

PRUEBA PILOTO DE EVALUACIÓN DE LA EXPRESIÓN ORAL EN LENGUA INGLESA EN LA PRUEBA DE ACCESO A LA UNIVERSIDAD (PAU) 2012

INFORME DE LA PRIMERA FASE

Noviembre, 27 2012

**M^a Ángeles Díez Santos, Madrid
Rubén Fernández Alonso, Asturias
Gillermo Gil Escudero, INEE
Stefano de Marco, INEE
Luis Sanz San Miguel, INEE**

RESULTADOS PRINCIPALES

Índice

I DESCRIPCIÓN DEL ESTUDIO PILOTO

I.1 Introducción	5
I.2 Objetivos	5
I.3 Procedimiento	6
<i>I.3.1 Prueba de evaluación</i>	6
<i>I.3.2 Criterios de evaluación</i>	6
<i>I.3.3 Examinadores</i>	7
<i>I.3.4 Datos de contexto de los alumnos</i>	7
<i>I.3.5 Datos adicionales</i>	7
I.4 Organización	8
I.5 Muestra	8

II RESULTADOS

II.1 Nivel de expresión oral en Lengua Inglesa	9
<i>II.1.1 Resultado total, por criterios y por partes de la prueba</i>	9
<i>II.1.2 Dimensionalidad de los criterios de calificación</i>	11
II.2 Relación del nivel de competencia con variables clave	12
<i>II.2.1 Características del alumno</i>	12
Sexo	12
Edad, repetición de curso	12
<i>II.2.2 Contexto del alumno</i>	14
Nivel socioeconómico y cultural de las familias de los estudiantes ISEC	14
Tamaño de la población	16
<i>II.2.3 Contexto del aprendizaje de la Lengua Inglesa</i>	17
En el centro escolar	17
Duración y frecuencia de las clases	17
Uso de la lengua inglesa hablada en clase	18
Recursos pedagógicos	18
Titularidad del centro	19
Tiempo dedicado al estudio y a tareas fuera del horario de las clases	20
Clases adicionales de inglés y su frecuencia	21
Grado de exposición a la Lengua Inglesa	21
Viajes de estudio e intercambios en países de habla inglesa y su duración	22
II.3 Relación con otras calificaciones	23
II.4 Fiabilidad de las calificaciones de los correctores	26
II.4.1 Concordancia entre correctores	26
II.4.2 Relación del nivel de concordancia de los correctores con otras variables	27
II.5 Capacidad de predicción del nivel de competencia en expresión oral	28
II.6 Opiniones del profesorado	29
III CONCLUSIONES Y RECOMENDACIONES	32

RESULTADOS PRINCIPALES

Diseño y estructura de la prueba

Los datos obtenidos muestran validez y fiabilidad estadística suficientemente altas y desde un punto de vista técnico, el diseño y la estructura de la prueba pueden mantenerse con las mismas características básicas que en el estudio piloto, en dos partes diferenciadas manteniendo aproximadamente la misma proporción de tiempo que en la prueba piloto, si bien reduciendo el tiempo de aplicación de 10 a 7 u 8 minutos. Se recomienda su aplicación a finales del segundo trimestre.

La división de la calificación en cinco criterios parciales calificados independientemente se ha manifestado conveniente para facilitar el trabajo de los correctores. Parece razonable que cada parte se califique calculando la media aritmética de los cinco criterios. Se propone utilizar una única puntuación global con una distribución de porcentajes del 30% para la primera parte y 70% para la segunda.

Se recomienda seguir formando equipos de tres correctores. Si, por razones económicas, se optase por utilizar dos correctores, debería consensuarse la calificación de ambos, dadas las diferencias encontradas.

Resultados en los diferentes contextos en los que se mueve el alumno.

La inclusión de la prueba de expresión oral en inglés en la PAU, dada la relación estadísticamente significativa y positiva encontrada entre los resultados y el ISEC, incrementaría el efecto de las diferencias de clase socio-económica y cultural de los estudiantes sobre las calificaciones que obtendrían en las PAU. Además, la titularidad del centro tiene influencia en las calificaciones de la prueba oral. Los alumnos de los centros privados obtienen mejores calificaciones que los de los centros públicos si no se tiene en cuenta el efecto del ISEC. Ahora bien, una vez descontado el efecto del ISEC la diferencia no es significativa.

La duración de las clases y su frecuencia semanal también influye en los resultados obtenidos en la prueba oral, mientras que el uso de la lengua hablada en clase y la utilización de recursos pedagógicos tienen una modesta influencia en los resultados, algo menos del 1% y del 1.2% respectivamente.

La exposición del estudiante al inglés mediante comunicaciones, orales o escritas, internet, etc., tienen una influencia de aproximadamente el 22% en la calificación de la prueba oral. Los viajes de estudio o intercambio en países de habla inglesa explican alrededor de un 12% de la variabilidad de la calificación final, en tanto que si el viaje es con la familia esa influencia se cifra en torno al 14%.

Relación de la calificación en la prueba oral con otras calificaciones de los estudiantes

La calificación en la prueba oral está altamente correlacionada con la calificación de inglés en el centro y con la calificación de inglés en la PAU. De manera que la puntuación de la prueba oral de inglés viene explicada, en buena parte, por las calificaciones otorgadas por los profesores de los centros y por el examen de inglés escrito. Las variables calificación de inglés en 2º de bachillerato, calificación de inglés en la PAU y repetición de curso explican el 50% de la variabilidad de la puntuación en expresión oral.

Debe considerarse si merece la pena el esfuerzo económico y organizativo necesario para implantar la prueba oral de inglés en la PAU, cuando el valor numérico añadido a la calificación de la PAU es muy pequeño. No obstante, la introducción de una prueba de estas características, condicionaría la enseñanza de la lengua inglesa que, necesariamente, se enfocaría a desarrollar aspectos más acordes con la expresión oral y resultaría beneficioso para alumnos.

Opiniones de los profesores

Los profesores participantes en la prueba valoran muy positivamente la metodología de la misma, el 87% de ellos la considera adecuada o muy adecuada. La organización también es valorada positivamente puesto que el 81.4% la considera adecuada o muy adecuada.

Nivel de expresión oral en lengua inglesa

Los resultados del estudio piloto muestran que aproximadamente 2 de cada 3 alumnos de 2º curso de Bachillerato (un 61,39%) tienen un nivel de expresión oral en Lengua Inglesa B1, según la definición del marco europeo de las lenguas y que es el que se ha establecido como nivel requerido para aprobar la prueba en las PAU.

I DESCRIPCIÓN DEL ESTUDIO PILOTO

I.1 Introducción

Con el objeto de llevar a cabo en las mejores condiciones posibles el cumplimiento de lo establecido en el Real Decreto 1892/2008 de 14 de noviembre, por el que se regulan las condiciones para el acceso a las enseñanzas universitarias oficiales de grado y los procedimientos de admisión a las universidades públicas españolas (BOE 283 de 24 de noviembre de 2008) en el que se establece una prueba de evaluación de la expresión oral en Lengua Extranjera se decidió (en las reuniones de la Conferencia Sectorial de Educación de 26 de enero de 2012 y de la Comisión General de Educación del 5 de marzo de 2012) llevar a cabo una prueba piloto de dicha evaluación.

La prueba piloto consiste en realizar un ensayo, durante los años 2012 y 2013, tanto de la prueba en sí misma como de su integración en el sistema completo de evaluación de la PAU, previamente a su implantación de modo generalizado. La prueba piloto se ha desarrollado en dos fases:

En la primera fase, realizada en el tercer trimestre de 2012, han participado 7 comunidades autónomas: Asturias, La Rioja, Comunidad Valenciana, Aragón, Castilla-La Mancha, Baleares y Madrid. Esta fase de la prueba piloto estaba dirigida principalmente al estudio de la prueba y la metodología de la evaluación.

La segunda fase, a realizar en el segundo trimestre de 2013, estará dirigida principalmente al estudio de los aspectos organizativos de la prueba y de su integración en la PAU.

El calendario de trabajos de la primera fase se desarrolló de la siguiente manera:

- Abril 2012. Diseño de la Prueba Piloto (muestreo, características del examen, organización de los exámenes, logística, etc.).
- Mayo 2012. Realización de los exámenes (del 3 de mayo al 1 de junio de 2012).
- Junio 2012. Codificación y grabación de los datos.
- Julio-Septiembre 2012. Análisis de datos.
- Octubre 2012. Informe sobre la primera fase de la prueba piloto.

I.2 Objetivos

El objetivo prioritario de la primera fase de la prueba piloto es probar la metodología de evaluación y las características de la prueba de cara a su posterior implantación de modo generalizado.

Para ello, era necesario analizar prioritariamente los siguientes aspectos de la prueba:

- la adecuación del nivel de exigencia de la misma.
- la adecuación de la duración de la prueba para cada alumno.
- la fiabilidad de las calificaciones de los examinadores.
- la adecuación de la escala de calificación.
- la adecuación de los estímulos utilizados para proporcionar contextos de referencia a la expresión oral de los estudiantes.

Adicionalmente al estudio de la metodología de evaluación, esta primera fase incluye también como objetivos secundarios:

- el análisis de la relación de los resultados de rendimiento obtenidos con:

- las calificaciones escolares en Bachillerato
 - las calificaciones en otras pruebas de la PAU
 - el nivel socioeconómico y cultural de las familias de los estudiantes.
- la consideración de los aspectos organizativos de la prueba.

I.3 Procedimiento

El procedimiento general de la primera fase de la prueba piloto preveía examinar a 25 alumnos de 56 centros educativos de Bachillerato, 1400 alumnos, para valorar su nivel de expresión oral en Lengua Inglesa. Adicionalmente, se recopilaron datos relevantes sobre el contexto personal de los alumnos, sobre la enseñanza recibida de la Lengua Inglesa y sobre los contextos de utilización de la misma.

I.3.1 Prueba de evaluación

Como punto de partida se consideró que el nivel promedio de la prueba de expresión oral en Lengua Inglesa debería aproximarse al nivel B1 definido en el marco común europeo para las lenguas (Anexo I).

En este nivel, el candidato deberá utilizar el idioma con cierta seguridad y flexibilidad, receptiva y productivamente, tanto en forma hablada como escrita, así como para mediar entre hablantes de distintas lenguas, en situaciones cotidianas y menos corrientes que requieran comprender y producir textos en una variedad de lengua estándar, con estructuras habituales y un repertorio léxico común no muy idiomático, y que versen sobre temas generales, cotidianos o en lo que se tiene un interés personal.

Más específicamente en cuanto a la Expresión e Interacción Oral los objetivos generales del nivel B1 se centran en la adquisición de la capacidad de producir textos orales bien organizados que traten temas personales y cotidianos y que estén adecuados al interlocutor y propósito comunicativo. Puede desenvolverse con una corrección, fluidez y espontaneidad que permitan mantener la interacción, aunque a veces resulten evidentes el acento extranjero, las pausas para planear el discurso o corregir errores y sea necesaria cierta cooperación por parte del examinador.

Para la prueba piloto se estableció una duración de 8 a 10 minutos de examen por alumno, consistiendo la prueba de evaluación de la expresión oral en dos partes:

- Una conversación inicial con el examinador de aproximadamente 2 ó 3 minutos de duración. Los temas de esta primera parte versaban sobre información personal: nombre, familia, trabajo, estudios, tiempo libre, vacaciones, experiencias en el pasado y planes para el futuro. En el Anexo II se incluyen diversas preguntas organizadas en bloques temáticos que pueden utilizarse para esta primera parte de la evaluación.
- Una conversación, de aproximadamente 6 ó 7 minutos, en la que se incluía la descripción, por parte del alumno, de unas fotos y la respuesta a preguntas sobre temas relacionados con las mismas, expresando opiniones, acuerdo o desacuerdo, sugerencias, etc. En el Anexo III se incluyen los estímulos utilizados en la prueba junto con un listado orientativo de posibles preguntas que pueden incluirse en la evaluación.

I.3.2 Criterios de evaluación

Cada parte de la prueba se evaluó utilizando cinco criterios y registrando las calificaciones concretas en la ficha de calificación. Los examinadores evaluaron cada criterio -alcance, corrección gramatical, fluidez,

interacción y coherencia-, con una de las calificaciones de la tabla, de 1 a 10 puntos. En el Anexo IV se incluyen los descriptores para la evaluación de cada uno de los criterios de la prueba de Expresión Oral y en el Anexo V la ficha de calificación utilizada por los examinadores de la prueba.

I.3.3 Examinadores

La prueba se realizó individualmente, con equipos de tres examinadores. El objetivo de utilizar tres examinadores es analizar el grado de fiabilidad de las calificaciones de las pruebas y estimar el número más adecuado de examinadores que deberán intervenir en la prueba final.

Los examinadores calificaron individualmente a cada alumno y evitaron compartir entre ellos las valoraciones y puntuaciones otorgadas, dado que uno de los objetivos de la prueba piloto es analizar el grado de concordancia entre las puntuaciones de los tres examinadores.

Cada comunidad autónoma decidió el número de equipos adecuado para llevar a cabo la primera fase de la prueba piloto en su comunidad. Asturias y la Comunidad Valenciana utilizaron 3 equipos de tres profesores, mientras que Aragón, Castilla-La Mancha, La Rioja y Madrid organizaron la prueba con 4 equipos. Baleares, dada su insularidad, empleó 7 equipos de profesores. En total participaron en la prueba piloto 90 profesores que conformaron 30 equipos de aplicación de la prueba piloto.

I.3.4 Datos de contexto de los alumnos

Para la recopilación de datos relevantes sobre el contexto personal de los alumnos, sobre la enseñanza recibida de la Lengua Inglesa y sobre los contextos de utilización de la misma, se utilizó un cuestionario para el alumnado que incluye dos partes diferenciadas: una primera parte dirigida a obtener datos del alumno y su contexto familiar y una segunda parte orientada a recopilar datos sobre el aprendizaje de la Lengua Inglesa. El cuestionario se recoge en el Anexo VI.

En la primera parte se recopilan datos básicos de identificación de los alumnos -edad, sexo-, de su trayectoria académica -repetición de curso- y datos necesarios para la estimación del nivel socioeconómico y cultural de su entorno familiar -trabajo, nivel de estudios, profesión y situación laboral de los padres, posesiones y número de libros en el hogar- y de otras características -comunidad autónoma o país de origen, lengua materna-. Para la codificación de las profesiones se utilizaron los códigos internacionales ISCO-2008 y para la de los países de origen de los estudiantes y sus familias la codificación ISO-3166.

En la segunda parte se recogen datos sobre el contexto del aprendizaje de la Lengua Inglesa tanto en el centro escolar -tamaño, duración y frecuencia de las clases, uso de la lengua hablada en clase, recursos pedagógicos- como sobre el tiempo dedicado al estudio y a tareas fuera del horario de las clases, así como datos sobre la utilización del inglés y el contacto con la Lengua Inglesa en horario extraescolar, incluyendo las clases adicionales de inglés y su frecuencia, junto con los viajes de estudio e intercambios en países de habla inglesa y su duración.

I.3.5 Datos adicionales

Se recogieron asimismo datos de rendimiento académico adicionales de los alumnos de la muestra, tanto del Bachillerato como de las calificaciones obtenidas en la Prueba de Acceso a la Universidad. Del Bachillerato, se recogieron la calificación final en Lengua Inglesa y la nota media del curso. De la PAU, la calificación en Lengua Inglesa y las calificaciones medias de las Fases General y Específica de la PAU.

I.4 Organización

Una vez seleccionada la muestra de Centros participantes en la prueba piloto, la unidad responsable de la evaluación de cada comunidad autónoma organizó la realización de los exámenes en su ámbito territorial.

I.5 Muestra

Se utilizó un diseño muestral dirigido con el objetivo de obtener datos que reflejasen la máxima variabilidad posible de los resultados. Para ello el diseño se fundamentó en la utilización de las variables de comunidad autónoma, tipo de centro -público, privado- y tamaño de la población -hasta 10.000 habitantes, de 10.000 a 100.000 y de más de 100.000 habitantes-. Adicionalmente, en cada comunidad autónoma se intentó incluir –siempre que fuese posible– al menos un centro bilingüe en cada una de las categorías de tipo de centro, así como al menos un centro privado en cada uno de los grupos de centros en poblaciones pequeñas y medianas. En la tabla I.5 se muestra el diseño muestral para cada una de las comunidades autónomas. Participaron en la primera fase del estudio piloto siete comunidades autónomas: Asturias, Aragón, Baleares, Castilla-La Mancha, Comunidad Valenciana, La Rioja y Madrid.

Tabla I.5: Diseño muestral para cada una de las CCAA participantes en el estudio.

DISEÑO MUESTRAL por CCAA			
Tamaño de la población	Tipo de Centro		Total
	Público	Privado	
Grande (> 100.000 hab.)	2 (a ser posible 1 centro Bilingüe)	2 (a ser posible 1 centro Bilingüe)	4
Mediano (< 100.000 hab. >10.000)	2 (a ser posible 1 centro Privado)		2
Pequeño (< 10.000 hab.)	2 (a ser posible 1 centro Privado)		2
Total			8

De manera que en cada comunidad autónoma se seleccionaron aleatoriamente 8 centros de educación secundaria y sus respectivos centros suplentes. Esta selección fue aleatoria por estratos definidos en función de las variables antes mencionadas -CCAA, titularidad, tamaño de la población-. Se excluyeron los centros con 10 alumnos o menos. En Baleares se seleccionó al menos un centro por isla.

Posteriormente, se seleccionaron aleatoriamente 25 alumnos en cada uno de los centros –así como suplentes en los casos en que fue posible–, de tal modo que la muestra planificada resultó, en principio, de 200 alumnos por cada CCAA, lo que supone un total de 1.400 alumnos.

Finalmente se evaluó a 1.194 alumnos, de los 1.400 planificados inicialmente, lo que supone una tasa de respuesta del 85.29 por ciento. Aunque esta tasa de respuesta no es alta, es suficiente para cumplir los objetivos del estudio. Las fechas de aplicación de la prueba, justo antes de la realización de la PAU es la razón básica por la que se ha obtenido esta tasa de respuesta.

II RESULTADOS

II.1 Nivel de expresión oral en Lengua Inglesa

La prueba de expresión oral en Lengua Inglesa consta de dos partes bien diferenciadas. La primera parte, cuya duración puede ser de 2 a 3 minutos, va dirigida fundamentalmente a establecer confianza en el alumno y a que éste se adapte al contexto de la prueba. En la segunda parte se pretende llevar a cabo una mayor profundización en la evaluación de la competencia del estudiante, mediante la descripción y comentario de estímulos fotográficos y su duración puede establecerse entre 6 y 7 minutos aproximadamente.

Tanto por la duración de cada parte como por la estructura de la prueba parece conveniente otorgar mayor peso a la segunda parte que a la primera. De esta manera, para la valoración inicial de los resultados se considera como puntuación del alumno en expresión oral en Lengua Inglesa el promedio ponderado de cada una de las dos partes de que consta la prueba: 20% para la primera parte y 80% para la segunda parte. La calificación de cada parte de la prueba es la media aritmética de las calificaciones otorgadas por los tres examinadores.

La relación entre las calificaciones por criterios entre la primera y la segunda parte de la prueba es muy alta, ya que la correlación entre la puntuación de cada criterio de la primera parte con el correspondiente de la segunda es muy elevada como se puede observar en la tabla II.1; donde se puede ver también el alto grado de correlación entre las puntuaciones total promedio de los estudiantes obtenidas en ambas partes.

Tabla II.1: Correlaciones entre las calificaciones de la primera y segunda parte de la prueba por criterios

	Criterio					Total
	Alcance	Corrección Gramatical	Fluidez	Interacción	Coherencia	
R	.968	.970	.969	.970	.969	.979
pvalor	.000	.000	.000	.000	.000	.000

II.1.1 Resultado total, por criterios y por partes de la prueba

La Tabla II.1.1 muestra las medias de las puntuaciones obtenidas por los alumnos en la primera parte -información personal- y en la segunda parte -descripción y comentario de fotografías- de la prueba, así como los promedios totales ponderados de cada criterio y de cada parte de la pruebaⁱⁱⁱ. Los resultados señalan que los estudiantes obtienen las puntuaciones más bajas en *Corrección Gramatical* y las más altas en *Interacción* y *Coherencia*, situándose los criterios de *Alcance* y *Fluidez* en un punto intermedio.

Tabla II.1.1: Promedio de las calificaciones de los estudiantes en la prueba de expresión oral en Lengua Inglesa, por partes de la prueba, criterios y su total ponderado.

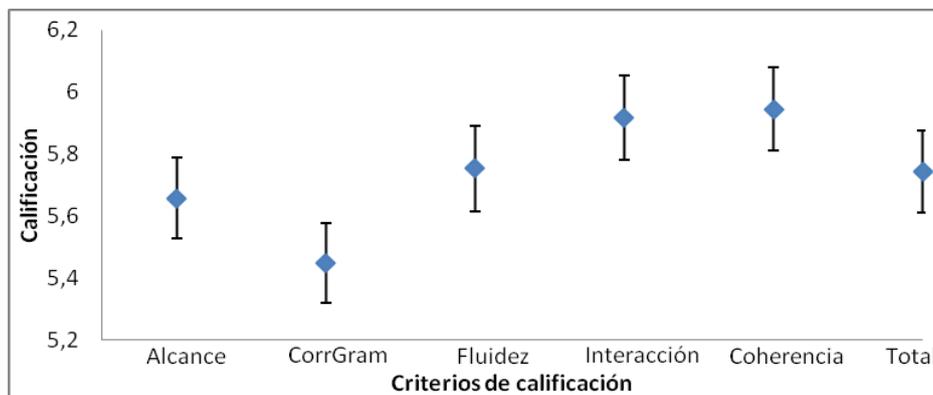
		Criterio					Total
		Alcance	Corrección Gramatical	Fluidez	Interacción	Coherencia	
1ª Parte	Media	5.78	5.56	5.86	6.04	6.04	5.86
	E.T.	0.0634	.06392	0.0676	0.0666	0.0658	0.0648
	D.T.	2.192	2.209	2.338	2.304	2.274	2.239
2ª Parte	Media	5.63	5.42	5.73	5.89	5.92	5.72
	E.T.	.06745	0.0670	.07162	.07056	.06935	.06850
	D.T.	2.330	2.313	2.474	2.438	2.396	2.366
Total Ponderado	Media	5.66	5.45	5.75	5.92	5.94	5.74
	E.T.	0.0663	0.0660	0.0705	0.0695	0.0683	.06754
	D.T.	2.292	2.282	2.436	2.400	2.361	2.333

El gráfico II.1.1 muestra el promedio de las calificaciones de los estudiantes en la prueba de expresión oral, por criterios de calificación y su total ponderado, junto con el intervalo de confianza al 95% para cada una de las calificaciones.

Si se comparan los resultados obtenidos entre los criterios respectivos de las dos partes de la prueba, se observa que existe diferencia estadísticamente significativa entre las calificaciones otorgadas por cada criterio en la primera y en la segunda parte de la prueba. La tabla A.VII.1 del anexo VII muestra los resultados de comparar la calificación de cada criterio de la primera parte con su respectivo de la segunda parte.

Estos resultados aconsejan mantener las dos partes de la prueba, siendo en la primera parte de la misma en la que los estudiantes obtienen, de manera significativa, mejores puntuaciones que en la segunda. En el Anexo VII se recogen los detalles de las comparaciones realizadas y el procedimiento estadístico utilizado para el cálculo de la significatividad de las diferencias. En el gráfico A.VII.1 se representan las medias de las diferencias de las calificaciones obtenidas por los alumnos en cada una de las partes de la prueba para cada criterio y el total.

Gráfico II.1.1: Promedio de las calificaciones de los estudiantes en la prueba de expresión oral en Lengua Inglesa, por criterios de calificación y su total ponderado



Una forma de considerar los resultados, más orientada a los aspectos prácticos de la prueba de expresión oral, es analizar el porcentaje de alumnos que obtienen calificaciones por encima y por debajo de 5 puntos, esto es que *aprueban* o *suspenden* en expresión oral teniendo como referencia el nivel B1 definido en el marco común europeo para las lenguas. La tabla A.VII.2 muestra el porcentaje de alumnos que obtienen 5 puntos o más en los cinco criterios valorados y en la puntuación total. Los resultados muestran que, en su conjunto, el 61.39% de los alumnos alcanzan el nivel B1 y se podrían considerar como aprobados, mientras que el resto, 38.61%, no alcanzan dicho nivel.

De los cinco criterios, el que tiene menor porcentaje de aprobados es el de *corrección gramatical*, 54.77%, mientras que el que alcanza un nivel más alto es el de *coherencia*, 64.57%. Ello puede deberse a que un fallo en gramática es considerado por los examinadores -por el profesorado en general- más grave y, por tanto, se penaliza más que un fallo en cualquier otro criterio de la prueba.

Al igual que sucede con la comparación de las puntuaciones de los criterios en cada parte de la prueba, puede observarse que los porcentajes de aprobados son más altos en la primera parte de la prueba que en la segunda parte en todos los criterios analizados y, por tanto, también en el total. Si bien esa diferencia resulta significativa, a nivel 0.05, solo en los criterios de *alcance* y *corrección gramatical*, como puede verse también en la tabla A.VII.2., del anexo VII. Finalmente, como era de esperar dada la ponderación realizada, la puntuación total se acerca más a la de la segunda parte de la prueba que a la de la primera parte.

II.1.2 Dimensionalidad de los criterios de calificación

Con el fin de estudiar la relación entre los cinco criterios contemplados en la prueba de expresión oral en Lengua Inglesa, en primer lugar se ha realizado una comparación por pares de criterios resultando que las puntuaciones difieren significativamente en todos los casos, como se muestra en la tabla A.VIII.a y en el gráfico A.VIII.1 del anexo VIII.

De manera que calificar independientemente los cinco criterios y posteriormente calcular su promedio, no solo resulta útil en cuanto que contribuye a facilitar el trabajo de los examinadores en el proceso de evaluación, sino que además es conveniente dadas las diferencias observadas.

Por otra parte, los cinco criterios contemplados en la prueba están altamente correlacionados, tanto dentro de cada parte de la prueba como en lo que respecta a la puntuación total ponderada. El anexo VIII recoge, en la tabla A.VIII.b, las correlaciones entre los cinco criterios evaluados. Todas las correlaciones obtenidas son superiores a .960, lo que indica que la consistencia de la escala es muy alta -a pesar de tratarse de una prueba muy corta-, dada la fuerte correlación lineal que existe entre cada pareja de criterios y, sobre todo, entre cada criterio y la puntuación total, ya sea de cada parte o de la calificación total ponderada.

En cuanto al peso que la puntuación de cada uno de los criterios tiene en la calificación final, la tabla A.VIII.c del anexo VIII muestra la matriz factorial en la que se reflejan los pesos de los distintos aspectos evaluados en relación al factor subyacente a la calificación en expresión oral. En términos generales, los pesos con los que los cinco criterios contribuyen al factor son similares, a pesar de que puede observarse una ligerísima menor contribución del criterio *corrección gramatical*.

Este resultado supone, dada la similitud de los pesos con que los cinco criterios contribuyen al factor común subyacente a los mismos, ver tabla A.VIII.c, que se puede utilizar la media aritmética de las puntuaciones de los cinco criterios para evaluar la expresión oral en Lengua Inglesa.

II.2 Relación del nivel de competencia con variables clave

El diseño de la prueba piloto considera dos grupos de variables que presumiblemente pueden tener influencia sobre el nivel de competencia de los alumnos en expresión oral en Lengua Inglesa: las que corresponden al contexto familiar del alumno y las relativas al contexto, escolar y extraescolar, en el que se ha desarrollado su aprendizaje.

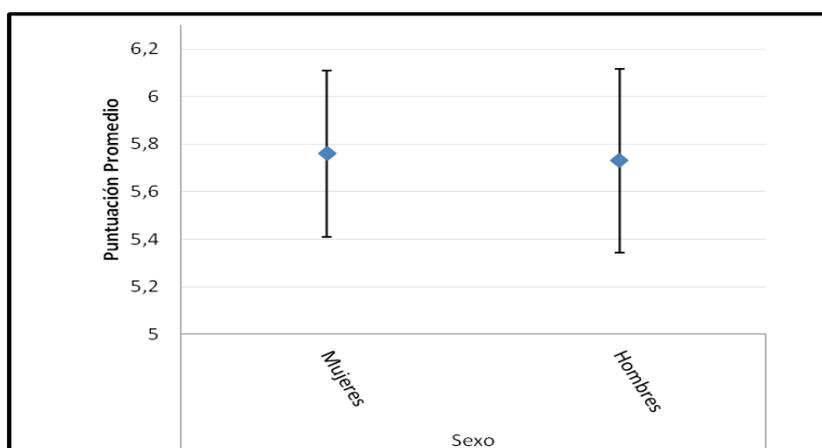
II.2.1 Características del alumno

II.2.1.a Sexo

No se encontraron diferencias estadísticamente significativas en el nivel de expresión oral en Lengua Inglesa entre alumnas y alumnos en la puntuación total ni en cada uno de los cinco criterios de evaluación por separado, como se puede ver en la Tabla A.IX.1 del anexo IX que recoge, además, los estadísticos básicos de las puntuaciones de cada criterio y la puntuación total para alumnas y alumnos.

El gráfico II.2.1 muestra los intervalos de confianza al 95% para la media de las puntuaciones totales de alumnas y alumnos, mientras que en el gráfico A.IX.1, del anexo IX, se representa esta misma información por criterios de calificación, concluyéndose que en ningún criterio existen diferencias significativas entre las puntuaciones obtenidas por alumnos y alumnas

Gráfico II.2.1.a Intervalos de confianza al 95% para la media de la puntuación total de alumnos y alumnas.

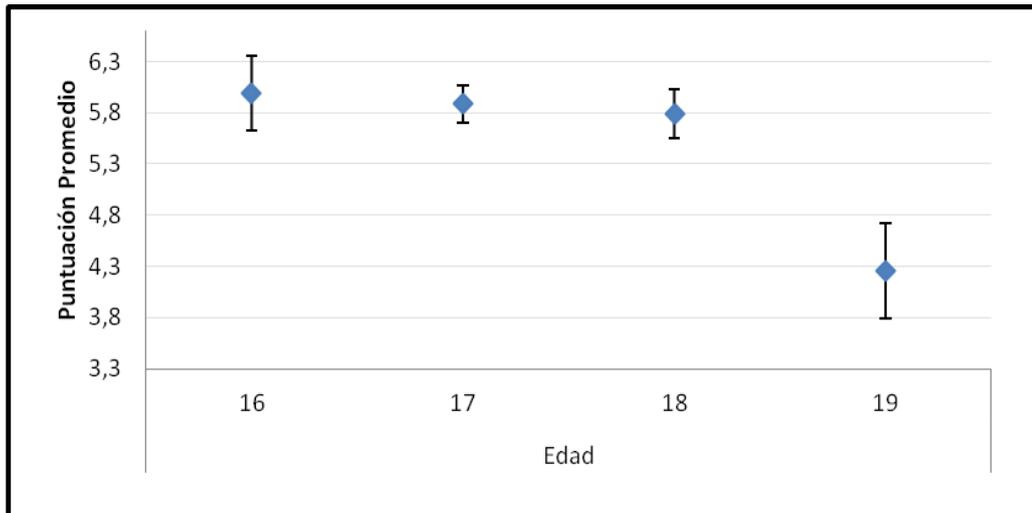


II.2.1.b Edad y repetición de curso

En el análisis general de los datos se encontraron diferencias significativas de rendimiento entre al menos dos de los tramos de edad considerados, tanto en la calificación global como en la de cada uno de los criterios evaluados en la prueba. Los resultados se pueden ver en la tabla A.IX.2 del anexo IX.

En el gráfico II.2.1.b, en el que se representan los intervalos de confianza para la media de las calificaciones totales obtenidas por los estudiantes según las categorías de edad consideradas, se observa que los alumnos con 19 a más años obtienen una calificación significativamente más baja que el resto de los estudiantes. Esa diferencia, no solo se observa en la puntuación total sino también en cada uno de los criterios contemplados en la prueba, tal como se pone de manifiesto en el gráfico A.IX.2 del anexo IX.

Gráfico II.2.1.b. Intervalos de confianza al 95% para la media de la puntuación total según las edades consideradas.



La cuantificación de las diferencias observadas puede verse en la tabla A.IX.3 del anexo IX donde se presentan las diferencias por pares, mediante la comparación de cada categoría con todas las demás. No se encontraron diferencias significativas entre los alumnos de 16, 17 y 18 años mientras que sí se encontraron diferencias significativas entre los alumnos de 19 años o más y las otras tres categorías.

Dado que los alumnos de 19 años o más que cursan 2º curso de bachillerato necesariamente son repetidores y, en buena proporción, repetidores de 2 cursos o más, edad y repetición están estrechamente asociadas. La Tabla II.2.1.c muestra la distribución de los alumnos en función de las cuatro categorías de edad -16, 17, 18 y 19 años o más- y las tres categorías de repetición de curso -no repetidores, repetidores de un curso, repetidores de dos o más cursos-. Como es evidente, los datos señalan alta correlación entre la edad y la repetición de curso.

Tabla II.2.1.c. Frecuencia de alumnos por categorías de edad y de repetición de curso

		Repetición de curso			Total
		No repetidores	Repetidores de 1 curso	Repetidores de 2 o más cursos	
Edad	16 años	96	0	0	96
	17 años	547	25	0	572
	18 años	375	63	3	441
	19 años o más	0	40	45	85
Total		1.018	128	48	1.194

El análisis de los datos relativos a la repetición de curso mostró que hay diferencias estadísticamente significativas entre los resultados de los alumnos en función de si habían repetido curso o no, como puede observarse en la tabla A.IX.4 del anexo IX.

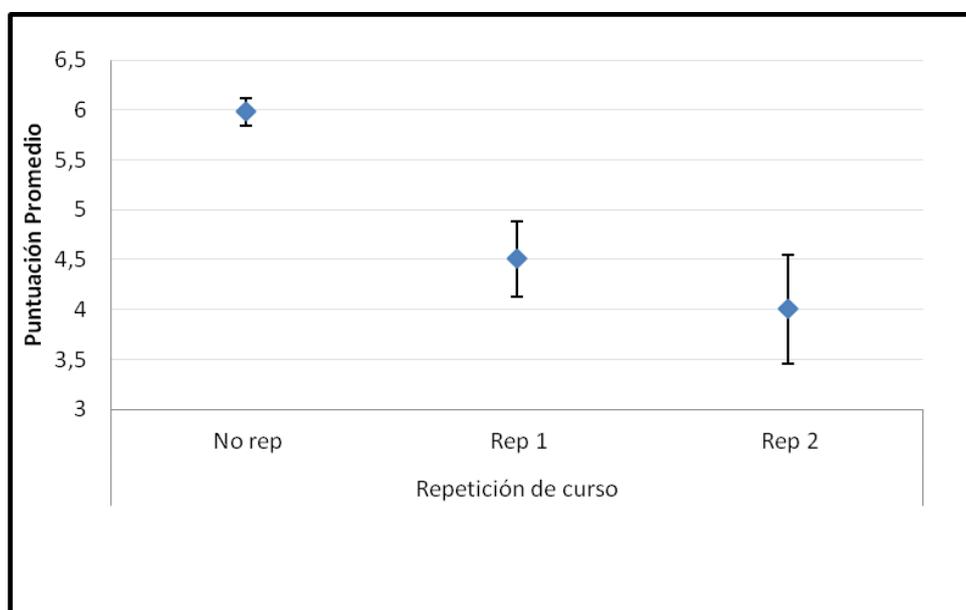
El análisis de la comparación por pares de categorías mostró que hay diferencias significativas entre los alumnos que no han repetido y los repetidores de un curso o de dos cursos o más mientras que no se encontraron diferencias significativas entre los alumnos que han repetido un curso y los que han repetido

dos o más, tal como puede observarse en la tabla A.IX.5 del anexo IX. Esta situación puede verse en el gráfico II.2.2.b.

En conclusión, los resultados en expresión oral están relacionados con la no repetición o repetición de curso, siendo poco relevante el que la repetición se haya producido en uno o más cursos, aunque se observa la tendencia de que los repetidores de dos cursos o más tiene un rendimiento menor que los repetidores de un curso.

Además, la diferencia en las calificaciones obtenidas entre los alumnos que no han repetido curso y el resto de estudiantes se produce no solo en la puntuación total sino también en cada uno de los criterios evaluados, como se puede observar en el gráfico A.IX.3 y su cuantificación en la tabla A.IX.5.

Gráfico II.2.2.b. Intervalos de confianza al 95% para la media de la puntuación total según las categorías de repetición de curso.



II.2.2 Contexto del alumno

II.2.2.a Nivel socioeconómico y cultural de las familias de los estudiantes –ISEC–

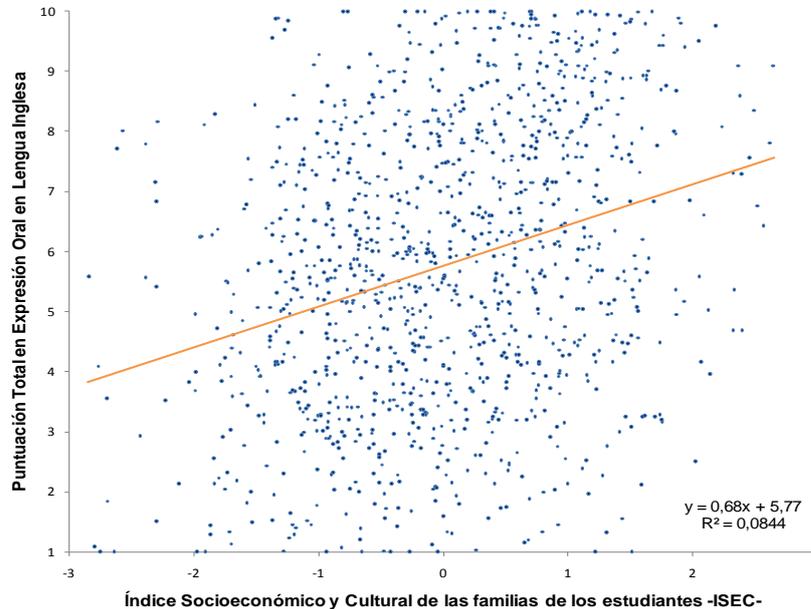
Para el análisis de la relación del nivel socioeconómico y cultural de las familias de los estudiantes con los resultados de competencia en expresión oral en inglés se estimó un índice de nivel socioeconómico y cultural –ISEC– a partir de los datos recopilados mediante el cuestionario de entorno de los alumnos. El ISEC se calculó a partir de tres subíndices de las familias de los estudiantes: un subíndice de nivel profesional –ISECProf–, un subíndice de nivel cultural –ISECCult– y un subíndice de nivel económico –ISECEco–. En el Anexo X se incluyen el procedimiento y detalles técnicos relativos al cálculo del ISEC.

Una vez estimado el ISEC, se transformó la puntuación factorial obtenida a una escala de media cero y desviación típica la unidad.

El coeficiente de correlación entre el ISEC y la puntuación total en expresión oral resulta $R=0.291$, $p \leq 0.000$, lo que implica que la puntuación en expresión oral está determinada en aproximadamente un 8%

por el nivel socioeconómico y cultural de las familias de los estudiantes, $R^2=0.084$. El gráfico II.2.2.a presenta el diagrama de dispersión de la relación entre el Índice de nivel Socioeconómico y Cultural y las puntuaciones en expresión oral en Lengua Inglesa.

Gráfico II.2.2.a: Diagrama de dispersión de la relación entre el Índice de nivel Socioeconómico y Cultural y las puntuaciones en expresión oral en Lengua Inglesa



Para valorar este resultado es conveniente compararlo con el grado en que el nivel socioeconómico y cultural de las familias de los estudiantes –ISEC– se relaciona con otros resultados educativos. En primer lugar, el ISEC muestra una correlación de .161 con la nota media de Bachillerato obtenida por los alumnos y .149 con la nota media obtenida en las PAU, lo que supone un grado de determinación de un 3% y un 2% respectivamente. En segundo lugar, la calificación en Lengua Inglesa en 2º curso de Bachillerato y en la prueba actual de Lengua Inglesa de las PAU muestran respectivamente unas correlaciones con el ISEC de .237 y .208, lo que supone que un 6% y un 4% de estas calificaciones se explican por el nivel socioeconómico y cultural de sus familias.

Estos resultados en su conjunto muestran que el área de la Lengua Inglesa se ve afectada en mayor medida por un efecto de clase social que la nota media de Bachillerato y la calificación obtenida por los alumnos en la PAU -en la que está ya incluida una prueba de Lengua Inglesa escrita-. En consecuencia, la incorporación de la prueba de expresión oral en Lengua Inglesa en las PAU, que muestra una mayor relación con el ISEC que la prueba escrita y que la calificación del centro en Lengua Inglesa, incrementaría aún más el efecto de las diferencias de clase socioeconómica y cultural de los estudiantes sobre las calificaciones finales que obtendrían en las PAU.

Lógicamente, el grado de relación entre el ISEC de las familias de los estudiantes y el rendimiento académico es menor al final de la educación secundaria post-obligatoria –Bachillerato- que al final de la educación obligatoria, debido a la selección de alumnos que se produce al final de la ESO, que está relacionada en gran medida con el nivel socioeconómico y cultural de las familias, lo que reduce el rango de variabilidad tanto del ISEC como del rendimiento académico. Esta selección al final de la educación obligatoria afecta en mayor medida a los estudiantes de niveles de ISEC y rendimiento más bajos.

Por ello, la relación ISEC-rendimiento académico es mayor en los estudios PISA situándose el grado en que el nivel socioeconómico y cultural explica las diferencias de rendimiento en torno al 14% para el

conjunto de los estudiantes españoles y entre el 14% y el 20% para el conjunto de los estudiantes de los países de la OCDE. En concreto, el porcentaje de varianza de los resultados explicada por el ISEC en lectura fue del 14% para el conjunto de los países de la OCDE en los años 2000 y 2009, y también para España se mantuvo constante en el 14% en ambos años, -OECD 2001, Tabla 8.1; OECD 2010, Tabla II.1.2-. En el año 2003, y en referencia a las matemáticas, estos porcentajes fueron del 14% para España y del 20% para la OCDE -OECD 2004, Tabla 4.3a-. En el área de Ciencias en el año 2006 los porcentajes fueron los mismos para los países de la OCDE y para España e iguales al 14% -OECD 2007, Tabla 4.4a-.

De modo similar, si se toma como punto de comparación la evaluación general de diagnóstico en Educación Secundaria -INEE, 2011- Las relaciones entre el nivel socioeconómico y cultural y el rendimiento académico son similares a las obtenidas por los estudios PISA. En lectura el 15% de la variación de los resultados se explica por el ISEC ($r=.383$); en el área del mundo físico esta capacidad explicativa es del 13% ($r=.363$); y en matemáticas y en el área social y ciudadana es del 10% ($r=.311$ y $r=.314$). En todos los casos, el grado de capacidad explicativa del ISEC sobre el rendimiento es superior al 8 por ciento observado en la expresión oral en Lengua Inglesa, probablemente atribuible al hecho de que el conjunto de los alumnos de Bachillerato es una selección de la población de los alumnos matriculados en los últimos cursos de la Educación Secundaria.

Los estudios PISA también han puesto de manifiesto la importante influencia sobre el rendimiento, no solamente del ISEC individual de los estudiantes sino del ISEC promedio de los centros educativos a los que asisten, de tal modo que la confluencia en los centros educativos de estudiantes provenientes de las familias con mayor nivel socioeconómico promueve un mayor rendimiento académico individual de los alumnos^{iv}.

El análisis de la influencia del ISEC promedio de los centros educativos sobre el rendimiento en Lengua Inglesa se mostró significativo siendo la correlación entre el ISEC promedio de los centros y las puntuaciones individuales de los alumnos en expresión oral igual a $.281$ ($p\leq.000$) lo que supone una capacidad explicativa de las puntuaciones de los alumnos del 8% a partir del ISEC promedio de sus centros.

En su conjunto, la capacidad explicativa del ISEC sobre el rendimiento en expresión oral, teniendo en cuenta simultáneamente el efecto del ISEC individual y el efecto del ISEC promedio de los centros educativos, es igual a un 11 por ciento ($R=.331$). Este efecto puede explicarse por la capacidad que los alumnos de familias con mayor poder económico tienen para realizar estancias en países de habla inglesa, ya sea en períodos cortos como las vacaciones de verano o en estancias de un curso completo.

II.2.2.b Tamaño de la población

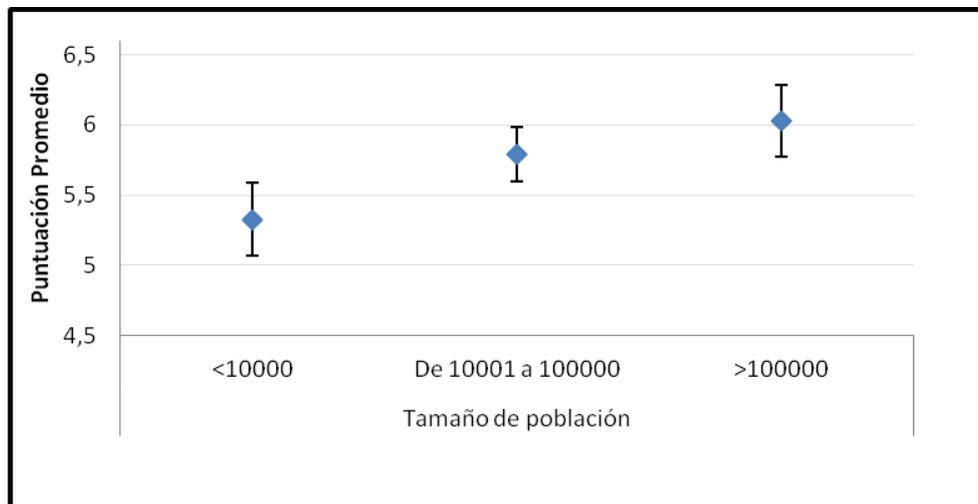
Las diferencias en los resultados de la puntuación en expresión oral en función del tamaño de la población en la que están situados los centros educativos al que asisten los alumnos, categorizado en tres niveles -hasta 10.000 habitantes, de 10.001 a 100.000 habitantes y más de 100.001 habitantes-, resultaron ser estadísticamente significativas.

La tabla A.XI7.1, del anexo XI, muestra los resultados básicos en los que se observa que el mejor rendimiento corresponde a los alumnos cuyos centros están situados en poblaciones entre 10.001 y 100.000 habitantes, siendo el grupo de alumnos situados en el grupo de hasta 10.000 habitantes el que logra una menor puntuación promedio. En la tabla A.XI.1 se recogen los datos relativos a los estadísticos básicos de las calificaciones obtenidas por los estudiantes por tamaño de población y por criterios de evaluación.

En la comparación por pares entre los tres tamaños considerados de población resultaron significativas las diferencias observadas entre el grupo de poblaciones pequeñas -hasta 10.000 habitantes- con los otros dos grupos, mientras que no se observaron diferencias significativas entre los grupos de

poblaciones medianas –de 10001 a 100000 habitantes- y grandes –más de 100000 habitantes-. Los resultados se muestran en la tabla A.XI.2 del anexo XI.

Gráfico II.2.2.b Intervalos de confianza al 95% para la media de las calificaciones de los estudiantes en la prueba de expresión oral en Lengua Inglesa, por tamaño de la población.



Finalmente, en el Gráfico II.2.2.b se muestran los promedios de las calificaciones totales de los estudiantes en la prueba de expresión oral por tamaño de la población, y el intervalo que contiene la media verdadera con un 95% de confianza. Estos resultados indican que el nivel de expresión oral en Lengua Inglesa es significativamente más bajo en los alumnos que asisten a centros educativos en poblaciones más pequeñas. Lo mismo sucede en los criterios evaluados en la prueba excepto en el de corrección gramatical en el que la significación no es tan destacada.

II.2.3 Contexto del aprendizaje de la Lengua Inglesa

II.2.3.a En el centro escolar

a.1 Duración y frecuencia de las clases

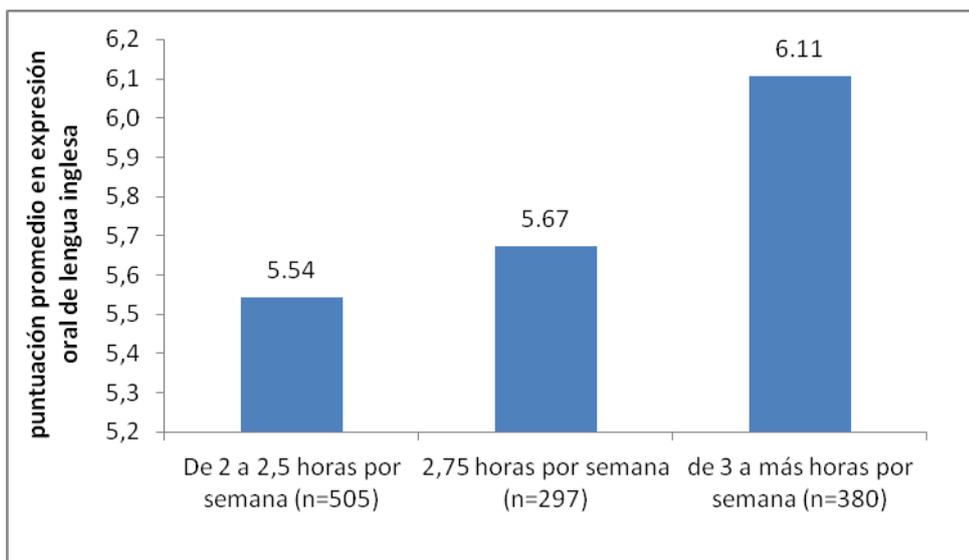
Para el análisis de los resultados obtenidos en función del número de horas semanales de clase recibidas por los estudiantes, teniendo en cuenta la distribución de frecuencias de las respuestas recibidas, se ha dividido la variable en tres categorías: estudiantes que dicen recibir entre *dos y dos horas y media* de inglés a la semana, los que afirman recibir *dos horas y cuarenta y cinco minutos* y, finalmente, los que reciben *tres o más horas* semanales de inglés.

En el gráfico II.2.3.1 se recogen las puntuaciones medias obtenidas en la prueba de expresión oral según las tres categorías consideradas. La puntuación más alta la recibe, claramente, el grupo de los que reciben *tres o más horas* semanales de inglés.

Con el fin de determinar si las diferencias observadas en la calificación de los tres grupos son o no significativas desde el punto de vista estadístico, se realiza un análisis de la varianza que muestra que existen diferencias significativas, con nivel superior al 99%, entre los grupos considerados en función del número de horas de clase de inglés recibidas por semana. Los resultados se muestran en la tabla A.XII.1 del anexo XII.

Mientras que en la comparación por pares de categorías se detecta que los alumnos que cursan *tres o más horas semanales* obtienen puntuaciones significativamente más altas, al 95%, que aquellos que cursan *entre dos y dos horas y media* y también de los que cursan *dos horas y cuarenta y cinco minutos semanales*. Sin embargo, no se aprecian diferencias estadísticamente significativas entre las dos categorías de alumnos que reciben menos de tres horas semanales de. La tabla A.XII.2 del anexo XII recoge los resultados de la comparación por pares entre las tres categorías.

Grafico II.2.3.1: Puntuaciones medias en función de las horas semanales de inglés recibidas por los estudiantes.



a.2 Uso de la lengua inglesa hablada en clase

Para evaluar la influencia que el uso habitual del inglés hablado en clase tiene en los resultados de la prueba oral, se llevó a cabo un análisis factorial en el que se agruparon aspectos tales como *si el profesor habla en inglés habitualmente en clase, si los alumnos hablan habitualmente en inglés durante la clase, si el profesor exige que sus alumnos realicen exposiciones y debates en inglés y si el profesor anima a que sus alumnos se comuniquen en inglés de forma activa*, descartándose el hecho de que *el profesor utilice libros música y películas en inglés*. Los detalles pueden verse en el Anexo XIII.

Mediante un análisis de correlación entre el índice construido, *uso del inglés hablado en clase*, con los aspectos mencionados y los resultados obtenidos por los alumnos en la prueba oral de Lengua Inglesa, se observa que existe correlación positiva, $r=0.097$ con $p \leq .01$, si bien ésta es muy pequeña, por lo que su influencia en los resultados es más bien modesta, quedando por debajo del 1%.

a.3 Recursos pedagógicos

El mismo procedimiento, que con el uso habitual del inglés hablado, se llevó a cabo con los recursos pedagógicos. De esta manera, se construye un índice en el que se incluyen aspectos como *la frecuencia de uso de medios audiovisuales (CDs, vídeos, DVD, etc.), revistas, periódicos, cómics o textos de canciones en inglés, internet, software específico y laboratorio de idiomas*; descartándose para el índice *el uso de libros de texto y libros de lectura en inglés*. Los detalles pueden verse en el Anexo XIII.

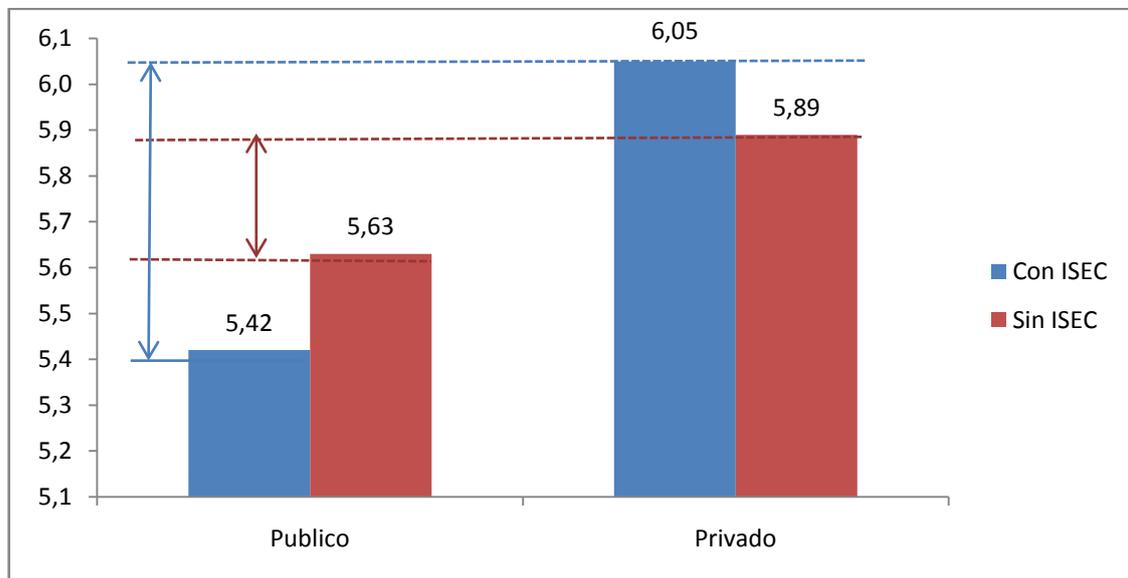
Al igual que sucede con el uso de la lengua hablada en clase, la correlación entre este índice y los resultados en la prueba oral es positiva, $r=.111$ con $p \leq .000$, por lo que su influencia en los resultados es también muy modesta, alrededor del 1.2%.

a.4 Titularidad del centro

Se ha evaluado el impacto de la titularidad del centro del alumno, público o privado, sobre su puntuación en la prueba oral de inglés. La tabla A.XIV.1 del anexo XIV contiene los estadísticos básicos de la *puntuación total* según la titularidad pública o privada del centro. En el gráfico II.2.3.2 se presentan las puntuaciones medias en función de la titularidad tanto contando como no con la influencia del Índice Socio-Económico (ISEC).

Los alumnos de los institutos públicos obtienen, en promedio, puntuaciones significativamente menores que los alumnos de los centros privados cuando no se tiene en cuenta la influencia del ISEC, como queda reflejado en la tabla A.XIV.2 del anexo XIV.

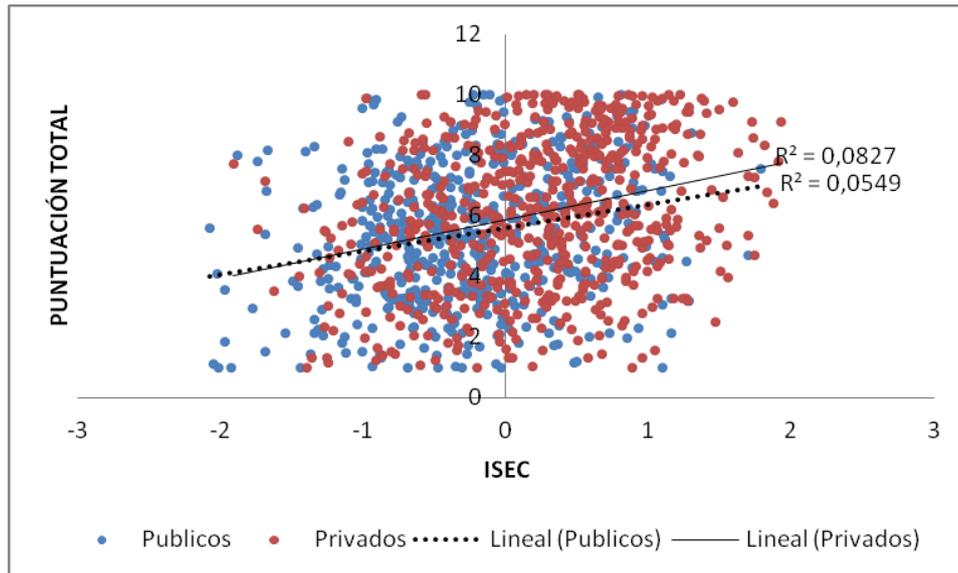
Gráfico II.2.3.2: Puntuaciones medias, sin corregir y corregidas por el ISEC, en función de la titularidad del centro.



Sin embargo, en el gráfico II.2.3.3 se observa que el ISEC influye de manera diferente en los dos tipos de centro considerados. Las tendencias lineales de las dos nubes de puntos, una para los centros públicos y otra para los privados, muestran que la condición socioeconómica del estudiante tiene más influencia sobre su puntuación en la prueba oral de inglés si el alumno cursa sus estudios en un centro privado.

Por esta razón, si corregimos el efecto de la titularidad del centro por el índice socioeconómico, observamos que las diferencias de puntuación entre centros públicos y privados ahora no son significativas, como se puede observar en la tabla A.XIV.3 del anexo XIV, donde muestra además, mediante el análisis de la covarianza, que el modelo corregido por el ISEC es significativo. El ISEC también tiene un efecto significativo sobre la puntuación media, tal como se refleja en la misma tabla.

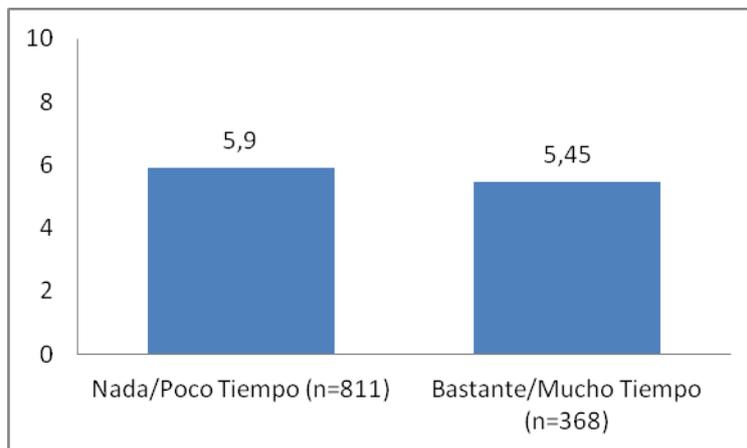
Grafico II.2.3.3: Coeficientes de determinación en función de la titularidad del centro.



II.2.3.b Tiempo dedicado al estudio y a tareas fuera del horario escolar

Para agrupar a los estudiantes en base a la auto-percepción que estos tienen acerca del tiempo dedicado al estudio del inglés y a las tareas fuera del horario de clase, Se ha utilizado una variable dicotómica. Las categorías utilizadas para esta variable son *Nada/Poco Tiempo* y *Bastante/Mucho Tiempo*. En el gráfico II.2.3.5 se presentan las puntuaciones medias por categoría.

Grafico II.2.3.5: Puntuaciones medias según la auto-percepción del alumno acerca del tiempo dedicado al estudio y a las tareas de inglés fuera del horario de clase.



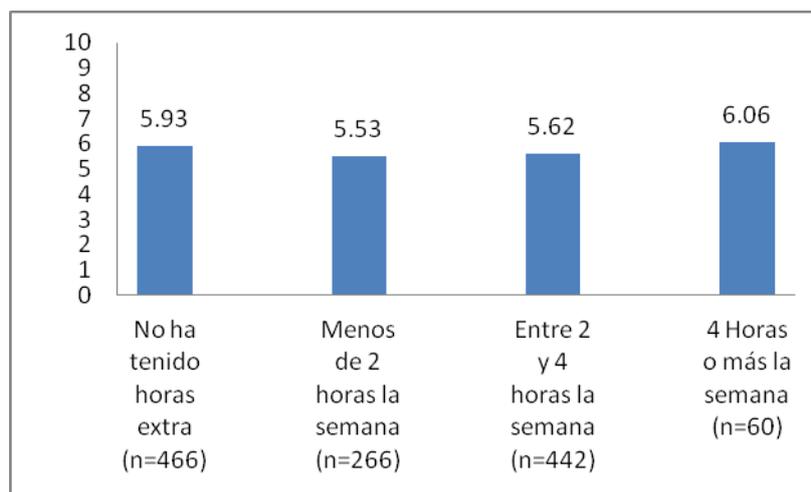
Los resultados obtenidos reflejan que los estudiantes que dedican menos tiempo al estudio y a la realización de tareas fuera del horario de clase obtienen puntuaciones significativamente mejores que aquellos que le dedican más tiempo, tal como se muestra en la tabla A.XV.1 del anexo XV.

Este resultado, en principio sorprendente, puede tal vez explicarse por el hecho de que los alumnos que menos dominio tienen del inglés son los que, para alcanzar el nivel adecuado, dedican más tiempo a las tareas fuera del horario de clase.

II.2.3.c Clases adicionales de inglés y su frecuencia

Para medir la influencia de clases adicionales inglés, horario extraescolar, en los resultados obtenidos por los alumnos en la prueba oral, se ha construido un índice para clasificar el número de horas extras de inglés recibidas por los alumnos. El índice toma los valores 0, *ninguna hora semanal*, 1, *menos de dos horas semanales*, 2, *entre dos y cuatro horas semanales*, y 3, *cuatro o más horas semanales*. Las puntuaciones medias obtenidas por los estudiantes incluidos en cada categoría se muestran en el gráfico II.2.3.6.

Grafico II.2.3.6: Puntuaciones en función del número de horas extra de inglés.



No obstante las diferencias observadas en la puntuación media de las distintas categorías del índice, aquellas no son significativas según se desprende del análisis de la varianza realizado, como se puede observar en la tabla A.XV.2, del anexo XV.

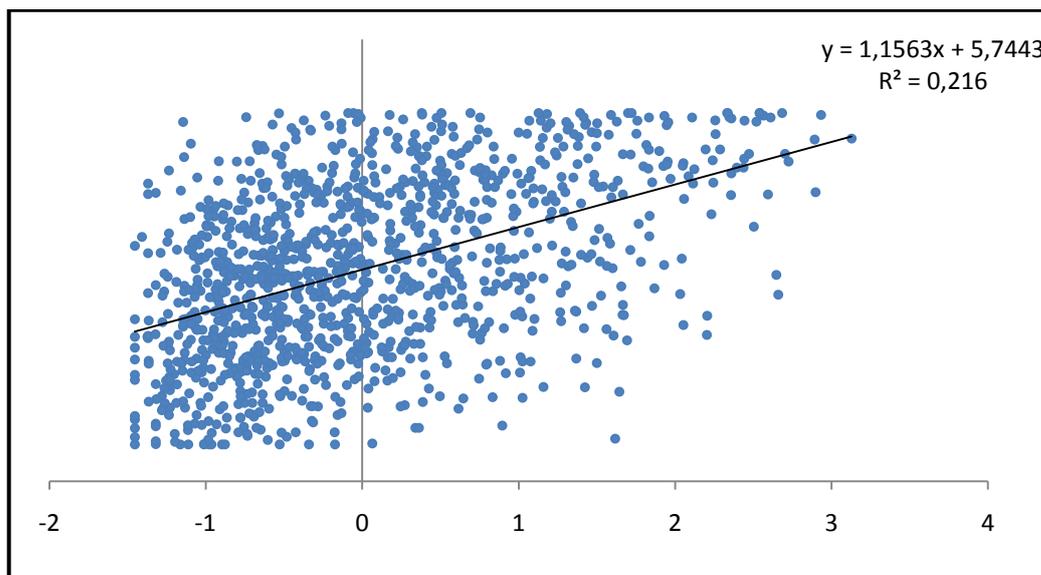
Las comparaciones por parejas, utilizando los test de Scheffé, Tukey o Bonferroni, no presentan diferencias significativas. Sin embargo, si utilizamos el test estándar LSD, aparecen diferencias significativas entre el grupo de los que no han recibido clases de inglés y los grupos de menos de dos horas y entre dos y cuatro horas. Los resultados se pueden ver en la tabla A.XV.3 del anexo XV.

II.2.3.d Grado de exposición a la Lengua Inglesa

Para medir el grado de exposición del estudiante a la Lengua Inglesa fuera del contexto escolar se han utilizado dos variables. Una -cuestión P22- mide la frecuencia con la que el estudiante utiliza el inglés como medio de comunicación bien por escrito -*chat, email, cartas, etc.*-, bien oralmente -*con amigos o familia, con turistas, etc.*- o por internet -*juegos online, por ejemplo*-.

La otra variable -cuestión P23- mide la frecuencia con la que el estudiante ha estado en contacto con el inglés a través de los medios de comunicación: escuchando canciones, viendo películas en inglés con subtítulos y sin subtítulos, lo mismo con programas de televisión, utilizando juegos de ordenador, leyendo libros en inglés, leyendo revistas o cómics en inglés y visitando sitios web en inglés.

Grafico II.2.3.6: Nube de puntos y recta de regresión de la calificación en la prueba oral respecto al *grado de exposición extraescolar al inglés*.



Con estas dos variables se ha construido un único índice como se expone con detalle en el Anexo XVI. Este índice, que mide el *grado de exposición extraescolar* del alumno a la Lengua Inglesa, tiene correlación positiva, $r=0.465$ y $p \leq 0.000$, con la puntuación total obtenida en la prueba oral, de forma que su influencia en la calificación se puede estimar en un 22% aproximadamente, véase el gráfico II.2.3.6. Este resultado parece razonable, ya que a mayor exposición al inglés más soltura tendrá el alumno a la hora de comunicarse en lengua inglesa.

II.2.3.e Viajes de estudio e intercambios en países de habla inglesa y su duración.

Para analizar la influencia que tienen las estancias en países de habla inglesa y su duración en los resultados obtenidos por los alumnos en la prueba se han construido dos índices, ambos con valores de 0 a 15, que cuantifican desde no haber estado nunca en un país de habla inglesa hasta estancias de larga duración una o más veces.

El primer índice, *viaje tiempo solo*, recoge la información de estancias individuales, de estudios o de intercambio, y tiene una correlación lineal positiva y significativa $-r=0.3449-$, aunque relativamente débil, con la puntuación total obtenida en la prueba. El modelo ajustado, que explica un 11.89% de la variabilidad de la puntuación total, se puede ver en la tabla A.XVI.4, del anexo XVI y el análisis de la varianza junto con el test de falta de ajuste del modelo se muestran en la tabla A.XVI.5.

Finalmente, el anexo XVI incluye también el gráfico de la regresión lineal junto con los intervalos de confianza al 95% para las predicciones individuales y medias. En este gráfico puede observarse la enorme variabilidad que tienen las predicciones individuales, lo que significa que el modelo no es adecuado para realizar predicciones acerca de la puntuación total que obtendría un estudiante en función del tiempo que haya empleado en estancias individuales en países de habla inglesa.

El segundo índice, *viajes tiempo en familia*, se refiere a estancias en países de habla inglesa con la propia familia del alumno, y también tiene correlación positiva y significativa, aunque también relativamente débil $-r=0.3749-$ con la calificación total obtenida por el alumno en la prueba oral. En este caso, el modelo ajustado explica el 14.05% de la variabilidad de la puntuación total y se puede ver en la

tabla A.XVI.6, mientras que el análisis de la varianza junto con el test de la falta de ajuste del modelo se muestran en la tabla A.XVI.7.

También en este caso hemos incluido en el anexo XVI el gráfico de la regresión lineal de la *puntuación total* sobre el índice viaje *tiempo en familia*, que incluye los intervalos de confianza al 95% para las predicciones individuales y medias. En este caso también se puede observar la gran variabilidad que tienen las predicciones individuales, de manera que el modelo no es adecuado para efectuar predicciones acerca de la calificación total que obtendría un estudiante en función del tiempo que haya estado con su familia en países de habla inglesa.

II.3 Relación con otras calificaciones

Con objeto de valorar la información adicional que aporta la evaluación de la expresión oral a otras informaciones disponibles sobre el nivel individual de competencia de los estudiantes en Lengua Inglesa y de rendimiento académico general, se recogieron datos de las calificaciones de los alumnos en el Bachillerato: su calificación en Lengua Inglesa al final de 2º curso de Bachillerato y su nota media de Bachillerato. Además, también sobre los resultados obtenidos en su PAU: calificación en Lengua Inglesa - escrita- y nota media de la Fase General.

Lógicamente, cuanto más alto sea el nivel de correlación entre la calificación en expresión oral y el resto de informaciones sobre el rendimiento de los alumnos, menor es la cantidad de información adicional que la evaluación de la expresión oral aporta; e inversamente cuanto más bajo sea el nivel de dicha correlación, mayor es la información adicional que proporciona la evaluación de la expresión oral.

La Tabla II.3.1 muestra los estadísticos descriptivos básicos de las calificaciones antes mencionadas que se han relacionado en este estudio con el nivel de expresión oral. Se observa que las calificaciones en inglés en la prueba escrita de las PAU son las más altas (7.40), seguidas de las calificaciones en inglés en segundo de Bachillerato (6.78) siendo la diferencia de medias entre ellas significativa ($F=210,21$; $g.l.=1, 740$; $p \leq .000$). Esto parece indicar que los profesores de los centros son algo más exigentes en sus calificaciones que los profesores que corrigen la prueba escrita de la PAU.

Además, como se ve en la tabla II.3.1, la variabilidad de las puntuaciones es mayor en las calificaciones que se obtienen en el centro de estudios que en la PAU, lo que parece indicar una mayor capacidad de discriminación entre los alumnos con más nivel de competencia y los de menor nivel en el centro escolar que en la PAU.

Tabla II.3.1 Estadísticos descriptivos básicos de diferentes calificaciones en Lengua Inglesa

Estadísticos básicos de las calificaciones de los estudiantes				
	Media	Error Típico	Desviación Típica	N
Puntuación en expresión oral	5,74	0,07	2,33	1.194
Calificación Inglés en 2º de Bachillerato	6,78	0,06	1,94	930
Calificación prueba escrita de inglés en las PAU	7,40	0,04	1,22	821
Nota Media de Bachillerato	6,39	0,07	1,95	772
Nota Media Fase General en las PAU	6,55	0,05	1,31	786

Por otro lado, debe observarse que las calificaciones obtenidas en la evaluación de la expresión oral son significativamente más bajas que el resto de calificaciones, como se observa en la tabla II.3.2, en la que

se muestran las comparaciones por pares de calificaciones. Esta diferencia puede ser un reflejo de que el nivel de expresión oral promedio de los estudiantes se encuentra por debajo de su nivel de conocimientos y destrezas escritas en Lengua Inglesa en comparación con lo esperado. En consecuencia, la inclusión de la evaluación de la expresión oral en la PAU, con el nivel de exigencia B1 utilizado en esta prueba piloto, reduciría en cierto grado el promedio de las notas de la fase general de las PAU, dependiendo este grado del peso que se asignase a esta prueba en el conjunto de las PAU.

Tabla II.3.2. Correlaciones entre la calificación total ponderada en expresión oral en Lengua Inglesa en la prueba piloto de las PAU y otras calificaciones

Criterio 1	Criterio 2	Diferencia de medias	E.T.	Límite inferior	Límite superior
Puntuación directa total	Calificación inglés centro	-0,969***	0,059	-1,084	-0,854
	Nota media bachillerato	-1,312***	0,073	-1,454	-1,170
	Calificación inglés PAU	-0,238***	0,066	-0,368	-0,108
	Nota media PAU	-0,397***	0,070	-0,535	-0,259

*** significativa por debajo del 0.001

La Tabla II.3.3 muestra las correlaciones entre la calificación en expresión oral en Lengua Inglesa en la prueba piloto y las calificaciones obtenidas por los alumnos en inglés en segundo de Bachillerato y en la prueba de inglés escrito en las PAU, así como con la nota media obtenida en el conjunto del Bachillerato y la nota media final en las PAU.

Tabla II.3.3. Correlaciones entre la calificación total ponderada en expresión oral en Lengua Inglesa en la prueba piloto de las PAU y otras calificaciones

	Correlaciones y nivel de significación			
	Nota en Inglés 2º de Bachillerato	Calificación en Inglés en la PAU	Nota media en Bachillerato	Nota media Fase General
Puntuación en expresión oral	.676	.637	.447	.513
p≤	.000	.000	.000	.000
R²	.46	.41	.20	.26
R² conjunta	.52		.28	
	.53			

Se observa que la calificación obtenida por los estudiantes en su centro educativo es la variable que predice en mejor medida su nivel de expresión oral, con una correlación igual .676, lo que implica que un 46% de la variabilidad de las puntuaciones en expresión oral se explica por las calificaciones del centro. En segundo lugar, la calificación que obtienen los alumnos en inglés en la prueba de inglés escrito en las PAU es el segundo mejor “predictor” del nivel de expresión oral con una correlación igual a .637 y una capacidad de explicación del 41%. En su conjunto, ambas calificaciones -Inglés en segundo de Bachillerato y calificación en la prueba escrita de las PAU- son capaces de predecir un 52% de la variabilidad de las puntuaciones en expresión oral.

De otra parte, la nota media de Bachillerato y la nota media de la Fase General en las PAU tienen unas correlaciones con la puntuación en expresión oral algo más bajas (.447 y .513, respectivamente) aunque también positivas y significativas, siendo su capacidad de predicción bastante menor (20% y 26% respectivamente). Consideradas conjuntamente, la capacidad de predicción del nivel de expresión oral a partir de estas calificaciones es de un 28%. Finalmente, las cuatro calificaciones consideradas explican el 53% de la variabilidad de la calificación en expresión oral, que es similar a la que explican las dos primeras

variables aisladamente, debido a la dependencia lineal que existe entre las cuatro calificaciones. En la Tabla II.3.3 se presentan las correlaciones existentes entre las cinco calificaciones analizadas.

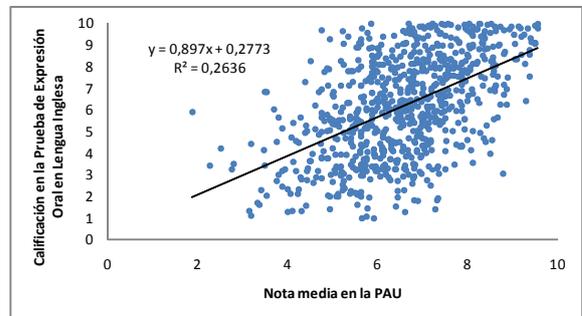
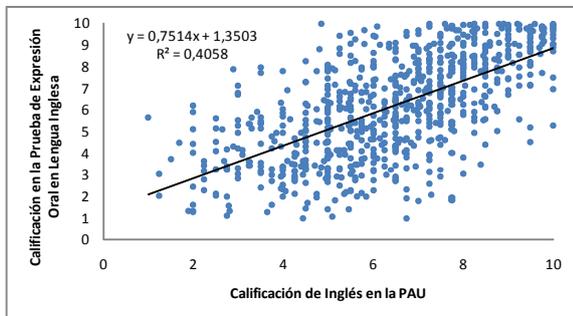
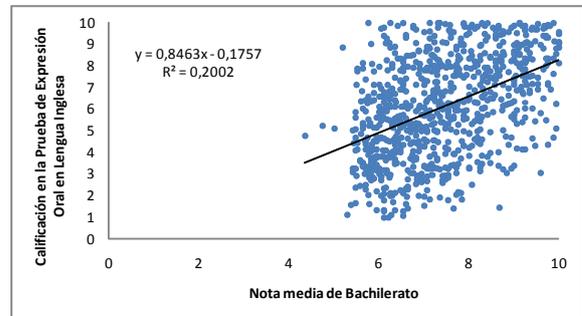
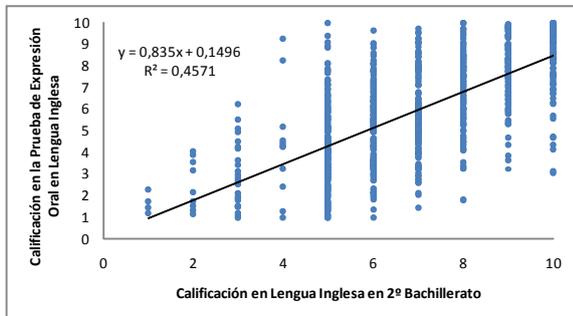
Tabla II.3.3 Correlaciones entre las calificaciones en Lengua Inglesa al final de 2º curso de Bachillerato, en la prueba escrita de Lengua Inglesa en las PAU, la Nota media de Bachillerato y la Nota media de la Fase General de las PAU, y sus correlaciones con la puntuación en expresión oral.

	Puntuación Total en Expresión Oral	Calificación Inglés Centro	Nota Media Bachillerato	Calificación Inglés PAU	Nota Media PAU
Puntuación Total en Expresión Oral	1,000	0,676	0,447	0,637	0,513
Calificación Inglés Centro		1,000	0,741	0,654	0,639
Nota Media Bachillerato			1,000	0,513	0,710
Calificación Inglés PAU				1,000	0,669
Nota Media PAU					1,000

En su conjunto estos resultados parecen indicar que la mayor parte de la información sobre el nivel de expresión oral de los estudiantes está contenida en las calificaciones de inglés que proporcionan los profesores de los centros educativos y en el examen escrito de inglés en la PAU.

En el Gráfico II.3.1 se muestran los diagramas de dispersión de la puntuación total en expresión oral en Lengua Inglesa de la prueba piloto respecto a las calificaciones obtenidas por los alumnos en Lengua Inglesa de 2º de Bachillerato, media del Bachillerato, Inglés en la PAU y nota media en la Fase General de la PAU, junto con la línea de regresión ajustada y el coeficiente de determinación.

Gráfico II.3.1: Dispersión de la Calificación obtenida en la Prueba de Expresión Oral en Lengua Inglesa y el resto de calificaciones consideradas.



II.4 Fiabilidad de las calificaciones de los correctores

Con el objetivo de estudiar la fiabilidad entre las calificaciones de los correctores se generó la puntuación media asignada por cada corrector a cada alumno. Para ello se utilizó la media ponderada, 20% para la primera fase y 80% para la segunda, de las dos fases del ejercicio para cada estímulo y se calculó una media general de la prueba. De esta forma, cada alumno tiene tres puntuaciones en la prueba, correspondientes a los tres correctores del tribunal.

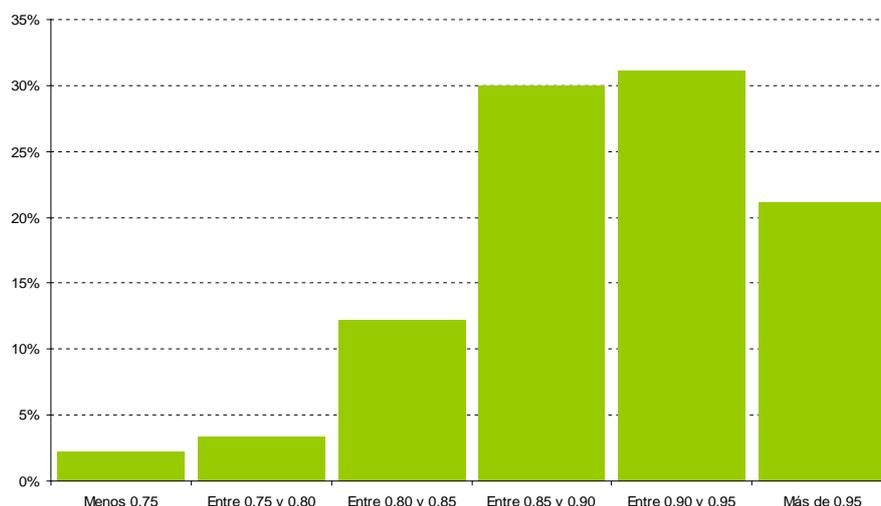
Los análisis se llevaron a cabo dentro de la matriz formada por los alumnos evaluados por el mismo tribunal de evaluadores. De esta forma la base de datos completa se subdivide en 30 matrices correspondientes a los 30 tribunales.

II.4.1. Concordancia entre correctores

Se calculó el coeficiente de correlación de Pearson para cada pareja de las tres puntuaciones asignadas por los tres evaluadores. Los resultados se recogen, en la tabla A.XVII.1 del anexo XVII, organizados por tribunales de tal forma que todos los sujetos fueron examinados por los mismos profesores. Como se puede observar la correlación mínima es de 0.67 y la máxima de 0.99, con una media de las correlaciones de 0.89, lo que indica que las puntuaciones asignadas por los distintos evaluadores mantienen una alta convergencia. Todas las correlaciones son estadísticamente significativas al nivel 0.01 (bilateral).

En el gráfico 4.II.1, se presenta la distribución de frecuencias –en porcentaje– de las correlaciones entre los correctores. Se observa que sólo en un 6% de los casos la correlación entre dos correctores es inferior a 0.80. Por el contrario en el 82% de las parejas la correlación supera el 0.85 y en el 52% de las ocasiones la correlación entre las puntuaciones de los dos correctores está por encima de 0.90.

Gráfico II.4.1. Distribución de frecuencias de las correlaciones entre correctores



Con el fin de estudiar las diferencias entre las puntuaciones medias otorgadas por los evaluadores, se realizó un análisis de la varianza (ANOVA) de medidas repetidas -1 factor intrasujeto- puesto que cada alumno dispone de tres puntuaciones correspondientes a los tres jueces de cada tribunal. Como se puede observar en la tabla A.XVII.2 del anexo XVII, en 13 de los 30 tribunales existen diferencias estadísticamente significativas con un nivel de significación del 5%. Además, en líneas generales, el tamaño del efecto es relativamente alto lo que indica que las diferencias encontradas resultarían relevantes. Por otro lado, la

potencia de cada una de las pruebas es alta lo que indica que con una alta probabilidad se está rechazando la hipótesis nula siendo falsa.

En definitiva, si bien los datos de la tabla A.XVII.1 indican que los correctores tienden a estar de acuerdo cuando se trata de ordenar el desempeño del alumnado, los resultados de la tabla A.XVII.2 señalan que existen más discrepancias a la hora de asignar una nota numérica a dicho desempeño. En otras palabras, se advierte una influencia significativa del efecto del corrector o calificador en casi la mitad de los tribunales estudiados.

Estos resultados parecen indicar que si bien las puntuaciones asignadas por los evaluadores tienen una alta correlación, ello no necesariamente significa que todos operen en la misma escala, ya que se encuentran diferencias entre las medias. En función de las diferencias de medias encontradas parece que, en casi la mitad de los tribunales, los correctores utilizaron niveles de exigencia distintos. A la vista de estos datos parece conveniente actuar en dos direcciones:

- Diseñar un protocolo de corrección más específico y ofrecer entrenamiento previo a los miembros de tribunal podría reducir las diferencias de medias encontradas entre las puntuaciones de los correctores.
- Sobre el número de correctores los datos indican que el tribunal debe estar compuesto como mínimo por dos personas, siendo tres miembros la opción más segura y conservadora. Esta conclusión se basa sólo en datos estadísticos, si bien la decisión sobre el número de miembros de tribunal depende de otros factores de tipo logístico y económico.
- Por ello si la decisión final fuese organizar tribunales con dos personas podría ser recomendable que ambos miembros consensuaran la nota antes que emplear calificaciones independientes. Ello permitiría fijar la media y evitaría los desajustes encontrados en esta parte del análisis.

II.4.2. Relación entre nivel de concordancia de los correctores con otras variables

Para estudiar el efecto que tiene el rol del profesor -examinador o corrector- se realiza un ANOVA de medidas repetidas, pero esta vez se incluye como medida intersujeto la variable de agrupamiento rol del profesor. A la vista de los resultados, que se pueden ver en la tabla A.XVII.3, solo se encuentran diferencias estadísticamente significativas en dos casos siendo además el tamaño del efecto prácticamente nulo y la potencia de la prueba muy baja. En conclusión, no existen diferencias estadísticamente significativas - NC=95%- entre el rol de examinador y el de corrector. Por lo tanto, se podría eliminar la rotación de los miembros del tribunal de tal forma que facilite la logística de aplicación.

Con la intención de esclarecer la interpretación de los resultados se realizaron los análisis con la matriz de datos completa -i.e. sin segmentar la base por tribunales-. La justificación es que al encontrarse altas correlaciones entre los correctores de cada tribunal pueden tomarse como intercambiables y, por tanto, asumir que son tres correctores del mismo tribunal los que evalúan a todos los sujetos. Como se puede observar en la Tabla II.4.1, de nuevo la correlación entre los correctores es muy alta, lo que confirma que las puntuaciones asignadas por los profesores tienen una alta covariación.

Tabla II.4.1. Matriz de correlaciones entre los correctores

N=1194	Corrector A	Corrector B
Corrector B	.890	1
Corrector C	.896	.899

Finalmente, considerando la matriz completa tampoco se encuentran diferencias estadísticamente significativas -NC=95%- en función del rol del profesor como se puede comprobar en la Tabla II.4.2.

Tabla II.4.2. Diferencia de medias en función del rol del profesor (examinador o corrector)

TRIBUNAL	F	Sig.	Tamaño del efecto	Potencia
N=1194	3.49	.062	.003	.463

II.5 Capacidad de predicción del nivel de competencia en expresión oral

Teniendo en cuenta los resultados de los análisis llevados cabo en este informe y de los análisis propios de un estudio de regresión decidimos ajustar un modelo de regresión para la Calificación en Expresión Oral en Inglés (CalExpOral) con las siguientes variables:

- Índice socioeconómico (*Isec*)
- Calificación en Inglés en 2º de Bac. (*Cal2Bac*) (centrada)
- Calificación en Inglés Prueba escrita PAU (*CalPau*) (centrada)
- Repetición de curso (*RepCur*)
- Tamaño de la población de ubicación del Centro (*TamPob*)
- Titularidad del Centro (*TitCen*)

El resultado es:

$$\begin{array}{l}
 \text{Cal ExpOral} = 5.3996 + 0.3904 \text{ Isec} + 0.5292 \text{ Cal2Bac} + 0.3920 \text{ CalPau} - 0.7196 \text{ RepCur} + 0.4294 \text{ TamPob} + 0.3819 \text{ Tit Cen.} \\
 \text{(e.s.)} \quad (0.1515) \quad (0.0889) \quad (0,0455) \quad (0.0402) \quad (0.2395) \quad (0.1429) \quad (0.1261) \\
 \text{(p-val)} \quad (0,0000) \quad (0,0000) \quad (0,0000) \quad (0,0000) \quad (0,0027) \quad (0,0027) \quad (0,0025)
 \end{array}$$

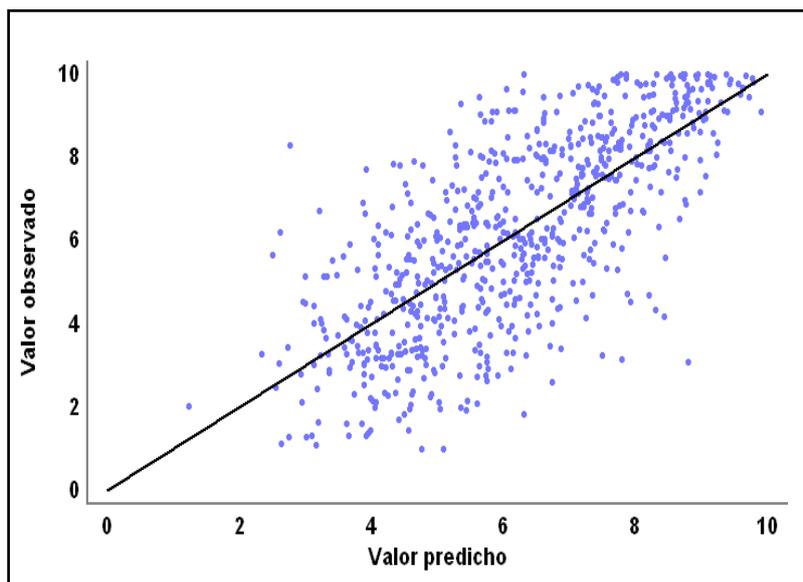
El análisis de la varianza indica que el modelo propone una relación estadísticamente significativa con un nivel de confianza superior al 95%. En este modelo todas las variables explicativas –regresoras- son significativas a nivel individual. Los detalles pueden verse en las tablas A.XVIII.1 y A.XVIII.2 del anexo XVIII. Además, el modelo propuesto posee el mejor -más pequeño- índice de información de Akaike (0.936483) -y también el de los demás criterios de información-; el mejor –menor- índice Cp de Mallows (7.0) y el menor Error Cuadrático Medio (2.50207) -y por tanto el mayor R cuadrado ajustado, 0.5323-.

Dada la correlación (0.6342) existente entre las variables *Cal2Bac* -Calificación en Inglés en 2º de Bachillerato- y *CalPau* -Calificación en Inglés en la Prueba de Acceso a la Universidad-, podría pensarse en simplificar ligeramente el modelo, no obstante hemos optado por incluir ambas variables dado que los índices obtenidos no indican serios problemas de multicolinealidad. En la tabla A.XVII.3 se muestran los índices que miden el grado de multicolinealidad, todos ellos dentro de límites aceptables.

En el gráfico A.XVIII.1 se observa el efecto individual que cada una de las variables explicativas tiene en la calificación total obtenida en la prueba de expresión oral. La calificación de inglés obtenida en 2º de bachillerato y la calificación de inglés obtenida en la PAU tienen un elevado efecto positivo, mientras que el índice socio-económico, la titularidad del centro y el tamaño de la población tienen un efecto positivo moderado o pequeño y, finalmente, la repetición de curso tiene un efecto negativo moderado.

El gráfico II.5.1 muestra la relación existente entre las observaciones realizadas, en ordenadas, y las predicciones obtenidas por el modelo de regresión ajustado.

Gráfico II.5.1: Dispersión de las calificaciones de la prueba oral en Lengua Inglesa respecto a las predicciones del modelo de regresión ajustado.



Como ya se ha comentado, el modelo propuesto explica el 53.23% de la variabilidad de la calificación obtenida por los alumnos en la prueba de expresión oral en inglés, siendo la calificación en inglés en 2º de Bachillerato, la calificación en inglés en la PA y la repetición de curso, las variables que más información aportan para predecir la calificación de expresión oral. Ajustando un modelo de regresión solo con estas tres variables regresoras resulta (modelo 2)

$$Cal\ ExpOral = 5.9943 + 0.5274\ Cal2Bac + 0.4306\ CalPau - 0.7405\ RepCur$$

(e.t.)	(0.0657)	(0.0461)	(0.0405)	(0.2444)
(p-val)	(0.0000)	(0.0000)	(0.0000)	(0.0025)

Este modelo explica el 50% de la variabilidad observada en la calificación de la Expresión Oral en Inglés. Los detalles de este modelo pueden verse en el Anexo XVIII, tablas A.XVIII.3 y A.XVIII.4, en el que se ha incluido el análisis de la adecuación de este último modelo ajustado.

II.6 Opiniones del profesorado

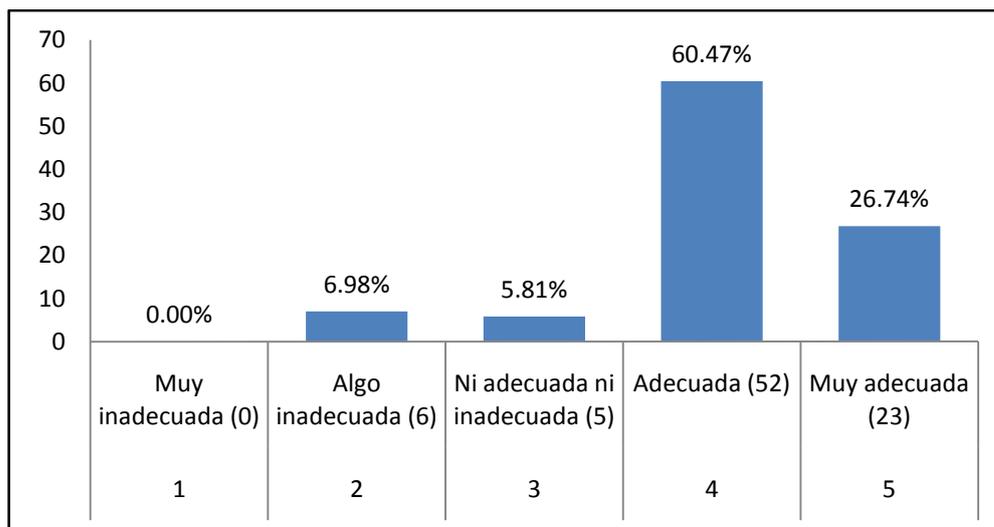
De los 90 profesores que aplicaron la prueba, se recabaron las opiniones de 86 distribuidos en las diferentes comunidades autónomas que participaron en este estudio: Aragón (12), Asturias (8), Baleares (23), Castilla La Mancha (12), La Rioja (10), Madrid (12) y Valencia (9).

El análisis se ha realizado teniendo en cuenta los dos grandes apartados sobre los que se pidió opinión a los profesores: Metodología y Organización. Dentro de estos apartados se han considerado los aspectos más significativos, tal como se expone a continuación. La escala de valoración va desde el 1, *muy inadecuada*, al 5, *muy adecuada*, pasando por el 2, *algo inadecuada*, el 3, *indiferente* y el 4, *adecuada*.

Metodología

Respecto a la valoración en conjunto de esta prueba piloto como método de evaluación de la expresión oral en Lengua Inglesa, el promedio de las valoraciones de los profesores ha sido del 4.07 (d.t. = 0.7790; e.t. = 0.084). Este dato muestra, en términos generales, una valoración positiva de la metodología empleada. De hecho, el 87.21% de los profesores entrevistados la considera muy adecuada o adecuada. El 5.81% se muestra indiferente y sólo el 6.98 % de los entrevistados considera la metodología adoptada algo inadecuada. El Gráfico II.6.1 muestra los resultados comentados.

Gráfico II.6.1: Resultados de la valoración de la metodología de la Prueba de Expresión Oral en Lengua Inglesa por parte de los profesores evaluadores.



Pese a ello, se han recogido 161 respuestas espontáneas sobre posibles mejoras al método de evaluación. En el Anexo XVI se recoge esta información resumida.

En primer lugar, es preciso destacar que casi el 20% indica que la prueba puede desarrollarse en menos tiempo, circunstancia que favorecería a los alumnos que tienen dificultad de expresión y al conjunto de la misma. El tiempo que se considera más apropiado sería entre 5-8 min.

En segundo lugar, con un 14% de los comentarios que lo señalan como aspecto a mejorar, se considera necesario revisar los descriptores, especialmente los de “interacción” y “coherencia” que son considerados poco claros. También consideran interesante valorar la comprensión oral y el lenguaje no verbal.

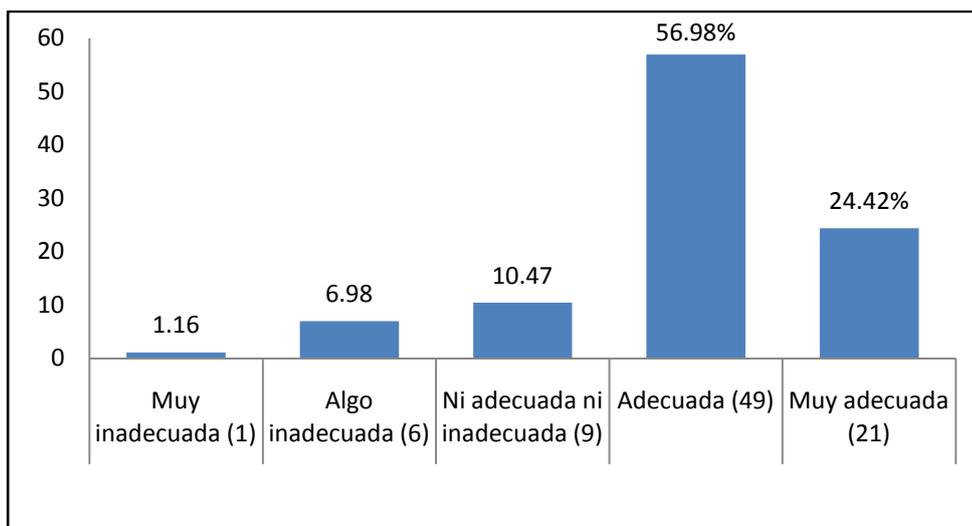
En cuanto a los estímulos fotográficos, el 28% del total de los comentarios sugieren que deben ser más variados, con temas cercanos a los alumnos (medio ambiente, otros deportes, objetos...). Así mismo deberían completarse con otros apoyos: títulos en inglés, preguntas al respecto, etc....

Finalmente, en el 6% de los comentarios se sugiere la necesidad de grabar la prueba y el 7,5% resalta la dificultad que implica la tarea de dirigir la prueba mientras que se evalúan los alumnos.

Organización

En cuanto a la organización de la prueba, la valoración de los profesores también ha sido positiva, alcanzándose una valoración media de 3.97 (d.t. = 0.8602; e.t. = 0.093). Expresado en porcentajes, el 81,4% de los profesores entrevistados ha considerado la organización adecuada o muy adecuada, mientras que el 10.47% se muestra indiferente y solo el 8.14 % la considera algo o muy inadecuada. En el Gráfico II.6.2 se muestran estos resultados.

Gráfico II.6.2: Resultados de la valoración de la organización de la Prueba de Expresión Oral en Lengua Inglesa por parte de los profesores evaluadores.



En el apartado de organización, se han recogido 101 comentarios o propuesta de los profesores. De estos, casi un 40% se refiere a cuestiones relacionadas con la información y las instrucciones a profesores, centros y alumnos. En este sentido, se pueden señalar como líneas de mejora:

- Elaborar un documento informativo en el que se detalle claramente la idoneidad del aula, la situación de los profesores y el orden de intervención.
- Concretar y clarificar los códigos de alumnos y profesores, de forma que no ofrezcan dudas respecto a la identificación de los mismos.
- Dependiendo de la comunidad autónoma, y por ello de la práctica establecida para distribuir la información, se propone una reunión informativa inicial, mayor comunicación con los centros, mejor selección de alumnos, disponer de un listado de los mismos, etc.
- Realizar una práctica sobre la corrección, previa a la realización de prueba, pues consideran que son muchos los ítems a valorar.

Finalmente, en cuanto a la organización de la prueba en el centro, la mayor dificultad destacada por los comentarios concierne a las fechas de aplicación. El 17% de los comentarios se refieren a la conveniencia de pasar la prueba en el 2º trimestre del curso escolar, para evitar que coincida con los exámenes finales de los alumnos. Entre los problemas mencionados por los profesores, el 9% de los comentarios se refiere a la desigual implicación de los centros (9% de los comentarios) y del absentismo del alumnado (8%). Este último punto sería solucionable, según los profesores, intentando motivar a los alumnos para que realicen la prueba.

III CONCLUSIONES Y RECOMENDACIONES

Diseño y estructura de la prueba

Desde un punto de vista técnico, teniendo en cuenta los resultados de la prueba piloto, el diseño y la estructura de la prueba pueden mantenerse con las mismas características básicas que en el estudio piloto. Si bien se recomienda su aplicación a finales del segundo trimestre para que no coincida con los exámenes finales de los alumnos ni con la prueba de acceso a la universidad, para que la tasa de participación sea lo más alta posible.

Esto es, dividida en dos partes manteniendo aproximadamente la misma proporción de tiempo que en la prueba piloto, puesto que se han observado diferencias estadísticamente significativas entre las calificaciones de la primera y la segunda parte de la prueba tanto en la calificación final como en cada uno de los criterios. Además, las calificaciones son más altas en la primera que en la segunda parte; de tal manera que en la primera parte el porcentaje de aprobados también resulta claramente superior al de la segunda.

La división de la calificación en cinco criterios parciales calificados independientemente se ha manifestado conveniente tanto desde el punto de vista práctico, para facilitar el trabajo de los correctores, como desde el punto de vista estadístico puesto que las diferencias encontradas son significativas en todas las comparaciones realizadas.

Por otra parte y puesto que cada criterio contribuye aproximadamente con el mismo peso a la calificación en cada parte de la prueba, parece razonable que cada parte se califique calculando la media aritmética de los cinco criterios.

Mientras que para la calificación final, dadas las diferencias observadas entre las calificaciones de las dos partes de la prueba, puede proponerse una distribución de pesos diferente a la utilizada en la prueba piloto. Como la calificación de la primera parte ha resultado ser significativamente más alta que la de la segunda, se propone una distribución de porcentajes del 30% para la primera parte y 70% para la segunda. Esta nueva distribución aumentaría la calificación final en expresión oral del estudiante, aproximándola a las que obtiene en las distintas calificaciones estudiadas de las que también difiere significativamente por abajo.

En cuanto al número de correctores que forman cada equipo, tres en el caso de la prueba piloto, se recomienda seguir formando equipos de tres correctores, porque si bien muestran acuerdo en cuanto a ordenar el desempeño del alumnado muestran diferencias en la asignación de una puntuación a ese desempeño. En efecto, se han encontrado diferencias significativas en las exigencias de los correctores al menos en la mitad de los tribunales. Sin embargo, no se han encontrado diferencias en cuanto al rol desempeñado por los profesores de examinador o de corrector.

Además se considera necesario, por la misma razón, establecer de forma más precisa el protocolo de corrección. Finalmente, si por razones económicas, se optase por utilizar dos correctores, debería consensuarse la calificación de ambos, dadas las diferencias encontradas.

Resultados en los diferentes contextos en los que se mueve el alumno.

Respecto al género, no se han encontrado diferencias estadísticamente significativas en los resultados obtenidos entre alumnos y alumnas, ni en la puntuación total ni en cada uno de los cinco criterios evaluados.

Sin embargo, sí que se encontraron diferencias en lo que respecta a la edad de los estudiantes. Los alumnos de 19 o más años de edad tienen unos resultados significativamente peores que los de 16, 17 y 18 años, no

observándose diferencias entre estos últimos grupos de edad. Relacionado con la edad, el hecho de repetir curso, una o más veces, influye de modo significativo en los resultados ya que los alumnos que no han repetido curso obtienen mejores calificaciones que aquellos que repiten o han repetido.

La inclusión de la prueba de expresión oral en inglés en la PAU, dada la relación estadísticamente significativa y positiva encontrada entre los resultados y el ISEC, incrementaría el efecto de las diferencias de clase socio-económica y cultural de los estudiantes sobre las calificaciones que obtendrían en las PAU.

En cuanto al tamaño de la población, éste también afecta a los resultados observados, encontrándose que los estudiantes de las poblaciones más pequeñas obtienen calificaciones significativamente más bajas que los del resto de poblaciones.

La duración de las clases y su frecuencia semanal también influye en los resultados obtenidos en la prueba oral, de tal manera que los estudiantes que dicen recibir más de tres horas semanales de inglés obtienen puntuaciones significativamente más altas que aquellos que dicen cursar menos de tres horas. Sin embargo, el uso de la lengua hablada en clase y la utilización de recursos pedagógicos tienen una modesta influencia en los resultados, algo menos del 1% y del 1.2% respectivamente.

La titularidad del centro tiene influencia en las calificaciones de la prueba oral. Los alumnos de los centros privados obtienen mejores calificaciones que los de los centros públicos si no se tiene en cuenta el efecto del ISEC. Ahora bien, una vez descontado el efecto del ISEC la diferencia no es significativa.

En cuanto al tiempo dedicado al estudio y a la realización de tareas en casa, los estudiantes que le dedican menos tiempo obtienen mejores calificaciones que aquellos que le dedican más tiempo. Una posible explicación es que los estudiantes que tienen más dominio del inglés dedican menos tiempo a las tareas propias de la lengua inglesa fuera del horario escolar. Por otra parte, respecto a las horas extra de inglés recibidas, no se observan diferencias significativas entre las categorías consideradas. Nuevamente es destacable el hecho de que no haya diferencia en los resultados obtenidos por los estudiantes que dedican 4 o más horas extra a la semana de clases de inglés respecto a los que no reciben ninguna clase extra.

Finalmente, la exposición del estudiante al inglés mediante comunicaciones, orales o escritas, internet, películas, videos, lecturas, etc., que miden el grado de exposición extraescolar del alumno a la lengua inglesa, tienen una influencia de aproximadamente el 22% en la calificación de la prueba oral. Mientras que los viajes de estudio o intercambio en países de habla inglesa explican alrededor de un 12% de la variabilidad de la calificación final, en tanto que si el viaje es con la familia esa influencia se cifra en torno al 14 %.

Relación de la calificación en la prueba oral con otras calificaciones de los estudiantes

Se han observado diferencias muy significativas entre la media de la calificación final en la prueba oral y las medias de las calificaciones obtenidas por los estudiantes tanto en el centro escolar, nota de inglés y nota media bachillerato, como en la PAU, nota de inglés y nota fase general. La calificación en la prueba oral de inglés es significativamente menor que la de las otras cuatro puntuaciones consideradas.

Por otro lado, la calificación en la prueba oral está altamente correlacionada con la calificación de inglés en el centro y con la calificación de inglés en la PAU y también lo está, aunque algo menos, con la nota media del bachillerato y la de la fase general de la PAU. De manera que la puntuación de la prueba oral de inglés viene explicada, en buena parte, por las calificaciones otorgadas por los profesores de los centros y por el examen de inglés escrito.

En la selección del modelo de regresión, de todas las variables consideradas resultan significativas el ISEC, la calificación de inglés en 2º de bachillerato, la calificación de inglés en la PAU y los índices de repetición de

curso, tamaño de población y titularidad del centro, que explican el 53.23% de la variabilidad de la puntuación en la prueba oral.

Finalmente, si se reduce el modelo a las variables calificación de inglés en 2º de bachillerato, calificación de inglés en la PAU y repetición de curso, estas explican el 50% de la variabilidad de la puntuación en expresión oral, esto es la mitad de la variación observada. Por ello, debe considerarse si merece la pena el esfuerzo económico y organizativo necesario para implantar la prueba oral de inglés en PAU, cuando el valor numérico añadido a la calificación de la PAU es muy pequeño. No obstante, la introducción de una prueba de estas características condicionaría la enseñanza de la lengua inglesa que, necesariamente, se enfocaría a desarrollar aspectos más acordes con la expresión oral y resultaría beneficioso para alumnos.

Opiniones de los profesores

Los profesores participantes en la prueba valoran muy positivamente la metodología de la misma, ya que el 87% de ellos la considera adecuada o muy adecuada. No obstante, sugieren que sería conveniente reducir el tiempo de la misma y clarificar los criterios de valoración, especialmente los de interacción y coherencia.

En cuanto a la organización, también es valorada positivamente puesto que el 81.4% la considera adecuada o muy adecuada. En este apartado la sugerencia más recurrentes se refieren a la mejora de las informaciones y de las instrucciones para centros, alumnos y profesores en lo que se refiere, sobre todo, a listados, códigos, etc., y a la necesidad de especificar más claramente los criterios de corrección.

Referencias

- INEE –Instituto Nacional de Evaluación Educativa– (2011) *Evaluación General de Diagnóstico 2010. Educación Secundaria Obligatoria. Segundo Curso. Informe de resultados*. Madrid, Ministerio de Educación.
- OECD (2001) *Knowledge and Skills for Life: First results from PISA 2000*. OECD, Paris.
- OECD (2004) *Learning for Tomorrow's World: First results from PISA 2003*. OECD, Paris.
- OECD (2007) *PISA 2006: Science Competencies for Tomorrow's World*. OECD, Paris.
- OECD (2010) *PISA 2009 Results: Volume II: Overcoming Social Background*. OECD, Paris.

Notas:

-
- ⁱ Se examinaron 25 alumnos de 2º curso de Bachillerato por centro excepto en la Comunidad Autónoma de Baleares que se examinaron alumnos de primer curso, debido a la imposibilidad de realizar la prueba piloto en el 2º curso en los centros seleccionados en las fechas establecidas.
- ⁱⁱ Excepto en la Comunidad Autónoma de Baleares al haber sido examinados alumnos de primer curso de Bachillerato.
- ⁱⁱⁱ Para el caso de los alumnos de la Comunidad autónoma de Baleares...
... se desarrolló un procedimiento específico de transformación de las puntuaciones de los estudiantes basado en el supuesto de un crecimiento anual de la competencia de 35 puntos en una escala de media 500 puntos y desviación típica de 100 puntos.
... no fue necesario realizar ninguna transformación dado que el nivel de competencia de los estudiantes no resultó estadísticamente diferente del resto de los estudiantes de las 6 comunidades autónomas restantes.
- ^{iv} Por otro lado, la relación entre el ISEC promedio de los centros y el rendimiento promedio de los centros educativos en expresión oral fue igual a .558 ($p \leq .000$), lo que implica que la nota media en expresión oral de los centros educativos está determinada en un 31 por ciento por el nivel socioeconómico y cultural promedio de las familias de sus alumnos.