



Elaboración de criterios uniformados que aseguren fiabilidad entre calificadoros en Lengua Inglesa. Segunda Parte. Proyecto de investigación

Massa, Agustín y Granata, Ana María

Fecha de recepción: 30/11/2018

Fecha de aprobación: 11/12/2018

Eje: Currículum, perspectivas didácticas y evaluación

Resumen

El presente trabajo tiene como objetivo mostrar los resultados y conclusiones de un proyecto de investigación llevado a cabo en el período 2016-17 por profesores de la asignatura Lengua Inglesa I de la Facultad de Lenguas, U.N.C. que formó parte de un proyecto mayor denominado “Hacia la elaboración de criterios uniformados de evaluación que aseguren fiabilidad entre los calificadoros de las pruebas escritas en las cátedras de Lengua Inglesa I, Práctica Gramatical del Inglés, Gramática Inglesa I, Práctica de la Pronunciación del Inglés, Lengua Castellana I, y Lecto comprensión en Lengua Extranjera IV (Inglés)” avalado por la Secyt. Lengua Inglesa I es una asignatura troncal de primer año de las carreras de Profesorado, Licenciatura y Traductorado Público Nacional de Inglés de la Facultad de Lenguas, UNC y su objetivo es lograr que los alumnos desarrollen las cuatro macro habilidades de la lengua a fin de alcanzar un nivel intermedio en su conocimiento de inglés. La escritura de párrafos académicos de aproximadamente 200 palabras es una de las secciones de las evaluaciones escritas que más tiempo de corrección demanda por parte de los docentes. Cada comisión cuenta con un número elevado de estudiantes y como el cuerpo docente a cargo suele fluctuar, se consideró necesario acordar ciertos criterios de que aseguraran que las correcciones de los párrafos se logaran de la manera más efectiva y objetiva posible. Por este motivo, cinco de los profesores de la asignatura formamos parte del subgrupo 1 del proyecto. La Profesora titular de la Cátedra B, actuó como evaluador externo. Con el presente trabajo intentaremos mostrar las dificultades que se presentan al momento de evaluar párrafos y trataremos de brindar algunas respuestas a este problema.



Introducción

El sistema educativo universitario requiere más y mejores niveles de educación y enseñanza, y a ello deben contribuir nuestras prácticas evaluativas. Es necesario entender que la evaluación es intrínseca a los procesos de enseñanza y de aprendizaje y por consiguiente mantiene una relación directa con la comprensión de la función pedagógica de la evaluación. Dicha función pedagógica debería actuar como un regulador continuo de los aprendizajes. Para que ello ocurra, el docente debe considerar la importancia de dicha regulación al momento de enseñar y de evaluar. Ya desde la planificación de la clase el docente debe tener en claro sus objetivos de enseñanza y de evaluación a fin de lograr aunar criterios entre lo que se enseña y lo que se evalúa. Debemos tener presente que como resultado de las evaluaciones los estudiantes son promovidos a niveles superiores o se ven obligados a permanecer en el que están. Esto genera mucha incertidumbre en el alumno ya que de su desempeño en una instancia evaluativa depende, en gran parte, su futuro en la universidad. Tanto alumnos como docentes deben entender la validez educativa de la evaluación con el objetivo de lograr que lo que se enseñe esté en concordancia con lo que se evalúe y esto, a su vez, provea del marco adecuado para que los alumnos logren autonomía de trabajo dentro de un marco demandante pero previsible

El presente trabajo se enmarca dentro del proyecto bianual del grupo de investigación que dirige el Magíster Fabián H. Negrelli en la Facultad de Lenguas, UNC denominado “*Hacia la elaboración de criterios uniformados de evaluación que aseguren fiabilidad entre los calificadores de las pruebas escritas en las cátedras de Lengua Inglesa I, Práctica Gramatical del Inglés, Gramática Inglesa I, Práctica de la Pronunciación del Inglés, Lengua Castellana I, y Lectocomprensión en Lengua Extranjera IV (Inglés)*” desarrollado a lo largo de los años 2016 - 2017. En nuestro caso, trabajamos en la asignatura Lengua Inglesa I. A su vez, y a los fines de la presentación en este congreso, nos dividimos en dos subgrupos: el primer grupo se encargará del planteo del problema, antecedentes y metodología y nuestra ponencia girará en torno a los resultados de los tres encuentros, grillas y conclusiones. Este proyecto surge por la necesidad de unificar criterios a la hora de evaluar a nuestros alumnos en una cátedra con numerosos docentes. Comprender que la evaluación es



intrínseca a los procesos de enseñanza y de aprendizaje tiene directa relación con la comprensión de la función pedagógica de la evaluación. Una de nuestras funciones docentes es monitorear en forma continua el proceso de aprendizaje de nuestros alumnos, para lo cual necesitamos herramientas que nos permitan evaluar mientras enseñamos, caso contrario deberíamos “sacrificar” tiempo dedicado a la enseñanza en sí, en pos de la evaluación. Este dilema suele presentarse a lo largo de nuestra carrera docente. Ante el interrogante ¿Qué es más importante: enseñar o evaluar? nos planteamos que ambos procesos son importantes, por lo que es necesario encontrar la forma más efectiva de evaluar mientras enseñamos. A la problemática que por sí misma presenta la evaluación, en nuestra cátedra se suma el hecho de que es una cátedra con un gran número de alumnos en diversas comisiones y un numeroso grupo de docentes que se reparten dichas comisiones. Un tema esencial que se enseña en esta materia es la redacción de párrafos académicos, el cual tiende a ser sumamente subjetivo en el momento de la corrección. El gran número de alumnos inscriptos en primer año de la Facultad de Lenguas lleva a que cada docente deba corregir muchos párrafos durante el año, así como en las distintas instancias evaluativas (exámenes parciales y finales). Además, en los exámenes finales, el párrafo constituye un ejercicio eliminatorio. Para lograr efectividad en la corrección y criterios de corrección uniformados se decidió aunar, una vez más, criterios de corrección y elaborar grillas para corroborar que la evaluación resulte más efectiva y objetiva.

Acerca de la fiabilidad y la objetividad en las correcciones de las evaluaciones escritas

Es importante proveer de fiabilidad y objetividad a las evaluaciones escritas a las que se enfrentan nuestros alumnos para poder aprobar una asignatura en los exámenes, tanto parciales como finales, de la universidad. Sabemos que la objetividad es intrínseca a la tarea de medir. Decimos por tanto, que una medición es objetiva si dos o más agentes, aplicando las mismas técnicas y procedimientos, llegan a resultados similares con un grado mínimo de desvío. Si bien existen otros instrumentos para realizar la evaluación de los aprendizajes de los estudiantes, las pruebas, tanto escritas como orales constituyen actualmente el único instrumento



que contempla el Reglamento de Exámenes vigente en la Facultad de Lenguas de la Universidad Nacional de Córdoba (en adelante FL, UNC) para evaluar a los alumnos que han obtenido la regularidad en una asignatura y determinar si están en condiciones de ser promocionados al nivel inmediato superior. El tipo de “prueba” que fue objeto de análisis en la presente investigación puede considerarse, siguiendo a Hyland (2003), como una evaluación sumativa, dado que se califica formalmente y se lleva a cabo al final de los procesos de enseñanza y de aprendizaje con el fin de determinar si se han logrado los objetivos planteados para todo el curso. En la asignatura Lengua Inglesa I la “prueba” se constituye en una herramienta de poder y de control y en una suma de puntos con el objetivo de generar una nota o calificación que impacta definitivamente en la decisión del docente de promover a los estudiantes a niveles superiores, lo cual, se torna, al menos, cuestionable.

Para concluir esta sección del informe, es necesario resaltar que en pos de llevar a cabo una evaluación justa en el aula, lo cual implica que sea equitativa, válida y transparente, el objetivo central de nuestro estudio fue asegurar la fiabilidad de las calificaciones de las pruebas escritas en nuestra asignatura. Luego de haber concluido nuestra tarea, confiamos en que los resultados de esta investigación aumentarán significativamente la fiabilidad en la corrección de las pruebas escritas en nuestra cátedra, ya que se tenderá a que los alumnos obtengan calificaciones sustancialmente más discriminatorias al haber una mayor concordancia entre evaluadores. Por otro lado, si bien nuestro proyecto fue concebido teniendo en cuenta algunas asignaturas de los planes de estudio de las carreras de Profesorado, Traductorado y Licenciatura en Inglés, los resultados muy probablemente sean de interés para mejorar el proceso de evaluación de otras materias en las que se observe la misma problemática.

Discusión

La forma de evaluar en la enseñanza superior es un aspecto muy importante que sin lugar a dudas, forma parte del proceso de formación de futuros profesionales. Como docentes e investigadores de la Universidad Nacional de Córdoba debemos reflexionar sobre la manera en que encaramos la evaluación en nuestros cursos.



Los criterios de evaluación son esenciales para determinar si se han alcanzado los objetivos de la asignatura, es decir, si los alumnos han logrado alcanzar el nivel propuesto por la cátedra. Al ser tan importantes estos criterios, es fundamental saber si el evaluado logra alcanzar el dominio del área en cuestión. Sin embargo, a veces esos criterios pueden fallar en su propósito, ya que al realizar una interpretación de la escala, los evaluadores pueden diferir en su interpretación. Autores como Watts y Garcia Carbonell, 2006 y Robb Singer y Lemahieu (2011) sugieren la adopción de criterios con descriptores de diferentes grados de corrección de las respuestas. De acuerdo con estos autores, esta clase de criterio aún a la rapidez de criterios holísticos que los evaluadores experimentados usan, así como también los criterios analíticos que favorecen a los evaluadores con menos entrenamiento.

Modalidad de trabajo

Se trabajó en grupos y en tres sesiones a fin de aunar criterios y de ver el progreso en el proceso de selección de los escritos.

Primera sesión

En esta primera instancia los docentes tomaron contacto con los párrafos a evaluar. Si bien dichos párrafos podrían resultarles conocidos a algunos de los evaluadores, debido a que fueron seleccionados de entre exámenes finales, el hecho de que se tipearan sin hacerles ninguna corrección hizo que los mismos resultasen totalmente nuevos para el equipo evaluador. Cada docente fue provisto de un set de siete párrafos a corregir con la idea de que corrigiera los párrafos según los lineamientos establecidos de antemano por la cátedra.

Los resultados de la primera sesión fueron un poco desalentadores. Cada docente evaluador se guió por los criterios de corrección consensuados por todos los miembros de la asignatura (cantidad de puntos que se descuentan en los casos en los cuales los alumnos escriben más de un párrafo, incluyen ideas no relevantes o cometen errores básicos de lengua). Se detectaron algunas diferencias en los puntajes otorgados por los distintos evaluadores; en este sentido, las discrepancias más notables fueron identificadas en relación a problemas con el título otorgado al párrafo, con la oración que presenta la idea principal de la temática que se desarrollará en el párrafo (*Topic Sentence*) y con el uso incorrecto de ciertas



estructuras gramaticales y el uso (no) relevante del léxico (*specific vocabulary*). Dichas diferencias resultaron muy significativas ya que, en algunos casos; las calificaciones oscilaban entre un 4 ó superior (aprobado) a un 2 ó 3 (desaprobado). Este fenómeno se observó en el caso de los párrafos 1, 3, 5, 6 y 7.

Gracias al debate generado entre los docentes evaluadores en esta primera sesión, se confeccionó una grilla analítica en donde se especificó el porcentaje de la nota que se restaría por cada uno de los distintos tipos de errores cometidos por los alumnos (Ver Apéndice 1). Esta grilla se utilizó para evaluar los mismos párrafos en la segunda sesión.

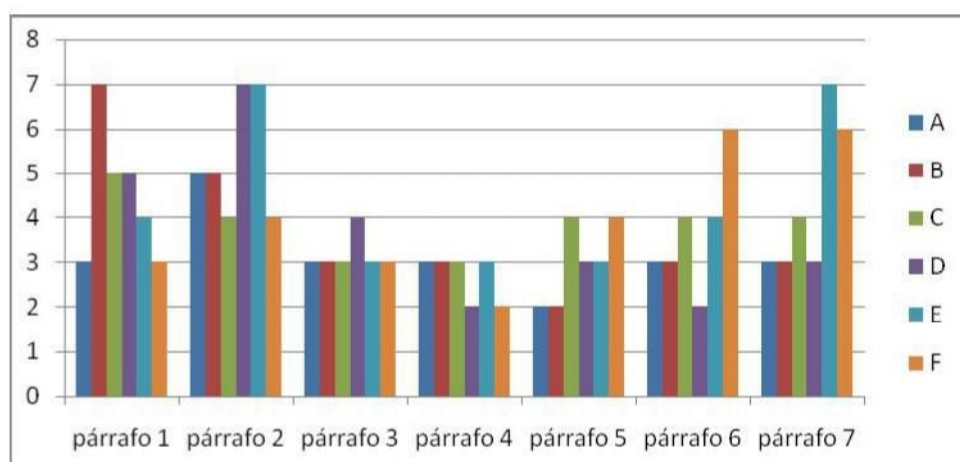


Figura 1: Resultados de la primera sesión

En la grilla elaborada luego de la primera sesión se tuvieron en cuenta los errores más frecuentes cometidos por los alumnos, tales como cantidad de párrafos escritos (este tipo de error es causa de aplazo, ya que se espera que el alumno produzca un solo párrafo académico siguiendo las convenciones enseñadas), relevancia de lo escrito, es decir que si se debe referir a las ventajas de un determinado tema, se refiera a eso y no incluya las desventajas y viceversa; cantidad y calidad de ideas soporte (*supporting ideas*), falta de título o inclusión de un título con errores (título que no sigue las convenciones académicas) o título irrelevante (que no se ajusta al contenido del párrafo), no hay oración que establece el tema (*topic sentence*), o ésta es incorrecta; no hay conclusión o la conclusión es incorrecta. También se tuvieron en cuenta otros errores como falta de vocabulario específico, no hay *idioms* o



collocations enseñadas en clase o relativos al tema evaluado, otros errores en el uso del léxico (preposiciones, ortografía, conectores), hay errores gramaticales o de *use of English* (sintaxis, errores de concordancia, orden lógico de las palabras en la frase, gerundios, infinitivos, plural de sustantivos, errores en los sustantivos contables o incontables, plural de sustantivos, uso u omisión de artículos, uso de la voz pasiva, verbos modales y tiempos verbales, oraciones incompletas o demasiado largas). También se tuvo en cuenta si el alumno incluyó el conteo de palabras al final del párrafo. En esta grilla, se incluyó en una segunda columna el porcentaje de puntos que serían restados en el caso de cada error.

Segunda sesión

En esta segunda instancia, se trabajó nuevamente con los mismos párrafos pero teniendo en cuenta la grilla elaborada.

Como se puede observar en la Figura 2, la grilla elaborada permitió consensuar los casos de los párrafos 4, 5, y 6, puesto que todos los docentes evaluadores les asignaron una nota inferior a 4. Sin embargo, la grilla no fue efectiva para el resto de los párrafos. Resulta llamativo el caso del párrafo 2 en donde el docente corrector D le asignó la nota 7 mientras que dos docentes correctores lo evaluaron con 6 y otros dos con 3 y 2 respectivamente. En el grupo de discusión se manifestó la dificultad de utilizar una grilla en donde se “descuentan” puntos y en donde no hay claridad en cuanto al puntaje a descontar en el caso de errores gramaticales y de vocabulario. Por estos motivos, se decidió elaborar una nueva grilla en donde se procediera de manera inversa, es decir, que, en vez de restar, se sumaran puntos según cada sección. Para determinar cada sección se adaptó el cuadro propuesto por Brown (2007, p. 413) teniendo en cuenta los objetivos de la asignatura Lengua Inglesa I y la grilla confeccionada previamente (Ver Apéndice 2). Esta adaptación redundaría en evaluar en base a los aciertos y no en base a los errores de los alumnos. Este cambio podría resultar sumamente beneficioso para los alumnos ya que ellos recibirían una devolución basada en sus aciertos y fortalezas y no en sus errores lo cual tiende a ser sumamente desalentador, especialmente en los alumnos de primer año de carrera universitaria donde todavía muchos de ellos no están seguros de su vocación.

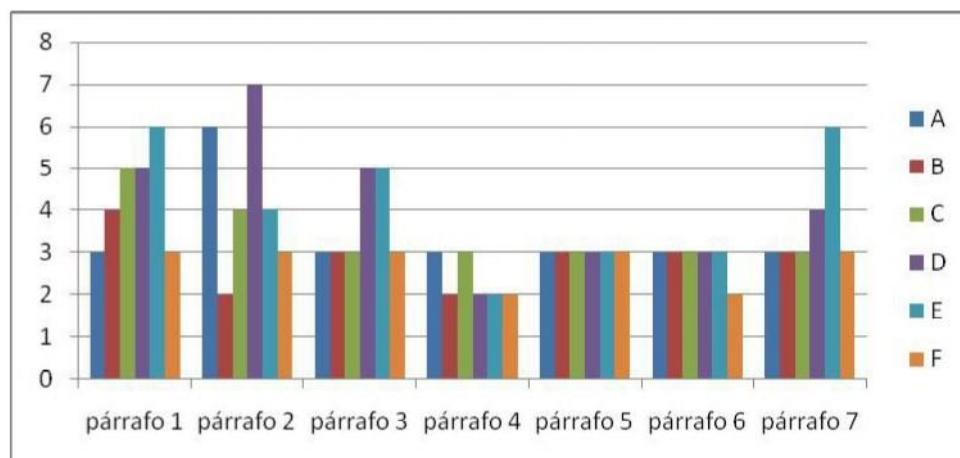


Figura 2: Resultados de la segunda sesión

En la grilla elaborada luego de esta sesión, se tuvieron en cuenta los encabezados, todos ellos con puntaje que se sumaría en base a los aciertos de los alumnos en lugar de restar puntos en base a los errores. (Sin embargo, y para una mayor claridad entre los docentes correctores, también se incluyó cuánto se restaría según cada error, aunque en la copia corregida, que recibirían los alumnos, solamente figurarían los aciertos y los puntos asignados y no los descontados). Las distintas grandes categorías que se incluyeron en esta segunda grilla fueron: contenido, organización y discurso, sintaxis / uso de la lengua (*use of English*), y vocabulario. Cada una de estas categorías incluía a su vez otras subcategorías que eran las siguientes: En contenido, se incluyó: la oración que establece el tema (*topic sentence*) es efectiva, ideas principales relevantes, buenas ideas secundarias, buena conclusión. Como se ve, estas rúbricas describen aciertos y no errores, aunque, y a sólo efecto de servir de guía entre los profesores correctores, también se incluyeron algunos errores frecuentes (que coinciden con la contraparte de los aciertos) y el puntaje a descontar por cada uno de ellos: el título es irrelevante, oración que establece el tema (*topic sentence*) incorrecta, 1 de 3 *support* es irrelevante, 1 de 2 *support* es irrelevante, *irrelevant supporting details*, conclusión incorrecta. En el rubro organización y discurso se incluyó: título, *topic sentence*, 2 o 3 ideas principales, conclusión, longitud (180-220 palabras), el alumno incluye el número de palabras, cohesión (referencia, conector, cambio en el sujeto), registro,



puntuación. La contrapartida de estos aciertos era la siguiente: no hay título o el título es una oración, Hay más de un párrafo, uso del modo imperativo, uso de contracciones y no se incluye el número de palabras del párrafo. En el rubro sintaxis y uso de la lengua (*use of English*), los aciertos que se incluyeron fueron: patrones verbales, concordancia, orden de las palabras, gerundio/infinitivo, formación del plural, uso del artículo, sustantivos contables o incontables, uso de la voz pasiva, verbos modales y tiempos verbales. En el caso de la contrapartida de estos aciertos, simplemente se dejó en claro que por cada caso incorrecto se descontaría un punto y que si el mismo error se repite varias veces solamente se descontaría una vez. En este rubro también se incluyeron los errores de escribir una frase larga e inconexa o de escribir el fragmento de una frase. Finalmente, en el rubro vocabulario, se incluyeron los siguientes aciertos: palabras y frases relevantes, colocaciones, preposiciones, ortografía, mayúsculas con sus correspondientes errores: no hay vocabulario relevante, Por cada pronombre indefinido (*something, anyone, anything, etc*). Se estableció que se restaría un punto por cada caso incorrecto.

Las evaluaciones del nuevo set de párrafos con la nueva grilla de corrección tuvieron muchas coincidencias (Figura 3). Hubo coincidencia total en cuanto a no aprobar a los alumnos en el caso de los párrafos 1, 3, 6 y 7. También hubo coincidencias con respecto al párrafo 4 aunque las notas variaron entre una calificación de 6 (70%) a 10 (100%). Con respecto al párrafo 5, cuando la profesora evaluadora D fundamentó las razones por las cuales le había asignado pocos puntos al párrafo en las secciones “contenido” y “organización y discurso” de la grilla, los evaluadores B, C y F estuvieron de acuerdo en que el puntaje que ellos habían asignado en esa sección tendría que haber sido menor. El único párrafo con el que no hubo acuerdo fue el párrafo 2.

La grilla confeccionada para la última sesión, la cual indicaba que se deberían sumar los puntos en virtud de los aciertos de los alumnos, fue más efectiva ya que permitió a los profesores evaluadores acordar criterios y poder determinar en qué aspectos había discrepancias, en el caso de haberlas. El subgrupo 1 va a seguir perfeccionando esta grilla para futuras instancias de evaluación y para colaborar en la formación de profesores asistentes, profesores adscriptos y de los nuevos docentes que en el futuro se sigan incorporando a la cátedra.

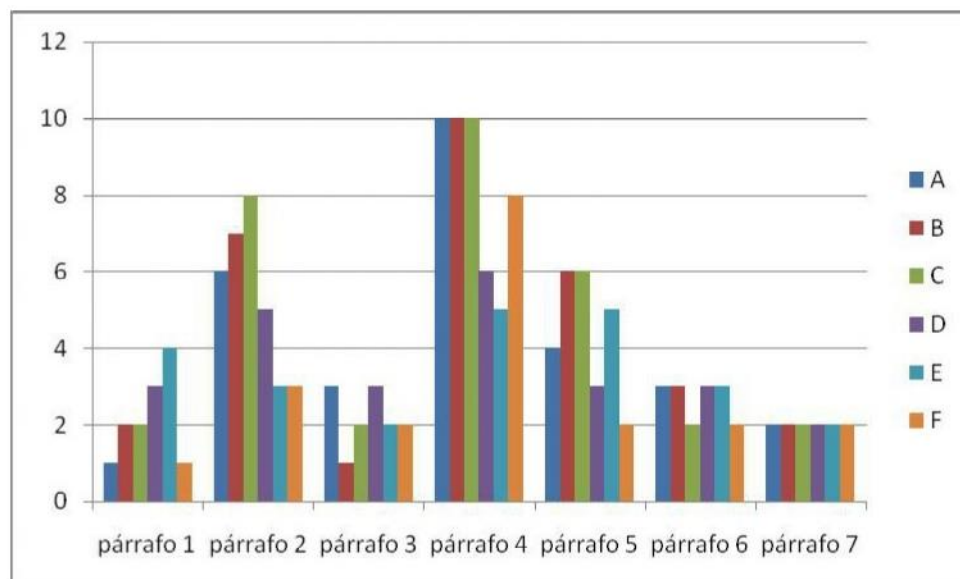


Figura 3: Resultados de la tercera sesión

Apéndice 1: Grilla elaborada en la primera sesión

Problema	% que se descuenta
Hay más de 1 párrafo	74%
2 de las 3 ideas secundarias (<i>supports</i>) son irrelevantes	70%
1 de las 2 ideas secundarias (<i>supports</i>) son irrelevantes	70%
1 de las 3 ideas secundarias (<i>supports</i>) es irrelevante	20%
No hay título o el título es incorrecto (es una oración, es irrelevante, etc.)	10%
No hay TS o la TS es incorrecta	20%
No hay conclusión o la conclusión es incorrecta	15%
No hay uso específico del léxico relacionado con el tema (words, semi-fixed expressions, fixed expressions, idioms, collocations)	30%
Hay errores en el uso del léxico (words, semi-fixed expressions, fixed expressions, idioms, collocations, prepositions, spelling, connectors)	30%
Hay errores en el uso de la gramática (syntax, agreement, WO, gerund / infinitive, plural of nouns, article, [C], [U], passive voice, modal verbs, verb tense, sentence fragment, run-on-sentence)	30%
No incluye el WC o hay un problema con el WC	5%



Apéndice 2: Grilla elaborada en la segunda sesión basada en Brown (2007)

CATEGORÍAS	PUNTAJE
Contenido	
<ul style="list-style-type: none"> • La TS es efectiva • Ideas principales relevantes • Buenas ideas secundarias • Buena conclusión 	10 p
<ul style="list-style-type: none"> • El título es irrelevante 	-4
<ul style="list-style-type: none"> • TS incorrecta 	- 8
<ul style="list-style-type: none"> • 1 de 3 “support” es irrelevante 	- 8
<ul style="list-style-type: none"> • 1 de 2 “support” es irrelevante 	-18
<ul style="list-style-type: none"> • Irrelevant supporting details 	- 4
<ul style="list-style-type: none"> • Conclusión incorrecta 	- 6
Organización y Discurso	
<ul style="list-style-type: none"> • Título • TS • 2 or 3 ideas principales • Conclusión • Longitud (180-220 palabras) • Se incluye el número de palabras • Cohesión (referencia, conector, cambio en el sujeto) • Registro • Puntuación 	10 p
<ul style="list-style-type: none"> • No hay título o el título es una oración 	- 4
<ul style="list-style-type: none"> • Hay más de un párrafo 	- 30
<ul style="list-style-type: none"> • Uso del modo imperativo 	-1
<ul style="list-style-type: none"> • Uso de contracciones 	-1
<ul style="list-style-type: none"> • No se incluye el número de palabras 	- 2
Syntax / Use of English	
<ul style="list-style-type: none"> • Patrones verbales • Concordancia • Orden de las palabras • Gerundio/infinitivo • Formación del plural • Uso del artículo • [C], [U] • Voz pasiva • Verbos modales • Tiempos verbales 	10 p



• Para cada caso incorrecto	-1
• Si el mismo caso se repite varias veces (e.j. “*people is”) se descuenta una sola vez	-1
• Frase larga e inconexa (<i>run-onsentence</i>)	-1
• Fragmento	-1
Vocabulario	
• Palabras y frases relevantes • Colocaciones • Preposiciones • Ortografía • Mayúsculas	10 p
• No hay vocabulario relevante	- 5
• Por cada pronombre indefinido (<i>something, anyone, anything, etc</i>)	- 1
• Por cada caso incorrecto	- 1
TOTAL	40p

Conclusión

A modo de conclusión, podemos decir que siempre es de suma importancia considerar a la evaluación como parte fundamental del proceso de aprendizaje y no separarlo de él. Para nuestros alumnos, el proceso de evaluación debe ser presentado como parte del proceso educativo; de esta manera se logrará ahuyentar los fantasmas del temor a ser observados y calificados de una forma subjetiva y alejada de lo enseñado. Para lograr esto, consideramos que la toma de conciencia de la subjetividad que podemos tener los correctores al momento la evaluación de un párrafo, por un lado, y en la elaboración de distintas grillas hasta encontrar aquella que, sin ser perfecta, permita que los distintos miembros de la cátedra corrijan de manera más uniforme, es un paso adelante hacia el objetivo de establecer criterios fiables de corrección de exámenes.

Referencias bibliográficas

- Brown, H. D. (2007). *Teaching by Principles. An Interactive Approach to Language Pedagogy*. New York: Pearson Education.
- Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*, 12, 17-29.



- Robb Singer, N, & Le Mahieu, P. (2011).The effect of scoring order on the independence of holistic and analytic scores.*Journal of Writing Assessment, 4*.
- Watts, F. & García Carbonell, A. (2006). *La evaluación compartida: investigación multidisciplinar*. Valencia: Servicio de Publicaciones de la Universidad Politécnica de Valencia.