoque holístico, que requiere asignar una nota al texto de manera global, el enfoque analítico implica la identificación de los diferentes aspectos (por ejemplo: contenido, organización, uso de la lengua, etc.) para determinar las fortalezas y debilidades de cada uno, y decidir una nota final. Los enfoques analíticos han sido adecuados y especialmente útiles al evaluar habilidades generales, tales como: la escritura, y textos extensos, por ejemplo, ensayos, en vez de tareas específicas (Caperucci, 2018; Vercellotti y McCormick, 2021; Vercellotti, 2021). El hecho de que los instrumentos de evaluación analíticos proporcionan información sobre los criterios de calidad relacionados a cada aspecto a evaluar los convierte en herramientas de aprendizaje muy valiosas, ya que revelan de manera clara las debilidades y fortalezas de un texto dado, lo que permite guiar a los alumnos para remediar dichas debilidades, revisar sus producciones después de la evaluación y, de esa manera, promueve el aprendizaje (Barkahoui y Hadidi, 2021; Brookhart, 2018; Gallardo, 2020; Huang y Han, 2013; Huff, 2021; Jonsson, 2014).

En cuanto a la validación de los instrumentos de evaluación, este estudio se enmarca en el modelo sociocognitivo de validación propuesto por Weir (2005) y O'Sullivan y Weir (2011), que combina las dimensiones sociales, cognitivas y de contenido, y su relación con el contexto educativo. Se trata de un proceso que implica proporcionar evidencia que apoye las inferencias que deriven de los instrumentos de evaluación en términos de resultados de exámenes y logros en un contexto específico (Purpura, 2016; Weir, 2005). La validación implica, por lo tanto, el análisis de la fiabilidad entre evaluadores

al aplicar tal instrumento, pero también las interpretaciones que hagan los evaluadores de dicho instrumento y su utilidad (Coombe, 2018; Panadero y Jonsson, 2013). Este proceso de análisis (y, en algunos casos, modificación) puede considerarse como una manera de hacer investigación científica que tiene como objetivo recolectar evidencia, consistente con un marco teórico, para determinar, además de la validez, la fiabilidad de tal instrumento (Parkes, 2007). Este proceso de validación es el que intentamos llevar a cabo como parte de nuestro proyecto de investigación para generar tanto evidencia cualitativa como cuantitativa que permita determinar la confiabilidad de una escala analítica para la evaluación de la escritura. Este artículo en particular se centra en los datos cualitativos obtenidos a partir de entrevistas con docentes evaluadores.

Estudios previos han demostrado que la producción de descriptores claros que conforman las bandas de escalas analíticas y holísticas desempeña un rol central en la efectividad, fiabilidad y validez de estos instrumentos de evaluación (Dawson, 2017; Faieza, 2019; Romano y Martínez, 2017; Sampson Anderson, 2017; Velasco-Martínez y Hurtado, 2018). Así, el análisis y la estandarización de tales descriptores es un paso esencial para determinar la confiabilidad entre evaluadores. Según Shaw y Weir (2007), la fiabilidad solo se logra a través de un proceso de entrenamiento de los evaluadores y de la estandarización de los instrumentos de evaluación, los cuales dependen principalmente de textos ilustrativos de referencia, es decir, “aquellos textos que han sido identificados por evaluadores expertos como representativos de un cierto nivel. De este modo, los estándares son comunicados por medio de ejemplos” (p.47).

Entre los procesos que han demostrado ser importantes para validar instrumentos de evaluación y lograr mejores niveles de fiabilidad, se encuentra la selección y el uso de textos ilustrativos de referencia, que ejemplifican los diferentes descriptores (Barkahoui y Hadidi, 2021; Shaw y Weir, 2007; Weir, 2005). La orientación de las escalas a habilidades o temas específicos, más que a productos o habilidades generales, y el entrenamiento de los evaluadores en la aplicación del instrumento (Cushing Weigle 2016, Erguvan y Aksu Dunya, 2020; Faieza, 2019; Gallardo, 2020) también han demostrado contribuir al logro de mayores niveles de fiabilidad.

Estas acciones se enfocan en los aspectos procedurales y cualitativos de la fiabilidad, y complementan los resultados cuantitativos proporcionados por pruebas estadísticas, por ejemplo, estudios correlacionales. Las concepciones más abarcadoras de validez, que incluyen aspectos de la fiabilidad, promueven este tipo de acciones integrales para determinar niveles de fiabilidad interevaluador en vez de emplear resultados de pruebas estadísticas como la única fuente de datos (Chapelle, 2020; Moss, 2003; Parkes, 2007; Wind y Peterson, 2017).

Independientemente del procedimiento empleado para determinar la fiabilidad, y del enfoque de evaluación que se adopte, en general, parece existir un acuerdo en la importancia de asegurar la consistencia y la equidad en la aplicación de criterios por parte de los evaluadores, y en la necesidad de asegurar su entrenamiento continuo.

En el proceso de validación de la escala analítica diseñada, se investigaron las percepciones de los docentes evaluadores obtenidas a través de entrevistas.

2. Metodología

2.1 Descripción del contexto

El estudio sobre las percepciones de los docentes forma parte de un proyecto de investigación mayor, cuyo objetivo es la validación de un instrumento de evaluación de la escritura en lengua extranjera diseñado especialmente para un contexto educativo específico: la cátedra de Lengua Inglesa II.

Este espacio curricular forma parte del segundo año de las carreras de Profesorado, Traductorado y Licenciatura en Lengua Inglesa de una universidad pública en Argentina. Se trata, por lo tanto, de una asignatura de los primeros años en el nivel de grado, caracterizada por la masividad (grupos de entre 80 y 115 estudiantes) y con una carga horaria total de 10 (diez) horas cátedra semanales. Los estudiantes presentan una edad promedio de entre 19 y 21 años y la mayoría de ellos cursó la asignatura correlativa correspondiente (Lengua Inglesa I) el año inmediatamente anterior.

La escritura académica -específicamente, la escritura de ensayos expositivos, de comparación y contraste, y de causa y efecto, de una extensión aproximada de 400 palabras- forma parte de los contenidos del programa, y el desarrollo de la habilidad de escritura es parte de los objetivos a lograr. Es por ello que en las instancias sumativas de evaluación (exámenes parciales y finales) se incluye la evaluación directa de la escritura, a través de una tarea de escritura de ensayo.

2.2. Participantes

En este estudio participaron siete docentes, todos ellos pertenecientes a la cátedra de Lengua Inglesa II. Cuatro de los docentes cuentan con más de diez años de antigüedad en la asignatura y amplia experiencia en evaluación de textos académicos, y los otros tres docentes cuentan con menos de diez años de experiencia docente en la cátedra.

2.3. Procedimiento

De acuerdo a los objetivos planteados y las preguntas de investigación que guían este proyecto, en el que se enmarca el estudio que aquí describimos, se propuso una

metodología mixta, que combina aspectos cuantitativos y cualitativos para determinar la validez y fiabilidad de un instrumento de evaluación de la escritura en lengua extranjera (inglés) previamente diseñado, una Escala analítica de calificación (*Analytic Scoring Scale*). Al seguir los lineamientos del *Common Scale for Writing Project* (CSW) que reportan Shaw y Weir (2007) para evaluaciones internacionales, se llevó a cabo un procedimiento en cuatro etapas.

Las primeras tres etapas comprendieron la búsqueda y selección de textos ilustrativos, la conformación de un corpus de referencia para la revisión y aplicación de la escala, y el diseño e implementación de un programa de entrenamiento de evaluadores en línea previo a la aplicación de la escala para la evaluación de ensayos académicos. En la cuarta y última etapa, se realizaron entrevistas a todos los docentes que habían participado en el programa de entrenamiento, a fin de recabar información cualitativa sobre sus percepciones en relación con dicho programa y con la escala implementada.

Se elaboró una guía de preguntas para llevar a cabo entrevistas semiestructuradas a docentes evaluadores (Ver Anexo A). Se realizaron siete, todas estas de manera virtual a través de videoconferencia durante el mes de mayo de 2022, con una duración promedio de 20 minutos. Se llevó a cabo la transcripción ortográfica de las entrevistas para luego analizar las respuestas de los docentes entrevistados.

En esta comunicación, presentamos los resultados de las entrevistas a docentes, focalizando especialmente en sus percepciones sobre el impacto, las fortalezas y debilidades del instrumento de evaluación diseñado.

Para el análisis de las respuestas, se aplicó un enfoque tipológico (Hatch, 2002); es decir, se tuvieron en cuenta las categorías predeterminadas para organizar el análisis, específicamente, las interpretaciones de los docentes evaluadores con respecto al impacto, las fortalezas y las debilidades de la Escala analítica de evaluación. Las preguntas mencionadas arriba e incluidas en la Guía para entrevistas a docentes indagan de manera directa sobre dichos aspectos. El análisis tipológico contempla la búsqueda de patrones o regularidades en el cuerpo de datos, y de relaciones (de similitud, diferencia, frecuencia, etc.) entre ellos y entre datos provenientes de la literatura en el campo (Hatch, 2002). Esto permite profundizar en la interpretación y el sentido de los datos recolectados en función de las preguntas de investigación. Finalmente, se pueden llegar a identificar temas o “conceptos integradores” para cada una de las categorías o tipologías abordadas (Hatch, 2002, p.156). Para este autor, una de las fortalezas de este tipo de análisis es su eficiencia, en función de las categorías ya establecidas, pero también la flexibilidad que ofrece al combinar procesos deductivos que se aplican al comienzo del análisis con procesos inductivos en etapas sucesivas de interpretación de los datos.

3. Resultados

La entrevista que se le realizó a los docentes constaba de siete preguntas, de las cuales cuatro (Preguntas 3, 4, 5, y 6) se relacionaban específicamente con la Escala analítica (el resto de las preguntas se refería al Programa de entrenamiento de evaluadores). Para este artículo, nos concentramos solamente en las preguntas sobre la Escala.

El análisis de las entrevistas a los docentes, más específicamente, sus respuestas a las preguntas tres a seis, relacionadas con sus percepciones sobre la Escala analítica de evaluación, indican una orientación muy positiva hacia dicho instrumento. A continuación, presentamos una síntesis de las respuestas obtenidas a las cuatro preguntas seleccionadas a los fines de este artículo:

- a. *Pregunta 3: ¿Cree que su modo de evaluar ensayos académicos en el contexto de Lengua Inglesa II ha cambiado luego de los encuentros y de haber utilizado la Escala analítica (Analytic Scoring Scale)? Si su respuesta es afirmativa, ¿podría describir brevemente en qué consiste el cambio percibido?*

La Pregunta 3 de la Guía para entrevistas a docentes les consulta a los entrevistados si se produjeron cambios en su manera de evaluar a partir del uso de la Escala analítica y cuáles fueron dichos cambios. La mayoría de los entrevistados respondieron que sí se produjeron cambios, en mayor o menor medida, en su manera de evaluar, y que esos cambios fueron favorables.

En primer lugar, la mayoría coincide en que las modificaciones se relacionan en gran parte con la ponderación de los distintos aspectos del texto que se consideran en la evaluación. Esto quiere decir que, según los entrevistados, la Escala analítica les permite asignarle un valor a cada aspecto del ensayo (y que ese valor sea luego considerado a la hora de asignar la nota final), a diferencia de la evaluación de corte más holístico que solía realizarse antes de utilizar la escala. Por ejemplo, un participante comenta: “Ahora con la rúbrica, la ponderación es diferente [...] también se consideran otros aspectos, que, aunque tienen distinto grado de importancia, al final suman” [E1]. Y otro docente agrega: “Por ahí me pasaba que le daba mucho a una de las categorías y menos peso a otra y no era tan equitativo, por así decirlo, en algún punto” [E4].

En segundo lugar, otros docentes destacan la utilidad de la Escala a la hora de fundamentar una nota. Esto se relaciona con un aspecto muy importante (e incluso determinante en algunos casos) de la evaluación: la calificación. Algunos entrevistados afirman que la Escala analítica les permite a los docentes brindarles a los estudiantes argumentos para respaldar la nota asignada. Un docente sostiene: “Los descriptores, es como que te ayudan a fundamentar y a encontrar el puntaje que se le puede dar a cada ensayo” [E2], y otro añade: “Tengo un elemento, una herramienta en la cual yo me puedo valer,

ya sea para justificar mi nota o por si el alumno me quiere hacer alguna pregunta” [E5]. Y en relación con la calificación, otro participante comenta que la Escala ofrece “Mayor precisión en la nota a la que se arriba” [E3]. Como podemos observar, y en relación con el cambio en la ponderación mencionado en el párrafo anterior, se detecta una mayor satisfacción por parte de la mayoría de los docentes con la calificación asignada a un ensayo a partir de la escala.

Finalmente, otros participantes mencionan un cambio positivo con respecto a la categorización de los elementos que se consideran a la hora de evaluar un ensayo académico. A partir del uso de la escala, algunos docentes pudieron ubicar en una determinada categoría un elemento que antes no sabían dónde ubicar o que antes pensaban que se relacionaba con una categoría, pero que, al intercambiar opiniones con otros colegas, descubrieron que ellos lo consideraban como perteneciente a otra. Por ejemplo, algunos docentes consideraban el uso de conectores como parte del uso de la lengua, mientras que en la Escala analítica esto se ubica en la categoría Organización. Esta categorización de los elementos o aspectos no es una cuestión menor, dado que contribuye a una mayor consistencia para un mismo evaluador y a un mayor acuerdo entre evaluadores, que son los objetivos de nuestro proyecto de investigación.

b. Pregunta 4: ¿Qué dificultades experimentó al momento de aplicar la Escala analítica (Analytic Scoring Scale) para la evaluación de ensayos académicos?

La cuarta pregunta de la Guía se enfoca en las dificultades encontradas al momento de aplicar la Escala. La mayoría de las respuestas a esta pregunta coinciden en dos grandes grupos de dificultades: por un lado, dificultades relacionadas con algunas bandas y/o descriptores y, por otro, dificultades para evaluar los casos ambiguos (*borderline*).

Con respecto a los descriptores y las bandas, algunos evaluadores expresan la necesidad de cuantificar errores para determinar a qué banda de una determinada categoría pertenece un ensayo, ya que la inclusión en los descriptores de palabras, tales como: “pocos”, “algunos” y “muchos” hace que esos descriptores sean poco precisos. Esto se menciona en particular con relación al uso de la lengua y, más específicamente, con referencia al vocabulario. Por ejemplo, un evaluador observa: “Creo que todos tenemos una vara distinta en cuanto a lo que se considera o no un excelente uso de vocabulario específico y un uso pobre o deficiente” [E1]. Del mismo modo, algunos docentes plantean como una dificultad la distinción entre “pobre”, “insuficiente” y “suficiente”; si bien los descriptores ayudan a despejar dudas, no siempre resulta sencillo hacerlo. Un participante comenta: “Por ahí la dificultad era con eso, de ver cómo lo interpreto yo, cómo lo interpreta otro” [E7].

En lo concerniente a dónde ubicar casos ambiguos (*borderline*), es decir, textos que podrían ubicarse en bandas diferentes de una misma categoría, algunos evaluadores

admiten que, si bien la escala aporta claridad en este sentido, siguen apareciendo casos en los que resulta más complejo ubicarlos. Por ejemplo, un participante afirma que: “Siempre hay esos grises que por ahí pueden generar alguna dificultad a la hora de catalogar algún error ... no solo error, las cosas buenas también” [E2], y otro agrega: “Esa diferencia entre *fair* e *inadequate* es como que ahí era difícil determinar hasta cuánto, sobre todo en lengua. Hasta cuánto era *fair*, hasta cuánto *inadequate*” [E3].

c. *Pregunta 5: ¿Qué categorías de la escala le han resultado más difíciles de evaluar? ¿Podría describir brevemente esas dificultades?*

La pregunta 5 de la Guía para entrevistas a docentes consulta sobre qué categorías fueron más difíciles de evaluar. Las respuestas obtenidas mencionan con mayor frecuencia a tres categorías como difíciles de evaluar: organización, contenido y formato.

Con respecto a la organización, las dificultades encontradas se relacionan con el puntaje asignado a esa categoría (un participante considera que es “muy elevado” [E7]) y con los descriptores (en el sentido mencionado en el análisis de la Pregunta 4, es decir, con relación a la confusión que pueden ocasionar palabras tales como *generally* y *mostly* etc.). Y en estrecha relación con esto, otro participante describe la dificultad para ubicar un ensayo en esta categoría cuando el texto cumple parcialmente con el descriptor de una banda, pero no con la totalidad del descriptor. Por ejemplo, un participante expresa: “Entonces, yo decía qué hago. Porque la otra parte la tiene bien [...] no sabía qué ponerle” [E4]. Esta misma dificultad también es referida en relación con la categoría Contenido, y el problema que surge cuando un texto no se ajusta a la descripción exacta de las distintas bandas. Por otra parte, un evaluador manifestó problemas a la hora de identificar si un error en el texto se relaciona con el contenido o con la organización, pero reconoce que se trata de una dificultad individual y no necesariamente relacionada con la escala.

En lo que respecta a formato (*Format and layout*), algunos evaluadores advirtieron dificultades para comprender inicialmente a qué se refería esta categoría, probablemente por no haberle prestado tanta atención a este aspecto anteriormente, y para identificar las características esenciales de cada banda de esta categoría. Por ejemplo, un evaluador se pregunta: “¿Cuál sería un ejemplo de algo que tenga más o menos puntos? Si bien no es la que tiene más puntaje, es la que me genera más dificultad a la hora de aplicarla” [E2].

d. *Pregunta 6: ¿Cuáles son, según su opinión, las fortalezas y debilidades de la Escala analítica (Analytic Scoring Scale)?*

La sexta pregunta de la Guía indaga sobre las fortalezas y debilidades de la Escala analítica. Entre las debilidades, se destacan, como se mencionó en el apartado anterior, la dificultad para ubicar ensayos que podrían ir en más de una banda [E2] y la falta de

especificidad de algunos descriptores [E4]. Además, los entrevistados mencionan como puntos débiles de la Escala analítica que la nota final asignada es mayor que la obtenida sin la escala (con evaluación holística) [E3], la distribución de los puntajes para las distintas categorías (demasiado puntaje asignado a contenido y organización en vez de al uso de la lengua) [E7] y el tiempo que lleva corregir textos con la escala (especialmente en situaciones de examen final, en las que el tiempo disponible es considerablemente menor al tiempo con el que se cuenta en una situación de examen parcial, por ejemplo) [E6]. Dos evaluadores observan que una debilidad de la escala es que, a su entender, no puede usarse sola (sin ningún otro comentario o anotación por parte del docente) porque el estudiante no llega a comprender la evaluación. Estos evaluadores afirman que el alumno de segundo año no está preparado aún para encontrar los problemas por su cuenta y que, por ende, necesita que el docente los marque en el texto (acompañados eventualmente de retroalimentación indirecta), además de la retroalimentación recibida con la Escala analítica [E1 y E5].

Con respecto a las fortalezas, los entrevistados se refieren, en primer lugar, a la utilidad de las descripciones detalladas de las distintas bandas de una categoría. Esto se considera un aspecto positivo tanto para docentes como para estudiantes. Para los docentes es positivo porque les permite llegar a un puntaje más “fidedigno” [E3] y para los estudiantes es ventajoso porque les permite comprender de mejor manera cómo son evaluados. En ese sentido, un evaluador señala que, tanto para docentes como para estudiantes, “está claro qué cosas se consideran al momento de evaluar y el puntaje que se le da a cada cosa” [E4]. Además, la Escala les permite identificar claramente tanto los logros como los problemas en su texto: “Hablando con los alumnos me decían que sí, que por lo menos ellos sabían que, por ejemplo, lo que es *organization* lo tenían bien” [E4]. Otro aspecto favorable de la Escala identificado por algunos entrevistados es la tendencia hacia la estandarización. Estos evaluadores indican que la Escala es un instrumento útil en tanto y en cuanto “permite estandarizar la forma de evaluación y reducir, en la medida de lo posible, estas subjetividades” [E1].

En un contexto académico universitario en el que se trabaja con comisiones de estudiantes muy numerosas y en el que la evaluación resulta determinante para acreditar saberes y avanzar en una carrera, la cuestión de la estandarización no es menor. Como agrega un entrevistado: “La evaluación a este nivel es crucial para la vida de las personas. Entonces, me parece que está bueno poder estandarizarla, poder tener mayores resguardos, poder explicar mejor las notas, el resultado que se obtiene en estas evaluaciones” [E6].

El análisis de las cuatro preguntas que anteceden nos permite inferir que, a pesar de algunas dificultades encontradas, los docentes evaluadores que utilizaron la Escala analítica de evaluación de ensayos académicos consideran que es una herramienta útil, con gran potencial pedagógico, y que contribuye a disminuir las subjetividades

relacionadas con la evaluación de producciones textuales, que pueden llegar a ser determinantes en un contexto de educación universitaria.

CONCLUSIONES

Los resultados del análisis de las entrevistas a docentes aportan información cualitativa fundamental sobre la utilidad del instrumento de evaluación diseñado y los posibles aspectos a mejorar a futuro. Las percepciones de los evaluadores nos permiten conocer sus interpretaciones con respecto al instrumento de evaluación y las maneras en las que lo valoran, y resultan, por lo tanto, fundamentales para determinar la validez de la Escala analítica en el contexto particular de evaluación en el que se pretende aplicar. Las respuestas de los docentes nos presentan un panorama bastante claro de las fortalezas y debilidades que perciben en el instrumento de evaluación.

Con respecto los aspectos positivos, la utilización de la Escala analítica parece favorecer la atención a diversos aspectos del texto a evaluar (contenido, organización, uso de la lengua y formato), logrando evitar que se sobrevaloren algunos aspectos sobre otros (por ejemplo, al considerar errores en el uso de la lengua sin tener en cuenta el contenido y la estructura del ensayo). En este sentido, las percepciones docentes sobre el instrumento de evaluación diseñado revelan una fortaleza de la escala, ya que apuntan a la posibilidad de lograr una evaluación más adecuada y precisa de todas las dimensiones de un texto y, por lo tanto, de los distintos contenidos relacionados con la escritura académica que forman parte del proceso de enseñanza (Vercellotti y McCormick, 2021).

Otro aspecto positivo que se destaca en las respuestas de los docentes se relaciona con la posibilidad de lograr, a través de las categorías incluidas en la escala y los descriptores específicos, mayor precisión al momento de identificar problemas u oportunidades de mejora en los textos que se evalúan. La posibilidad de identificar y acordar sobre los componentes específicos de cada una de las categorías contribuye, además, a lograr mayor confiabilidad inter e intra evaluadores.

Finalmente, algunos evaluadores (especialmente E4) destacaron el valor formativo que puede tener un instrumento analítico de evaluación (Brookhart, 2018; Huff, 2021), ya que le permite a los estudiantes identificar, con mayor claridad que a través de otros instrumentos, qué aspectos específicos relacionados con la escritura deben mejorar para continuar desarrollando la habilidad de escritura. Esta es una de las fortalezas que más se destaca en el contexto de la asignatura, ya que se trata de un espacio curricular en el que se espera que los estudiantes desarrollen estrategias y autonomía en el aprendizaje, y el uso sistemático de la Escala analítica puede fomentar procesos de autoevaluación (Huff, 2021).

En este sentido, y teniendo en cuenta también una de las debilidades de la Escala analítica identificadas por algunos de los evaluadores (Pregunta 6, [E1] y [E5]), se agregó una tercera columna, a la derecha de cada uno de los descriptores, para que los docentes evaluadores puedan agregar comentarios explicativos o sugerencias concretas sobre cómo resolver problemas recurrentes relacionados con cada una de las dimensiones o categorías de la escala.

Con respecto a las dificultades reportadas por los docentes, se destacan algunos problemas inherentes a todo tipo de evaluación, específicamente, la calificación de casos ambiguos. Esta problemática particular fue abordada en encuentros posteriores entre docentes y se incluyeron ejemplos textuales en el material disponible para los docentes de manera de poder contar con casos orientadores al momento de aplicar determinadas bandas y puntajes. De todos modos, se trata de una problemática compleja que requiere reflexión y puesta en común sistemáticas entre el equipo docente.

Por otro lado, las dificultades relacionadas con las categorías específicas de la escala revelan la importancia de socializar interpretaciones de los descriptores, de manera de poder seleccionar las bandas de forma adecuada y persistente. En este sentido, el material de referencia fue actualizado una vez finalizado el programa de entrenamiento de evaluadores con el objetivo de aclarar confusiones relacionadas con aspectos que deben ser relacionados con el contenido del texto y aquellos que corresponden a la organización textual.

Las percepciones de docentes evaluadores sobre el impacto, las fortalezas y debilidades de una Escala analítica de evaluación nos aportan información fundamental para la implementación del instrumento diseñado, ya que proveen datos cualitativos que nos permitirá avanzar en la validación de dicho instrumento. El proceso mismo de investigación ha contribuido a fomentar la discusión y reflexión grupales sobre las decisiones que informan las prácticas de evaluación de la escritura que sin duda se han enriquecido y potenciado a partir del diálogo.

REFERENCIAS BIBLIOGRÁFICAS

- Alderson, J.; Brunfaut, T. y Harding, L. (2017). Bridging assessment and learning: a view from second and foreign language assessment. *Assessment in Education: Principles, Policy & Practice*, 24(3), 379-387. <https://doi.org/10.1080/0969594X.2017.1331201>
- Barkahoui, K. y Hadidi, A. (2021). *Assessing Change in English Second Language Writing Performance*. Routledge.
- Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*, 3 (22). <https://doi.org/10.3389/feduc.2018.00022>

- Brookhart, S. M. y Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, 67(3), 343-368. <https://doi.org/10.1080/00131911.2014.929565>
- Caperucci, D. (2018). Assessment and certification of foreign language learning through rubrics: A methodological perspective. En A. Lopes y R. Ruiz Cecilia (Eds.), *New Trends in Foreign Language Teaching: Methods, Evaluation and Innovation* (294-307). Cambridge Scholars Publishing.
- Chapelle, C. A. (2020). *Argument-based Validation in Testing and Assessment*. SAGE Publications.
- Coombe, C. (2018). *An A to Z of Second Language Assessment: How Language Teachers Understand Assessment Concepts*. British Council. www.britishcouncil.org/exam/aptis/research/assessment-literacy
- Crusan, D. (2010). Assess thyself lest others assess thee. En T. Silva y P. Matsuda (Eds.), *Practicing theory in second language writing* (pp. 245-262). Parlor Press.
- Cushing Weigle, S. (2016). Second language writing assessment. En R. Manchón y P. K. Matsuda, *Handbook of Second and Foreign Language Writing* (pp. 473-496). De Gruyter.
- Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research, and practice. *Assessment and Evaluation in Higher Education*, 42 (3), 347-360.
- Erguvan, I. y Aksu Dunya, B. (2020). Analyzing rater severity in a freshman composition course using many-facet Rasch measurement. *Language Testing in Asia*, 10(1). <https://doi.org/10.1186/s40468-020-0098-3>
- Faieza, C. (2019). Application of rubrics in the classroom: A vital tool for improvement in assessment, feedback and learning. *International Education Studies*, 12(1), 61-68.
- Ferris, D. y Hedgcock, J. (2023). *Teaching L2 Composition. Purpose, process and practice* (Fourth edition). Routledge.
- Gallardo, K. (2020). Competency-based assessment and the use of performance-based evaluation rubrics in higher education: challenges towards the next decade. *Problems of Education in the 21st Century*, 78(1), 61-79.
- Goodman, S. y Swann, J. (2003). Planning the assessment of student writing. En C. Coffin, M.J. Curry, S. Goodman, A. Hewings, T.M. Lillis y J. Swann. (Eds), *Teaching academic writing. A toolkit for higher education* (pp. 73-100). Routledge.
- Hatch, J. A. (2002). *Doing qualitative research in education settings*. State University of New York Press.

- Huang, J. y Han, T. (2013). Holistic or analytic – A dilemma for professors to score EFL essays? *Leadership and Policy Quarterly*, 2(1), 1-18.
- Huff, J. (2021). College Writing Instructors Using Rubrics to Drive Instruction. *Academia Letters*, Article 2639. <https://doi.org/10.20935/AL2639>
- Hyland, K. (2019). *Second language writing*. Cambridge University Press.
- Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment and Evaluation in Higher Education*, 39(7), 840–852. <https://doi.org/10.1080/02602938.2013.875117>
- Moss, P. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement Issues and Practice*, 22 (4), 13-25. <http://journalofwritingassessment.org/article.php?article=54>
- O’Sullivan, B. y Weir, C. (2011). Language testing and validation. En B. O’Sullivan (Ed.), *Language Testing: Theory & Practice* (pp. 13–32). Palgrave.
- Panadero, E. y Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129-144. <https://doi.org/10.1016/j.edurev.2013.01.002>
- Parkes, J. (2007). Reliability as argument. *Educational Measurement: Issues and Practice*, 26 (4), 2-10. [http://www.jwalkonline.org/docs/Grad%20Classes/Spring%2008/modmeas/general%20articles/Parkes%20\(2007\).pdf](http://www.jwalkonline.org/docs/Grad%20Classes/Spring%2008/modmeas/general%20articles/Parkes%20(2007).pdf)
- Purpura, J. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100(1), 190-208. <https://doi.org/10.1111/modl.12308>
- Romano, M. E. y Martínez, J. I. (2017). Evaluación de la escritura en lengua extranjera (inglés): Análisis de criterios docentes. En F. Ávalos, H. Gargiulo, J. Rodríguez y E. Villanueva de Debat. (Eds.), *Currículum, perspectivas didácticas y evaluación. Didáctica del Español Lengua Segunda y Extranjera. Lengua de señas* (pp. 78-91). Universidad Nacional de Córdoba. https://repositoriosdigitales.mincyt.gob.ar/vufind/Record/RDUUNC_47a9b641ffd7a0880d006acdf60f4a3
- Ruecker, T. y Crusan, D. (2018). *The Politics of English Second Language Assessment in Global Contexts*. Routledge.
- Sampson Anderson, S. (2017). Differences across levels in the language of agency and ability in rating scales for large-scale second language writing assessments. *Studia Anglica Posnaniensia*, 52 (2), 147-172. <https://repozytorium.amu.edu.pl/bitstream/10593/21251/1/Sampson%20Anderson.pdf>
- Shaw, S.D. y Weir, C.J. (2007). *Examining writing. Research and practice in assessing second language writing*. Cambridge University Press.

- Velasco-Martínez, L. C. y Hurtado, J. C. T. (2018). The use of rubrics in higher education and competences evaluation. *Profesorado*, 22(3), 183–208. <https://doi.org/10.30827/profesorado.v22i3.7998>
- Vercellotti, M. (2021). Beyond the rubric: Classroom assessment tools and assessment practice. *TESL-EJ*, 25 (3). <https://www.tesl-ej.org/wordpress/issues/volume25/ej99/ej99a9/>
- Vercellotti, M. y McCormick, D. (2021). Constructing analytic rubrics for assessing open ended tasks in the language classroom. *TESL-EJ*, 24(4). <https://www.teslej.org/wordpress/issues/volume24/ej96/ej96a2/>
- Weir, C. J. (2005). *Language testing and validation. An evidence-based approach*. Palgrave Macmillan.
- Wind, S., y Peterson, M. (2017). A systematic review of methods for evaluating rating quality in language assessment. *Language Testing*, 35(2), 161-192.

¹ Es Doctora en Ciencias del Lenguaje, Facultad de Lenguas, UNC. Se desempeña como Profesora Titular Regular de Lengua Inglesa II y como docente en cursos de posgrado en la Facultad de Lenguas, UNC. Dirige proyectos de investigación relacionados con la temática de la escritura en lengua extranjera desde el año 2014.

² Es Magister en Inglés, Facultad de Lenguas, UNC. Se desempeña como Profesora Titular Regular de Lengua Inglesa II en la Facultad de Lenguas, UNC, y como Profesora Adjunta Regular de Lengua Inglesa II y Lengua Inglesa II en el Departamento de Lenguas, Facultad de Ciencias Humanas, UNRC. Co-dirige proyectos de investigación relacionados con la escritura en lengua extranjera desde el año 2014.

³ Es Doctora en Ciencias del Lenguaje, Facultad de Lenguas, UNC. Se desempeña como Profesora Titular Regular de Lengua Inglesa I, Profesora Titular Interina de Lengua Inglesa IV, Profesora Adjunta Regular de Lengua Inglesa II y como docente en cursos de posgrado en la Facultad de Lenguas, UNC. Integra proyectos de investigación relacionados con la escritura en lengua extranjera desde el año 2002.

ANEXO A

ANALYTIC SCORING SCALE¹

Criteria	Score	Descriptor
Content Appropriateness How well does the writing accomplish its purpose/ address the essay question?	35 -31	Excellent/Very good: Thorough and complete response to the essay question; highly focused on the issue; information is relevant and accurate and/or appropriate (no section within any of the paragraphs is inaccurate or off topic).
	30-26	Good: Complete response to the essay question; adequately focused on the issue; information is mostly relevant and accurate and/or appropriate (one section within any of the paragraphs may be inaccurate or off topic).
	25-21	Fair: Fairly complete response to the essay question (probably with some minor omissions); mostly focused on the issue, but some inaccurate, and/or inappropriate information (two sections within any of the paragraphs may be inaccurate or off topic).
	20-10	Inadequate: Quite incomplete response to the essay question (major omissions); slightly focused on the issue, but information is often irrelevant, inaccurate, and/or inappropriate (one out of two body paragraphs may be inaccurate or off topic).
	9-0	Unacceptable: Incomplete response to the essay question; totally out of focus on the issue; information is irrelevant, inaccurate, and/or inappropriate (more than one body paragraph may be inaccurate or off topic).
Organization How well does the organization of the text adapt to the type of essay and method of development required?	30-26	Excellent/Very good: Fully consistent with method of development/ organization of the particular type of essay required; topic and subtopics are appropriately organized and sequenced according to the thesis statement and topic sentences; relations among text sections are clear to the reader. Thesis statement and topic sentences are appropriate and accurately focused.
	25-22	Good: Generally consistent with method of development/ organization of the particular type of essay required; topic and subtopics are appropriately organized and sequenced according to the thesis statement and topic sentences; relations among text sections are mostly clear to the reader (a few transition signals may be absent or misused). Thesis statement and topic sentences are adequate.
	21-18	Fair: Mostly consistent with method of development/ organization of the particular type of essay required; relations among text sections are sometimes unclear (some transition signals may be absent or misused). Some topic sentences may be absent or incomplete and/or the thesis statement may display some problems in focus.
	17-9	Inadequate: Mostly inconsistent with method of development/ organization of the particular type of essay required (a different type of essay, an extremely short essay, or an essay with extremely short paragraphs); relations among text sections are seldom clear. Some topic sentences may be absent or wrongly focused and/ or thesis statement may be incomplete or unsuitable.
	8-0	Unacceptable: Totally inconsistent with method of development/ organization of the particular type of essay required (the text is not an essay); relations among text sections are frequently unclear. Most topic sentences may be absent or wrongly focused and thesis statement is absent or unsuitable.

¹ Esta escala de evaluación ha sido adaptada y analizada en el marco del Proyecto de investigación: Evaluación de la escritura en lengua extranjera (inglés) en el nivel superior: validación de una escala analítica de evaluación. (SeCyT – UNC).

<p>Quality of expression & Language Accuracy</p> <p>How well does language use suit the context of communication?</p>	25-22	Excellent to very good: Appropriate use of language for the context of communication. Great variety and specificity in the use of vocabulary. Accurate language/grammatical structures. No errors in spelling and punctuation.
	21-18	Good: Generally appropriate use of language for the context of communication. Some variety and specificity in the use of vocabulary. Mostly accurate language/grammatical structures. A few errors in spelling and punctuation.
	17-15	Fair: Barely appropriate use of language for the context of communication. Attempt at variety and specificity in the use of vocabulary. Some inaccurate grammatical structures. Some errors in spelling and punctuation.
	14-8 7	Inadequate: Inappropriate use of language for the context of communication. Little variety and some specificity in vocabulary. Several/Many inaccurate grammatical structures (some basic mistakes in grammar, e.g.: verb forms, agreement). Several errors in spelling or punctuation.
	7-0	Unacceptable: Totally inappropriate use of language for the context of communication. Lack of variety and specificity in the use of vocabulary. Too many inaccurate grammatical structures. Too many errors in spelling and punctuation.
<p>Formal Appropriateness / Layout</p> <p>How well does the writing meet established norms for formatting, capitalization, neatness?</p>	10-9	Excellent to very good: Appropriate use of English academic writing conventions. ² Layout (title format, margins, spacing and indentation), and neatness fully meet established norms.
	8-7	Good: Generally appropriate use of academic English writing conventions. Layout and neatness may show some problems, but generally meet established norms.
	6	Fair: Acceptable use of academic English writing conventions. Layout and neatness partially meet established norms.
	5-3	Inadequate: Mostly inappropriate use of academic English writing conventions. Layout and neatness rarely meet established norms.
	2-0	Unacceptable: Totally inappropriate use of academic English writing conventions. Layout and neatness never meet established norms.

(Adapted from Tribble, 1996 in Kern, 2000, p. 287-289)

² By this, we mean absence of contractions, as well as acknowledgement of sources (in our case this means reference to the rubrics if there is any kind of literal transcription from the quote or instructions to the task and/or reference to background reading, listening or audiovisual material).

ANEXO B

GUÍA PARA ENTREVISTAS A DOCENTES

1. En la encuesta que completó al finalizar el *Rater-training programme* (2021), expresó que el programa le había sido útil. ¿Podría ampliar/comentar sobre qué actividades y/o materiales le resultaron de mayor utilidad?
2. ¿Considera que sería deseable o necesario realizar encuentros como los que se llevaron a cabo en el *Rater-training programme*? ¿Podría referirse brevemente a las razones sobre las que basa su respuesta?
3. ¿Cree que su modo de evaluar ensayos académicos en el contexto de Lengua Inglesa II ha cambiado luego de los encuentros y de haber utilizado la Escala analítica (*Analytic Scoring Scale*)? Si su respuesta es afirmativa, ¿podría describir brevemente en qué consiste el cambio percibido?
4. ¿Qué dificultades experimentó al momento de aplicar la Escala analítica (*Analytic Scoring Scale*) para la evaluación de ensayos académicos?
5. ¿Qué categoría/s de la escala le ha/n resultado más difícil/es de evaluar? ¿Podría describir brevemente esas dificultades?
6. ¿Cuáles son, según su opinión, las fortalezas y debilidades de la *Analytic Scoring Scale*?
7. ¿Desea agregar algún comentario u opinión sobre este tema?