

EVALUACIÓN DE LA ESCRITURA EN LENGUA EXTRANJERA: APLICACIÓN DE UN INSTRUMENTO DE EVALUACIÓN HOLÍSTICA

María Elisa Romano

Romano.mariaelisa@gmail.com

Julia Inés Martínez

juliamartinez@gmail.com

Natalia Verónica Dalla Costa

natidc@yahoo.com

**Facultad de Lenguas
Universidad Nacional de Córdoba**

Eje temático: Prácticas docentes y profesionales

Resumen

Si bien los procesos de enseñanza y de aprendizaje abarcan diversos aspectos, lo cierto es que la evaluación ocupa un lugar muy importante que no siempre es tenido en cuenta a la hora de fundamentar, valorar y proyectar prácticas docentes. Esta importancia se ve notablemente incrementada en contextos académicos, donde la evaluación es determinante para la acreditación de los diversos actores de la comunidad educativa. Es por esto que desde la cátedra *Lengua Inglesa II* se propuso, a partir de 2014, estudiar los procesos de evaluación de la escritura académica, a fin de determinar pautas a seguir para lograr mayor precisión y calidad en dichos procesos. En este trabajo comunicaremos los resultados preliminares de una etapa de la investigación, la cual consistió en la aplicación, por parte de todos los docentes de la cátedra, de una escala holística para calificar ensayos académicos en inglés. Luego de la aplicación, se procedió a analizar los resultados de la calificación, en función de la concordancia entre los distintos evaluadores.

Introducción

Las instancias y los procesos de evaluación, que constituyen un aspecto integral en todo proceso de enseñanza y de aprendizaje, cobran especial importancia en el contexto académico, dado que determinan la acreditación necesaria para que todos los participantes puedan funcionar apropiadamente en la comunidad educativa. Por este motivo, el impacto de la evaluación y los modos de evaluar en las prácticas áulicas es una problemática que merece ser analizada y explorada en detalle, ya que, en muchos casos, los rasgos característicos de los procesos de evaluación determinan las prácticas pedagógicas concretas. La importancia de la evaluación es una realidad que docentes e investigadores no podemos desconocer y sobre la cual debemos reflexionar sistemáticamente.

En lo que respecta a la evaluación de la escritura en lengua extranjera en particular, diversos especialistas coinciden en que se trata de un campo de acción complejo, que suele generar ansiedad y aprehensión entre docentes a raíz de los efectos e implicancias tanto a nivel pedagógico como ético y político (Crusan, 2010a; Cushing Weigle, 2002; Hyland, 2003), y que continúa siendo un campo de permanente investigación, especialmente debido a la variedad y complejidad de contextos en los que se lleva a cabo. En uno de los últimos volúmenes de la revista especializada *Assessing Writing*, la editora destacó la necesidad de contar con más investigaciones situadas en contextos de escritura en lengua extranjera (específicamente, en inglés

como lengua extranjera) que analicen prácticas concretas de evaluación y retroalimentación de la escritura (Lee, 2014).

La evaluación de la escritura es un aspecto constitutivo de la asignatura *Lengua Inglesa II* de las carreras de Profesorado, Traductorado y Licenciatura de inglés de la Facultad de Lenguas, UNC, y se lleva a cabo de diversas maneras y en diferentes instancias. Durante el ciclo lectivo, los estudiantes escriben ensayos como parte de las tareas para realizar fuera del aula y estos ensayos son luego corregidos por los docentes. Si bien estas evaluaciones son informales en el sentido de que no llevan una nota ni determinan la aprobación del curso, sirven a los estudiantes como instancias de práctica previas a las evaluaciones formales. Con respecto a este último tipo de evaluación, el alumno regular de *Lengua Inglesa II* debe aprobar tres evaluaciones parciales, dos de las cuales incluyen un ejercicio de escritura de un ensayo. Por otro lado, el examen final escrito de la asignatura consta de una sección de escritura de ensayo eliminatoria y una sección de escritura adicional para alumnos libres, que también es eliminatoria. Como se puede observar, la evaluación de la escritura es un aspecto fundamental de la materia y, como tal, demanda índices altos de confiabilidad y validez.

En el bienio 2014-2015, llevamos a cabo un proyecto de investigación cuyo objetivo fue explorar los principios y las creencias que guían la aplicación de criterios de evaluación de ensayos por parte de los docentes de la cátedra. Los resultados de dicha investigación demostraron diferencias no sólo en la naturaleza de los criterios de evaluación que utilizan los docentes, sino también, y especialmente, en las prioridades que establecen en el proceso de evaluación de las producciones escritas de los estudiantes. Por esta razón, para el período 2016-2017 nos propusimos diseñar y analizar la aplicación de dos instrumentos diferentes de evaluación de ensayos académicos en lengua extranjera. El objetivo a largo plazo continúa siendo el mismo que nos habíamos fijado en el período anterior: lograr la unificación de criterios que garantice la confiabilidad del sistema de calificación de esta prueba de escritura, que constituye una de las instancias de acreditación obligatoria para aprobar la materia. Tomando como punto de partida los datos recolectados en 2014-2015, nos planteamos diseñar dos instrumentos de evaluación de la escritura, uno de naturaleza holística y otro de características analíticas, y analizar su aplicación por parte de los docentes de la cátedra con el fin de evaluar su confiabilidad y utilidad. En este trabajo en particular presentamos los resultados preliminares de la aplicación de uno de estos instrumentos, diseñado en el marco del proyecto de investigación: una escala holística de evaluación de la escritura (ver Anexo I)

Marco teórico y antecedentes

La evaluación de la escritura objeto de la presente investigación (evaluación de ensayos en exámenes parciales y finales) puede considerarse, siguiendo a Goodman & Swann (2003) y Hyland (2003), como una *evaluación sumativa*, dado que se califica formalmente y se lleva a cabo, generalmente, al final del proceso de enseñanza con el fin de determinar si se han logrado los objetivos planteados para todo el curso. Este tipo de evaluación es identificada, además, como una *evaluación directa* de la escritura; es decir, se evalúa la escritura a través una producción textual individual y no a través de ejercicios de componentes lingüísticos que hacen a la escritura (como, por ejemplo, ejercicios de vocabulario, gramática, etc.), lo que se define como *evaluación indirecta* de la escritura (Williams, 2005, p. 120). La *evaluación directa* de muestras concretas de escritura suele ser considerada como una de las maneras más efectivas de lograr una apreciación de la habilidad escrituraria de un estudiante (Hamp-Lyons, 2001; Williams, 2005), especialmente por tratarse de una instancia real de escritura, en la que existe

una interacción entre escritores y evaluadores; en otras palabras, se trata de un tipo de evaluación en el que el componente humano es central (Ferris & Hedgecock, 2014; Hamp-Lyons, 2001).

Dos términos de uso frecuente en el área de evaluación (no sólo de evaluación en escritura) y fundamentales para este estudio son confiabilidad (*reliability*) y validez (*validity*). El primero hace referencia a la medida en la cual una determinada respuesta (un texto, en el caso de la escritura) será evaluada siempre de la misma manera con los mismos resultados (Ferris & Hedgecock, 2014). Este punto es especialmente controversial en el caso de la escritura, ya que implica que, ante un mismo ensayo, dos evaluadores diferentes deberían arribar a la misma calificación. Una tarea de escritura puede considerarse confiable si mide lo que intenta medir de manera consistente tanto en lo que respecta a los estudiantes (el mismo estudiante en diferentes ocasiones) como en lo que respecta a los evaluadores (la misma tarea a la vista de distintos evaluadores) (Hyland, 2003).

En cuanto al segundo concepto, implica que la prueba debe evaluar lo que se pretende evaluar, no otras habilidades y/o contenidos (Williams, 2005). Hamp-Lyons (1990) identifica cuatro tipos de validez: validez aparente (*face validity*), validez de contenido (*content validity*), validez de criterio (*criterion validity*) y validez de constructo (*construct validity*), pero sostiene que la validez de constructo (la evaluación refleja la realidad psicológica del comportamiento en el área que se evalúa) es un tipo de validez que engloba a todas las demás. Por su parte, la autora (Hamp-Lyons, 2003) ha propuesto una nueva visión de la validez que es más integradora y que incluye la validez de contenido, el control de la calidad, la validez de constructo, la utilidad de la evaluación, la habilidad de la evaluación de tener significado y relevancia en más de un contexto (generalización de validez) y la validez consecuencial (*consequential validity*). Messick (1996) destaca la importancia de que las evaluaciones tengan “representación de constructo” (*construct representation*) y “relevancia de constructo” (*construct relevance*). La primera característica se refiere al hecho de que la evaluación debería ser tan similar a la habilidad real como sea posible, mientras que la segunda se refiere al hecho de que la evaluación debe evaluar sólo las habilidades que son parte del constructo (no debería haber dificultades en la evaluación que no sean parte de la habilidad/destreza misma). Finalmente, la posibilidad de llevar a cabo la evaluación en un marco institucional es un punto muy importante a tener en cuenta, ya que la disponibilidad o limitación de tiempo, recursos, infraestructura, etc. suelen determinar las prácticas de evaluación. Estos aspectos son identificados por algunos autores como componentes de la validez contextual (*context validity*) de la evaluación (Shaw & Weir, 2007).

En lo que respecta a abordajes de evaluación de un texto, una distinción básica (en la cual enmarcamos el diseño de instrumentos de evaluación) es la que existe entre el *enfoque holístico* y el *enfoque analítico*. La *evaluación holística* aborda al texto como un todo y conlleva la asignación de una nota o evaluación final general. Esta calificación se basa en la premisa de que el constructo de la escritura es una entidad única que puede capturarse mediante una escala única que integra las cualidades inherentes de la escritura y que la calidad sólo puede ser reconocida por lectores expertos (y no mediante instrumentos objetivos). Entre los problemas típicos de este tipo de calificación se destacan la aplicación de diferentes criterios por parte de los evaluadores y las diferencias en el contexto en el que se llevan a cabo las evaluaciones, que suelen derivar en resultados disímiles (Hamp-Lyons & Kroll, 1996; Hamp-Lyons, 2003).

El *enfoque analítico* implica la identificación de distintos aspectos del texto, de modo de determinar las fortalezas y debilidades de cada uno de los aspectos, para luego establecer una evaluación final. Este enfoque implica la calificación de rasgos múltiples y considera el constructo de la escritura como un fenómeno complejo y multifacético.

Una de las ventajas de esta forma de evaluar los textos es que les permite a los evaluados analizar las fortalezas y debilidades en sus propios textos, y de esta manera, orientar estrategias de revisión y mejora de los propios productos y procesos de escritura (Deane et al., 2008; Hamp-Lyons, 2003; Williams, 2005). Hamp-Lyons (2003) afirma que, a la hora de evaluar la escritura en una segunda lengua, este tipo de calificación resulta más conveniente. Asimismo, sostiene que sirve como herramienta para investigar lo que sucede durante el proceso de evaluación y que es esencial para los docentes de escritores en lengua extranjera, ya que les brinda información valiosa que facilitará la toma de decisiones. Sin embargo, algunos investigadores afirman que la evaluación analítica detallada a través de una lista de criterios requiere más tiempo, esfuerzo y entrenamiento, sin evidenciar mayor efectividad que la evaluación holística en la calificación de textos escritos (Freedman, 1981; Hunter, Jones y Randhawa, 1996; Mei, 2010; Robb Singer & Lemahieu, 2011). Algunos estudios recientes en evaluación de la escritura parecen comprobar uno de los principios expresados por Perkins (1983): la evaluación holística es más efectiva o útil cuando el objeto de evaluación es la capacidad o habilidad general del estudiante para escribir (en lengua extranjera o materna), más que una producción textual en particular (Panou, 2013; Rezaei & Lovorn, 2010).

Independientemente del enfoque que se favorezca, parece existir un acuerdo generalizado sobre la importancia de asegurar la consistencia y equidad en la aplicación de criterios por parte de los evaluadores; tal como afirma Crusan (2010b), “es muy importante seguir estándares rigurosos para garantizar que nuestras evaluaciones sean válidas, justas y apropiadas al contexto y los propósitos para los cuales son diseñadas” (p. 32). Teniendo en cuenta estas investigaciones, las cuales dan cuenta de la complejidad inherente a la evaluación de la escritura, y considerando la naturaleza particular de nuestro contexto educativo, planteamos como uno de los objetivos de nuestra investigación, diseñar un instrumento holístico para la evaluación de la escritura en lengua extranjera y determinar la confiabilidad de dicho instrumento en virtud del contexto de enseñanza y aprendizaje propio de *Lengua Inglesa II*.

Metodología

En esta investigación participaron todos los docentes de la cátedra *Lengua Inglesa II* de las carreras de Profesorado, Traductorado y Licenciatura de inglés de la Facultad de Lenguas, UNC, que estuvieron a cargo de la evaluación de ensayos en instancias de evaluación (profesores asistentes, adjuntos y titulares): ocho evaluadores en total.

Se propuso una metodología que combine aspectos cuantitativos y cualitativos, de modo de poder determinar qué instrumentos de evaluación son más efectivos en el contexto particular de escritura en lengua extranjera en el que nos encontramos.

En 2016, se diseñó la escala holística de evaluación. Para el diseño de este instrumento se tuvieron en cuenta investigaciones previas (Ferris y Hedgecock, 2014; Kern, 2000; Mei, 2010; Shaw y Weir, 2007), así como también los resultados obtenidos en el proyecto de investigación 2014-2015 sobre criterios docentes y ponderación en la evaluación de la escritura.

En 2017, una evaluadora experta seleccionó diez (10) ensayos escritos por estudiantes en la segunda evaluación parcial escrita del ciclo lectivo inmediatamente anterior (septiembre de 2016). Durante el mes de abril, se llevó a cabo una sesión de entrenamiento en la que se aclaró y acordó la implementación de la escala holística de evaluación para la evaluación de los 10 ensayos seleccionados. Cada docente evaluó dichos textos de manera individual en el término de una semana. Luego de haber aplicado la escala holística, se le solicitó a cada uno de los evaluadores que seleccionara, dentro de cada una de las bandas que había utilizado para calificar los ensayos, el descriptor que consideraba prioritario o determinante en la elección de la

banda. Una vez recolectados los 80 ensayos (10 por cada evaluador) se llevó a cabo un análisis de correspondencia entre evaluadores.

Resultados

La aplicación de la escala holística reflejó variedad en las calificaciones otorgadas a los distintos ensayos por cada uno de los evaluadores.

Tabla 1: Bandas seleccionadas por los evaluadores para la calificación de los ensayos

| | E 1 | E 2 | E 3 | E 4 | E 5 | E 6 | E 7 | E 8 | E 9 | E 10 |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| Banda 5 | | | | | | | | | | 1 |
| Banda 4 | 1 | | | | | 1 | | 1 | 1 | 2 |
| Banda 3 | 6 | 1 | 1 | 1 | | 4 | 1 | 4 | 5 | 5 |
| Banda 2 | 1 | 7 | 5 | 5 | 7 | 2 | 5 | 3 | 2 | |
| Banda 1 | | | 2 | 2 | 1 | 1 | 2 | | | |

Esta tabla contiene parte de los resultados obtenidos, específicamente, el número de evaluadores que seleccionó cada una de las bandas de la escala holística aplicada para calificar cada uno de los 10 ensayos. En negrita se indica el mayor número de evaluadores que coincidieron en la selección de cada una de las bandas. Como podemos observar, en ningún caso encontramos coincidencia entre los 8 evaluadores; el número mayor es 7, en el caso de los ensayos 2 y 5, y la correspondencia se da en la banda 2, es decir, en el caso de un ensayo en el que la mayoría de los evaluadores consideró que el texto estaba desaprobado.

En lo que respecta a las bandas que describen ensayos aprobados, la mayor coincidencia se observa en el caso del ensayo 1, al cual 6 de los 8 evaluadores calificaron de acuerdo a la banda 3; otros casos de coincidencia se evidencian en los ensayos 9 y 10, en los que 5 evaluadores coincidieron en aplicar la banda 3. De estos dos últimos casos, el que nos indica mayor uniformidad entre evaluadores es el correspondiente al ensayo 10, ya que los tres evaluadores que no aplicaron la banda 3 aplicaron las dos bandas superiores respectivamente, es decir, en este caso, todos los evaluadores consideraron que el ensayo estaba aprobado.

Los ensayos 10 y 5 se destacan como los únicos dos casos en los que hay acuerdo entre todos los evaluadores en lo respecta a la aprobación o aplazo de la tarea, ya que, aunque no calificaron con las mismas bandas, todas indican que el ensayo estaría aprobado (E 10) o desaprobado (E 5).

En la tabla siguiente se muestran los mismos resultados en porcentajes. Tal como describimos anteriormente, en los exámenes 2 y 5 es donde hubo mayor acuerdo en la banda escogida, mientras que en el 6 y en el 8 es donde se produjo mayor disparidad.

Tabla 2: Porcentaje de acuerdo entre evaluadores para la selección de bandas

| Ensayo | Banda | Porcentaje de evaluadores |
|--------|-------|---------------------------|
| 1 | 3 | 75% |
| 2 | 2 | 88% |
| 3 | 2 | 63% |
| 4 | 2 | 63% |
| 5 | 2 | 88% |
| 6 | 3 | 50% |
| 7 | 2 | 63% |
| 8 | 3 | 50% |
| 9 | 3 | 63% |
| 10 | 3 | 63% |

Como puede observarse, las bandas 2 y 3 fueron las más empleadas; es decir, la mayoría de las calificaciones oscilaron entre la primera banda de aplazo (Banda 2) y la banda más baja de aprobación (Banda 3).

Aunque la comparación de estos resultados con los de estudios previos se dificulta aquí debido a que no siempre se utilizan los mismos métodos de análisis, podemos mencionar algunos datos relevantes de investigaciones en las que se aplicaron instrumentos holísticos de evaluación en contextos similares. Por ejemplo, Robb Singer y LeMahieu (2011) reportan un 94% de acuerdo entre evaluadores, un porcentaje mayor al obtenido en nuestro estudio.

Más en detalle, los descriptores de este instrumento de evaluación empleados con mayor frecuencia, como mínimo, en cuatro instancias, fueron los siguientes:

Tabla 3: Descriptores aplicados con mayor frecuencia por los evaluadores

| Descriptor | Banda | Nº Aplicaciones |
|---|-------|-----------------|
| <i>Many content points included with some omissions (some subtopics are underdeveloped; some points lack supporting ideas).</i> | 3.1 | 16 |
| <i>Some major content points inadequately covered, omitted, and/or irrelevant.</i> | 2.1 | 16 |
| <i>Several ideas inadequately organized with an unsuitable method of development.</i> | 2.2 | 13 |
| <i>Some ideas inadequately organized, but with a suitable overall method of development.</i> | 3.2 | 8 |
| <i>Too many content omissions and/or irrelevance due to misinterpretation of the essay question.</i> | 1.1 | 4 |
| <i>Little control of language with several errors which obscure communication.</i> | 2.4 | 4 |
| <i>Some control of language with some errors which may impede communication.</i> | 1.4 | 4 |

Como podemos observar, son siete los descriptores que fueron identificados con mayor frecuencia por los evaluadores como determinantes en la elección de una determinada banda. Aquellos que sólo fueron seleccionados sólo una, dos o hasta tres veces no han sido incluidos en esta tabla. Entre estos descriptores, la mayoría hace referencia al contenido del texto (3.1; 2.1; 1.1) o a la organización del ensayo (2.2; 3.2). Sólo en ocho casos se identificaron como determinantes descriptores relacionados con el uso

de la lengua (2.4 y 1.4). Estos resultados coinciden con los resultados de un estudio anterior, en el que los docentes evaluadores expresaron que priorizan estos aspectos del texto (contenido y organización) sobre la corrección gramatical y el uso de vocabulario al momento de evaluar y calificar un texto (Romano & Martínez, 2017).

Es interesante observar que los descriptores que presentan mayor frecuencia (3.1 y 2.1) pertenecen a dos bandas diferentes: una de aprobación (Banda 3) y otra de aplazo (Banda 2); es decir, el contenido del texto fue considerado como determinante tanto en casos en los que los evaluadores decidieron aplazar como en casos en los que aprobaron los ensayos. Ahora bien, los descriptores relacionados con uso de la lengua fueron seleccionados como determinantes sólo en los casos de aplazo (2.4 y 1.4); es decir, hubo situaciones en las que los evaluadores expresaron que los errores o problemas relacionados con aspectos léxico-gramaticales orientaron la selección de la banda.

Si bien se puede inferir, a partir de la selección de descriptores, un acuerdo generalizado en lo que respecta a qué aspectos del texto deben ser priorizados al momento de evaluar (específicamente, contenido y organización del texto), la forma en que se califica cada uno de estos aspectos a través de la selección de bandas varía sustancialmente. Las razones que pueden explicar estas variaciones pueden tener que ver no solo con diferencias individuales entre evaluadores, sino también y principalmente con las concepciones que cada uno de los evaluadores tiene sobre los indicadores de habilidad escrituraria en relación con cada uno de los aspectos mencionados anteriormente. Asimismo, y tal como lo comprobaron Rezaei y Lovorn (2010), las diferencias también pueden deberse a impresiones subjetivas de los evaluadores, que se conforman durante la primera lectura del texto y que tienen gran influencia en la manera en la que se aplica la escala de evaluación para calificar un texto.

Conclusiones

Estos resultados preliminares nos indican que, a pesar de que existen ciertos puntos de acuerdo, se debe continuar trabajando para lograr mayores niveles de confiabilidad entre evaluadores.

Especialmente, parecería que es necesario realizar un trabajo conjunto para establecer los criterios de aprobación y aplazo, a fin de que esa calificación tan significativa y de alto impacto para los estudiantes no presente variaciones, como hemos encontrado en la mayoría de los textos evaluados en esta etapa de nuestra investigación.

Con respecto a los aspectos determinantes que orientaron la selección de bandas, concluimos que, si bien el contenido y organización de los textos han demostrado ser prioritarios, los aspectos relacionados con la corrección gramatical y el uso del léxico parecerían jugar un papel importante al momento de seleccionar las bandas más bajas, es decir, al momento de reprobar un texto.

Este tipo de trabajos nos permite, además de indagar sobre modos e instrumentos de evaluación más confiables, reflexionar sobre las propias prácticas evaluativas, lo que constituye el primer paso hacia la incorporación de nuevas pautas.

Referencias bibliográficas

- Crusan, D. (2010a). "Assess thyself least others assess thee". En Silva, T. y Matsuda P. (Eds.), *Practicing theory in second language writing*. West Lafayette, Indiana: Parlor Press, pp. 245-262.
- Crusan, D. (2010b). *Assessment in the second language writing classroom*. Michigan: Ann Arbor.
- Cushing Weigle, S. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

- Deane, P; Odendhal, N.; Quinlan, T.; Fowles, M.; Welsh, C. & Bivens-Tatum, J. (2008). *Cognitive models of writing: Writing proficiency as a complex integrated skill*. Princeton, NJ: ETS.
- Ferris, D. & Hedgcock, J. (2014). *Teaching L2 Composition. Purpose, process and practice*. New York: Routledge.
- Freedman, S.W. (1981). "Influences on evaluators of expository essays: Beyond the text." *Research in the Teaching of English*, 13 (3), pp. 245–255.
- Goodman, S. & Swann, J. (2003). "Planning the assessment of student writing". En Coffin, C; Curry, M.J.; Goodman, S.; Hewings, A.; Lillis, T.M. & Swann, J. (Eds), *Teaching academic writing. A toolkit for higher education*. USA: Routledge, pp. 73-100.
- Hamp-Lyons, L. (2001). "Fourth generation writing assessment." En Silva, T. & Matsuda, P. (Eds.), *On second language writing*. New Jersey: Lawrence Erlbaum Associates, pp. 117-128
- Hamp-Lyons, L. (1990). "Second language writing: assessment issues". En Kroll, B (ed.) *Second Language Writing. Research insights for the classroom*. Estados Unidos: Cambridge University Press, pp.69-87.
- Hamp-Lyons, L. & Kroll, B. (1996). "Issues in ESL writing assessment: An overview". *College ESL*, 6 (1), pp. 52-72.
- Hunter, D. M., Jones, R., & Randhawa, B.S. (1996). "The use of holistic versus analytic scoring for large-scale assessment of writing". *The Canadian Journal of Program Evaluation*, 11 (2), pp. 61-85.
- Hyland, K. (2003). *Second language writing*. Nueva York: Cambridge University Press.
- Kern, R. (2000). *Literacy and language teaching*. Oxford: Oxford University Press.
- Lee, I. (2014). "Feedback in writing: Issues and challenges". *Assessing Writing*, 19, pp. 1-5.
- Mei, W. (2010). "Investigating raters' use of analytic descriptors in assessing writing". *Reflections on English Language Teaching*, 9 (2), pp. 69-104.
- Messik, S. (1996). "Validity and washback in language testing". *Educational Testing Service Research Report*. Princeton: ETS. Recuperado de <http://files.eric.ed.gov/fulltext/ED403277.pdf>
- Panou, D. (2013). "L2 writing assessment in the Greek School of Foreign Languages". *Journal of Language Teaching and Research*, 4, pp. 649-654.
- Perkins, K. (1983). "On the use of composition scoring techniques, objective measures, and objective tests to evaluate ESL writing ability". *TESOL Quarterly* 17, pp. 651-671.
- Rezaei, A. & Lovorn, M. (2010). "Reliability and validity of rubrics for assessment through writing". *Assessing Writing*, 15, pp. 18-39.
- Robb Singer, N. & Lemahieu, P. (2011). "The effect of scoring order on the independence of holistic and analytic scores". *Journal of Writing Assessment*, 4 (1), pp. 1-13.
- Romano, M.E. & Martínez, J.I. (2017). "Evaluación de la escritura en lengua extranjera (inglés): Análisis de criterios docentes". En Ávalos, F.; Gargiulo, H.; Rodríguez, J. & Villanueva de Debat, E. (Eds.), *Currículum, perspectivas didácticas y evaluación. Didáctica del Español Lengua Segunda y Extranjera. Lengua de señas*. Córdoba: Universidad Nacional de Córdoba, pp. 78-91. Recuperado de <https://rdu.unc.edu.ar/handle/11086/5574>
- Shaw, S.D. y Weir, C.J. (2007). *Examining writing. Research and practice in assessing second language writing*. Cambridge: Cambridge University Press.
- Williams, J. (2005). *Teaching writing in second and foreign language classrooms*. Chicago: McGraw-Hill.

ANEXO I
ESCALA HOLÍSTICA PARA LA EVALUACIÓN DE ENSAYOS ACADÉMICOS

| BAN D | MARK S | DESCRIPTORS |
|------------------|------------------------|--|
| 5 | 10 9 [40-36] | <p><u>Full completion of the essay writing task</u></p> <ul style="list-style-type: none"> ▪ All content points included (with appropriate expansion). ▪ Ideas effectively organized, with an appropriate method of development. ▪ Wide range of accurate structures, content-specific vocabulary and transition signals. ▪ Complete control of language with minimal errors. ▪ Register and format consistently appropriate to purpose and audience. |
| 4 | 8 7 6 [35-28] | <p><u>Very good completion of the essay writing task</u></p> <ul style="list-style-type: none"> ▪ Most content points included with minor omissions (some examples or expansions may be omitted; some subtopics could have been further elaborated or developed). ▪ Ideas clearly organized, with an appropriate method of development. ▪ Very good range of accurate structures, content-specific vocabulary and transition signals. ▪ Very good control of language with some errors when attempting more complex language. ▪ Register and format on the whole appropriate to purpose and audience. |
| 3 | 5 4 [27-24] | <p><u>Good completion of the essay writing task</u></p> <ul style="list-style-type: none"> ▪ Many content points included with some omissions (some subtopics are underdeveloped; some points lack supporting ideas). ▪ Some ideas inadequately organized, but with a suitable overall method of development. ▪ Good range of accurate structures, content-specific vocabulary and transition signals. ▪ Some control of language with some errors which may impede communication. ▪ Inconsistent attempts at register and format appropriate to purpose and audience. |
| 2 | 3 [23-16] | <p><u>Fair completion of the essay writing task</u></p> <ul style="list-style-type: none"> ▪ Some major content points inadequately covered, omitted, and/or irrelevant. ▪ Several ideas inadequately organized with an unsuitable method of development. ▪ Limited range of structures, content-specific vocabulary and transition signals. ▪ Little control of language with several errors which obscure communication. |

| | | |
|----------|---------------------------|---|
| | | <ul style="list-style-type: none"> ▪ Unsuccessful attempts at appropriate register and format. |
| 1 | <p>2 1 [15-1]</p> | <p><u>Poor completion of the essay writing task</u></p> <ul style="list-style-type: none"> ▪ Too many content omissions and/or irrelevance due to misinterpretation of the essay question. ▪ Lack of organization and no clear method of development. ▪ Narrow range of structures, content-specific vocabulary and transition signals. ▪ No control of language with several errors which obscure communication. ▪ Little or no awareness of appropriate register and format. |
| 0 | | <ul style="list-style-type: none"> ▪ No completion of the task. |