

GUÍA EXPLICATIVA A LOS FORMULARIOS DE SOLICITUD

MODALIDAD 1: PRIMERA ACREDITACIÓN Y CENTRO PREVIAMENTE ACREDITADO

- El objetivo de las recomendaciones que se hacen en esta guía explicativa es, exclusivamente, mejorar la calidad de los exámenes CertAcles.
- La misión de la Comisión de Acreditación es verificar que, al elaborar sus exámenes, los centros respeten las especificaciones y recomendaciones de los exámenes CertAcles.
- Desde la Comisión de Acreditación pretendemos que estas recomendaciones sean incorporadas en los exámenes a corto y medio plazo (acciones de mejora a 2 o 4 años, es decir entre 1 y 2 períodos de acreditación), dentro de las posibilidades de cada Centro.
- Debido al alcance de algunas de las propuestas, es importante y aconsejable que los Centros planifiquen la incorporación de las mejoras, así como la formación necesaria de sus equipos de evaluadores para llevarlas a cabo.
- En beneficio de nuestras propias Universidades (en nombre de quienes los Centros expiden las certificaciones) y en el de todos los Centros certificadores, es importante tener la certeza de que todos los exámenes incorporan paulatinamente las mejoras sugeridas.
- Las recomendaciones de la Comisión de Acreditación responden a la diferente casuística observada en las solicitudes de acreditación de exámenes. Por ello es importante revisar periódicamente las recomendaciones que figuran en la guía explicativa.

1. SOLICITUD DE ACREDITACIÓN COMO INSTITUCIÓN CERTIFICADORA SEGÚN MODELO CERTACLES

Página web de la institución:

- Debe indicarse la URL del centro de lenguas, no de la universidad.
- En la web del centro debe estar publicada la guía del candidato o toda aquella información pertinente para los candidatos incluidos modelos de las tareas.

Convocatorias:

- A efectos de publicación en la Web de ACLES, cada mes de enero, el/la director/a del centro deberá comunicar al responsable de su zona, los meses de las convocatorias de exámenes CertAcles.

2. ESPECIFICACIONES DEL EXAMEN

2.1 Aspectos generales

2.1.2 Perfil examinandos

- La edad mínima de los candidatos es de 16 años.
- Deberá indicarse el tipo de población al que está destinado el examen (universitarios, PAS, PDI, sociedad en general, etc.)

2.1.3 Ámbitos de uso

- Aquellos recomendados en el MCER: Personal, Social o público, Académico, Profesional.

2.1.4 Destrezas evaluadas

- Enumerar las destrezas evaluadas e indicar si se incluye o no el “uso de la lengua” como prueba independiente.
- Si se trata de un examen con destrezas integradas, se deberá indicar qué destrezas se integran: comprensión lectora/auditiva o expresión escrita/oral (sólo a partir del nivel C1) y será necesario presentar el pilotaje de las diferentes tareas además de los análisis estadísticos requeridos para las pruebas discretas, un análisis cualitativo de los descriptores del marco a los que se enfoca la evaluación y la forma en la que esta está diseñada, además de las parrillas de evaluación de destrezas productivas (que deberán ser específicamente diseñadas para tareas integradas). Asimismo, se solicitará al centro al menos un ejemplo de la producción (oral y escrita) de candidato con corrección de acuerdo a la parrilla de destreza integrada y con comentarios cualitativos de varios correctores del centro de lo que se considera un estándar mínimo de aprobado.

2.1.5 Autenticidad

- Por autenticidad (Newby, D. 2002) se entiende que los textos (tanto orales como escritos) que se empleen en el examen que no hayan sido escritos y diseñados con fines pedagógicos. Por ejemplo: textos ya editados aparecidos en métodos de lengua, exámenes etc.
- Los textos auténticos se pueden adaptar en longitud o duración, pero no se debe realizar otro tipo de adaptación, por ejemplo, modificando estructuras o vocabulario, cortando párrafos en mitad del texto o, en los audios, ralentizar la velocidad.
- En todo caso, si los textos han sido adaptados deberá indicarse así en la tarea.
- Se podrán crear textos ex novo, siempre y cuando resulten naturales y verosímiles. Esta circunstancia también deberá indicarse.

2.2 Especificaciones de la prueba de comprensión lectora

2.2.1 Tipos de tareas

- Se recomienda utilizar la mayor variedad de tareas posibles con objeto de aumentar el número de elementos de juicio del candidato. Ver estructura de exámenes CertAcles
- Las diferentes tareas deben servir para evaluar distintos aspectos del contenido del texto (comprensión global, comprensión detallada etc.).
- En caso de utilizar un mismo tipo de tareas se deben justificar los motivos (por ej. porque se utilizan hojas de marca para la corrección de esta prueba del examen).
- Se debe evitar el uso de tareas con respuesta dicotómicas (si/no; verdadero/falso). No obstante, en caso de incluirse este tipo de tarea de verdadero/falso deberá implicar justificación. Se ha de evitar la opción “no consta en el texto” sustituyéndola por “justificación”.
- En las tareas de respuesta abierta conviene limitar el número de palabras.
- Se aconseja no utilizar tareas de uso de lengua (gramática, vocabulario etc) en esta prueba.
- Se debe utilizar un mismo tipo de ítems por tarea. Por ejemplo: no mezclar ítems de elección múltiple y de respuesta corta en una misma tarea, de acuerdo con las teorías de buenas prácticas en elaboración de exámenes
- Se recuerda que los modelos de tareas deben estar publicados en la web del centro para conocimiento de los candidatos con la suficiente antelación.

2.2.2 Número de tareas

- Consultar 'Estructura de los exámenes CertAcles'.
- Se recuerda que debe haber una única tarea por texto.
- Se recuerda que cuantos más ítems tenga la tarea, más elementos de juicio para la evaluación del candidato se tendrán.

2.2.3 Número de ítems por tareas

- Consultar 'Estructura de los exámenes CertAcles'.
- Se recuerda que cuantos más ítems tenga la tarea, más elementos de juicio para la evaluación del candidato se tendrán.

2.2.4 Número de palabras por texto y total de la prueba

- Consultar 'Estructura de los exámenes CertAcles'.
- Se podrá aumentar hasta un 25% el número de palabras aconsejado en la tabla según la lengua, con la debida justificación (según idioma).

2.2.5 Fuente de los textos

- En el examen siempre deberá indicarse la fuente de los textos de cada tarea; la fuente puede indicarse de forma corta (ej: www.lemonde.fr)
- Siempre deberá indicarse si el texto es de creación propia o si ha sido adaptado.

2.2.6 Instrucciones

- Las instrucciones deben de ser claras, detalladas y concisas.
- El nivel de lengua utilizado en las instrucciones debe ser un nivel inferior al nivel del examen en cuestión
- En la solicitud debe incluirse un ejemplo de las instrucciones de una tarea en la lengua meta (lengua meta a partir del nivel A2).

2.2.7 Modelo de tareas

- En la solicitud deben incluirse dos ejemplos de tareas.
- Se recomienda dar a los candidatos un ejemplo de respuesta a un ítem en cada tarea.

2.3 Especificaciones de la prueba de comprensión auditiva

2.3.1 Tipos de tareas

- Se recomienda utilizar la mayor variedad de tareas posibles con objeto de aumentar el número de elementos de juicio del candidato.
- Se aconseja que las diferentes tareas sirvan para evaluar distintos aspectos del contenido del texto (comprensión global, comprensión detallada etc.).
- En caso de utilizar un mismo tipo de tareas explicar los motivos (por ej. porque se utilizan hojas de marca para la corrección de esta prueba del examen).
- Se debe evitar el uso de tareas con respuesta dicotómicas (si/no; verdadero/falso). No obstante, en caso de incluirse este tipo de tarea de verdadero/falso deberá implicar justificación. Se ha de evitar la opción “no consta en el texto” sustituyéndola por “justificación”.
- En las tareas de respuesta abierta conviene limitar el número de palabras.
- Se aconseja no utilizar tareas de uso de lengua (gramática, vocabulario etc) en esta prueba.
- Se recomienda utilizar un mismo tipo de ítems por tarea. Por ejemplo: no mezclar ítems de elección múltiple y de respuesta corta en una misma tarea.
- Se recuerda que los modelos de tareas deben estar publicados en la web del centro para conocimiento de los candidatos.

2.3.2 Número de tareas

- Consultar ‘Estructura de los exámenes CertAcles’.
- Se recuerda que cuantas más tareas, más elementos de juicio para evaluar del candidato se poseen.
- Se recuerda que se debe utilizar una única tarea por texto.

2.3.3 Número de ítems por tareas

- Consultar 'Estructura de los exámenes CertAcles'.
- Se recuerda que cuantos más ítems tenga la tarea, más elementos de juicio para la evaluación del candidato se tendrán.
- Se recomienda utilizar un mismo tipo de ítems por tarea. Por ejemplo: no mezclar en una misma tarea ítems de elección múltiple y de respuesta corta.
- Se recomienda en cada tarea proporcionar a los candidatos un ejemplo de respuesta a un ítem

2.3.4 Duración de los audios y total de la prueba

- Consultar 'Estructura de los exámenes CertAcles'.

2.3.5 Número de escuchas por audio

- Cada audio o vídeo se escuchará o visionará 2 veces.

2.3.6 Instrucciones

- Las instrucciones deben de ser claras, detalladas y concisas.
- Se recomienda que el nivel de lengua utilizado en las instrucciones sea inferior al nivel del examen.
- En el formulario debe incluirse un ejemplo de las instrucciones de una tarea en la lengua meta a partir del nivel A2.

2.3.7 Modelo de tareas

- En el formulario deben incluirse dos ejemplos de tareas.
- Se recomienda dar a los candidatos un ejemplo de respuesta a un ítem en cada tarea.

2.4. Especificaciones de la prueba de expresión escrita

2.4.1 Tipos de tareas

- Consultar 'Estructura de los exámenes CertAcles'.

2.4.2 Número de tareas

- Consultar 'Estructura de los exámenes CertAcles'.

2.4.3 Tipo de discurso

- Se debe adaptar a lo prescrito en el MCER para cada nivel

2.4.4 Número de palabras por tarea

- Consultar 'Estructura de los exámenes CertAcles'.

- Se podrá aumentar hasta un 25% el número de palabras aconsejado en la tabla con la debida justificación (según idioma).

2.4.5 Número de palabras total de la prueba

- Consultar 'Estructura de los exámenes CertAcles'.

2.4.6 Instrucciones

- Es importante incluir pautas explícitas del contenido del texto que el candidato debe elaborar y proporcionar un contexto suficiente para que sea una tarea verosímil (quién escribe, para qué, a quién, qué debe decir etc.)
- Las instrucciones deben de ser claras, detalladas y concisas.
- Se recomienda que el nivel de lengua utilizado sea inferior al nivel del examen.
- En el formulario debe proporcionarse un ejemplo de las instrucciones de una tarea en la lengua meta a partir del nivel A2.

2.5. Especificaciones de la prueba de expresión e interacción oral

2.5.1 Instrucciones

- Es importante incluir pautas explícitas para la interacción y proporcionar un contexto suficiente para que sea una tarea verosímil.
- Las instrucciones deben de ser claras, detalladas y concisas.
- Se recomienda que el nivel de lengua utilizado sea inferior al nivel del examen.
- En caso de proporcionar un input escrito para la tarea, éste debe ser breve.
- En el formulario debe incluirse un ejemplo de las instrucciones de una tarea en la lengua meta a partir del nivel A2.

2.5.2 Procedimiento de grabación

- Grabación audio (no tiene por qué ser vídeo)
- A efectos legales, se deberá solicitar la autorización de los candidatos para ser grabados (antes de la entrevista, en la matrícula en el examen, etc) para fines de revisión de exámenes, pedagógicos y de formación de profesorado (estandarización de evaluadores).

2.6. Especificaciones de la prueba del uso de la lengua

En caso de que se incluya esta prueba, se recomienda que el uso de la lengua sea evaluada de forma independiente al resto de las pruebas.

3. Criterios de evaluación

3.1 Peso de las distintas pruebas

Se recomienda que todas las pruebas del examen tengan el mismo peso en la calificación final.

3.2 Criterios de superación del examen

Para considerar adquirido el nivel global se recomienda que la calificación total sea superior o igual al 60% sobre la totalidad de la puntuación posible (siendo el mínimo el 50%) y, además, haber alcanzado un mínimo del 50% de la puntuación posible en cada prueba.

La escala de calificaciones se establece entre 0 y 10 puntos

- Entre 5,0 y 6,9 puntos (50%-69% de la puntuación posible) = APROBADO
- Entre 7,0 y 8,9 puntos (70%-89% de la puntuación posible) = NOTABLE
- Entre 9,0 – 10 puntos (90%-100% de la puntuación posible) = SOBRESALIENTE

4. Criterios y procedimientos de corrección

- Se recuerda que, en el caso de exámenes con destrezas integradas, se solicitará al centro un ejemplo de corrección de las destrezas integradas que especifique los criterios que se han tomado para dicha corrección.
- El centro deberá proporcionar un ejemplo de la producción (oral y escrita) de candidato con corrección de acuerdo a la parrilla de destreza integrada y con comentarios cualitativos de varios correctores del centro de lo que se considera un estándar mínimo de aprobado.

■ 4.1.1 Puntuación de las tareas y de los ítems

- Todos los ítems deben recibir la misma puntuación.

4.1.2 Método/procedimiento de corrección

- En las tareas de respuesta abierta no se tendrán en cuenta los errores ortográficos y gramaticales, siempre y cuando no afecten a la comprensión.

■ 4.2.1 Puntuación de las tareas y de los ítems

- Todos los ítems deben recibir la misma puntuación.

4.2.2. Método/procedimiento de corrección

- En las tareas de respuesta abierta no se tendrán en cuenta los errores ortográficos y gramaticales, siempre y cuando no afecten a la comprensión.

■ 4.3.1 Categorías evaluadas (ej. adecuación, cohesión, etc.)

- Indicar y describir las categorías evaluadas (según el MCER)

4.3.2 Procedimiento de corrección

- Se recomienda altamente la doble corrección en todos los casos, si es posible. De no ser así la fiabilidad de los correctores debe ser muy alta, y se tendrá que aportar evidencias de ello.
- Se debe realizar la doble corrección en al menos un 25% de los candidatos y siempre en los casos dudosos.

4.3.3 Descriptores y parrillas de corrección

Se recuerda que, en el caso de exámenes con destrezas integradas, deberán presentarse las parrillas de evaluación de las destrezas productivas (que deberán ser específicamente diseñadas para tareas integradas)

- Indicar el descriptor que corresponde a cada valor numérico de la escala de las diferentes categorías (ej: fluidez, cohesión, etc.) evaluadas.
- El equivalente al apto/aprobado debe estar a mitad de la escala.

4.4.1. Categorías evaluadas (ej. fluidez, pronunciación, etc.)

- Indicar y describir las categorías evaluadas (según el MCER)

4.4.2. Procedimiento de corrección

- Se recomienda altamente que en la prueba haya un examinador y un evaluador, aportando ambos su criterio final de evaluación. De no ser así, la fiabilidad de los correctores debe ser muy alta y se tendrá que aportar evidencias de ello.
- Se debe realizar la doble corrección en al menos un 25% de los candidatos y siempre en los casos dudosos.

4.4.3- Descriptores y parrillas de corrección – Incluir en Anexo II

- Indicar el descriptor (según el MCER) que corresponde a cada valor establecido en la escala de las diferentes categorías (ej: fluidez, cohesión, etc.) evaluadas

5. Elaboración de exámenes

5.5.1. Descripción del equipo de elaboración de exámenes (número y perfil)

- El número de personas puede variar de una lengua a otra; se recomienda, siempre que sea posible, que no sea una única persona quien los elabore.
- El equipo de elaboración de exámenes puede tener un perfil variado (profesor del centro, licenciado, doctor, contratado externo...)
- Se recomienda la estabilidad de los equipos de elaboración de exámenes y la continuidad en el tiempo de, al menos, dos de sus miembros ya formados.

5.5.2. Formación recibida por el equipo de elaboración

- Este apartado incluye cualquier tipo de formación recibida (talleres, seminarios, formación interna, formación externa, congresos etc.).
- Se recomienda que los centros inicien o continúen los planes de formación de sus equipos de evaluación.
- En el formulario deberá indicarse el número de horas de formación recibida por cada persona responsable de la elaboración de los exámenes.

5.5.3. Proceso de elaboración de exámenes

- Explicar brevemente cómo se elaboran los exámenes (criterios de selección de textos, si se realiza textmapping, si tiene pilotaje, etc. o cualquier otra información que se considere pertinente para explicar el proceso).

6. Guía de administración de exámenes

6.6.1 Aspectos generales

- Se recomienda que el centro disponga de una guía del candidato, una guía de administración y una guía para los examinadores y evaluadores.
- En la web del centro debe estar publicada la guía del candidato o toda aquella información referida al examen pertinente para el candidato

6.6.2 Certificado

- El certificado que se emita debe reflejar que se trata de la acreditación de dominio de un determinado nivel de lengua (independiente de la formación recibida).

7. Examinadores

- Establecer claramente las tareas que debe desempeñar todo el personal que participe en la administración del examen (personal de la administración, vigilantes, examinadores, correctores etc.)
- En la medida de lo posible se debe evitar que un profesor examine a sus propios alumnos.

7.1 Formación (estandarización de correctores)

- Se debe hacer una sesión de homogeneización de evaluadores por año, como mínimo, y realizar un breve informe de seguimiento de la misma.

8. Custodia de exámenes y destrucción

- En lo referido a la custodia y destrucción de exámenes, cada centro se regirá por lo dispuesto en su propia normativa o en la de su universidad.

9. Características del examen

- La Comisión de Acreditación considera que el pilotaje del examen y los análisis posteriores a la convocatoria de dicho examen son fundamentales para garantizar la calidad de los mismos. Es por ello que, con vistas a implantarlo en un futuro de forma generalizada, se recomienda y anima a los centros que aún no lo hayan hecho, a iniciar la formación de sus equipos de evaluadores.
- En esta solicitud es conveniente incluir cualquier tipo de análisis que ya se esté realizando en este sentido, por sencillo que sea (por ejemplo, juicio de expertos, valoración entre pocos candidatos, etc.).

9.1 Buenas prácticas de pilotaje

(Método, número de ítems, niveles, descripción de la muestra, etc.)

NOTA: Sin perjuicio de que puedan realizarse análisis más complejos dependiendo del tipo de prueba y de la decisión y posibilidades de cada centro, al finalizar esta guía explicativa se facilita un documento que puede consultarse, a modo de ejemplo, sobre cuáles son los análisis estadísticos imprescindibles.

- Debe pasar por un proceso de prepiloteo y de pilotaje posterior (para establecer los puntos de corte)
- Se recomienda una muestra de 100 sujetos (una muestra mínima de dos personas por ítem del examen).
- La población muestra debe tener el mismo perfil que la población destinataria del examen. (Si por razones logísticas no es posible, se pueden tomar sujetos de perfiles lo más cercanos posible).
- La muestra debe incluir sujetos del nivel inferior y superior a los evaluados en el examen.
- Si los participantes en la muestra son alumnos de un nivel del MCER, es necesario que hayan realizado al menos un 75% del curso del nivel correspondiente. De ser menos, se le considerará del nivel inferior para la muestra.
- Se recomienda realizar el pilotaje con la misma muestra de sujetos. Si por razones logísticas no es posible, se puede realizar en fases distintas o en colaboración con otras universidades.
- Se debe explicar con claridad la manera en que han sido recogidos los datos del pilotaje y el proceso de control aplicado.

9.1.2-Análisis de las propiedades de los ítems

(Teoría clásica, teoría de respuesta al ítem - dificultad, discriminación,...-, análisis de distractores, etc. Comentar proceso de selección de los ítems).

- Especificar tipos de análisis de ítems aplicados
- Especificar criterios de selección de ítems con análisis de dificultad, de discriminación y de distractores.
- Será necesario evidenciar una justificación de los resultados de los análisis con una muestra.

9.2- Buenas prácticas de Validez 9.2.1- Contenido y proceso de respuesta

(Ítems representativos de los requisitos del MCER, experiencia de los elaboradores de ítems, consulta a expertos, proceso de recogida de datos).

- Especificar el proceso de creación de ítems
- Especificar la validación con expertos y el método de calibración de ítems.

9.2.2-Estructura interna

(Diseño, muestra y tipo de análisis: análisis factorial, análisis de componentes principales, teoría de respuesta al ítem, etc.)

- Si por razones logísticas no todas las partes del examen estuvieran pilotadas por los mismos sujetos, es necesario establecer ítems de anclaje.
- No serán obligatorios (aunque sí recomendados) ítems de anclaje si se aplican teorías psicométricas como Teoría de respuesta al ítem (*IRT*) con programas como *Winsteps*, *Bilog* o *JMetrix*.
- Será necesario evidenciar una justificación de los resultados de los análisis indicando los índices estadísticos usados y cómo deben interpretarse para darlos por buenos.

9.2.3-Relación con otras variables

(Correlación entre las pruebas del examen, comparación de medias, determinación de los puntos de corte -curvas ROC, Angoff, CCM,...- etc.).

- Deben ser claramente explicados los procesos, métodos y criterios utilizados para establecer los puntos de corte.
- Será necesario evidenciar una justificación de los resultados de los análisis con una muestra.

9.3- Buenas prácticas de Fiabilidad

(método y datos aportados: equivalencia, consistencia interna, estabilidad, etc.)

Especificar los métodos utilizados y resultados (Alfa de Cronbach, Error estándar de estimación (*SEM*), etc.). También se debe indicar como se debe interpretar el índice estadístico.

Será necesario evidenciar una justificación de los resultados de los análisis con una muestra.

10. Buenas prácticas de Control de Calidad

10.1- Encuestas de satisfacción

- Mostrar la encuesta y sus resultados.

10.2- Postanálisis estadístico

- Indicar métodos utilizados y resultados.
- Será necesario evidenciar una justificación de los resultados de los análisis con una muestra.

10.3- Análisis de correctores

- Indicar métodos utilizados (Homogeneización de correctores, Coeficientes de correlación, ICC, etc.) y resultados.

Guía de ejemplo para análisis estadístico

Esta guía no tiene carácter prescriptivo y se elabora a modo de ejemplo de cuáles son los análisis estadísticos *mínimos* (sin perjuicio de que puedan realizarse análisis más complejos dependiendo del tipo de prueba y de la decisión y posibilidades de cada centro).

La función del análisis estadístico es proporcionarnos datos sobre el funcionamiento de nuestra prueba y por tanto su importancia radica en la interpretación que se hace de los mismos y los pasos que se toman como consecuencia de estas interpretaciones. Por esto, todo análisis debe acompañarse de una explicación de cómo se ha interpretado y de cuáles son las decisiones que se han tomado a raíz de estas interpretaciones (ejemplos a continuación).

A continuación, proporcionamos una estructura de análisis que puede utilizarse **a modo de ejemplo**. Los datos proporcionados en el ejemplo provienen de los análisis de pilotaje de varias pruebas reales.

Los ejemplos proporcionados han sido elaborados utilizando el programa spss, pero es posible realizarlos con otros programas como, por ejemplo: Excel, TAP (Test Analysis Program) de la universidad de Ohio, TiaPlus, CITAS, etc. En caso de teoría de respuesta al ítem (para universidades que quieran ir más allá o disponga de un número de alumnos por convocatoria superior a 150-200 candidatos) Winsteps, Facets o R pueden utilizarse dependiendo del propósito del análisis.

ANÁLISIS A NIVEL DE LA PRUEBA

Estadística descriptiva

N. de candidatos valido (cuantas respuestas se han analizado)	70
Candidatos faltantes (cuantos candidatos estaban en la prueba, pero no contestaron a las preguntas)	132
Media	16,8000
Mediana	16,0000
Modo	9,00 ^a
Desviación estándar	7,15987
Rango	33,00

Análisis cualitativo de los resultados: los candidatos se desvían 7.15 de la media con un rango de notas de 33 y una media de 16,8. Analizando los resultados generales se puede ver como funcionó el test en general, lo que queda representado también en el histograma. Idealmente, buscaremos que la media y la mediana tengan valores similares pues significaría que la prueba está bien dirigida al conjunto de los candidatos. Una desviación estándar menor que la media indicará que la media refleja el nivel de la candidatura. Si la desviación estándar es mayor que la media, la variedad en puntuaciones hará que la media no refleje el nivel real de la candidatura.

Fiabilidad

Alpha de Cronbach	N de ítems
,896	38

El Alpha de Cronbach mide la consistencia interna de la prueba y es uno de los indicadores de fiabilidad más comúnmente utilizados, pero es muy susceptible al número de ítems, cuantos más ítems, más fiabilidad. La fiabilidad también puede medirse mediante la fórmula Kuder-Richardson, la fórmula de Spearman Brown, etc.

En cualquier caso, lo importante sería la interpretación de los datos. En el ejemplo citado, diríamos que el rango de Alpha de Cronbach aceptable para pruebas de competencia estandarizadas sería

$0.9 \leq \alpha$	excelente
$0.8 \leq \alpha < 0.9$	bueno
$0.7 \leq \alpha < 0.8$	aceptable

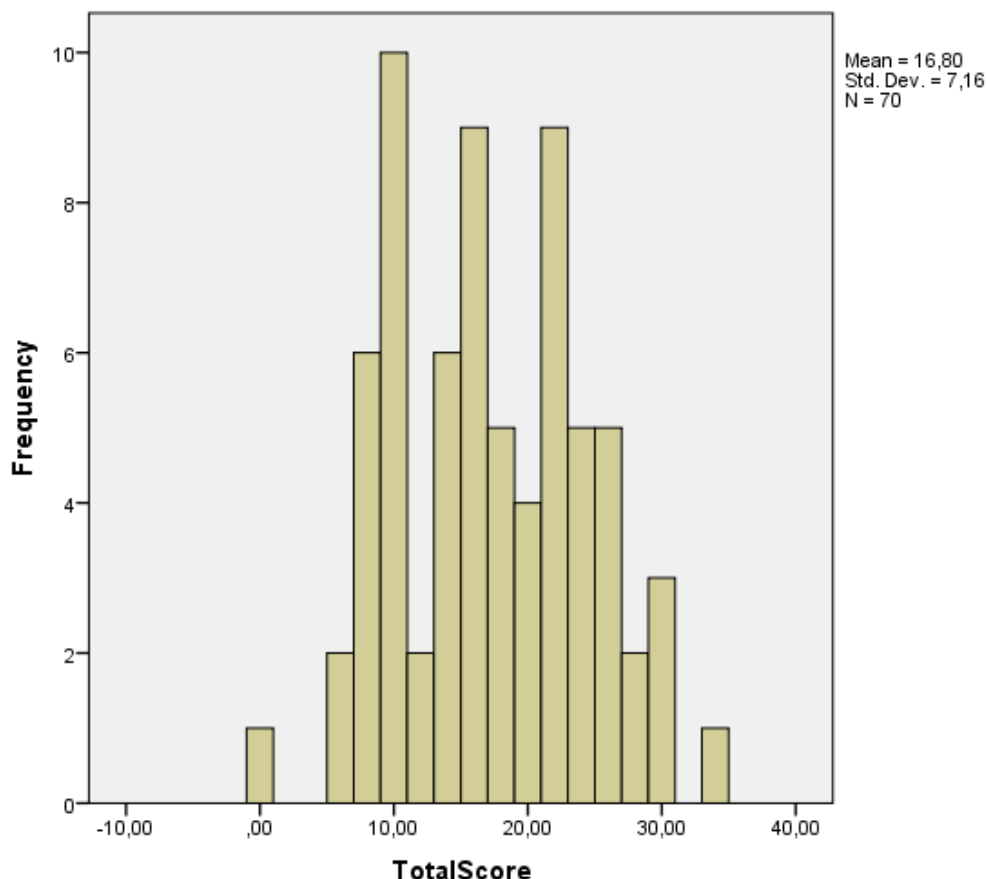
Diríamos por tanto que la prueba muestra una consistencia interna buena acercándose a excelente.

Correlación entre pruebas de examen

La correlación entre las pruebas nos indica la relación entre los resultados de los candidatos entre las diferentes partes de la prueba. Aunque la relación no indica causalidad, que las partes estén relacionadas en una prueba de lenguas debería ser lo deseable (siempre y cuando la prueba esté diseñada para ello). En el ejemplo, en verde, se pueden ver los comentarios sobre los tipos de correlación encontrados.

		Listening	Reading	Writing	Speaking
Listening	Correlation Coefficient Sig. (2-tailed) N				
Reading	Correlation Coefficient Sig. (2-tailed) N	,530** ,000 41 <i>Correlación positiva moderada estadística mente significativa</i>			
Writing	Correlation Coefficient Sig. (2-tailed) N	,671** ,000 41 <i>Correlación positiva moderada estadística mente significativa</i>	,853** ,000 41 <i>Correlación positiva fuerte estadística mente significativa</i>		
Speaking	Correlation Coefficient Sig. (2-tailed) N	,726** ,000 41 <i>Correlación positiva fuerte y estadística mente significativa</i>	,791** ,000 41 <i>Correlación positiva fuerte y estadística mente significativa</i>	,933** ,000 41 <i>Correlación positiva fuerte y estadística mente significativa</i>	

		Listening	Reading	Writing	Speaking
Listening	Correlation Coefficient Sig. (2-tailed) N				
Reading	Correlation Coefficient Sig. (2-tailed) N	,530** ,000 41 <i>Correlación positiva moderada estadísticamente significativa</i>			
Writing	Correlation Coefficient Sig. (2-tailed) N	,671** ,000 41 <i>Correlación positiva moderada estadísticamente significativa</i>	,853** ,000 41 <i>Correlación positiva fuerte estadísticamente significativa</i>		
Speaking	Correlation Coefficient Sig. (2-tailed) N	,726** ,000 41 <i>Correlación positiva fuerte y estadísticamente significativa</i>	,791** ,000 41 <i>Correlación positiva fuerte y estadísticamente significativa</i>	,933** ,000 41 <i>Correlación positiva fuerte y estadísticamente significativa</i>	



El histograma proporciona una visión general del comportamiento de los candidatos en la prueba. Resulta útil para proporcionar un vistazo rápido de los resultados y para comparar comportamientos de la prueba con otras destrezas o con administraciones anteriores. En este caso, vemos que ningún candidato ha obtenido la puntuación más alta, lo cual en principio y si el análisis de los ítems es correcto, es positivo, pues indica que hasta el candidato con mayor competencia ha tenido la posibilidad de mostrar su habilidad.

ANÁLISIS A NIVEL DE ÍTEM

Índices de facilidad y discriminación. Análisis cuantitativo y cualitativo

Es una buena idea presentar el análisis al lado del ítem o tarea analizado para facilitar su interpretación y guiar a los desarrolladores de la prueba en las modificaciones que pueden ser necesarias.

A modo de guía, los índices de facilidad y discriminación pueden interpretarse de la siguiente manera:

IF

.30 -.70	Ítems óptimos
.20-.80	ítems buenos atendiendo a la discriminación (el índice de discriminación tiene que ser bueno) y a que aporten consistencia a la prueba

* los ítems demasiado fáciles o demasiado difíciles sólo son útiles si discriminan bien, lo que será complicado puesto que no generan suficiente variabilidad entre los resultados

I.D

>40	óptima
.30-.39	Razonablemente buena

.20-29	Marginal (debe mejorarse el ítem)
<19	A mejorar o descartar

*la discriminación negativa indica que los alumnos más competentes contestan incorrectamente y los alumnos menos competentes lo hacen correctamente. Esto suele ser debido a errores en las claves de respuestas, la existencia de más de una respuesta correcta o preguntas « tramposas »

A continuación, mostramos un ejemplo del análisis de una prueba de comprensión oral. Vemos la tarea, los índices de facilidad y discriminación, y las consideraciones hechas por los responsables. A continuación, se presentarían las modificaciones realizadas por los redactores.

Tarea 1. Aquí introduciríamos la tarea y sus ítems	Item	IF	Consideraciones Facilidad	ID	Consideraciones discriminación
Q1.	Q1	,46	Ok	,493	óptimo
Q2.	Q2		Ok	-	A eliminar.
Q3.		,73		,048	Discrimina negativamente
Q4.	Q3	,66	Ok	,242	Marginal, necesita mejorarse
Q5.	Q4	,44	Ok	,201	Marginal, necesita mejorarse
Q6.	Q5	,52	Ok	,244	Marginal, necesita mejorarse
Q7.	Q6	,51	Ok	,507	óptimo
	Q7	,02	Excesivamente difícil	,227	Marginal, necesita mejorarse, puede ser por la dificultad

Cambios propuestos o ajustes de puntuación (en caso de ser resultados de administración real). Cambios realizados (en caso de ser resultados de pilotaje)

Aquí incluiríamos el análisis de los ítems con resultados problemáticos y los cambios que se proponen para su mejora (de eliminación del ítem, reformulación, corrección de errores, etc). La idea es usar los datos estadísticos para mejorar la prueba.

Fiabilidad. Análisis cuantitativo y cualitativo

Identificaríamos los ítems problemáticos a nivel de fiabilidad y cómo se comportaría la prueba en caso de eliminar este ítem.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q1	28,54	43,614	,120	,896
Q2	28,59	42,449	,431	,893
Q3	28,59	42,652	,364	,894
Q4	28,56	43,120	,304	,895
Q5	28,57	42,886	,333	,894
Q6	28,54	43,411	,249	,895
Q7	28,53	43,818	,000	,896
Q8	28,67	42,630	,231	,896
Q9	28,79	40,895	,485	,892
Q10	28,77	41,396	,402	,893
Q11	28,84	40,453	,529	,891
Q12	28,83	40,434	,540	,891
Q13	28,76	40,766	,532	,891
Q14	28,74	40,745	,550	,891
Q15	28,64	41,914	,434	,893
Q16	28,54	43,469	,212	,896
Q17	28,54	43,469	,212	,896
Q18	28,57	43,321	,170	,896
Q19	28,67	41,644	,451	,893
Q20	28,87	40,751	,465	,892
Q21	28,76	42,013	,297	,895
Q22	28,81	41,719	,322	,895
Q23	28,54	43,469	,212	,896
Q24	28,90	39,396	,685	,888
Q25	28,97	41,072	,389	,894
Q26	28,86	40,095	,584	,890
Q27	29,21	42,519	,177	,898
Q28	28,84	39,613	,677	,888
Q29	28,83	41,043	,433	,893
Q30	28,81	40,994	,449	,892
Q31	28,73	41,128	,489	,892
Q32	29,19	41,139	,400	,893
Q33	28,79	41,852	,311	,895
Q34	28,70	40,706	,613	,890
Q35	28,87	40,114	,574	,890
Q36	28,90	40,613	,478	,892
Q37	28,76	41,230	,444	,893
Q38	28,93	40,502	,489	,892

Modificaciones realizadas

Aquí podríamos introducir los ítems problemáticos (subrayados en rojo) tal cual venían redactados en el examen –con clave de respuesta incluida- para poder analizar con facilidad cual ha podido ser el problema (mala redacción, clave de respuestas equivocada, pregunta “tramposa”, más de una opción correcta, etc.). A continuación deberíamos describir cual ha sido la decisión al respecto por parte del equipo desarrollador (eliminar el ítem, mejorarlo, subsanar errores, etc.)

Análisis cuantitativo y cualitativo de distractores

Q1 Aquí introduciríamos el ítem con las respuestas posibles a. b. c. d.	% A	1		Q6	% A	2
	% B	41			% B	17
	% C	49			% C	68
	% D	0			% D	2
Q2	% A	53		Q7	% A	25
	% B	25			% B	66
	% C	2			% C	5
	% D	2			% D	0
Q3	% A	2			% A	2
	% B	24			% B	24
	% C	46			% C	46
	% D	0			% D	0
Q4	% A	2			% A	2
	% B	36			% B	36
	% C	44			% C	44
	% D	0			% D	0
Q5	% A	32			% A	32
	% B	41			% B	41
	% C	25			% C	25
	% D	2			% D	2

Modificaciones realizadas

Las modificaciones realizadas y la razón de la modificación serían incluidas aquí, ej.

Q4. XXXXXXXX

- a. Se modifica el distractor porque es muy evidente y se reformula**
- b. el distractor funciona**
- c. Es la respuesta correcta**
- d. El distractor no funciona, se decide eliminarlo**

Lo óptimo sería repilotar si hay cambio muy significativos

** Los distractores marcados en rojo no están funcionando como distractores, en los casos en los que sólo quedan dos opciones, se corre el peligro de que la pregunta sea contestado por suerte (50% de posibilidades). Dos distractores (además de la respuesta correcta) son suficientes si funcionan, por lo que se aconseja eliminar después del pilotaje el de peor funcionamiento siempre y cuando los otros dos están funcionando.*

ANÁLISIS DE CORRECTORES DE TAREAS PRODUCTIVAS

Los datos óptimos que nos permiten analizar si la prueba se está corrigiendo de manera fiable son la fiabilidad intra-corrector y la fiabilidad inter-corrector. En caso de que, por diseño de la corrección, por número de correctores o por circunstancias del centro, esto no fuera posible, se requerirá al menos el envío de un Excel con las puntuaciones otorgadas por cada corrector a las producciones que ese corrector haya corregido.

Ejercicio 1 de expresión escrita (ejemplo de envío de notas de correctores)

	Media de las notas dadas por el corrector	Número de producciones corregidas por el corrector
Corrector 1		
Corrector 2		
Corrector 3		
Corrector 4		
Corrector 5		
Corrector 6		
Corrector 7		