

Factores que intervienen en la evaluación de la escritura ELE en estudiantes de secundaria

Factors Involved in the Assessment of Spanish Second Language Writing in Secondary School Students

Diana Acevedo-Whitney

Escuela Secundaria de Katy (EEUU)

DianaEWhitney@katyisd.org

RESUMEN

El presente estudio surgió por la necesidad de descubrir y revisar los efectos del uso del baremo en los procesos de autoevaluación y la retroalimentación en el mejoramiento de las habilidades de escritura de estudiantes adolescentes de E/LE A1 y A2 en la escuela secundaria. En esta publicación solo se presenta el desarrollo de la segunda pregunta e hipótesis de la investigación original, que tratan sobre los factores y variables que intervienen o influyen en los procesos de evaluación de los alumnos sobre la habilidad escrita, como auto-evaluadores y de las evaluaciones del profesor. Se determinó que el nivel de aprendizaje y los grupos de evaluación (pares y el profesor) no fueron los únicos factores que intervinieron en los procesos de evaluación llevados a cabo en la secuencia de redacciones y re-ediciones que se realizaron como tratamiento durante el estudio en cuestión.

Palabras clave: autoevaluación, baremos, evaluación, nivel de aprendizaje, reconocimiento de errores

ABSTRACT

The present study wanted to discover, and review the effects of the use of a rubric in the process of self-evaluation and feedback. Also, for the improvement of the writing skills in adolescent students of Spanish as a Second Language, in levels A1 and A2 in high school. This article presents the development of the second research question, and hypothesis of the original study, which dealt with factors, and variables that intervene or are involved in the evaluation of student's writing ability, during self-evaluation, and the evaluations performed by the teacher. It was determined that the level of SSL learning, and assessment groups (peer and teacher) were not the only influencing variables when evaluating a sequence of written and edited compositions, that were given as treatment during this research.

Keywords: self-evaluation, rubrics, evaluation, level of learning, error recognition

1. INTRODUCCIÓN

La evaluación puede ser definida de forma simple, como “un juicio de base” (Boud y Falchikov, 2007), y de manera compleja, como un proceso que sigue unas etapas: “la intención, la medida, el juicio y la decisión” según Cuq (2007: 163); de acuerdo a este autor, dichas etapas incluyen las metas, la selección de la medida para juzgar y tomar decisiones, según la modalidad de la tarea y el tiempo en el cual se realiza.

Actualmente existe una cultura de aprendizaje formativo en donde se enfatiza la existencia de una estrecha relación entre el aprendizaje y la evaluación. El modelo de Dochy, Segers, Gijbels y Struven (2007) representa a la evaluación como un instrumento para el aprendizaje, en vez de un instrumento de aprendizaje”, y en la cual son partícipes, los alumnos y el profesor. En el aula de ELE la evaluación en manos de los alumnos ha adquirido una gran fuerza, hasta el punto de considerarse tan importantes como la evaluación docente. Sin embargo este tipo de didáctica, donde se trabaja con la retroalimentación por pares y la autoevaluación, no se ha implementado del todo a nivel global, como sostienen Serrano, Pons y Ruiz (2007) y Berridge (2007),

ya que aún hay instituciones y cuerpo docente que continúan en la línea tradicional de enseñanza y no permiten que los estudiantes compartan la responsabilidad de evaluar dentro de las aulas. Hay adicionalmente una necesidad importante de realizar cambios legislativos que promuevan el uso de estrategias de evaluación por pares y autoevaluación en las clases de ELE de secundaria, por lo menos a nivel estatal en algunas partes de Estados Unidos.

Se han dado mayores pasos a nivel investigativo en relación a la retroalimentación por pares en las aulas de educación superior, con estudiantes nativo-hablantes y aprendientes de una lengua extranjera, pero aún es necesario investigar a fondo los aspectos relacionados con la retroalimentación y la autoevaluación en las asignaturas de ELE en las escuelas de educación secundaria. No se han explicado claramente cuáles son los factores o variables que intervienen o influyen en las evaluaciones de la habilidad escritora en este tipo de contexto escolar. A partir de este precedente, se implementó el estudio del cual se hace referencia a continuación.

2. MÉTODO

El objetivo fundamental de la investigación fue descubrir y revisar los efectos del uso del baremo en los procesos de autoevaluación y retroalimentación en el mejoramiento de las habilidades de escritura de estudiantes adolescentes de ELE A1 y A2 en la escuela secundaria.

En este artículo, solo se presentará el desarrollo de una de las preguntas e hipótesis planteadas en la investigación. La pregunta fue:

- ¿Qué elementos y variables caracterizan la evaluación de una secuencia de composiciones escritas por estudiantes norteamericanos de ELE, cuando dicha evaluación se desarrolla por autoevaluación, por pares o cuando la realiza un docente?

La hipótesis planteaba que el nivel de competencia en la escritura del estudiante, va en incremento no solo debido a la retroalimentación que recibe como input, sino debido a las destrezas de autoevaluación que va paulatinamente desarrollando con la activación, durante las revisiones de la habilidad de saber aprender y de procesos metacognitivos, en respuesta a la interacción del alumno con el instrumento de evaluación. También a través de la influencia de factores sobre el desempeño del estudiante durante los procesos de escritura y edición.

En este estudio participaron estudiantes de 9º a 11º grado, registrados en seis aulas de español de una escuela de secundaria norteamericana. Se mantuvieron los grupos intactos; la muestra se constituyó con 145 estudiantes, 70 de nivel A1 y 75 de nivel A2, que asistieron a una clase diaria de 40 minutos con el mismo profesor, quien fue el mismo investigador. Las aulas de nivel A1 tenían 33 mujeres y 37 hombres, y las de nivel A2 33 mujeres y 42 hombres. Sus edades oscilaban entre los 14 y 17 años.

Todos los estudiantes recibieron un entrenamiento inicial sobre el manejo del baremo mixto modificado de Jacobs et al (1981), dada su relativa vigencia desde que se elaboró y que fue aplicado en estudios recientes como el de Matsuno (2009) y Tamjid y Birjandi (2012). Este baremo se usó en un estudio anterior en Acevedo (2013). Se realizaron dos sesiones para entrenar a los estudiantes en la revisión de dos textos diferentes trabajando por pares (ver Anexo 1). Los alumnos asignaron puntuaciones parciales y holísticas de los textos con el uso de las categorías del baremo. Posteriormente, se inició la fase de pre-escritura, donde los estudiantes cumplimentaron un primer cuestionario sobre la habilidad general de escritura según la evaluación subjetiva del estudiante. Este cuestionario constaba de enunciados simples que conducían a la selección de una respuesta cerrada de sí o no. Este modelo de cuestionario se modificó de Oscarson (2009: 104). El cuestionario 1, en la Tabla 1, se aplicó también después de las fases del tratamiento al final del estudio.

Cuestionario 1		
<i>Por favor, marque SÍ o NO si está de acuerdo o no con los enunciados siguientes:</i>	Sí	No
1. Pienso que puedo escribir frases en español.		
2. Pienso que puedo escribir un párrafo en español.		
3. Pienso que puedo escribir una composición corta en español (entre 50-75 palabras).		

4. Pienso que puedo escribir una composición de más de 75 palabras.		
---	--	--

Tabla 1. Cuestionario 1 con preguntas auto-evaluativas sobre la habilidad compositiva del alumno

Durante la investigación, los estudiantes escribieron tres redacciones, evaluaron por pares y autoevaluaron sus composiciones. Las fases y los procesos de evaluación y recolección de los datos se sintetizan en la tabla a continuación.

Un segundo cuestionario, en la Tabla 2, fue respondido por los estudiantes en el momento de terminar cada redacción; éste contenía las mismas categorías del baremo lo cual buscaba facilitar el proceso de calificación parcial y total, de las categorías y de la composición en general.

En esta tarea, estoy satisfecho con... (Marcar más de uno si es necesario)	Pero pienso que podría mejorar, o puedo tener errores en... (Marcar más de uno si es necesario)	Mi puntuación basada en el baremo, para cada categoría es...
contenido	contenido	
vocabulario	vocabulario	
ortografía	ortografía	
gramática	gramática	
puntuación	puntuación	
Puntuación total		

Tabla 2. Cuestionario 2 para la autoevaluación de la composición post-redacción

Se aplicó un tercer cuestionario, en la Tabla 3, para recoger las reflexiones o apreciaciones de los estudiantes respecto a la retroalimentación dada por los pares; si estaban de acuerdo o en desacuerdo, y si reconocían los errores marcados o comentados, antes de la re-edición de la composición.

Cuestionario 3
Mire su composición, la puntuación y las correcciones que los revisores sugieren. Tome su tiempo para leer y responder a continuación a todas las preguntas relativas a la retroalimentación recibida.
1. ¿Está de acuerdo con las correcciones sugeridas por el revisor? Si responde <i>sí</i> o <i>no</i> , explique por qué.
2. ¿Está de acuerdo con la puntuación recibida? Si responde <i>sí</i> o <i>no</i> , explique por qué.
3. ¿Está consciente de algunos errores que pasó por alto antes de la revisión? De algunos ejemplos y explique por qué están mal.
4. ¿Ha aprendido algo nuevo con la retroalimentación recibida en relación a sus errores, si los tuvo? Explique.

Tabla 3. Cuestionario 3, aplicado post-revisión y tras la retroalimentación del revisor

3. RESULTADOS

Para entender los factores y variables que influyeron en la autoevaluación durante el estudio, los análisis se enfocaron solo en los estudiantes que cumplimentaron los cuestionarios y redactaron las tres composiciones. El número final de informantes fue de 76, 34 en el aula A1 y 42 en el aula A2. Para iniciar, se realizó una prueba de Chi cuadrado sobre los datos encontrados en el cuestionario 1 de pre-escritura. Con un intervalo de confianza del 95% se pudo determinar que las aulas eran homogéneas al comienzo del estudio. Posteriormente, los registros obtenidos de las calificaciones de autoevaluación de los estudiantes en cada fase de evaluación se compararon con las del profesor-investigador durante y al final del estudio. Las diferencias encontradas entre los puntajes de autoevaluación (AE) y del profesor-investigador (EP) fueron

denominadas delta. Este dato se calculó para cada fase del estudio y se denominaron: delta 1, delta 2 y delta 3.

Para indagar sobre la hipótesis de la investigación, acerca de los elementos que intervienen en la evaluación, se procedió a realizar comparaciones entre las variables: nivel de aprendizaje, género, grupos de evaluación (pares y el profesor), frecuencia de errores pre-edición y post-edición, retroalimentación y el reconocimiento de los errores; además de la influencia de los errores de pre-edición sobre los errores post-edición. Cabe aclarar que la variable de género no se había estimado al comienzo de los cálculos, pero cuando se realizaron algunos análisis, se vio que era importante incluirlos.

Primero se analizó si el género y el aprendizaje estuvieron relacionados con las diferencias delta de los puntajes. En la prueba 1, se omitieron 2 casos donde las puntuaciones de autoevaluación fueron demasiado bajas con respecto al resto del grupo y afectaban la normalidad del histograma de esta variable, ya que mostraban una diferencia entre los puntajes de AE y EP muy alta. La media (M) del valor delta para el aula A1 fue de 1.74 y la desviación estándar (D.E) de 8.76, y en el aula A2, la media fue de 2.90 y la desviación estándar de 2.90. La comparaciones entre el valor delta y el nivel de aprendizaje a través de la prueba ANOVA fue: $F(1,71) = 0.382$, $p = 0.584$, $R^2 = 0.010$. Esto fue evidencia significativa para afirmar que el nivel de aprendizaje de los estudiantes influyó en la variabilidad de la diferencia entre las puntuaciones de AE y EP. De otro lado, con respecto al género, la prueba ANOVA arrojó una diferencia significativa: $F(1, 71) = 8.045$, $p = 0.006$, $R^2 = 0.089$; el género es causante del 9% aproximadamente de las variaciones en la AE de los estudiantes en esta primera fase. La media de delta para el sexo femenino fue más alta, lo que sugiere que las adolescentes tuvieron mayores discrepancias de puntuación con el profesor que los chicos adolescentes. Luego se analizó si la frecuencia de errores pre-edición y el nivel de aprendizaje fueron factores influyentes. Con la prueba de ANCOVA se muestra que la frecuencia de errores influye en alrededor del 6% en las variaciones de los valores de AE: $F(1, 71) = 4.219$, $p = 0.044$, $R^2 = 0.056$. También se hicieron cálculos para determinar si la simultaneidad de las variables, nivel de aprendizaje y frecuencia de errores pre-edición (Fep1), influyó en la AE. El resultado mostró que la Fep1 genera variaciones significativas en los puntajes de AE, $F(1, 71) = 4.219$, $p = 0.044$, $R^2 = 0.056$. El valor R^2 muestra que la Fep1 influye en el 4% de las variaciones en los puntajes de AE. Asimismo, se analizaron, los grupos de evaluación (pares y profesor) y la Fep1, se encontró que la influencia de los grupos de evaluación no fue significativa, $p = 0.132$. Sin embargo, la Fep1 mostró lo contrario: $F(1, 71) = 4.833$, $p = 0.031$, $R^2 = 0.064$; es decir, el 6% de las variaciones en las autoevaluaciones fueron causadas por la frecuencia de errores antes de la edición de las composiciones, cuando se considera simultáneamente con los grupos de evaluación.

De igual forma se indagó sobre las variaciones del puntaje del profesor en relación al nivel de aprendizaje de los estudiantes y la Fep1. Se encontró que cuando se consideran estas dos variables de forma simultánea, con un intervalo de confianza del 95%, se puede afirmar que el nivel de aprendizaje influyó en un 12% en la variación del puntaje del profesor, y la Fep1 en un 32%.

Adicionalmente, se hizo el cálculo de la frecuencia de errores post-edición (Fept1) teniendo en cuenta los grupos de evaluación, la retroalimentación (directa, indirecta o mixta), el reconocimiento de errores luego de la retroalimentación correctiva (Re1) y la frecuencia de errores pre-edición (Fep1), de todas estas variables, solo el Re1, $p = 0.044$, y la Fep1, $p < 0.001$, tienen efecto en la frecuencia de errores post-corrección; la primera en un 6%, y la segunda en un 67%. Las variables y sus efectos en la prueba 1, se resumen en la Tabla 4.

Variables y su influencia en los resultados de la prueba 1			
Variables	AE	EP	Fept1
Fep1 + Aula	Fep1 6%	32% y 12%	
Fep1 + Ge1	Fep1 6%		
Fep1 + Ge1 + RC1 + Re1			Fep1 67% y Re1 6%

Tabla 4. Relación de las variables y su influencia en la prueba 1

En el caso de la prueba 2, se efectuó el mismo procedimiento para el manejo de los datos calculados y variables. El histograma de los puntajes de AE mostró un caso extremo que se omitió, por lo cual, se trabajó

con 75 autoevaluaciones. El valor delta 2 mostró variaciones mayores a 24.00 puntos entre los puntajes del profesor y del auto-evaluador, por lo cual, fue necesario suprimir 6 casos, quedando 69. Se inició con la prueba ANOVA entre el valor delta 2 y el nivel de aprendizaje. Los resultados mostraron que la relación entre estos dos factores no fue significativa, $p = 0.404$. En cuanto al género, se encontró evidencia que la variabilidad entre los puntajes de auto-evaluación con los del profesor, fueron causadas por el género por lo menos en un 9%, $p = 0.011$.

En esta segunda fase se tuvieron en cuenta también las variables de la prueba 1, como los grupos de evaluación (Ge1) y el reconocimiento de los errores post-evaluación (Re1). Se observó que la distribución de los grupos de evaluación no influyeron en la AE holística en la prueba 2. Sin embargo, cuando se analizó la interacción entre las variables Ge1 y el nivel de aprendizaje en la AE en esta segunda fase, se encontró que sí fue significativa, $p = 0.008$. Adicionalmente, con el test de MANOVA, considerando los puntajes holísticos dados por los grupos de evaluadores en la prueba 1 y el Re1, se vio evidencia de que el Re1 influyó significativamente ($p = 0.031$) fue significativo; su injerencia en los puntajes de auto-evaluación fue de alrededor del 6%. Luego se consideraron los grupos de evaluación de la prueba 1 con la frecuencia de errores post-edición en la prueba 1 (Fept1). Se determinó que la variable Fept1, en presencia de la variable Ge1 influyó en los puntajes de AE en un 7% de manera significativa, $p = 0.022$.

Cuando se intentó analizar si los factores de nivel de aprendizaje y la frecuencia de errores pre-edición en esta segunda fase (Fep2) fueron variables que influyeron también en el resultado holístico de AE en la prueba 2, con el test de Levene se mostró que no había homogeneidad de varianzas entre estas variables, por lo tanto se realizó la prueba de Kruskal Wallis y se encontró que las diferencias en las frecuencias de error no fueron significativas sobre el nivel de aprendizaje, pero sí sobre el puntaje de auto-evaluación ($p = 0.046$), aunque el porcentaje de influencia no se pudo determinar.

En cuanto a la variación del puntaje del profesor con respecto a la simultaneidad de las variables, aula (nivel de aprendizaje) y a Fep2, mostró que la Fep2 influyó significativamente ($p < 0.001$) en los puntajes dados por el profesor, por lo menos en un 34%. Con respecto a la frecuencia de errores post-edición (Fept2), en relación con los Ge2, la retroalimentación, el reconocimiento de errores (Re2) y la Fep2, con un modelo de regresión, se determinó que tuvieron influencia importante en la Fept2, la Fep2, $p = 0.0001$, en un 8%; la retroalimentación, $p = 0.026$, en un 12% y la variable Ge2, $p = 0.024$, en un 7%. Cuando se observan los efectos de la retroalimentación (RC) de forma discriminada, solo la retroalimentación directa ($p = 0.008$) es significativa, con una injerencia de un 10% aproximadamente sobre la frecuencia de errores post-edición. Las variables y sus efectos en la prueba 2 se resumen en la Tabla 5.

Variables y su influencia en los resultados de la prueba 2			
VARIABLES	AE	EP	Fept2
Ge1	No		
Ge1 + Aula	Ge y A 9%		
Ge1 + Re1	Re1 6%		
Ge1 + Fept1	Fept 7%		
Fep2 + Aula	Fep2 Sí	Fep2 34%	
Ge2 + RC2 + Re2 + Fep2			Ge2 7%, RC 12% RC directa 10% Fep2 80%

Tabla 5. Relación de las variables y su influencia en la prueba 2

En la prueba 3, la prueba de ANOVA no mostró evidencia que el nivel de aprendizaje haya tenido injerencia en las diferencias entre los puntajes de AE y EP, es decir, en delta 3. Con respecto al género, las diferencias tampoco fueron significativas en esta fase del estudio, $p = 0.092$. La prueba ANCOVA para determinar la influencia de las variables de nivel de aprendizaje y grupos de evaluación en la prueba 2 (Ge2), no se pudo aplicar, ya que el test de Levene mostró una diferencia significativa entre las varianzas de ambos factores, $p = 0.042$. Se realizó en su defecto la prueba Kruskal Wallis, y se observó que en ninguno de los casos, las diferencias en las distribuciones de los puntajes de AE se debieron al nivel de aprendizaje o a los grupos de evaluación de la prueba anterior. Cuando se consideraron las variables Ge2 y Re2 simultáneamente, se

encontró que ni Ge2, $p = 0.410$, ni Re2, $p = 0.285$, ejercieron influencia en los puntajes holísticos de AE en los alumnos en la tercera prueba. Se analizaron también los Ge2 y la frecuencia de errores post-edición en la prueba 2 (Fep2), se encontró que estas variables actuaron como co-dependientes a los puntajes de AE en esta fase, en un 12%. Posteriormente, las variables nivel de aprendizaje y la frecuencia de errores pre-edición en la prueba 3 (Fep3), mostraron que interactuando simultáneamente, influyen significativamente en los puntajes de AE, $p = 0.036$, por lo menos en un 7%. El efecto es mayormente causado por la variable Fep3 en un 11%. De hecho, cuando se tuvo en cuenta esta variable solamente, su influencia fue también significativa, $p = 0.017$, con una injerencia en los puntajes de AE de un 8%. Por otro lado, la variabilidad de los puntajes del profesor teniendo en cuenta la simultaneidad del nivel de aprendizaje y la variable Fep3, se determinó que la Fep3 tuvo una injerencia significativa, $p < 0.001$, del 40% en estos puntajes. La variabilidad de la frecuencia de errores post-edición en la prueba 3, comparando las variables grupos de evaluación (Ge3), retroalimentación (RC3), reconocimiento de los errores (Re3) y Fep3, en esta fase, se debió a la retroalimentación, $p = 0.010$ y a la variable Fep3, $p < 0.001$, en un 70%. El tipo de retroalimentación significativa fue la indirecta, $p = 0.003$, en un 12%. Las variables y sus efectos en la prueba 3 se resumen en la Tabla 6.

Variables y su influencia en los resultados de la prueba 3			
Variables	AE	EP	Fep3
Ge2 + Re2	No		
Fep2 + Ge2	12%		
Fep3	8%		
Fep3 + Aula	7%	Fep3 40%	
Ge3+ RC3 + Re3 + Fep3			RC 15%, RC indirecta 12% Fep3 70%

Tabla 6. Relación de las variables y su influencia en la prueba 3

4. CONCLUSIONES

Las respuestas a la pregunta de la investigación acerca de los elementos y variables que caracterizan la evaluación en una secuencia de composiciones escritas por estudiantes norteamericanos de ELE, cuando se desarrolla por auto-evaluación, por pares o por un docente, se encontró que hubo varias variables que fueron estadísticamente significativas en los procesos de evaluación llevados a cabo en el estudio. Con las pruebas realizadas, se determinó que el nivel de aprendizaje y los grupos de evaluación fueron factores intervinientes en las variaciones de los puntajes holísticos de AE. La variable de género, por su parte, mostró que las mujeres presentaron mayores diferencias entre los puntajes de AE y del profesor que los hombres en las dos primeras pruebas. Esto podría indicar que las estudiantes adolescentes, sobreestimaron sus puntuaciones con respecto al puntaje del profesor. Andrade (2009) reportó en su momento que el género influye en el uso de los baremos. Sin embargo, en este estudio, se vio que esta tendencia se superó en la tercera prueba. Esto muestra que con la práctica, los estudiantes en general, aprendieron o aplicaron mejores destrezas el uso e interpretación del baremo hasta lograr evaluar de forma más comparativa entre los grupos, a pesar de las diferencias de género; es decir, la variable de género perdió su influencia en la autoevaluación, mostrando que al término del tratamiento, los adolescentes en general, obtuvieron puntajes de AE similares entre los grupos.

En la segunda fase, las variables que influyeron en el proceso de AE, además del género fueron: la frecuencia de errores post-edición vistas en la prueba 1, el reconocimiento de errores de la prueba precedente, la frecuencia de errores pre-edición de la prueba 2; mientras que las variables grupos de evaluación y el nivel de aprendizaje fueron intervinientes; esto muestra que los procesos de evaluación con el tiempo reciben influencias de forma multifactorial que tienen que ver con los agentes evaluadores y la retroalimentación que ofrecen, la respuesta a los comentarios de los revisores, la previa experiencia de redacción y re-escritura de la composición anterior y la revisión de los errores en la composición actual.

En la tercera prueba, ni el nivel de aprendizaje ni el género fueron influyentes en las diferencias entre los puntajes de AE y EP; el puntaje holístico de AE no recibió influencia del reconocimiento de los errores en la prueba anterior, en tanto que la simultaneidad de las variables: grupos de evaluación 2 y la frecuencia de errores post-edición 2, el nivel de aprendizaje y la frecuencia de errores pre-edición 3; fueron co-dependientes al factor de AE. Estas variables tienen que ver con la habilidad y conocimiento que tiene el estudiante de la lengua para poder procesar la retroalimentación recibida e incorporarla en la reedición de la composición.

En general, la influencia de la variable frecuencia de errores pre-edición en presencia de la variable nivel de aprendizaje fue evidente en las tres fases de escritura, a partir de un 6% en la primera, hasta un 7% en la tercera fase. Esto puede indicar que los estudiantes pudieron dar un mejor uso e interpretación del baremo para valorar y puntuar sus composiciones. Asimismo, el efecto simultáneo de la frecuencia de errores pre-edición y el nivel de aprendizaje fue más notable sobre los puntajes del profesor; su influencia tendió a incrementar en cada fase, 32%, 34% y 40% respectivamente.

En general, se mostró que el nivel de aprendizaje por sí mismo, no ejerce influencia en las diferencias entre los puntajes de AE con los del evaluador; sin embargo, sí influye en el puntaje de AE cuando el alumno realiza el procesamiento del input ofrecido por los revisores y las correcciones de los errores en la edición de las composiciones. Por su parte, los puntajes del profesor tuvieron como variables intervinientes el nivel de aprendizaje de los estudiantes y la frecuencia de errores pre-edición.

La frecuencia de errores post-edición, fue determinada por la frecuencia de errores pre-edición y el reconocimiento de los errores pre-edición en la prueba 1; en tanto que en la prueba 2, la retroalimentación se agregó como variable independiente, específicamente la RC directa, y los grupos de evaluación 2. La RC directa, fue otorgada principalmente por los pares, en ambas aulas. La efectividad de la retroalimentación directa se debe principalmente a que es más sencillo para los estudiantes incorporar la retroalimentación directa en las re-ediciones, que otro tipo de retroalimentación. Esto no implica necesariamente que los estudiantes hayan reconocido los errores; pudo ser que sólo los reemplazaron con las correcciones sugeridas, con ello mejorando la frecuencia de errores post-edición.

Sin embargo, dado que en la prueba 2, el reconocimiento de los errores fue una de las variables influyentes en los puntajes de AE, se puede inferir que la respuesta a la retroalimentación en la primera prueba, impactó en el mejoramiento de los errores, post-edición, para la segunda prueba. Es decir, los estudiantes procesaron la retroalimentación recibida y en consecuencia, lograron la superación de los errores. En la prueba 3, la retroalimentación indirecta y la frecuencia de errores pre-edición, influyeron en la frecuencia de errores post-edición. Sorpresivamente la utilización del cuestionario 3 resultó ser un instrumento también muy importante en la reflexión metacognitiva de los estudiantes; mientras el cuestionario 1 introdujo a los estudiantes en el proceso de autoevaluación, el baremo, por su parte, permitió que los alumnos hicieran el monitoreo de sus habilidades y desempeños en cada prueba; el cuestionario 3 congregó tanto los aspectos evaluativos de las composiciones, con el desempeño y el procesamiento de la retroalimentación para cada prueba. Este instrumento fue importante para los receptores de la retroalimentación y pudo ayudar a mejorar sus destrezas de evaluación posteriores.

Finalmente, cuando se espera que los estudiantes adolescentes aprendan a evaluar la habilidad escritora en ELE en las aulas de secundaria, son importantes la práctica y la continuidad en las tareas evaluativas que permitan que los estudiantes tengan oportunidades de perfeccionar sus destrezas para evaluarse y evaluar a otros. Como docentes, es importante estar conscientes de que las experiencias relativas a la evaluación y al uso de instrumentos para tal fin, se suman y en consecuencia, las influencias de los factores aquí encontrados, sobre la evaluación, pueden permanecer, disminuir o desaparecer a medida que los alumnos se van perfeccionando autoevaluando o evaluando a otros en cada fase de composición y re-edición.

REFERENCIAS BIBLIOGRÁFICAS

- Acevedo, D. (2013). La negociación de baremos para la evaluación colaborativa: un estudio de casos con alumnos norteamericanos de escuela secundaria [en línea]. *Revista Nebrija de Lingüística Aplicada*. Recuperado de: <http://www.nebrija.com/revista-linguistica/la-negociacion-de-baremos-para-la-evaluacion-colaborativa>

- Andrade, H., Wang, X., Du, Y & Akawi, R. (2009). Rubric-reference self-assessment and self-efficacy for writing. *The Journal of Educational Research*, 102 (4), pp. 287-302.
- Berridge, 2009. Peer Interacting and Writing Development in a Social Studies High School Classroom. California: University of California, pp.1-42.
- Boud, D. & Falchikov, K. (2007). Developing assessment for informing judgement. *Rethinking Assessment in Higher Education*. Nueva York: Routledge, pp. 181-197.
- Jacobs, H., Zingraf, S., Wormuth, D., Hartfiel, F. & J. Hughey, (1981). *Testing ESL Composition: A Practical Approach*. Massachusetts: Newbury House Publishers.
- Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26 (1) pp. 75-100.
- Oscarson, D. (2009). Self-assessment of Writing in Learning English as a Foreign Language. A Study at the Upper Secondary Level [en línea]. Recuperado de: <http://gupea.ub.gu.se/handle/2077/19783>
- Serrano, J. M., Pons, R. M., & Ruiz, M. G. (2007). Perspectiva Histórica del Aprendizaje Colaborativo: Un largo y tortuoso camimo através de cuatro siglos. *Revista Española de Pedagogía*, 236, pp. 125-138.
- Tamjid, N. & Birjandi, P. (2012). The role of self-, peer and teacher assessment in promoting Iranian EFL learner's writing performance. *Assessment and Evaluation in Higher Education*, 37 (5), pp. 513-533.

ANEXO

BAREMO MIXTO. A partir de Jacobs et al. (1981) y Oscarson (2009)		
Categorías	Puntaje	Descriptorios
Contenido	20	Excelente: Comunica efectivamente en la lengua objeto, es relevante. Hay descripciones detalladas, claras, coherentes y organizadas.
	15	Bueno: Algunos errores en la comunicación, es en su mayor parte relevante. Hay descripciones con poco detalle. Ideas incompletas, lógicas pero desorganizadas.
	10	Regular: Desarrollo inadecuado del tema, fallos en la comunicación. Ideas confusas, desconectadas, secuenciadas en forma ilógica.
	5	Pobre: No comunica, no es relevante. No se comprenden las ideas, no hay organización; o no hay texto suficiente para evaluar.
Vocabulario	28	Excelente: Selección adecuada y efectiva de palabras y frases. Demuestra conocimiento del vocabulario enseñado; pertinente uso del mismo.
	23	Bueno: Pocos errores en la selección de palabras y frases. Demuestra moderado conocimiento del vocabulario enseñado; con pocas repeticiones de palabras.
	19	Regular: Errores frecuentes en la selección de palabras y frases. Repeticiones frecuentes que demuestran poco conocimiento del vocabulario enseñado.
	15	Pobre: Palabras aisladas o en inglés. Demuestra poco conocimiento del vocabulario enseñado; o no hay texto suficiente para evaluar.
Ortografía	12	Excelente: Ninguno o muy pocos errores en el texto, en el uso de las mayúsculas o acentos ortográficos.
	9	Bueno: Errores ortográficos en menos del 50% del texto, en el uso de las mayúsculas o acentos ortográficos.

	6	Regular: Errores ortográficos en todo el texto, en el uso de las mayúsculas y acentos ortográficos.
	3	Pobre: Escritura ilegible; o no hay texto suficiente para evaluar.
Gramática	20	Excelente: Construcciones adecuadas y efectivas, con errores en una categoría, de correspondencias de género, número, verbo-sujeto, conjugaciones, sintaxis, artículos, pronombres o preposiciones.
	15	Bueno: Construcciones efectivas, con errores en dos categorías, como correspondencias de género, número, verbo-sujeto, conjugaciones, sintaxis, artículos, pronombres o preposiciones.
	11	Regular: Construcciones inadecuadas en todo el texto, con errores en tres categorías, como correspondencias de género, número, verbo-sujeto, conjugaciones, sintaxis, artículos, pronombres o preposiciones.
	7	Pobre: Errores en más de cuatro categorías. Incomprensible; o no hay texto suficiente para evaluar.
Puntuación	20	Excelente: Ninguno o muy pocos errores aislados de puntuación, como comas.
	15	Bueno: Algunos errores de puntuación, como puntos, comas; algunas ideas u oraciones sin pausas, en menos del 50 % del texto.
	11	Regular: Errores múltiples de puntuación y continuación de ideas sin pausas en la oración, en más del 50% del texto.
	7	Pobre: Mala escritura; o no hay texto suficiente para evaluar.