



Número 12
Diciembre de 2019

Diseño y validación de la batería de competencia literaria (BCL)

Xavier Mínguez-López

<https://orcid.org/0000-0001-6361-9263>

Vicente Alfonso Benlliure

<https://orcid.org/0000-0001-6010-0474>

Universidad de Valencia

Recibido: 01-11-2018

Aceptado: 16-12-2019

Pág. 35 a la 49

Resumen:

Palabras clave:

Competencia literaria,
Evaluación Competencia
Literaria, Conceptos literarios,
Procedimientos Literarios,
Actitudes hacia la Literatura.

El objetivo de este trabajo es presentar el proceso de diseño y validación de la Batería de Competencia Literaria (BCL). La Batería está formada por tres escalas: Escala de Conceptos Literarios (ECL), Escala de Procedimientos Literarios (EPL) y Escala de Actitudes hacia la literatura (EAL). La adquisición de la Competencia Literaria (CL) es el principal objetivo de la llamada Educación Literaria. Sin embargo, existen pocos métodos para medir esta competencia y la mayoría presentan problemas prácticos que limitan su estandarización. Un total de 221 adolescentes participaron en el proceso de diseño y validación. Se llevaron a cabo estudios sobre las propiedades de las pruebas incluyendo medidas de fiabilidad, dificultad de los ítems, ajuste, etc. Finalmente, el estudio de la validez de las escalas se llevó a cabo mediante el análisis de las correlaciones entre las puntuaciones de las pruebas y dos medidas externas de CL: la valoración de un ejercicio práctico realizado por los alumnos por parte de jueces expertos y la valoración llevada a cabo por los profesores de la asignatura de Literatura sobre el rendimiento del alumno en dicha materia. Los resultados del estudio muestran que la BCL puede ser un instrumento fiable y válido para la evaluación de la CL en adolescentes.

Introducción

El concepto de competencia literaria (CL) tiene su origen en los estudios literarios. Culler (1976) y Schmidt (1985) se basaron en el concepto de competencia lingüística de Chomsky (1966) para describir el sistema literario, aunque los primeros autores no se inscribían en un marco educativo. Tras la introducción del término “competencia comunicativa” por Dell Hymes (2000) muchos autores (Bachman, 1990; Canale, Michael Swain y Merryll, 1980; Celce-Murcia, Dörnyei y Thurrell, 1995) han matizado las subcompetencias que entran en juego en el total de la competencia comunicativa¹.

El concepto de CL desde una perspectiva educativa fue acuñado por Bierwisch (1970), quien afirmaba que era “the human ability to produce poetic structures and understand their effect”. Van Dijk (1972) prefiere hablar de una habilidad o capacidad de los seres humanos para producir e interpretar textos literarios. En la misma línea, Salvador (1984) y Mendoza (2004) sugieren que el lector modelo de Umberto Eco no sería otra cosa que la postulación de la competencia literaria del receptor ideal.

Coenen (1992, p.73) entiende que un sujeto con competencia literaria es capaz de comunicar con y acerca de la literatura. En una línea similar, Torell (2011) añade que “what motivates competent reading [...] is a constant will to understand our own selves, meeting ‘the other’ in a literary work”.

Pero ¿qué elementos constituyen la competencia literaria? Mendoza (1998; 2008) propone una clasificación exhaustiva que incluye la mayor parte de los elementos imprescindibles para su definición. El citado autor da una especial importancia a los procedimientos del lector en el camino a la descodificación de los textos y enfatiza el componente pragmático, el que ayuda al lector a conectarlo con el resto del universo literario. Así, el lector es quien reconstruye la situación enunciativa y adopta la actitud adecuada al tipo de texto y a su intención, identificando claves y estímulos. Estas estrategias implican conocer las características propias que convierten un texto en literario y los aspectos pragmáticos que ayudan al lector a ajustarse al texto mientras lee. Estos saberes se podrían incluir en tres categorías: los referidos al saber cultural-enciclopédico, los referidos a las modalidades del discurso y los saberes estratégicos. También Reyes-Torres (2015) propone una tríada de elementos basándose en los trabajos de Torell (2011) y Kucer (2001). Según este autor habría una dimensión constitutiva y cognitiva, una dimensión performativa y una tercera sociocultural y estética.

En definitiva, consideramos que la competencia literaria se podría sintetizar en torno a tres grandes bloques (Mínguez-López, 2015):

1. Procedimientos que tienen que ver con la decodificación específica del texto a través de la lectura, y que están también relacionadas con un código concreto como es la literatura. Este bloque también incluye la capacidad del lector competente de activar su *intertexto*, (Mendoza, 2001, p.97).
2. Conocimientos: el conocimiento enciclopédico necesario para entender el texto, el conocimiento de la literatura, el léxico, la pragmática, etc. Los conocimientos desempeñan un papel relevante pues son la materia prima con la que trabajan las operaciones mentales en torno al suceso literario.

¹ Para un resumen de la evolución de las subcompetencias de la competencia comunicativa se puede consultar (Cenoz Iragui, 2004)

3. Actitudes hacia la lectura y la literatura. Incluye la predisposición inicial, pero también interés por nuevos lenguajes y temáticas, una actitud crítica ante la literatura, etc.

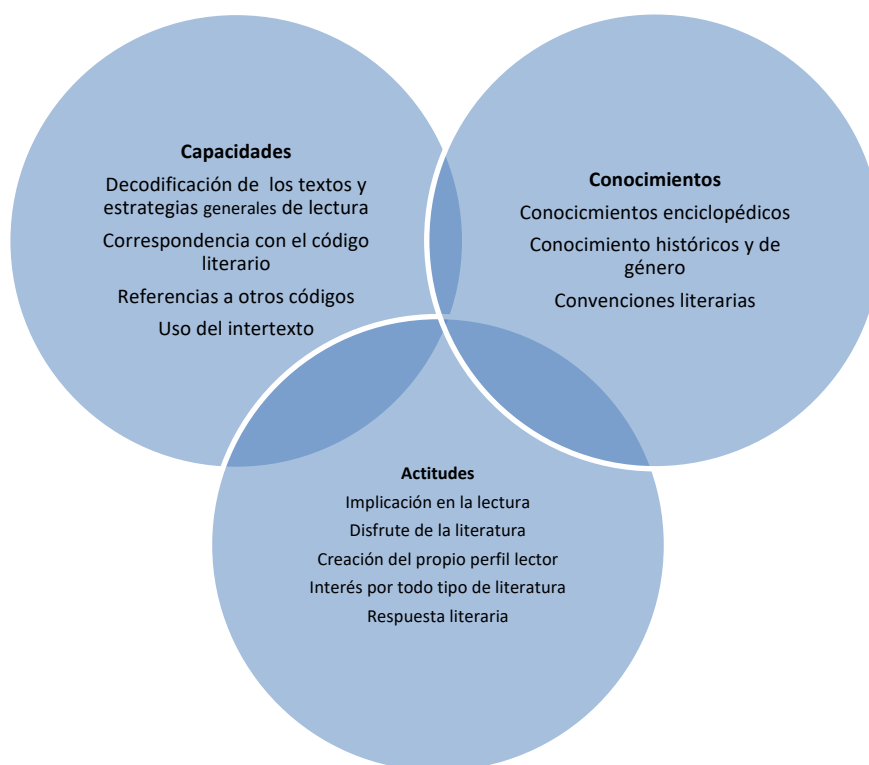


Figura 1. *Los tres grandes bloques de la Competencia Literaria*

Medir la competencia literaria

A pesar de que el concepto de competencia literaria suscita interés, su traducción a datos medibles contrastables ha supuesto no pocos quebraderos de cabeza. De entre los métodos más utilizados destacan el *Foregrounding*, los comentarios de texto y las entrevistas personales. Los cuestionarios se aplican habitualmente para la evaluación de la materia de literatura, pero no tanto para la medición de la competencia literaria.

En el *Foregrounding* (lenguaje en primer plano) se marcan los rasgos lingüísticos que confieren al texto mayor carga semántica, fónica y sintáctica. Estos rasgos atraerían la atención del lector formado al encontrar en ellos las claves para descodificar, pero también disfrutar del texto. Miall (2006) analiza algunos experimentos realizados con este método tanto en poesía como en narrativa. Las evaluaciones basadas en este método arrojan luz respecto del proceso de lectura, aunque no pueden ser utilizados como un indicador de esta, dado que un bajo grado de CL no influye en los resultados. Además, en palabras de Salgado (2015, p. 245): “Literariness cannot be considered a textual feature only but is rather the effect of a multifaceted process integrating textual features and cognitive operations”. Por otro lado, la utilización del *Foregrounding* para medir la competencia literaria ignora uno de los componentes que diversos autores consideran fundamental como es el uso del intertexto (Mendoza, 2001; Witte, Rijlaarsdam y Schram, 2010).

Los comentarios de texto presentan también algunos problemas metodológicos. Aun con las pautas proporcionadas por Hanauer (1996), presentan el inconveniente de su componente subjetivo ligado a la interpretación de algunos rasgos. Por otro lado, la calidad de un comentario tiene mucho que ver con la textualización, con lo cual se está midiendo también de manera directa

la competencia escrita de los sujetos. Finalmente, también existe otro problema de orden práctico pues supone una gran dificultad a la hora de estandarizar la prueba y limitarla en el tiempo.

Las entrevistas personales permiten obtener información a partir de la comunicación directa con el sujeto evaluado, pero requieren la participación de un experto. Si bien esta metodología puede ser rigurosa y exhaustiva, conlleva ciertos problemas de estandarización del procedimiento, coste y subjetividad por parte de los entrevistadores. Es el caso de la investigación de Witte et al (2012) o de Amo y Ruiz (2014).

Paran (2010) considera que existen cinco dilemas para quien se enfrenta a la medición de la CL: la necesidad de evaluar la literatura que implica estandarización y focalización en la evaluación y no en la lectura; la dificultad de diferenciar entre la evaluación lingüística y la literaria; la dificultad de evaluar la propia percepción o aquello que se considera conocimiento público sobre el texto; la necesidad de decidir entre tareas reales o tareas de tipo pedagógico; y la conveniencia de usar metalenguaje. De la resolución de estos cinco dilemas, Paran (2010) deduce algunos principios a la hora de evaluar la literatura: incluir las interpretaciones públicas y la apreciación privada de las obras, usar diversas tareas en la evaluación y diseñar test con diferentes partes, incluir la posibilidad de elegir y ofrecer al alumnado criterios transparentes, facilitar textos para evitar que las pruebas sean memorísticas y minimizar el peso del lenguaje.

Partiendo de estos principios, el objetivo del presente trabajo es construir y validar la Batería de Competencia Literaria (BCL) para la evaluación de este complejo constructo.

Método

Participantes

Un total de 221 adolescentes participaron en alguna de las diferentes fases del presente estudio. El 51.9 % eran chicos, el 48.1% restante chicas con edades comprendidas entre los 12 y los 14 años ($\bar{x}=12.29$). Todos los participantes asisten regularmente a clases de primero de Educación Secundaria Obligatoria (ESO) en tres centros públicos de la provincia de Valencia (España). Los tres centros tienen características muy similares ya que están localizados en ciudades pequeñas del entorno de la ciudad de Valencia, son públicos y ofrecen tanto el Programa Plurilingüe de Educación en Valenciano como el paralelo en castellano. La procedencia socioeconómica es también muy similar según las memorias de los centros.

Los participantes formaron parte de uno de los siguientes dos grupos: un primer grupo de 52 sujetos que respondieron a la versión piloto de la Batería, un segundo grupo de 169 sujetos a los que les fue administrada una segunda versión y un ejercicio de CL.

Procedimiento

Diseño de la BCL: Fase previa y prueba piloto

Para diseñar los instrumentos de conforman la BCL, se partió de la distinción entre los tres campos que hemos considerado imprescindibles para determinar la competencia literaria del alumnado y que agrupan las aportaciones más relevantes expuestas anteriormente: conocimientos, procedimientos y actitudes. Siguiendo esta estructura teórica, la BCL está compuesta por tres escalas: la Escala de Conceptos Literarios (ECL), que evalúa el conocimiento enciclopédico de acuerdo al nivel escolar que se supone por el currículo de lengua y literatura; la Escala de Procedimientos Literarios (EPL) que evalúa la capacidad de descodificar los textos desde una

perspectiva multidimensional; y finalmente, la Escala de Actitudes hacia la literatura (EAL) que evalúa cuál es el posicionamiento respecto del hecho literario.

Cada uno de estos ítems se midió con diferentes pruebas tal como sugiere Paran (2010), incluyendo respuestas cerradas como en el caso del conocimiento de obras, autores, recursos literarios y otros.; escalas Likert como en el caso de la valoración de las actitudes; algunas respuestas abiertas como el recuerdo del último libro leído; y alguna pregunta con una valoración variable dependiendo de la mayor o menor adecuación de la respuesta. En definitiva, las dos primeras escalas son fundamentalmente cuestionarios tipo examen (correcto/incorrecto) mientras que la tercera es un cuestionario tipo Likert con cuatro opciones de respuesta.

Para desarrollar las escalas, se identificaron una serie de constructos teóricos a partir de la revisión de la literatura especializada. Por ejemplo, para la escala de Conceptos Literarios (ECL), se partió de constructos previos como tipos de obras, recursos literarios, conocimiento de obras y autores claves junto con otros conceptos medidos habitualmente en enseñanza secundaria.

Para la evaluación de los conocimientos en procedimientos literarios se partió de ideas previas como la capacidad de regramatización, la construcción del intertexto, la ubicación histórica de los textos, la identificación del género, etc. Los contenidos específicos de las pruebas de conceptos y procedimientos se ajustaron al currículum oficial, según lo trabajado en 6º de primaria. Finalmente, por lo que respecta a la evaluación de las actitudes se trabajó en torno a conceptos previos como el autoconcepto como lector, la respuesta literaria, el interés por la lectura, etc.

Los constructos identificados permitieron elaborar un primer banco de ítems. Los comentarios generales de los investigadores fueron positivos y se incorporaron algunos matices que mejoraron los cuestionarios iniciales. Dicho banco de ítems fue revisado por un grupo de 6 expertos en literatura (4 nacionales y 2 internacionales) para hacer una primera selección. Sus comentarios y matices fueron de gran utilidad para escoger aquellos ítems con mejores cualidades.

Esta primera fase de diseño y construcción desembocó en la versión piloto de la batería compuesta por dos cuestionarios unidimensionales (ECL y EPL) con respuestas tipo examen (acierto/error) y un tercer cuestionario compuesto inicialmente por 5 factores teóricos (autoconcepto como lector, respuesta literaria, interés por la lectura, conexión con la vida real y relaciones personales) sobre actitudes hacia la literatura con respuestas en escala tipo Likert.

La prueba de Conceptos Literarios ECL quedó inicialmente formada por un total de 36 ítems, la EPL por 52 y la ACL por 20. En total, la BCL está compuesta por 108 ítems. La aplicación de esa primera versión piloto a un grupo de 52 sujetos tuvo lugar a través de la plataforma Google Forms. El análisis de los resultados se realizó utilizando los programas de análisis estadísticos Mplus y Ministep.

Los primeros análisis permitieron explorar las propiedades de las pruebas y sus ítems, dando lugar a la eliminación de aquellos que no presentaban garantías psicométricas y depurando las escalas. Para analizar dichas propiedades tomamos como referencia la Teoría de Respuesta al Ítem (TRI) y el Modelo de Rasch. La TRI parte de los supuestos de Dimensionalidad, Independencia Local y la Curva Característica de Ítems. Por su parte, el Modelo de Rasch postula que un enunciado eficiente sólo debe ser contestado correctamente por aquellos alumnos que posean la habilidad requerida por las demandas cognitivas del ítem (Rasch, 1980). Los análisis basados en este modelo permiten identificar como anomalías todos los casos que se apartan de la expectativa del modelo. En este marco teórico, se realizaron pruebas de estructura factorial, consistencia interna, discriminación y dificultad de los ítems y Ajuste (Infit y Outfit).

La experiencia con la prueba piloto permitió explorar las propiedades de las pruebas y sus ítems y eliminar aquellos que no presentaban garantías psicométricas. Fundamentalmente se eliminaron aquellos ítems que atentaban contra la unidimensionalidad de las escalas ECL y EPL y aquellos que no cumplieron los estándares considerados adecuados (Índice de dificultad: entre .20 y .85; Índice de discriminación ≥ 25 ; Correlación biserial puntual ≥ 30). Aplicando estos criterios fueron eliminados un total de 25 ítems (6 ítems de la ECL, 16 de la EPL y 3 de la EAL).

Estos análisis permitieron disponer de una segunda versión más depurada de las tres escalas. La prueba de conceptos quedó formada, en esta segunda versión, por un total de 30 ítems, la de procedimientos por 36 y la de actitudes por 17. Esta versión de la Batería fue administrada a un total de 169 alumnos de primero de ESO con la presencia del profesor/a y de un investigador. Paralelamente, estos alumnos llevaron a cabo un ejercicio complementario a la BCL. Dicho ejercicio consistió en la realización de un relato corto en el que debían incorporar todas o algunas de las siete palabras de carácter polisémico que se le adjuntaban (e.g., sierra, banco, ala, entre otras). Los textos realizados por los sujetos durante este ejercicio fueron evaluados por tres jueces expertos en literatura. Los jueces valoraron de forma independiente el grado de competencia literaria del relato en una escala de 0 a 10. La media de las puntuaciones de los tres jueces se utilizó como criterio externo. Por otro lado, también se les pidió a los profesores de las materias de lengua y literatura de los sujetos participantes en el estudio que aportaran una valoración del nivel de CL de cada alumno en una escala de 0 a 10.

Resultados

Propiedades de la versión final de la Batería

Tras el último pase de la prueba se repitieron los análisis siguiendo los supuestos de la TRI. Los análisis permitieron disponer de una versión más depurada de las tres escalas. En su versión final las tres escalas pasaron a contar con 25, 30, y 15 ítems cada una. La puntuación global de la BCL es de 100 (conceptos y procedimiento 40 %, actitudes el 20 %)². Se espera que el alumnado con una CL extremadamente alta oscile entre puntuaciones de 75 y 100; el alumnado con una CL alta, entre 65 y 74, el alumnado con una CL moderada, en valores entre 50 y 64, el alumnado con una CL baja con valores entre 37.5 y 49 y el alumnado con CL muy baja entre 25 y 37. Por debajo de 25 se consideró que la prueba era nula ya que esta puntuación se considera extremadamente baja, incluso para alumnado con escasa CL.

Escala de Conceptos Literarios (ECL)

Unidimensionalidad y Fiabilidad

Para poner a prueba la dimensionalidad de la prueba de conceptos, se llevó a cabo un análisis factorial exploratorio (EFA) de máxima verosimilitud con rotación oblimin. El Screen Test mostró un claro protagonismo del primer factor con respecto al resto (con un eigenvalue del primer factor hasta cuatro veces superior al del segundo factor). El segundo y resto de factores mostraron valores residuales similares, sugiriendo que los ítems son indicadores de un factor latente común.

²Aunque se puede considerar que algunos ítems pueden pertenecer a más de una escala, dado que nuestro interés residía sobre todo en el valor final de la CL, no creemos que esta circunstancia altere las conclusiones.

Con el fin de poner a prueba la consistencia interna de la prueba procedimos al cálculo del índice de correlación biserial puntual entre cada ítem y la prueba. En la tabla 1 se puede comprobar que los 25 ítems tienen valores aceptables. Solamente el ítem 6 se encuentra en el límite de lo deseable. En general, las puntuaciones indican una buena consistencia interna de la prueba. Paralelamente, se calculó el coeficiente de confiabilidad (.94) avalando este la fiabilidad de esta.

Tabla 1. *Correlaciones ítem-escala BCL (r_{bp})*

ÍTEM	r_{bp}	ÍTEM	r_{bp}
C1	.41	C14	.40
C2	.47	C15	.50
C3	.41	C16	.49
C4	.41	C17	.47
C5	.33	C18	.44
C6	.29	C19	.40
C7	.45	C20	.39
C8	.37	C21	.53
C9	.48	C22	.49
C10	.45	C23	.30
C11	.43	C24	.36
C12	.44	C25	.38
C13	.41		

Discriminación y dificultad de los ítems

En la tabla 2 encontramos los índices de discriminación y dificultad de los ítems de la escala de conceptos. La capacidad discriminativa del ítem aumenta según se aleja de cero, bien sea hacia 1 o -1. La puntuación 1, indica que todos los sujetos con los conocimientos pertinentes aciertan el ítem y los que no los tienen los fallan, dando lugar a una discriminación perfecta. Observando la citada tabla podemos comprobar como ciertos ítems como el 1, 10, 18 Y 19 tienen los índices más elevados mientras que otros ítems como el 4 o el 15 se encuentran al límite de lo deseable en este parámetro.

Por lo que respecta a los índices de dificultad, las puntuaciones más altas hacen referencia a ítems más fáciles, mientras que las más bajas a ítems más difíciles. Los enunciados más fáciles corresponden a los ítems 2, 4, 8 y 15. Los más difíciles corresponden a los ítems 6, 13, 17 y 23.

Tabla 2. *Índices de dificultad y discriminación de la prueba ECL*

ÍTEM	ÍNDICE DISCRIMINACIÓN	ÍNDICE DIFICULTAD	ÍTEM	ÍNDICE DISCRIMINACIÓN	ÍNDICE DIFICULTAD
C1	.93	.43	C14	.36	.28
C2	.40	.85	C15	.27	.73
C3	.29	.29	C16	.28	.41
C4	.27	.78	C17	.40	.26
C5	.55	.61	C18	.86	.57
C6	.66	.22	C19	.85	.54

C7	.33	.29	C20	.37	.26
C8	.78	.77	C21	.29	.48
C9	.31	.31	C22	.38	.28
C10	.93	.37	C23	.48	.22
C11	.35	.28	C24	.60	.62
C12	.93	.68	C25	.83	.41
C13	.41	.20	TOTAL	.52	.46

Medidas de Ajuste (Infit y Outfit)

Por último, se obtuvieron las medidas de ajuste INFIT y OUTFIT. Estas medidas ponen a prueba el principio básico del modelo de Rasch. Los índices infit y outfit son estadísticos basados en los residuos. En las tablas 3 y 4 podemos ver los valores de los índices infit y outfit y comprobar que muestran un buen ajuste, tanto de la escala global (1.00 y 1.04 respectivamente) como de todos los ítems de la escala (el único ítem con una puntuación ligeramente por encima de lo esperado es el ítem 4 con un outfit de la media cuadrática sin estandarizar de 2.16). La tabla 4 incluye también el nivel de dificultad en términos de lógits que confirman lo expuesto en la tabla de índices de dificultad. En definitiva, podemos concluir que las propiedades son aceptables pues los alumnos que poseen los conocimientos sobre conceptos literarios son los que responden acertadamente a los ítems.

Tabla 3. *Medidas de ajuste para la Media y la Desviación Típica*

ECL	INPUT		MEASURED		INFIT		OUTFIT	
	Total	Contados	Medida	Sep real	IMSQ	ZSTD	IMNQ	ZSTD
X	36.4	25	0	.30	1.00	-.1	1.06	.1
S.D	15.9	0	1.27	.05	.14	1.0	.37	1.1

Índice de Separación: 4.04; Índice de confiabilidad: .94

Tabla 4. *Valores de Infit y Outfit de la ECL*

Ítem	Dificultad en logits	Error estándar	INFIT		OUTFIT	
			MNSQ	ZSTD	MNSQ	ZSTD
C13	2.13	.35	1.09	.5	1.16	.5
C6	1.68	.32	.71	-1.8	.51	-1.2
C23	1.49	.30	.95	-.3	.86	-.2
C17	1.23	.30	1.08	.6	1.36	1.0
C20	1.23	.29	.99	.0	.97	.0
C11	1.15	.29	.97	-.2	.92	-.2
C14	1.07	.28	.79	-1.0	.79	-.6
C22	.68	.27	.99	-1.9	.68	-1.4
C3	.61	.37	.99	.0	.86	-.5
C7	.54	.27	.99	-.1	.92	-.3
C9	.54	.27	.99	.0	1.14	.7
C10	.12	.26	.98	-.2	.92	-.4
C16	.12	.26	.98	-.1	.93	-.3

C25	-.02	.26	.88	-1.1	.78	-1.2
C1	-.02	.26	1.03	.3	1.00	.1
C21	-.29	.26	.87	-1.3	.79	-1.2
C19	-.43	.27	.97	-.3	1.05	.3
C18	-.72	.27	1.24	2.0	1.40	1.9
C5	-.80	.27	1.02	.2	1.12	.6
C24	-.87	.28	1.26	2.0	1.71	2.1
C12	-1.03	.28	.85	-1.2	.76	-1.0
C15	-1.45	.30	.89	-.7	.71	-1.0
C8	-2.32	.37	1.23	1.0	1.36	.9
C4	-2.46	.38	1.08	.4	1.65	1.3
C2	-2.46	.38	1.22	.9	2.16	1.9

INFIT: Ajuste cercano; OUTFIT: Ajuste lejano; MNSQ: Media cuadrática sin estandarizar; ZSTD: Media cuadrática estandarizada

Escala de Procedimientos Literarios (EPL)

Unidimensionalidad y Fiabilidad

El EFA rotación oblimin confirmó la unidimensionalidad de la prueba. El factor explica hasta el 41 % de la varianza. La tabla 5 expone los índices de correlación de cada ítem con la escala. Todos ellos muestran valores adecuados. Los índices más elevados los muestran los ítems 1, 5, 6, 28 y 30. Los ítems con puntuaciones más discretas son el 20 y el 22. Por su parte, el índice de confiabilidad del EPL es de .94

Tabla 5. *Correlaciones ítem-escala TPL (r_{bp})*

Items	r_{bp}	Items	r_{bp}
P1	.58	P16	.32
P2	.37	P17	.27
P3	.34	P18	.41
P4	.29	P19	.42
P5	.54	P20	.41
P6	.42	P21	.32
P7	.30	P22	.24
P8	.31	P23	.30
P9	.28	P24	.20
P10	.26	P25	.34
P11	.30	P26	.40
P12	.27	P27	.38
P13	.45	P28	.49
P14	.31	P29	.40
P15	.38	P30	.46

Discriminación y dificultad de los ítems

En la tabla 6 encontramos los índices de discriminación y dificultad de los ítems de la escala de procedimientos. La prueba tiene adecuados índices de discriminación con un par de ítems en el límite de lo deseable (12 y 21). Por su parte, el índice de dificultad de toda la prueba PCL es .55. Los ítems de mayor dificultad son el 9, 11, 22 y 4. Los ítems de menor dificultad son 15, 3, 20 y 26.

Tabla 6. *Índices de dificultad y discriminación de la prueba EPL*

ÍTEM	ÍNDICE DISCRIMINACIÓN	ÍNDICE DIFICULTAD	ÍTEM	ÍNDICE DISCRIMINACIÓN	ÍNDICE DIFICULTAD
P1	.63	.79	P16	.71	.37
P2	.78	.44	P17	.39	.47
P3	.41	.86	P18	.81	.58
P4	.31	.31	P19	.35	.82
P5	.44	.73	P20	.41	.85
P6	.30	.35	P21	.26	.45
P7	.43	.44	P22	.48	.30
P8	.56	.43	P23	.30	.45
P9	.42	.29	P24	.30	.39
P10	.41	.77	P25	.92	.70
P11	.37	.30	P26	.75	.82
P12	.25	.49	P27	.46	.69
P13	.93	.63	P28	.26	.58
P14	.27	.57	P29	.61	.41
P15	.39	.86	P30	.80	.49
			TOTAL	.49	.55

Medidas de Ajuste (Infit y Outfit)

Por lo que respecta a las medidas de ajuste cercano y lejano (Infit y Outfit) de la EPL podemos observar en la tabla 7 que dichos índices indican un buen ajuste. Las estadísticas de media cuadrática muestran la distorsión del sistema de medición utilizado. Se esperan valores en torno a 1. Los valores sustancialmente inferiores son indicativos de que las observaciones son demasiado predecibles y los valores sustancialmente por encima indican impredecibilidad (González, 2008). El criterio para considerar buenos índices de ajuste se encuentra entre 1 y 1.30 para la media cuadrática sin estandarizar y entre -2 y .2 para la estandarizada. Podemos observar en la citada tabla que los ítems que en este sentido requieren de futuros análisis son el 11, 19 y 28.

Tabla 7. *Medidas de ajuste para la Media y la Desviación Típica*

EPL	INPUT		MEASURED		INFIT		OUTFIT	
	Total	Contados	Medida	Sep Real	IMSQ	ZSTD	IMNQ	ZSTD
X	43.0	30	0	.29	.99	.0	1.02	.1
S.D	16.6	0	1.21	.04	.11	.9	.23	1.1

Índice de Separación: 3.99; Índice de confiabilidad: .94

Tabla 8. Valores de Infit y Outfit de la ECL

Ítem	Dificultad en logits	Error stándar	INFIT		OUTFIT	
			MNSQ	ZSTD	MNSQ	ZSTD
P9	2.40	.34	.99	.0	1.12	.5
P11	1.74	.28	1.12	.8	1.40	1.7
P22	1.44	.27	1.12	1.1	1.31	1.7
P4	1.29	.26	1.03	.3	1.02	.2
P6	1.09	.26	1.15	1.5	1.32	2.1
P16	.98	.26	.97	-.2	.96	-.2
P24	.96	.25	.89	-1.2	.86	-1.1
P29	.84	.25	1.13	1.5	1.23	1.8
P8	.65	.25	1.01	.1	.99	-.1
P2	.59	.25	.98	-.2	.97	-.3
P7	.53	.25	1.07	.9	1.09	.9
P21	.47	.25	.99	-.1	1.01	.1
P23	.41	.25	1.00	.1	1.02	.2
P17	.22	.25	1.01	.2	1.02	.3
P12	.22	.25	.87	-1.6	.88	-1.2
P30	.10	.25	.97	-.3	.96	-.3
P14	.05	.25	.89	-1.5	.87	-1.3
P18	.04	.25	1.12	1.3	1.11	.9
P28	-.25	.25	.95	-2.1	.76	-2.1
P13	-.44	.26	.94	-.6	.94	-.4
P27	-.96	.29	.94	-1.7	.63	-1.8
P25	-.96	.29	.87	-.8	.75	-1.1
P5	-1.23	.31	1.07	.4	.99	.1
P10	-1.33	.32	.77	-1.2	.61	-1.5
P1	-1.46	.32	1.00	.0	.91	-.2
P19	-1.55	.34	1.08	.4	1.54	1.5
P26	-1.67	.35	.96	-.1	1.07	.3
P20	-1.67	.35	.92	-.8	.99	-1.3
P3	-1.67	.35	.99	-.0	.95	.0
P15	-1.95	.39	.96	-.1	1.02	.2

INFIT: Ajuste cercano; OUTFIT: Ajuste lejano; MNSQ: Media cuadrática sin estandarizar; ZSTD: Media cuadrática estandarizada

Escala de Actitudes Hacia la Literatura (EAL)

A diferencia de las anteriores, en el EAL no hay respuestas correctas o incorrectas, sino que se trata de un cuestionario en el que los alumnos responden en una escala tipo Likert de 1 a 5. También, a diferencia de las anteriores, no se trata de una escala unidimensional, sino que incluye diversos factores. Para explorar esa estructura factorial se llevó a cabo un EFA de máxima verosimilitud con rotación Oblimin, pues se asume que los factores están relacionados. Dicha rotación permitió la depuración final de la prueba con una solución final que explica un 49 % de la varianza y que redujo la estructura original de cinco a tres factores. Los tres factores del EAL son: 1. Lectura, vida real y relaciones personales 2. Interés por la lectura y autopercepción como

lector y 3. Lectura como fuente de experiencias y creatividad.

Para analizar la fiabilidad de la escala se calculó el Alpha de Cronbach de los factores y de la escala global. En las tablas 9 y 10 podemos observar los índices por factores e ítems. La puntuación total oscila entre 15 y 75 y su consistencia interna puede considerarse adecuada con un Alpha de Cronbach de .81.

Tabla 9. *Consistencia interna de la escala EAL*

	F1	F2	F3	Total
Número de ítems	6	4	5	15
Varianza explicada	29.6 %	11.2 %	7.5 %	48.3 %
Alpha de Cronbach	.78	.74	.70	.81

Tabla 10. *Índices de consistencia interna y fiabilidad ítem a ítem de la escala EAL*

Ítem	Media	D.T.	Correl. Ítem total	Alpha Cronbach excl. ítem	Ítem	Media	D.T.	Correl. Ítem total	Alpha Cronbach excl. ítem
A1	3.70	1.110	.435	.71	A9	4.51	.69	.422	.72
A2	3.68	1.258	.491	.70	A10	3.51	1.167	.444	.71
A3	3.46	1.445	.439	.71	A11	3.55	1.28	.395	.79
A4	4.52	.820	.336	.72	A12	3.02	1.26	.316	.82
A5	3.18	1.282	.483	.70	A13	4.51	.69	.422	.72
A6	3.93	1.264	.528	.70	A14	3.51	1.167	.444	.71
A7	2.59	1.153	.411	.71	A15	3.01	1.275	.323	.72
A8	3.66	1.263	.509	.70					

Validez de la Bateria

La validez de criterio se analizó mediante el estudio de las correlaciones de Pearson entre las puntuaciones de los cuestionarios y dos medidas utilizadas como criterio: la CL valorada de forma independiente por jueces expertos a partir de un ejercicio práctico y la CL valorada por los profesores de la asignatura de Literatura en el correspondiente Centro Educativo a partir de su rendimiento regular en el aula.

Para analizar el grado de acuerdo entre los tres jueces evaluadores del criterio externo de CL se calculó el Coeficiente de Correlación Intraclase (CCI). Este índice para medidas promedio fue de .95, valor equiparable al alfa de Cronbach, resultado que avala la fiabilidad de la medida.

Tabla 11. *Correlaciones entre las valoraciones de los jueces, los profesores y el BCL*

	ECL	EPL	EAL	BCL
Jueces (media)	.25(**)	.15(*)	.42(**)	.32(**)
Profesores	.76(**)	.83(**)	.31(**)	.93 (**)

* p <.05, ** p<.01

Como se observa en la Tabla 11, las tres escalas muestran correlaciones significativas en la dirección esperada con cada una de las variables de criterio. Tanto las puntuaciones globales como las específicas de cada escala muestran relaciones directas y significativas tanto con la valoración subjetiva del CL de cada alumno llevada a cabo por los profesores en base a su conocimiento por la interacción diaria con el alumno en clase como en las valoraciones hechas por los jueces del ejercicio práctico sobre CL. Destaca la alta correlación entre la valoración de los profesores y la puntuación global en la BCL. En dicha valoración, cobran protagonismo los procedimientos literarios, con la correlación más alta. Las correlaciones con las valoraciones de los jueces también son significativas, aunque no tan elevadas. Precisamente, la correlación más baja la encontramos entre la escala de procedimientos y la valoración de los jueces, quizás porque en el ejercicio práctico (evaluación de un producto del adolescente) es más fácil apresarse los conocimientos y las actitudes que el proceso para llegar a ese producto final.

Discusión

Si bien los sistemas de evaluación en el aula deben diferir de la prueba que aquí se presenta, no se puede negar que, como sugieren Paran y Sercu (2010), la literatura se antoja a multitud de docentes como una materia imposible de ser evaluada. Esta circunstancia ha llevado a sistemas de evaluación erráticos centrados tan solo en los contenidos (Badia y Cassany, 1994). Nuestra intención con la BCL es, no tan solo ofrecer una herramienta de medición de la CL, sino también sugerir qué aspectos se deben medir de ella. Si la CL es una competencia poliédrica, su evaluación también lo debería ser.

En este trabajo se han presentado las propiedades psicométricas de esta prueba, que en líneas generales y con algunos aspectos todavía por pulir, puede considerarse una opción plausible a la hora de abordar la evaluación de este complejo constructo. Creemos, como apunta Miall (2006), que en ocasiones la consideración de la literatura como una materia indefinible ha impedido que se lleven a cabo estudios de tipo empírico como este, en el que se trata de objetivar conceptos que a menudo se han dejado en una calculada ambigüedad. Si bien la literatura contiene un alto grado de subjetividad, esto no significa que no se pueda estudiar desde parámetros científicos.

Desde la educación literaria, consideramos que contar con un instrumento de evaluación de estas características abre la puerta a multitud de investigaciones que certifiquen la validez de las propuestas didácticas de educación literaria. La posibilidad de medir previa y posteriormente el nivel de CL del alumnado con una herramienta estandarizada, validada y fiable que arroja un valor numérico final (aunque con la posibilidad de segmentarlo en tres valores de acuerdo con las tres escalas utilizadas), resulta un avance considerable en los estudios de didáctica de la literatura pero también en los estudios empíricos sobre literatura.

La BCL no tiene un valor absoluto sino relativo ya que va ligado a elementos subjetivos y ambivalentes y también a circunstancias concretas difíciles de prever. Por poner un ejemplo, uno de los centros estaba situado en la localidad natal de un autor incluido en la prueba, por lo que el porcentaje de alumnado que lo conocía era mucho mayor que en los otros dos centros. Sin embargo, el hecho de contar con un gran número de ítems permite minimizar los efectos de estas circunstancias.

También existen limitaciones. La primera es la propia definición de CL que, como hemos visto, no ha sido consensuada de manera consistente por la literatura especializada de manera internacional. Aunque desde nuestro punto de vista, las aportaciones de Mendoza (2004) y de

Witte, Janssen y Rijlaarsdam (2006) son suficientemente sólidas como para respaldar la BCL. Por otro lado, existe una cierta dificultad en la traducción de la BCL a otras lenguas ya que es necesario encontrar textos de las mismas características que los utilizados en la batería para poder realizar comparaciones. Si bien es posible encontrar textos formalmente similares, resulta más complicado determinar su lugar en el canon de cada literatura o su popularidad entre el alumnado de cada nivel. Creemos que estas desviaciones no son excesivamente relevantes tanto por la cantidad de ítems que se utiliza en la BCL como por lo relativo de estos datos. Por otro lado, las diferencias de resultado entre los grupos en valenciano y castellano no son significativas, cosa que indica que la correspondencia entre las lenguas era correcta. Una última limitación es, al ser la competencia literaria un proceso en desarrollo, la prueba solo es válida en su forma actual para alumnado de 1º de ESO.

Como líneas de futuro, creemos que la más importante es adaptar la batería a los diferentes niveles educativos tanto de secundaria como primaria e incluso infantil. También creemos que debe ser una línea de investigación futura encontrar y/o validar otros sistemas de medición de la competencia literaria como el mismo *foregrounding*, cuestionarios más breves, redacción de textos, etc. Con la BCL como referente, puede ser más sencillo determinar el alcance de estos otros sistemas de medición.

En conclusión, consideramos que la BCL cumple con los objetivos para los que fue diseñado y resulta un instrumento fiable y válido para medir la CL en la población adolescente. Es por esto por lo que se propone como una herramienta de gran utilidad tanto para los investigadores interesados en realizar estudios empíricos sobre la enseñanza-aprendizaje de la literatura, como para docentes que quieran establecer un tipo de medida objetiva entre su alumnado y detectar así posibles carencias en sus métodos didácticos

Referencias

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Badia, J. y Cassany, D. (1994), La classe de literatura, avui. *Articles*, 1, 7-14.
- Bierwisch, M. (1970). Poetics and linguistics. In D. C. Freeman (Ed.), *Linguistics and literary style*. New York: Holt, Rinehart and Winston.
- Canale, M. y Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Celce-Murcia, M., Dörnyei, Z. y Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6(2), 5-35.
- Cenoz Iragui, J. (2004). El concepto de competencia comunicativa. In J. Sánchez Lobato, e I. Santos Gargallo (Eds.), *Vademécum para la formación de profesores: Enseñar español como segunda lengua (L2) lengua extranjera (LE)*, 449-465. Madrid: Sociedad General Española de Librería.
- Chomsky, N. (1966). *Topics in the theory of generative grammar*. The Hague: Mouton.
- Culler, J., Eco, U., Gandy, R., Robey, D. y Varela, P. (1976). *Introducción al estructuralismo*. Madrid: Alianza.
- de Amo, J.M. y Ruiz, M^a. M. (2014) Niveles de adquisición de la competencia literaria en estudiantes de secundaria: análisis de resultados. En Osvaldo Cleger y Jose M. de Amo, *La educación literaria y la e-literatura desde la minificción*. ICE: Barcelona.

- González, M. J. (2008). *El análisis de reactivos con el Modelo Rasch*. Manual Técnico A. México D.F. INEE.
- Hymes, D. (2000). Acerca de la competencia comunicativa. In M. Llobera (Ed.), *Competencia comunicativa. documentos básicos en la enseñanza de lenguas extranjeras* (pp. 13-37). Madrid: Edelsa.
- Kucer, S. B. (2001). *Dimensions of literacy :A conceptual base for teaching reading and writing in school settings*. Mahwah NJ: Lawrence Erlbaum Associates.
- Mendoza, A. (1998). *Conceptos clave en didáctica de la lengua y la literatura*. Barcelona: Universitat de Barcelona.
- Mendoza, A. (2008). Leer y comparar: Notas sobre las posibilidades del comparativismo en el aula de educación secundaria. *Lenguaje y textos*, 28, 19-41.
- Mendoza, A. (2001). *El intertexto lector*. Cuenca: EUCL.
- Mendoza, A. (2004). *La educación literaria: Bases para la formación de la competencia lecto-literaria*. Archidona Málaga: Aljibe.
- Miall, D. S. (2006). Empirical approaches to studying literary readers: The state of the discipline. *Book History*, 9, 291-311.
- Mínguez-López, X. (2015). El papel de la animación en el desarrollo de la competencia literaria. *Elos. Revista de Literatura Infantil e Xuvenil*, 2, 157-172. doi:<http://dx.doi.org/10.15304/elos.2.2670>
- Paran, A. y Sercu, L. (2010). *Testing the untestable in foreign language education*. Buffalo, NY: Multilingual Matters.
- Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: Univ Chicago.
- Salvador, V. (1984). *El gest poètic :Cap a una teoria del poema*. València: Universitat de València, IFV & ICRTV.
- Torell, O. (2011). Literary competence beyond conventions. *Scandinavian Journal of Educational Research*, 45(4), 369-379.
- van Dijk, T. A. (1972). *Some aspects of text grammars : A study in theoretical linguistics and poetics*. The Hague: Mouton.
- Witte, T., Janssen, T. y Rijlaarsdam, G. (2006). *Literary competence and the literature curriculum*. Unpublished manuscript.
- Witte, T., Rijlaarsdam, G. y Schram, D. (2012). An empirically grounded theory of literary development. Teachers' pedagogical content knowledge on literary development in upper secondary education. *LI-Educational Studies in Language and Literature*, 12, 1-33.



Number 12
2019 December

Design and validation of the literary competence battery (LBC)

Xavier Mínguez-López

<https://orcid.org/0000-0001-6361-9263>

Vicente Alfonso Benlliure

<https://orcid.org/0000-0001-6010-0474>

Universidad de Valencia

Received: 01-11-2018

Accepted: 16-12-2019

Pag. 50 to 64

Abstract:

Keywords

Literary competence, Literary Competence Assessment, Literary concepts, Literary Procedures, Attitudes towards Literature,

The objective of this work is to present the process of design and validation of the Literary Competence Battery (BCL). The Battery consists of three scales: Literary Concepts Scale (ECL), Literary Procedures Scale (EPL) and Attitudes towards Literature Scale (EAL). The acquisition of Literary Competence (CL) is the main objective of the so-called Literary Education. However, there are few methods to measure this competence and most of them present practical problems that limit their standardization. A total of 221 adolescents participated in the design and validation process. Studies were conducted on the properties of the tests including measures of reliability, difficulty of the items, adjustment, etc. Finally, the study of the validity of the scales was carried out by analyzing the correlations between the test scores and two external CL measures: the assessment of a practical exercise performed by the students by expert judges and the assessment carried out by the professors of the Literature subject on the student's performance in said subject. The results of the study show that BCL can be a reliable and valid instrument for the evaluation of CL in adolescents.

Introduction

The concept of literary competence (LC) has its origin in literary studies. Both Culler (1976) and Schmidt (1985) relied on Chomsky's concept of linguistic competence (1966) to describe the literary system, although they did not sign up to an educational framework. Since the introduction of the term "communicative competence" by Dell Hymes (2000), many authors (Bachman, 1990; Canale & Swain, 1980; Celce-Murcia, Dörnyei, & Thurrell, 1995) have sought to determine the sub-skills that come into play in total communicative competence³.

The concept of LC from an education perspective was coined by Bierwisch (1970), who claimed that it was "the human ability to produce poetic structures and understand their effect". Van Dijk (1972) preferred to talk of an ability of human beings to produce and interpret literary texts. Along the same lines, Salvador (1984) and Mendoza (2004) suggested that Umberto Eco's model reader is nothing other than the nomination of the literary competence of the ideal recipient.

Coenen (1992, p. 73) understood an individual with literary competence as being someone able to communicate with and about literature. In a similar vein, Torell (2011) added that "what motivates competent reading [...] is a constant will to understand our own selves, meeting 'the other' in a literary work".

But what elements constitute literary competence? Mendoza (1998, p. 2008) proposed an exhaustive classification that includes most of the elements essential to its definition. He gave special importance to the reader's procedures for decoding a text and emphasised the pragmatic component that helps the reader connect it with the rest of the literary universe. Thus, the reader is one who reconstructs the enunciative situation and adopts the attitude appropriate to the type of text and its intention, identifying keys and stimuli. These strategies involve knowing the characteristics that make a text literary and the pragmatic aspects that help the reader to adjust to the text while reading. This knowledge can be broken down into three categories: cultural-encyclopaedic knowledge; knowledge of the modalities of discourse; and strategic knowledge. Based on the works of Torell (2011) and Kucer (2001), Reyes-Torres (2015) also proposed a triad of elements: a constitutive and cognitive dimension; a performative dimension; and a third, sociocultural and aesthetic, dimension.

In short, we believe that literary competence can be synthesized around three large blocks (Mínguez-López, 2015):

1. Procedures: which have to do with the specific decoding of a text through reading and which are also related to a specific code such as literature. This block also includes the ability of the competent reader to activate his or her intertext (Mendoza, 2001, p. 97);
2. Knowledge: the encyclopaedic knowledge necessary to understand a text, e.g., knowledge of literature, vocabulary, pragmatics, etc. Knowledge plays a pertinent role because it is the raw material with which mental operations work around the literary event.
3. Attitudes to reading and literature: this includes initial predisposition but also interest in new languages and themes, a critical attitude to literature, and so on.

³ For a summary of the evolution of the sub-skills of communicative competence, see Iragui (2004).

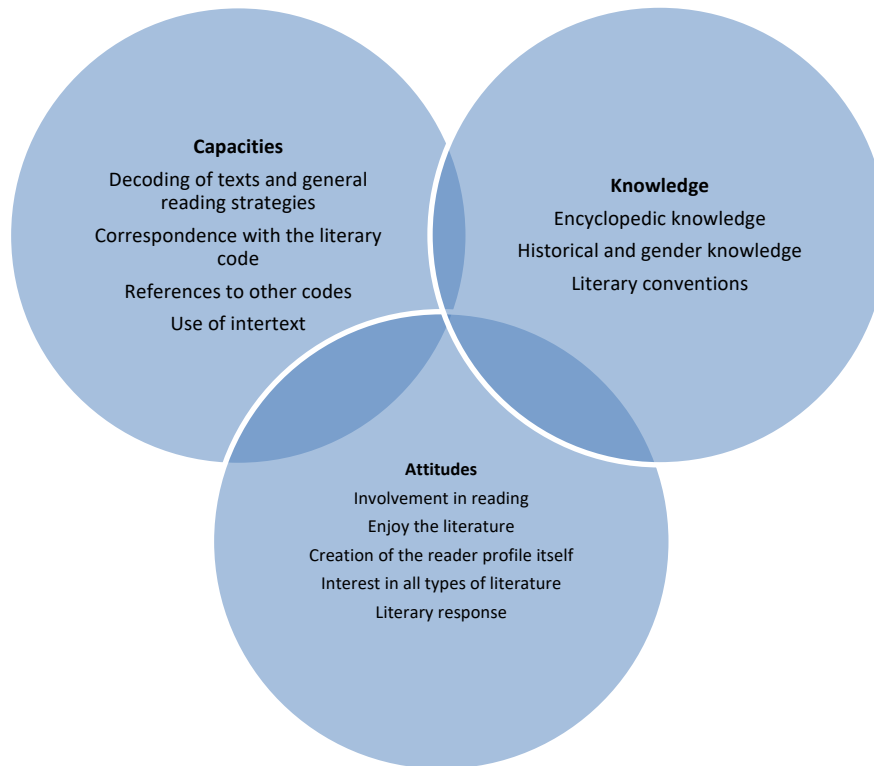


Figure 1. *The three big blocks of the Literary Competition*

Measuring literary competence

Although the concept of literary competence is of interest, its translation into verifiable, measurable data has caused more than a few headaches. Among the most used methods, foregrounding, text commentaries, and personal interviews stand out; questionnaires are usually applied to the evaluation of the subject of literature and less so to the measurement of literary competence.

In foregrounding (foreground language), the linguistic features that give a text a greater semantic, phonic, and syntactic load are marked. These features attract the attention of the reader, through which he or she finds the key to decode, as well as enjoy, the text. Miall (2006) analysed a number of studies using this method on both poetry and narrative texts. He found that although evaluations based on this method shed light on the reading process, they could be used as an indicator of the same since a low level of LC did not influence the results. In addition, in the words of Salgado (2015, p. 245): “Literariness cannot be considered a textual feature only but is rather the effect of a multifaceted process integrating textual features and cognitive operations”. Moreover, the use of foregrounding to measure literary competence ignores one of the components that various authors consider fundamental: the use of intertext (Mendoza, 2001; Witte, Rijlaarsdam, & Schram, 2010).

Text comments also present some methodological problems. Even with the guidelines provided by Hanauer (1996), it presents the disadvantage of its subjective component linked to the interpretation of some features. On the other hand, the quality of a comment has a lot to do with textualization, which is also directly measuring the written competence of the subjects. Finally, there is also another practical problem because it is a great difficulty in standardizing the test and limiting it in time.

Personal interviews can obtain information from direct communication with the evaluated subject, but they have the participation of an expert. interviewers. This is the case of the investigation by Witte et al (2012) or by Amo and Ruiz (2014).

Paran (2010) argued that there are five dilemmas for those grappling with LC measurement: 1) the need to standardize literature evaluation and the focus on evaluation rather than reading; 2) the difficulty of differentiating between linguistic and literary evaluation; 3) the difficulty of extrapolating individuals' perceptions versus what is considered public knowledge about the text; 4) the need to decide between real tasks and pedagogical tasks; and 5) the convenience of using meta-language. From the resolution of these five dilemmas, Paran (2010) deduced a number of principles for evaluating literature: include public interpretation and private appreciation of works; use various tasks to perform the evaluation and design tests that have various elements; include the option of offering students transparent criteria; use texts that prevent tests from being memorised; and, minimise the density of the language.

Based on these principles, the objective of this study was to build and validate a Literary Competence Battery (LBC) for evaluating what is a complex construction. The guidelines provided by Hanauer (1996) highlighted the disadvantage of subjective components linked to the interpretation of certain features. Furthermore, the quality of a commentary has a lot to do with textualization, which also directly measures the writing competence of subjects. Finally, there is another problem of a practical nature: it is very difficult to standardize a test and assign it an appropriate time limit.

Personal interviews allow information to be obtained through direct communication with the subject being evaluated. However, such interviews require the participation of an expert. Although this methodology can be rigorous and exhaustive, it entails certain other problems, including that of standardizing the procedure and the cost and subjectivity of the interviewers. This was the case in the study by Witte et al (2012) and that by Amo and Ruiz (2014).

Methodology

Participants

A total of 221 adolescents participated across the various phases of the study. Over half (51.9 %) were boys and all were between the ages of 12 and 14 years (=12.29). All participants were regular attendants of classes in their first year of compulsory secondary education (ESO) in three public centres in the province of Valencia (Spain). The three centres had very similar characteristics: all were located in small cities around the city of Valencia, all were public institutions, and all ran the Multilingual Education Programme in Valenciano and a parallel programme in Spanish. The socioeconomic origin of participants was also very similar, according to the planning of the centres.

The participants were allocated to one the following two groups: a first group of 52 subjects, who were given the pilot version of the test battery; and a second group of 169 subjects, to whom a second version and a LC exercise was administered.

Procedure

LBC design: Preliminary phase and pilot test

The design of the LBC instruments was based on the three distinct fields that we consider essential for determining the literary competence of students, as outlined above: knowledge, procedures, and attitudes. Thus, in accordance with this theoretical structure the battery was composed of three scales: the Literary Concepts Scale (ECL), which assessed encyclopaedic knowledge at the appropriate level of the language and literature curriculum; the Literary Procedures Scale (EPL), which assessed students' ability to decode texts from a multidimensional perspective; and finally, the Attitudes to Literature Scale (EAL), which evaluated what the position of the reader is regarding the literary fact.

Each of these three elements was measured in various ways, as suggested by Paran (2010), including the following: closed-response measures, for testing participants' knowledge of works, authors, literary resources, and so on; Likert scales, to assess their attitudes; some open-response questions about, for example, the last book that participants had read; and various follow-up questions where necessary to clarify participants' initial answers. In short, the first two scales were essentially exam-type questionnaires (correct/incorrect), while the third was a Likert-type questionnaire with four response options.

To develop the scales, a series of theoretical constructs was identified from a review of the scientific literature. For example, the scale of Literary Concepts (ECL) was based on previously identified constructs, such as types of works, literary resources, knowledge of works, and key authors, along with other concepts usually measured in secondary education.

To evaluate knowledge and literary procedures, we began with previously identified relevant concepts, such as the capacity for re-pragmatisation, the construction of intertext, the historical location of texts, the identification of gender, etc. The specific contents of the tests were then adjusted to suit the official curriculum for sixth-grade students (Spanish school). To evaluate attitudes, we constructed our measures around existing concepts such as self-concept as a reader, response to literature, interest in reading, and so on.

The constructs identified allowed us to develop a first bank of items. General comments on these by the researchers were positive, although some small adjustments were made to improve the initial questionnaires. This item bank was then reviewed by a group of six literature experts (four national and two international) to inform our first selection. Their comments and suggested changes were very useful in helping us to identify the items with the best qualities.

This first phase of design and construction resulted in a pilot version of the battery composed of the following: two one-dimensional questionnaires (ECL and EPL) with exam-type answers (success/error); and a third Likert-scale questionnaire on attitudes to literature, initially composed of five theoretical factors (self-concept as reader, literary response, interest in reading, connection with real life, and personal relationships).

The ECL test was initially formed of 36 items, the EPL of 52, and the EAL of 20. Thus, in total the LBC was composed of 108 items. This pilot version was administered to a group of 52 subjects through the Google Forms platform. Analysis of the results was performed using the statistical analysis programmes Mplus and Ministep.

The first analyses allowed us to explore the properties of the tests and their items, leading to the elimination of those that did not show psychometric promise and the debugging of the

scales. For these analyses, we took as our reference points the Theory of Response to the Item (TRI) and the Rasch Model. The TRI starts from assumptions of dimensionality, local independence, and the characteristic curve of items. For its part, the Rasch Model postulates that an efficient statement will be answered correctly only by students who possess the skill required by the cognitive demands of the item (Rasch, 1980). Analyses based on this model allow all cases that depart from the expectations of the model to be identified as anomalies. In this theoretical framework, tests of factorial structure, internal consistency, discrimination, and difficulty of the items, as well as adjustment (infit and outfit), were performed. Essentially, those items that attempted against the unidimensionality of the ECL and EPL scales were eliminated, as were those that did not meet the standards considered adequate (difficulty index: between .20 and .85; discrimination index ≥ 25 ; punctual biserial correlation ≥ 30). Applying these criteria, a total of 25 items were eliminated (6 items from the ECL, 16 from the EPL, and 3 from the EAL).

These analyses resulted in a second, more refined, version of the three scales. The proof of concepts was formed, in this second version, of a total of 30 items, the procedures of 36 items, and the attitudes of 17 items. This version of the battery was then administered to a total of 169 students in their first ESO, in the presence of their teacher and a researcher. These same students also carried out an exercise that was complementary to the LBC. The exercise consisted of writing a short story in which the students had to incorporate all or some of seven polysemic words attached to it (e.g., saw, bank, wing). The literary competence of the texts was then independently evaluated by three judges, who are experts in literature, on a scale of 0 to 10. The average of the scores of the three judges was used as one of two external criteria. In addition, the language and literature teachers of the participating students were also asked to provide an assessment of the LC of each student, again on a scale of 0 to 10. This was the second of the external criteria used.

Results

Properties of the final version of the battery

Once this round of testing was completed, analyses were repeated in accordance with the assumptions of the TRI, resulting in a more refined version of the three scales. In their final version, the three scales had 25 (ECL), 30 (EPL), and 15 (EAL) items each. The highest possible LBC score was 100 (concepts and procedures 40 %, attitudes 20 %)⁴. Students with extremely high LC were expected to score between 75 and 100, while those with high LC were expected to score between 65 and 74, those with moderate LC between 50 and 64, those with low LC between 37.5 and 49, and those with very low LC between 25 and 37. Scores of below 25 were considered void, since this score is considered extremely low even for students with low LC.

Literary Concepts Scale (ECL)

Unidimensionality and reliability

To test the dimensionality of the concepts scale, an exploratory factor analysis (EFA) of maximum likelihood with oblimin rotation was carried out. The screen test showed a clear role of the first factor with respect to the others (with an eigenvalue for the first factor of up to four times

⁴ Although it may be considered that some items may belong to more than one scale, since our interest resided primarily in the final value of the CL, we do not believe that this circumstance alters the conclusions

that of the second factor). The second and subsequent factors showed similar residual values, suggesting that they are indicators of a common latent factor.

In order to test the internal consistency of the test, we proceeded to calculate the point biserial correlation index between each item and the test. Table 1 shows the 25 items to have acceptable values. Only item 6 is at the limit of what is considered desirable. In general, the scores indicate good internal consistency of the test. In addition, the reliability coefficient was calculated at .94, confirming the test's reliability.

Table 1. Item-scale correlations LBC (r_{bp})

ÍTEM	r_{bp}	ÍTEM	r_{bp}
C1	.41	C14	.40
C2	.47	C15	.50
C3	.41	C16	.49
C4	.41	C17	.47
C5	.33	C18	.44
C6	.29	C19	.40
C7	.45	C20	.39
C8	.37	C21	.53
C9	.48	C22	.49
C10	.45	C23	.30
C11	.43	C24	.36
C12	.44	C25	.38
C13	.41		

Discrimination and difficulty of the items

Table 2 shows the indices of discrimination and difficulty of items in the concepts scale. The discriminative capacity of the item increases as it moves away from zero, towards either 1 or -1. A score of 1 indicates that all the subjects with relevant knowledge were correct on the item and all those without were incorrect, reflecting perfect discrimination. Looking at Table 2, we can see that certain items, such as 1, 10, 18, and 19, have the highest rates, while others, such as 4 and 15, are at the limit of what is considered desirable for this parameter.

With regard to the indices of difficulty, higher scores reflect easier items, lower scores more difficult ones. Thus, the easiest statements were items 2, 4, 8, and 15 and the most difficult ones, items 6, 13, 17, and 23.

Table 2. Indexes of difficulty and discrimination of the test ECL

ITEM	INDEX	INDEX	ITEM	INDEX	INDEX
C1	.93	.43	C14	.36	.28
C2	.40	.85	C15	.27	.73
C3	.29	.29	C16	.28	.41
C4	.27	.78	C17	.40	.26

C5	.55	.61	C18	.86	.57
C6	.66	.22	C19	.85	.54
C7	.33	.29	C20	.37	.26
C8	.78	.77	C21	.29	.48
C9	.31	.31	C22	.38	.28
C10	.93	.37	C23	.48	.22
C11	.35	.28	C24	.60	.62
C12	.93	.68	C25	.83	.41
C13	.41	.20	TOTAL	.52	.46

Adjustment measures (infit and outfit)

Finally, INFIT and OUTFIT adjustment measures were obtained. These measures test the basic principle of the Rasch model. The two indices were statistically based on the residuals. In Tables 3 and 4 we can see that their values confirm a good fit, both of the global scale (1.00 and 1.04 respectively) and of all the scale's items (the only item with a score slightly higher than expected was item 4, with an outfit of an unstandardized quadratic mean of 2.16). Table 4 also includes level of difficulty, in terms of logits that confirm what is stated in Table 2's difficulty indices. In short, we can conclude that the properties are acceptable because the students who had knowledge of the literary concepts were the ones who responded correctly to the items.

Table 3. Adjustment measures for the Mean and Typical Deviation

ECL	INPUT		MEASURED		INFIT		OUTFIT	
	Total	Counted	Medida	Sep real	IMSQ	ZSTD	IMNQ	ZSTD
X	36.4	25	0	.30	1.00	-.1	1.06	.1
S.D	15.9	0	1.27	.05	.14	1.0	.37	1.1

Separation Index: 4.04; Reliability Index: .94

Table 4. Infit and Outfit values of the ECL

Item	Logits Difficulty	Standar error	INFIT		OUTFIT	
			MNSQ	ZSTD	MNSQ	ZSTD
C13	2.13	.35	1.09	.5	1.16	.5
C6	1.68	.32	.71	-1.8	.51	-1.2
C23	1.49	.30	.95	-.3	.86	-.2
C17	1.23	.30	1.08	.6	1.36	1.0
C20	1.23	.29	.99	.0	.97	.0
C11	1.15	.29	.97	-.2	.92	-.2
C14	1.07	.28	.79	-1.0	.79	-.6
C22	.68	.27	.99	-1.9	.68	-1.4
C3	.61	.37	.99	.0	.86	-.5
C7	.54	.27	.99	-.1	.92	-.3

C9	.54	.27	.99	.0	1.14	.7
C10	.12	.26	.98	-.2	.92	-.4
C16	.12	.26	.98	-.1	.93	-.3
C25	-.02	.26	.88	-1.1	.78	-1.2
C1	-.02	.26	1.03	.3	1.00	.1
C21	-.29	.26	.87	-1.3	.79	-1.2
C19	-.43	.27	.97	-.3	1.05	.3
C18	-.72	.27	1.24	2.0	1.40	1.9
C5	-.80	.27	1.02	.2	1.12	.6
C24	-.87	.28	1.26	2.0	1.71	2.1
C12	-1.03	.28	.85	-1.2	.76	-1.0
C15	-1.45	.30	.89	-.7	.71	-1.0
C8	-2.32	.37	1.23	1.0	1.36	.9
C4	-2.46	.38	1.08	.4	1.65	1.3
C2	-2.46	.38	1.22	.9	2.16	1.9

INFIT: Close setting; OUTFIT: Far setting; MNSQ: Unstandardized quadratic mean; ZSTD: Standardized quadratic mean.

Literary Procedures Scale (EPL)

Unidimensionality and reliability

The EFA rotation oblimin confirmed the one-dimensionality of the test. The factor explained up to 41 % of the variance.

Table 5 shows the correlation rates of each item with the scale. All of them show adequate values. The highest indices can be seen for items 1, 5, 6, 28, and 30. The items with the most discrete scores are 20 and 22. In addition, the EPL's reliability index was .94

Table 5. Item-scale correlations TPL (r_{bp})

Items	r_{bp}	Items	r_{bp}
P1	.58	P16	.32
P2	.37	P17	.27
P3	.34	P18	.41
P4	.29	P19	.42
P5	.54	P20	.41
P6	.42	P21	.32
P7	.30	P22	.24
P8	.31	P23	.30
P9	.28	P24	.20
P10	.26	P25	.34
P11	.30	P26	.40
P12	.27	P27	.38
P13	.45	P28	.49
P14	.31	P29	.40
P15	.38	P30	.46

Discrimination and difficulty of the items

Table 6 shows the indices of discrimination and difficulty for the items of the procedural scale. The test has adequate discrimination rates, with a couple of items at the limit of what is considered desirable (12 and 21). For its part, the difficulty index for the entire PCL test was .55. The most difficult items were 9, 11, 22, and 4, while the least difficult items were 15, 3, 20, and 26.

Table 6. Indices of difficulty and discrimination of the EPL test.

ITEM	DISCRIMINATION INDEX	DIFFICULTY INDEX	ITEM	DISCRIMINATION INDEX	DIFFICULTY INDEX
P1	.63	.79	P16	.71	.37
P2	.78	.44	P17	.39	.47
P3	.41	.86	P18	.81	.58
P4	.31	.31	P19	.35	.82
P5	.44	.73	P20	.41	.85
P6	.30	.35	P21	.26	.45
P7	.43	.44	P22	.48	.30
P8	.56	.43	P23	.30	.45
P9	.42	.29	P24	.30	.39
P10	.41	.77	P25	.92	.70
P11	.37	.30	P26	.75	.82
P12	.25	.49	P27	.46	.69
P13	.93	.63	P28	.26	.58
P14	.27	.57	P29	.61	.41
P15	.39	.86	P30	.80	.49
			TOTAL	.49	.55

Adjustment measures (infit and outfit)

With regard to the near and far adjustment measures (infit and outfit) of the EPL, we can observe in Table 7 that these indices indicate good adjustment. The quadratic mean statistics show the distortion of the measurement system used. Values around 1 were expected: substantially lower values would indicate that the observations were too predictable, and values substantially above that, that they were too unpredictable (González, 2008). The criterion for considering good adjustment rates is between 1 and 1.30 for the unstandardized quadratic mean and between -2 and .2 for the standardized one. Table 7 shows that the items requiring future analysis in this regard are 11, 19, and 28.

Table 7. Adjustment measures for the Mean and Typical Deviation

EPL	INPUT		MEASURED		INFIT		OUTFIT	
	Total	Counted	Measured	Sep Real	IMSQ	ZSTD	IMNQ	ZSTD
X	43.0	30	0	.29	.99	.0	1.02	.1
S.D	16.6	0	1.21	.04	.11	.9	.23	1.1

Separation Index: 3.99; Reliability Index: .94

Table 8. Infit and Outfit values of the ECL

Item	Logits Difficulty	Error standar	INFIT		OUTFIT	
			MNSQ	ZSTD	MNSQ	ZSTD
P9	2.40	.34	.99	.0	1.12	.5
P11	1.74	.28	1.12	.8	1.40	1.7
P22	1.44	.27	1.12	1.1	1.31	1.7
P4	1.29	.26	1.03	.3	1.02	2
P6	1.09	.26	1.15	1.5	1.32	2.1
P16	.98	.26	.97	-.2	.96	-.2
P24	.96	.25	.89	-1.2	.86	-1.1
P29	.84	.25	1.13	1.5	1.23	1.8
P8	.65	.25	1.01	.1	.99	-.1
P2	.59	.25	.98	-.2	.97	-.3
P7	.53	.25	1.07	.9	1.09	.9
P21	.47	.25	.99	-.1	1.01	.1
P23	.41	.25	1.00	.1	1.02	.2
P17	.22	.25	1.01	.2	1.02	.3
P12	.22	.25	.87	-1.6	.88	-1.2
P30	.10	.25	.97	-.3	.96	-.3
P14	.05	.25	.89	-1.5	.87	-1.3
P18	.04	.25	1.12	1.3	1.11	.9
P28	-.25	.25	.95	-2.1	.76	-2.1
P13	-.44	.26	.94	-.6	.94	-.4
P27	-.96	.29	.94	-1.7	.63	-1.8
P25	-.96	.29	.87	-.8	.75	-1.1
P5	-1.23	.31	1.07	.4	.99	.1
P10	-1.33	.32	.77	-1.2	.61	-1.5
P1	-1.46	.32	1.00	.0	.91	-.2
P19	-1.55	.34	1.08	.4	1.54	1.5
P26	-1.67	.35	.96	-.1	1.07	.3
P20	-1.67	.35	.92	-.8	.99	-1.3
P3	-1.67	.35	.99	-.0	.95	.0
P15	-1.95	.39	.96	-.1	1.02	.2

INFIT: Close fit; OUTFIT: Far setting; MNSQ: Unstandardized quadratic mean; ZSTD: Standardized quadratic mean.

Attitude to literature scale (EAL)

Unlike the previous scales, the EAL had no right or wrong answers; rather, students responded on a Likert scale of 1 to 5. Also, unlike the previous scales, it was not a one-dimensional scale but included several factors. To explore this factor structure, a maximum likelihood EFA with Oblimin rotation was carried out, since it was assumed that the factors were related. This rotation allowed the final refinement of the test, with a final solution that explained 49% of the variance and reduced the original structure from five to three factors. The three factors of the EAL were: 1.- reading, real life and personal relationships; 2.- interest in reading and self-perception as a reader; and 3.- reading as a source of experiences and creativity.

To assess the reliability of the scale, Cronbach's alpha of the factors and the global scale were calculated. In Tables 9 and 10 we can see the indices by factor and item. Total score ranged from 15 to 75. Furthermore, with a Cronbach's alpha of .81, the scale's internal consistency could be considered adequate.

Table 9. Internal consistency of the scale EAL

	F1	F2	F3	Total
Number of items	6	4	5	15
Variance Explained	29.6 %	11.2 %	7.5 %	48.3 %
Alpha de Cronbach	.78	.74	.70	.81

Table 10. Internal consistency indices and reliability item by item of the scale EAL

Item	Media	D.T.	Correl. Item total	Alpha Cronbach excl. ítem	Item	Media	D.T.	Correl. Item total	Alpha Cronbach excl. ítem
A1	3.70	1.110	.435	.71	A9	4.51	.69	.422	.72
A2	3.68	1.258	.491	.70	A10	3.51	1.167	.444	.71
A3	3.46	1.445	.439	.71	A11	3.55	1.28	.395	.79
A4	4.52	.820	.336	.72	A12	3.02	1.26	.316	.82
A5	3.18	1.282	.483	.70	A13	4.51	.69	.422	.72
A6	3.93	1.264	.528	.70	A14	3.51	1.167	.444	.71
A7	2.59	1.153	.411	.71	A15	3.01	1.275	.323	.72
A8	3.66	1.263	.509	.70					

Battery validity

The validity of the criterion was analysed using Pearson's correlations between scores on the questionnaires and the two external criteria measures: the LC of students as independently assessed by expert judges on the basis of a practical exercise; and the LC of students as assessed by their literature teachers on the basis of students' regular performance in the classroom.

To calculate the degree of agreement between the three judges assessing the external LC criteria, the Intraclass Correlation Coefficient (ICC) was calculated. The index for average measures was .95, a value comparable to Cronbach's alpha and one that confirms the reliability of the measure.

Table 11. Correlations between the assessments of judges, teachers and the BCL.

	ECL	EPL	EAL	BCL
Judges (average)	.25(**)	.15(*)	.42(**)	.32(**)
Teachers	.76(**)	.83(**)	.31(**)	.93 (**)

* p < .05, ** p < .01

As can be seen in Table 11, the three scales show significant correlations in the expected direction with each of the criteria variables. Both the global and the specific scores of each scale show direct and significant relationships with both the subjective assessments of CL of the teachers and those of the judges. The analyses highlighted the strong correlation between teacher assessment and overall scores on the LBC. Here, literary procedures took centre stage with the highest correlation. Correlations with the assessments of the judges were also significant, although not as high. Specifically, the lowest correlation was found between the procedures scale and the assessment of the judges, perhaps because in the practical exercise it was easier to capture knowledge and attitudes than it was the process used to generate the written product.

Discussion

Although evaluation systems used in the classroom differ from the test battery presented here, it cannot be denied that, as Paran and Sercu (2010) suggest, literary competence seems to a multitude of teachers to be an impossible subject to evaluate. This circumstance has led to erratic evaluation systems focused only on the contents (Badia & Cassany, 1994). Our intention with the LBC was not only to provide a LC measurement tool, but also to suggest what aspects should be measured by it. If LC is a polyhedral competence, so too should its evaluation be.

Although some aspects remain to be polished, the psychometric properties of the test battery presented suggest that in general it can be considered a plausible option for evaluating the complex construct that is literary competence. We believe, as Miall (2006) pointed out, that the view of literature as an indefinable subject has to some extent prevented empirical studies like this one, which is concerned with rendering objective concepts that have often been left in calculated ambiguity, from being carried out. Although literature does contain a high degree of subjectivity, this does not mean that it cannot be studied using scientific parameters.

We believe that having an instrument that can evaluate the characteristics of literary education opens the door to a multitude of investigations that can validate didactic proposals for literary education. The possibility of being able to measure the LC level of students with a standardized, validated, and reliable tool that yields a final numerical value (albeit with the option of segmenting it into three values according to the three scales used here), denotes considerable progress not only in the didactic but also the empirical study of literature.

The LBC has not an absolute value but a relative one, as it is linked to subjective and ambiguous elements as well as to certain circumstances whose impact is difficult to determine. For example, one of the participating centres was located in the hometown of an author included in the test, so the percentage of students there who knew him was much higher than that in the other two centres. However, having a large number of items should have served to minimise the effects of these circumstances.

There are also some limitations with the study. The first is the definition of LC itself, which, as we have seen, is not consistently agreed upon in the international literature. However, from our point of view the contributions of Mendoza (2004) and Witte, Janssen, and Rijlaarsdam (2006) are strong enough to support the LBC. Furthermore, there is a certain difficulty in translating the LBC into other languages, since it is necessary to find texts of the same characteristics in order to make comparisons. Although it would be possible to find formally similar texts, it would be more difficult to determine their place in the canon of other countries' literature or their popularity among students at each level. We believe that such divergences are not overly relevant either for the number of items used in the LBC or for the relative data of this investigation. In this study, differences in the results between Valencian- and Spanish-speaking students were not significant, indicating that correspondence between these two languages was strong. A final limitation is that since literary competence is a competence that develops, and so our test in its current form is valid only for students in their first ESO.

As for future lines of enquiry, we believe that the most important is to adapt the battery to the different education levels of both secondary and primary-school children. Another future line of research should be to find or validate other literary competence measurement systems, such as foregrounding, shorter questionnaires, text writing, and so on. With the LBC as a reference, it may be easier to determine the scope of these other measurement systems.

In conclusion, we believe that the LBC meets the objectives for which it was designed and is a reliable and valid instrument for measuring LC in the adolescent population. That is why it is proposed as a very useful tool both for researchers interested in conducting empirical studies on the teaching-learning of literature, and for teachers who want an objective method of assessing their students and thus detecting possible deficiencies in their own teaching methods.

References

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Badia, J., & Cassany, D. (1994). La classe de literatura, avui. *Articles*, 1, 7-14.
- Bierwisch, M. (1970). Poetics and linguistics. In D. C. Freeman (Ed.), *Linguistics and literary style*. New York: Holt, Rinehart and Winston.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6(2), 5-35.
- Cenoz Iragui, J. (2004). El concepto de competencia comunicativa. In J. Sánchez Lobato, & I. Santos Gargallo (Eds.), *Vademécum para la formación de profesoras: Enseñar español como segunda lengua (L2) lengua extranjera (LE)*, 449-465. Madrid: Sociedad General Española

de Librería.

- Chomsky, N. (1966). *Topics in the theory of generative grammar*. The Hague: Mouton.
- Culler, J., Eco, U., Gandy, R., Robey, D., & Varela, P. (1976). *Introducción al estructuralismo*. Madrid: Alianza.
- de Amo, J.M., Ruiz, M^a.M. (2014) Niveles de adquisición de la competencia literaria en estudiantes de secundaria: análisis de resultados. En Osvaldo Cleger y Jose M. de Amo, *La educación literaria y la e-literatura desde la minificción*. ICE: Barcelona.
- González, M. J. (2008). *El análisis de reactivos con el Modelo Rasch*. Manual Técnico A. México D.F. INEE.
- Hymes, D. (2000). Acerca de la competencia comunicativa. In M. Llobera (Ed.), *Competencia comunicativa. documentos básicos en la enseñanza de lenguas extranjeras* (pp. 13-37). Madrid: Edelsa.
- Kucer, S. B. (2001). *Dimensions of literacy: A conceptual base for teaching reading and writing in school settings*. Mahwah NJ: Lawrence Erlbaum Associates.
- Mendoza, A. (1998). *Conceptos clave en didáctica de la lengua y la literatura*. Barcelona: Universitat de Barcelona.
- Mendoza, A. (2008). Leer y comparar: Notas sobre las posibilidades del comparativismo en el aula de educación secundaria. *Lenguaje y textos*, 28, 19-41.
- Mendoza, A. (2001). *El intertexto lector*. Cuenca: EUCL.
- Mendoza, A. (2004). *La educación literaria: Bases para la formación de la competencia lecto-literaria*. Archidona Málaga: Aljibe.
- Miall, D. S. (2006). Empirical approaches to studying literary readers: The state of the discipline. *Book History*, 9, 291-311.
- Mínguez-López, X. (2015). El papel de la animación en el desarrollo de la competencia literaria. *Elos. Revista de Literatura Infantil e Xuvenil*, 2, 157-172. doi:<http://dx.doi.org/10.15304/elos.2.2670>
- Paran, A., & Sercu, L. (2010). *Testing the untestable in foreign language education*. Buffalo, NY: Multilingual Matters.
- Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: Univ Chicago.
- Salvador, V. (1984). *El gest poètic: Cap a una teoria del poema*. València: Universitat de València, IFV & ICRTV.
- Torell, O. (2011). Literary competence beyond conventions. *Scandinavian Journal of Educational Research*, 45(4), 369-379.
- van Dijk, T. A. (1972). *Some aspects of text grammars: A study in theoretical linguistics and poetics*. The Hague: Mouton.
- Witte, T., Janssen, T., & Rijlaarsdam, G. (2006). *Literary competence and the literature curriculum*. Unpublished manuscript.
- Witte, T., Rijlaarsdam, G., & Schram, D. (2012). An empirically grounded theory of literary development. Teachers' pedagogical content knowledge on literary development in upper secondary education. *L1-Educational Studies in Language and Literature*, 12, 1-33.