




Universidad de Concepción
Dirección de Postgrado
Facultad de Humanidades y Arte-Programa de Doctorado en Lingüística



Criterios evaluativos para el diseño de un prototipo de una prueba de Español como Lengua Extranjera con fines específicos Académicos (EFA)

Tesis para optar al grado de Doctor en Lingüística

ANA ENRIQUETA VINE JARA

CONCEPCIÓN-CHILE

2015

Profesora guía: Dra. Anita Ferreira Cabrera

Depto. de Español, Facultad de Humanidades y Arte

Universidad de Concepción

ÍNDICE

Índice general.....	1
Índice de figuras.....	5
Índice de gráficos.....	6
Índice de tablas.....	6
Resumen.....	10
Introducción.....	11

CAPÍTULO I: Fundamentos teóricos

1. La evaluación en las segundas lenguas.....	13
1.1. La evaluación de segundas lenguas con fines específicos.....	15
1.1.1. La evaluación del Español con Fines específicos Académicos (EFA).....	17
1.1.2. El discurso académico en EFA.....	19
1.1.2.1. Los géneros académicos en EFA.....	20
1.2. Tipos de evaluación.....	21
1.3. Tipos de pruebas.....	24
1.4. Fases en la elaboración de pruebas.....	26
1.5. La Evaluación de Lenguas a través del Computador (CALT).....	28
1.5.1. La implementación en CALT.....	29
1.5.2. La evaluación en CALT.....	30
1.5.3. Ventajas de la implementación de pruebas de EFA a través del computador.....	33
1.5.4. El uso de Moodle en la elaboración de pruebas de segundas lenguas	33
1.6. Pruebas de proficiencia en el Español como Lengua Extranjera (ELE).....	35
1.6.1. Pruebas de ELE con fines generales.....	35
1.6.2. Pruebas de de proficiencia en Español con Fines Específicos Académicos.....	46
1.7. Formatos de aplicación de las pruebas de ELE y de EFA.....	48

2.	Directrices para la elaboración de pruebas en Segundas Lenguas.....	49
2.1.	Consideraciones éticas	49
2.2.	Especificaciones de las pruebas.....	51
2.3.	Selección de materiales.....	69
2.4.	Diseño de tareas e ítems.....	72
2.4.1.	Directrices para la elaboración de tareas académicas.....	72
2.4.2.	Directrices para la redacción de los ítems.....	75
2.5.	Cualidades de los instrumentos de evaluación.....	76
2.6.	Análisis de los ítems.....	79
2.6.1.	La Teoría Clásica de los Test (TCT).....	79
2.6.1.1.	Índice de dificultad.....	80
2.6.1.2.	Índice de discriminación.....	81
2.7.	Sociedades internacionales de evaluación de lenguas.....	83
2.7.1.	Criterios evaluativos de las sociedades de evaluación de lenguas.....	86

CAPÍTULO II: METODOLÓGICO

3.	Metodología.....	90
3.1.	Preguntas de investigación.....	90
3.2.	Objetivos generales.....	90
3.2.1.	Objetivos específicos.....	91
3.3.	Tipo de estudio.....	91
3.4.	Procedimiento.....	92
3.4.1.	Etapa 1: Modelo de criterios evaluativos para el diseño del prototipo de prueba de EFA en línea.....	92
3.4.2.	Etapa 2: Aplicación del modelo de criterios evaluativos en la elaboración del prototipo de prueba de EFA en línea de nivel B2.....	94
3.4.2.1.	Criterios éticos y de equidad.....	94
3.4.2.2.	Criterios de diseño.....	95
3.4.2.3.	Criterios de elaboración.....	100
3.4.2.4.	Criterios de validación y pilotaje.....	106

CAPÍTULO III: RESULTADOS Y DISCUSIÓN

4. Resultados	111
4.1. Resultados de la validez de apariencia.....	111
4.2. Resultados de la validez de contenido.....	125
4.2.1. Análisis estadístico de concordancia entre los jueces (W de Kendall)...	129
4.3. Resultados de la validación del prototipo de prueba de EFA en línea según los principios de CALT.....	130
4.4. Resultados del pilotaje.....	133
4.4.1. Porcentaje de aprobación para determinar el nivel de EFA de la muestra.....	135
4.4.2. Resultados aplicación de la prueba T- Student.....	140
4.4.3. Resultados de las correlaciones.....	141
4.4.4. Análisis y validación de los ítems del prototipo de prueba.....	142
4.4.4.1. Análisis de los ítems de competencia lingüística.....	148
4.4.5. Confiabilidad de las secciones de expresión e interacción oral y escrita.....	150
4.4.6. Análisis de las dimensiones del prototipo de prueba.....	155
4.4.6.1. Análisis del comportamiento de los textos de la sección de comprensión lectora.....	159
4.4.6.2. Análisis del comportamiento de los textos de la sección de comprensión auditiva.....	161
4.4.7. Resultados de la escala Likert aplicada a la muestra.....	162
4.4.8. Análisis de los resultados de la muestra por área de estudio.....	165

CAPÍTULO IV: CONCLUSIONES Y PROYECCIONES DE ESTA INVESTIGACIÓN

5. Conclusiones	169
Bibliografía.....	179
Anexos.....	187
Anexo 1: Tabla de especificaciones nivel B2 (MCER, 2002).....	188
Anexo 2: Rúbricas para evaluar las tareas de expresión escrita y oral.....	189
Anexo 3: Escala para la evaluación de textos.....	193
Anexo 4: Implementación del prototipo de prueba de EFA en línea nivel B2.....	196
Anexo 5: Muestra de un prototipo de prueba de EFA en línea nivel B2 respondida....	202
Anexo 6: Cuestionario de validez de apariencia.....	206
Anexo 7: Pauta para la evaluación del juicio de expertos.....	210
Anexo 8: muestras de textos escritos por la muestra.....	215
Anexo 9: Transcripción de una entrevista.....	218
Anexo 10: Cuestionario de evaluación sobre la prueba EFA para los estudiantes extranjeros.....	220

Índice de figuras

Figura 1: Clasificación de los tipos de exámenes en función del uso de los resultados (Bordón, 2006).....	25
Figura 2: Fases en la elaboración de una prueba.....	26
Figura 3: Exámenes de proficiencia en ELE.....	35
Figura 4: Pantalla de acceso a la prueba CELE.....	43
Figura 5: Acceso al Aula Virtual de ELE.....	44
Figura 6: Muestra de la sección de comprensión auditiva prueba CELE.....	45
Figura 7: La competencia comunicativa (MCER, 2002).....	53
Figura 8 Ponderación y calificación de los DELE.....	67
Figura 9: Modelo metodológico para la elaboración de una prueba de EFA.....	93
Figura 10: Ejemplo de tabla de especificaciones del prototipo de prueba EFA en línea B2	97
Figura 11: Criterios para evaluar los textos.....	101
Figura 12: Ejemplo de elaboración de preguntas.....	104
Figura 13: Tarea de escritura.....	105
Figura 14: Criterios de validación y pilotaje.....	107
Figura 15: Pauta de evaluación de contenido para los jueces.....	108
Figura 16: Pregunta 3 del texto 1 de comprensión lectora.....	112
Figura 17: Pregunta 3 del texto 2 de comprensión lectora.....	113
Figura 18: Pregunta 3 corregida.....	113
Figura 19: Pregunta 4 corregida.....	113
Figura 20: Pregunta 5 corregida.....	114
Figura 21: Preguntas 2 y 4 corregidas.....	115
Figura 22: Preguntas 1 y 2 del audio 2 modificadas.....	115
Figura 23: Pregunta 2 del audio 3 corregida.....	116
Figura 24: Ejemplo de texto escrito por el sujeto 13 de la muestra.....	151

Índice de gráficos

Gráfico 1: Resultados generales prototipo de prueba EFA y prueba CELE.....	138
Gráfico 2: Resultados de la muestra en las secciones de interacción oral y escrita.....	154
Gráfico 3: Comportamiento dimensiones de corrección objetiva del prototipo de prueba EFA.....	155
Gráfico 4: Dimensiones del prototipo de la prueba EFA.....	157
Gráfico 5: Criterios de corrección en la expresión e interacción escrita.....	157
Gráfico 6: Criterios de corrección expresión e interacción oral.....	158
Gráfico 7: Comportamiento de los textos e ítems de comprensión lectora.....	160
Gráfico 8: Comportamiento de los textos e ítems de comprensión auditiva.....	161

Índice de tablas

Tabla 1: Distintos tipos de evaluación según el MCER (2002).....	22
Tabla 2: Fase I de diseño de un interfaz para una prueba de lengua basada en computador.....	31
Tabla 3: Fases II y III de usabilidad y de ensayos de pruebas en CALT.....	32
Tabla 4: Equivalencia entre diplomas y niveles del MCER.....	36
Tabla 5: Descripción del DELE B2 (versión septiembre, 2012).....	37
Tabla 6: Descripción del CELU (versión n°3).....	39
Tabla 7: Descripción del EPLE.....	40
Tabla 8: Descripción del CELA Intermedio (B2).....	41
Tabla 9: Estructura de la prueba CELE.....	43
Tabla 10: Descripción del EXELEAA.....	47
Tabla 11: Descripción del eLADE.....	48
Tabla 12: Niveles comunes de referencia definidos por el MCER (2002).....	52
Tabla 13: Duración de las pruebas de ELE y de EFA.....	68
Tabla 14: Áreas de estudios de los estudiantes del Programa ELE-UdeC.....	96
Tabla 15: Secciones del prototipo de prueba de EFA en línea B2.....	97
Tabla 16: Evaluación de los textos de comprensión lectora.....	101
Tabla 17: Evaluación de textos orales de comprensión auditiva.....	102

Tabla 18: Textos y audios que componen el prototipo de prueba.....	103
Tabla 19: Descripción de la muestra.....	109
Tabla 20: Área de estudios de la muestra piloto.....	110
Tabla 21: Número y porcentaje de aciertos en comprensión lectora.....	111
Tabla 22: Número y porcentaje de aciertos en comprensión auditiva.....	114
Tabla 23: Número y porcentaje de aciertos en la competencia lingüística.....	117
Tabla 24: Actitudes de los estudiantes sobre el prototipo de prueba de EFA.....	119
Tabla 25: Codificación de las preguntas abiertas.....	121
Tabla 26: Dimensión I. Identificación de información específica en textos especializados.....	125
Tabla 27: Dimensión II. Inferencia de las ideas principales de los párrafos.....	126
Tabla 28: Dimensión III: inferencia de la idea central del texto.....	127
Tabla 29: W de Kendall.....	129
Tabla 30: Resultados dimensión evaluación de segundas lenguas.....	131
Tabla 31: Resultados de la validación en la dimensión técnica del prototipo de prueba.....	132
Tabla 32: Porcentajes de logro de la muestra en el prototipo de la prueba EFA de nivel B2.....	133
Tabla 33: Puntajes re-escalados prototipo de prueba EFA en línea nivel B2.....	136
Tabla 34: Puntajes re-escalados prueba CELE nivel B1.....	137
Tabla 35: Valor p en las distintas secciones de las pruebas.....	140
Tabla 36: Coeficiente de correlación de Pearson.....	141
Tabla 37: Correlaciones entre las diferentes partes del prototipo de la prueba EFA y el total.....	141
Tabla 38: Muestra del número de aciertos por ítem en comprensión lectora.....	142
Tabla 39: Dificultad y discriminación comprensión lectora.....	143
Tabla 40: Índice de dificultad ítems de comprensión lectora.....	144
Tabla 41: Índice de discriminación ítems comprensión lectora.....	145
Tabla 42: Dificultad y discriminación comprensión auditiva.....	145
Tabla 43: Índice de dificultad ítems comprensión auditiva.....	146
Tabla 44: Índice de discriminación ítems comprensión auditiva.....	146

Tabla 45: Índice de dificultad y discriminación competencia lingüística.....	147
Tabla 46: Índice de dificultad ítems competencia lingüística.....	147
Tabla 47: Índice de discriminación ítems competencia lingüística.....	148
Tabla 48: Puntajes de la expresión e interacción escrita.....	150
Tabla 49: Puntajes de la expresión e interacción oral.....	152
Tabla 50: Porcentajes de apreciación de la muestra sobre el prototipo de la prueba EFA B2.....	162
Tabla 51: Codificación de las respuestas abiertas.....	164
Tabla 52: Resultados de la comprensión lectora por área de estudio.....	166
Tabla 53: Resultados de cada texto por áreas de estudio.....	166
Tabla 54: Resultados de la comprensión auditiva por área de estudio.....	167
Tabla 55: Resultados de cada audio por áreas de estudio.....	167



Agradecimientos

Al finalizar esta tesis doctoral no puedo dejar de agradecer a todos quienes han contribuido de manera relevante en esta etapa de mi formación.

En primer lugar, agradezco a Dios por concederme vida, salud e inteligencia para poder realizar este trabajo de tesis, que en momentos parecía imposible de lograr.

También le doy gracias a mi esposo, a mis padres y a mis hermanos por su paciencia, amor y apoyo incondicional, instándome siempre a superar las dificultades y lograr la meta.

Agradezco de manera muy especial a la Dra. Anita Ferreira, quien ha guiado este trabajo de tesis con rigurosidad, con sabias orientaciones y con una dedicación personal generosa, que ha ido más allá de lo que una profesora guía debe a un tesista, sin tener en cuenta horarios, noches y fines de semana. Muchísimas gracias por todo su apoyo y entrega.

Debo agradecer también al Programa de Español como Lengua Extranjera de la Universidad de Concepción (ELE-UdeC), a su Directora, Dra. Anita Ferreira, quien siempre me facilitó los medios para aplicar el prototipo de prueba a los estudiantes extranjeros. Así como también agradecer a mis colegas, Jessica Elejalde y Catherine Cáceres, por apoyarme en el proceso de pilotaje.

Mis sinceros agradecimientos a la Dra. Katia Sáez Carrillo por su amabilidad, disposición, paciencia y entrega en todo el procesamiento estadístico de los datos de la tesis.

Agradezco a la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) por la beca de Doctorado que ha financiado mis estudios doctorales durante estos cuatro años.

Agradezco también al Programa de Doctorado en Lingüística de la Universidad de Concepción por la formación y los conocimientos entregados.

Por último, quiero agradecer a todos los profesores que participaron como jueces expertos en las distintas etapas de este trabajo. Muchas gracias por su tiempo y compromiso.

Resumen

La evaluación del Español como Lengua Extranjera con Fines Específicos Académicos (EFA) presenta una serie de problemáticas que evidencian la falta de claridad respecto de lo que se debería evaluar en una prueba de EFA (Llorián, 2012). En este sentido, se necesita que las instituciones que han desarrollado pruebas de EFA muestren cuáles han sido los criterios considerados para el diseño, elaboración y validación de estos instrumentos.

Cada año, un gran número de estudiantes extranjeros llega a las universidades de habla hispana a realizar un periodo de intercambio. Sin embargo, no existen pruebas de EFA con un reconocimiento nacional o internacional que permitan medir sus competencias en el uso del español en un ámbito académico (Mendoza, 2015). Acorde con esto, se requiere de instrumentos en el ámbito del EFA que logren evaluar de manera válida y confiable el nivel de proficiencia de los estudiantes. Dichas propuestas deben adaptarse a los nuevos formatos de aplicación de pruebas de lenguas extranjeras, considerando las ventajas de los instrumentos automatizados (García y Navarro, 2008).

Teniendo en cuenta la problemática señalada, este trabajo de tesis tiene como objetivos: 1) Proponer un modelo de criterios evaluativos para el diseño de un prototipo de una prueba de EFA, en línea, para un nivel B2, 2) Implementar los criterios evaluativos del modelo en la elaboración del prototipo de prueba en una plataforma en línea y 3) Evaluar el prototipo de prueba de EFA, en línea, en términos de la validez de apariencia, de contenido y de los índices de discriminación y de dificultad del instrumento en una aplicación piloto.

Los resultados de este trabajo evidencian que el modelo de criterios evaluativos permitió elaborar un prototipo de prueba de EFA en línea para un nivel B2 con validez de apariencia y de contenido en lo referido a los ítems y a los aspectos tecnológicos. Los datos cualitativos y cuantitativos de este prototipo de prueba de EFA sugieren los criterios y procedimientos necesarios para que, en un futuro, se diseñe y elabore una prueba de proficiencia de EFA que permita precisar el dominio de los estudiantes extranjeros de intercambio en el Español como Lengua Extranjera con fines específicos Académicos.

INTRODUCCIÓN

Este trabajo de tesis se circunscribe dentro de la interdisciplina de la Lingüística Aplicada, específicamente, en lo que corresponde a la adquisición, enseñanza y evaluación de segundas lenguas o lenguas extranjeras.

En materia de adquisición y enseñanza de segundas lenguas, la evaluación es un área que ha experimentado un desarrollo importante desde sus inicios, en los años sesenta, hasta ahora (Bordón, 2008). En sus comienzos, el modelo dominante eran exámenes que evaluaban el dominio del estudiante en cuanto al sistema lingüístico de la L2 (enfoque estructural), sin considerar el uso de la lengua. En los años ochenta, con el surgimiento del enfoque comunicativo se cuestiona el hecho de que este modelo, que está bien asumido en la enseñanza, se debe adoptar en la evaluación. En los años noventa el gran desafío son los exámenes de tipo comunicativo, con todo lo que implica adoptar una perspectiva comunicativa de la lengua para su evaluación. También es una época de preocupación respecto de cómo calificar la actuación de los estudiantes. En este sentido, surgen propuestas de escalas que pretenden otorgar una puntuación lo más objetiva posible.

En la actualidad, los temas que preocupan se refieren a la evaluación de la lengua con fines específicos, la profesionalización de los evaluadores, la definición de estándares de práctica y criterios éticos que aseguren la equidad en las pruebas y que permitan elaborar instrumentos evaluativos válidos, confiables y viables (Bordón, 2008; Llorián, 2012).

En cuanto a la evaluación de lenguas con fines específicos existen cuestionamientos respecto de lo que se considera general y específico en una prueba de lengua. En este sentido, la Evaluación del Español como Lengua Extranjera con Fines Específicos Académicos (EFA) presenta una serie de problemáticas relacionadas con la falta de acuerdo entre los especialistas respecto de lo que se está midiendo en una prueba de EFA y la necesidad de contar con estudios que muestren cómo se han diseñado, desarrollado y validado las pruebas existentes (Llorián, 2012). A esto se suma que, actualmente, no hay pruebas de EFA que tengan un reconocimiento nacional o internacional (Mendoza, 2015), lo cual sitúa al español en un escenario desfavorable en el panorama de la certificación lingüística internacional. Del mismo modo, se necesita avanzar hacia formatos actualizados

de aplicación de pruebas, acorde con las nuevas realidades educativas (García y Navarro, 2008).

Teniendo en cuenta que la evaluación del EFA es una necesidad, tanto a nivel nacional como internacional, dado que no existen pruebas reconocidas ni válidas en este ámbito (Mendoza, 2015) este trabajo de tesis doctoral se plantea como objetivos: 1) Proponer un modelo de criterios evaluativos para el diseño de un prototipo de una prueba de EFA, en línea, para un nivel B2, 2) Implementar los criterios evaluativos del modelo en la elaboración del prototipo de prueba en una plataforma en línea, 3) Evaluar el prototipo de prueba en términos de la validez de apariencia, de contenido y de los índices de discriminación y de dificultad del instrumento al aplicarlo a un grupo de estudiantes extranjeros de intercambio en la Universidad de Concepción, Chile.

El modelo de criterios que se propone se sustenta en las directrices que proporcionan los especialistas en elaboración de pruebas de proficiencia y en las propuestas de las sociedades lingüísticas de evaluación de lenguas. Este prototipo de prueba se orienta a estudiantes extranjeros de intercambio en la Universidad de Concepción, por lo que el formato de aplicación en línea representa una ventaja, puesto que en un futuro podría permitir evaluar a los estudiantes aún antes de que lleguen a la universidad, y así proponerles cursos de español con fines académicos a quienes lo requieran.

Este trabajo de tesis se circunscribe en los Proyectos de investigación FONDECYT 1110812, 1140651, dirigidos por la Dra. Anita Ferreira Cabrera. Específicamente, en lo que se refiere a los lineamientos metodológicos para evaluar niveles de competencia en el Español como Lengua Extranjera.

La tesis se organiza en cuatro capítulos: En el capítulo I se aborda los fundamentos teóricos que sustentan el modelo de criterios evaluativos del prototipo de prueba, referidos a la elaboración, implementación y validación. En el capítulo II se presenta el diseño metodológico de este trabajo, en el capítulo III se presenta y analiza los resultados obtenidos. Finalmente, en el capítulo IV se presenta las conclusiones y proyecciones de este trabajo.

CAPÍTULO I: Fundamentos Teóricos

1. La evaluación en las segundas lenguas

En un sentido general, la evaluación consiste en la recogida sistemática de información con la finalidad de obtener datos que permitan tomar decisiones acertadas (Bachman, 1990; Alderson, Clapham y Wall, 1998; Pastor, 2003; Bordón, 2006). En el contexto académico es frecuente que la evaluación (y consiguiente toma de decisiones) se realice a través de exámenes. Aun cuando algunos utilicen indistintamente los términos evaluación y examen es necesario precisar la diferencia entre ellos.

La evaluación utiliza métodos cuantitativos (exámenes) y cualitativos (observaciones, valoraciones). Bordón (2006) considera que evaluar es un proceso más amplio que examinar, ya que la evaluación está asociada a la toma de decisiones y este no siempre es el objetivo de un examen, cuya finalidad puede ser el de servir de repaso o incentivo. Aun cuando, en ocasiones, puede ser evaluativo en tanto que sus resultados lleven a la toma de decisiones. En este sentido, se refiere a los exámenes que permiten asignar a un estudiante en un determinado curso de lengua y a aquellos que determinan si el candidato ha alcanzado el nivel de proficiencia establecido por el examen.

Para Bachman (1990) una prueba es un instrumento de medición diseñado para obtener una muestra específica de un comportamiento del individuo. Como un tipo de medición, una prueba se administra en circunstancias controladas (Huhta, 2008) y cuantifica necesariamente características de la persona de acuerdo con procedimientos explícitos. De modo que el usuario de la prueba pueda interpretar los resultados como la evidencia de las habilidades que presenta un individuo.

Los términos evaluación y examen se aplican a diversas disciplinas y en distintas actividades. Sin embargo, en el campo de la lengua es más complejo porque el objeto de la evaluación o examen, la lengua, es difícil de precisar. A esto se suman las necesidades derivadas acerca de la complejidad que involucra el proceso de examinar, como es el cumplimiento de los tres requisitos básicos de todo instrumento evaluativo: la validez, la confiabilidad y la viabilidad de los instrumentos utilizados.

De acuerdo con Bordón (2006), en el campo de la enseñanza de segundas lenguas (SL o L2) se necesita de la evaluación por dos razones:

- a) Cuestiones prácticas: A partir de la evaluación se pueden tomar decisiones acertadas. Se evalúa a los aprendientes para averiguar si han cumplido los objetivos de un programa determinado, asignarlos a un curso, o bien para otorgarles un diploma que avale la posesión de un nivel de lengua específico. En la mayoría de estos casos, la evaluación se realizará a través de exámenes.
- b) Valor pedagógico: La evaluación es un componente esencial del currículo de enseñanza de una L2.

En el ámbito de la evaluación de segundas lenguas, el MCER (2002) define la evaluación como la valoración del grado de proficiencia lingüístico que logra el usuario. Todas las pruebas son una forma de evaluación, aun cuando hay otras formas de evaluar que no podrían considerarse pruebas, como las listas de cotejo, la observación del profesor, etc. Por lo tanto, evaluar es un concepto más amplio que medir la proficiencia de la lengua. El MCER se refiere, específicamente, a la evaluación de la proficiencia lingüística y no al sentido amplio de lo que implica el concepto. Pastor (2003) señala que la evaluación es una operación indisociable de todo proceso de enseñanza-aprendizaje, independiente de su materialización o no en algún tipo de prueba. No obstante, afirma la autora que la evaluación ha sido la parte menos considerada del proceso de aprendizaje y que falta un modelo de evaluación acorde con el enfoque comunicativo de la lengua.

La evaluación de la lengua, específicamente de las segundas lenguas, ha experimentado un notable desarrollo desde sus inicios en los años sesenta hasta ahora. Bordón (2008) realiza un interesante resumen cronológico sobre el área de evaluación a través de los distintos enfoques de enseñanza de lenguas. Al respecto señala que en la década de los sesenta domina en el área de evaluación de lenguas un tipo de exámenes en los que se usa un procedimiento estructuralista, que revela el dominio que tiene el estudiante del sistema lingüístico de la L2, pero no ahonda más allá. Por lo tanto, no se preocupa del uso de la lengua. Se utilizan pruebas de respuesta cerrada, principalmente, de selección múltiple. Mientras que en los años ochenta con el surgimiento del enfoque comunicativo aparecen

algunas inquietudes. Se cuestiona el hecho de que este modelo comunicativo, que está bien asumido en la enseñanza, se adopte en la evaluación, ya que medir el uso de la lengua es una tarea más compleja que hacerlo a través de elementos discretos.

A fines de los ochenta señala Bachman (2000) que está claro que la evaluación de la lengua se ha posicionado en el ámbito de la Lingüística Aplicada, ya que amplió la visión de la habilidad lingüística y definió una línea de trabajo orientada a asuntos referidos tanto al análisis del discurso como a la adquisición de segundas lenguas.

En los años noventa el gran desafío son los exámenes de tipo comunicativo, con todo lo que implica adoptar una concepción comunicativa de la lengua para su evaluación. También es una época de preocupación respecto de cómo calificar la actuación de los estudiantes, en este sentido surgen propuestas de escalas que pretenden otorgar una puntuación lo más objetiva posible. En la actualidad, los temas que preocupan están referidos a: la necesidad de determinar la naturaleza de la lengua y contar con un modelo para evaluarla, la determinación del nivel de especificidad de las pruebas, la profesionalización de los evaluadores, los tres requisitos fundamentales que todo instrumento de evaluación debe cumplir (validez, confiabilidad y viabilidad) y la aplicación del desarrollo de la tecnología informática al diseño y calificación de exámenes (Bachman 2000, Bordón, 2008; Llorián, 2012).

Definido el contexto general de la evaluación de segundas lenguas es pertinente revisar la situación de la evaluación de las segundas lenguas con fines específicos y de manera focalizada, del español con fines específicos académicos.

1.1. La evaluación de segundas lenguas con fines específicos

La enseñanza del español con fines específicos surge en los años ochenta y en la actualidad presenta dos variantes: los fines profesionales y los fines académicos. Acorde con Vázquez (2004), el español con fines académicos se puede caracterizar como una disciplina que “investiga las características de los géneros académicos -en un amplio sentido de la palabra” (Vázquez, 2004: 1130) con la finalidad de facilitar a las personas no nativas de una lengua la adquisición de destrezas que le permitan cumplir de manera exitosa las tareas propias de los ámbitos universitarios.

La evaluación de lenguas con fines específicos se diferencia de los fines generales porque considera el uso de la lengua en ámbitos especializados. De acuerdo con lo señalado por Llorián (2012) existe una serie de cuestiones no resueltas en el ámbito de la evaluación con fines específicos.

Las principales problemáticas que enfrenta la evaluación con fines específicos se relacionan con: 1) El constructo que se está midiendo en una prueba con fines específicos, 2) El diseño y desarrollo de pruebas que permitan evaluar de manera válida, 3) La definición de lo que se considera específico en una prueba de lengua y 4) La orientación de los fines. Estas problemáticas se relacionan con la validez que debe mostrar todo instrumento evaluativo, para lo cual es relevante contar con un modelo de criterios que permita dar garantías de lo que se está evaluando y cómo se está llevando a cabo.

En relación con la primera problemática, no se tiene claridad respecto de lo que está midiendo en una prueba con fines específicos. Esto ha llevado a que algunos duden si realmente este tipo de pruebas son necesarias (Douglas, 2000), puesto que lo que se busca evaluar es la habilidad del estudiante de usar la lengua con propósitos comunicativos. Si se evalúa al candidato en función de los conocimientos que tiene sobre algún tema específico, entonces la medición podría estar influida por otras variables no lingüísticas. Por otra parte, quienes están a favor de las pruebas con fines específicos argumentan que la especificidad debe basarse en la noción de *autenticidad*, por lo que las tareas de la prueba serán específicas en la medida que se basen en situaciones a las que se verá enfrentado el estudiante en su vida real. En consecuencia, lo que se busca es que el candidato haga uso de manera integrada de sus conocimientos y de su dominio lingüístico-comunicativo.

Lo expuesto en esta primera problemática lleva a la cuestión de qué tan generalizables pueden llegar a ser estas pruebas. En este sentido, hay que tener en cuenta que no existe una prueba buena o mala, y que en la medida que se restrinja más el grado de especificidad de una prueba servirá a un menor número de estudiantes (Llorián, 2012).

El segundo cuestionamiento se refiere a que los contextos de uso de la lengua varían, por lo que, no está claro cómo se debería elaborar una prueba con fines específicos. En este sentido, se sugiere que se trabaje con tareas auténticas que muestren los aspectos claves considerados en un contexto específico, dado que lo que se evalúa son las habilidades y no el dominio disciplinar sobre un tema en concreto (Llorián, 2012).

La tercera problemática se relaciona con la definición de lo que se considera específico en una prueba de lengua, en contraste con una prueba general. Para ello, Dudley-Evans y St. John (1998) sugieren que se considere una línea de continuidad, por lo que en un extremo se sitúe los fines generales y en el otro los fines específicos. Así, una prueba podría ser más o menos específica. Otra sugerencia para subsanar esta dificultad se relaciona con la tendencia en la búsqueda de las bases comunes, que sean transversales a las áreas de los fines específicos (Hyland, 2002). En este sentido, habría que observar aspectos recurrentes o generalizables, por ejemplo, en el ámbito académico.

Finalmente, la cuarta complejidad tiene que ver con los distintos tipos de fines, entre ellos, se reconocen los fines específicos: de negocios, de turismo, de las ciencias de la salud, de ingeniería, de ciencia y tecnología, académicos, etc. Sin embargo, la línea divisoria entre estos distintos fines no está definida y no es fácil conseguirlo.

1.1.1. La evaluación del Español con Fines Específicos Académicos (EFA)

En relación con la evaluación del español con fines académicos, no está claro lo que una prueba tendría que evaluar en este ámbito, dado que se necesita delimitar los contextos de uso que debería abordarse en las pruebas, los cuales tienen que dar cuenta de las distintas habilidades que necesitan demostrar los candidatos en este ámbito. Es sabido que no se puede desarrollar un examen individualizado o dirigido a grupos restringidos, motivo por el cual se requiere de investigaciones que contribuyan en la determinación de los contextos de uso de lengua para fines académicos. Lo que hay a nivel investigativo en cuanto a elaboración y resultados de pruebas con fines específicos no cumple con los estándares definidos en los códigos de práctica por las distintas sociedades de evaluación (Llorián, 2012).

En la evaluación del Español como Lengua Extranjera (ELE), a diferencia del inglés, hay un desarrollo muy incipiente en cuanto a la evaluación para fines específicos, y especialmente, para el ámbito académico. Aun cuando se observa un aumento en el número de interesados en certificar su dominio en los fines específicos. En el caso de los fines académicos, pocos estudiantes extranjeros de intercambio en universidades de países donde se habla el español presentan algún tipo de certificado que acredite su dominio del español con fines académicos. Esto refleja que los actuales sistemas de admisión universitaria no siempre

exigen una certificación del español en este ámbito, a diferencia del inglés o de otras lenguas extranjeras (LE). En este sentido, se requiere de esfuerzos desde las universidades y de los organismos gubernamentales para aunar criterios que ayuden a mejorar esta situación (Llorián, 2012).

Desde la perspectiva evaluativa, Llorián (2012) plantea algunas interrogantes relacionadas con: la necesidad de contar con un modelo evaluativo para el español con fines específicos, la definición de lo que se considera específico, los criterios para evaluar dicho lenguaje y las garantías de evaluar este tipo de pruebas de manera válida y confiable. A esto se suma, que no está claro en qué medida estos instrumentos se pueden vincular a un marco de referencia externo, en este caso, el MCER (2002). En este contexto, además se requiere de equipos de redactores y calificadores que tengan un dominio disciplinario en las temáticas planteadas en el examen, para así hacer preguntas que eliciten las habilidades que tendrá que mostrar el estudiante en su desempeño académico y los ítems no queden reducidos a lo léxico o lo gramatical. Finalmente, concluye Llorián (2012) que en la evaluación del español como segunda lengua o lengua extranjera con fines generales y específicos se necesita que las instituciones den a conocer, a través de artículos o informes, el proceso que han llevado a cabo para calificar y validar las pruebas que han desarrollado. Se trata de problemáticas no resueltas en la evaluación del español con fines académicos, y cuyas respuestas son necesarias para lograr la profesionalización de la evaluación de proficiencia en los países de habla hispana (Martínez, 2011).

De acuerdo con Regueiro y Saéz (2013) el lenguaje académico es aquel que se utiliza en los textos orales y escritos del ámbito universitario y que son elaborados por los miembros de la comunidad académica. Sin embargo, se necesita de investigaciones que aporten cuáles son las competencias que se necesita en este ámbito, el tipo de tareas, el grado de especificidad de estas, etc. En el lenguaje académico se utiliza distintos tipos de estilos académicos: *estilo académico o especializado*, *estilo académico formativo* y *estilo académico divulgativo*. El estilo académico se caracteriza por un estilo especializado que tiene como meta la objetividad y es utilizado por la comunidad científica en géneros discursivos como el artículo científico, la tesis, la reseña, etc. El estilo académico formativo corresponde al que usan los investigadores y científicos en su papel de profesores ya sea en las aulas, en los

manuales, en las guías, etc. y el estilo académico divulgativo corresponde al que se utiliza para comunicar resultados a la sociedad en la prensa, en la radio, en la televisión, etc.

1.1.2. El discurso académico en EFA

El discurso académico se entiende como el conjunto de producciones textuales en el ámbito de la enseñanza e investigación científica. Estas producciones textuales tienen como objetivo difundir un conocimiento de la disciplina y contribuir en la formación de los miembros de la comunidad de especialistas (Parodi, 2012).

Los estudios sobre el discurso académico surgen a principios de los años ochenta (Vázquez, 2001), motivados por la idea de que el lenguaje debe ser estudiado dentro de sus ámbitos naturales de uso. Sin embargo, la primera investigación en este ámbito en el Español como Lengua Extranjera (ELE) data de 1997 y corresponde a los resultados del proyecto ADIEU (Discurso Académico en la Unión Europea). En dicho proyecto el concepto de discurso académico se entiende de una manera práctica como la comunicación que tiene lugar entre estudiantes y docentes en el contexto de las clases magistrales, las lecturas universitarias, los criterios de bondad al momento de evaluar no solo los exámenes sino también las horas de consulta (Vázquez, 2001). Este proyecto tiene como objetivo explorar las características del discurso académico (universitario) español para así orientar la inserción de estudiantes de otras lenguas en las universidades españolas. De esta forma se garantiza que los estudiantes puedan aprobar sus asignaturas y obtener los créditos exigidos.

El proyecto ADIEU centra su atención en el estudio de algunos géneros del discurso académico oral y escrito en español. Específicamente, desde una perspectiva didáctica se centra en la clase magistral, la monografía y el artículo científico, como los géneros académicos más utilizados en el contexto de los intercambios estudiantiles en universidades españolas. No obstante, existe un amplio número de otros géneros académicos que habría que considerar, tanto orales como escritos: las exposiciones y presentaciones orales, las pruebas, las reseñas, los resúmenes, los apuntes, etc. (Regueiro y Sáez, 2013).

El objetivo principal del Proyecto ADIEU es aportar con evidencia empírica respecto de cómo se concibe la clase magistral, la lectura y escritura de artículos y de monografías en el

contexto académico español. De esta manera se pretende apoyar a los estudiantes extranjeros que realizan intercambios, pues aunque estos manejen estos géneros en su lengua materna (L1) existen algunas diferencias en la lengua extranjera que deben tenerse en cuenta para lograr un buen desempeño académico. Por ello, lo que pretende la evaluación del español con fines específicos académicos es medir lo que el estudiante es capaz de hacer en cuanto al manejo del discurso académico en la lengua meta (Pastor, 2006).

1.1.2.1. Los géneros académicos en EFA

Los géneros son formas discursivas convencionales que se van constituyendo históricamente en una cultura determinada. Corresponden a productos socioculturales, reconocibles entre los miembros de una comunidad (Alexopoulou, 2010). No obstante, están sujetos a variación dado que responden también a los cambios sociales.

Regueiro y Sáez (2013) mencionan una lista de géneros académicos y los dividen en función de las habilidades de comprensión y de expresión. Así, distinguen:

- Géneros académicos de comprensión: clases magistrales, manuales, monografías, artículos científicos y capítulos.
- Géneros académicos de expresión:
 - De estudio propio: apuntes, resúmenes.
 - En clases: exámenes, reseñas, *abstract*, monografías, exposición, presentación.
 - Fin de ciclo: tesis.

1.2. Tipos de evaluación

Refiriéndose a los tipos de evaluación, Pastor (2003) menciona aquella distinción que se da entre la evaluación cuantitativa y cualitativa. La primera se lleva a cabo mediante pruebas objetivas por lo que se concreta mediante cifras y análisis estadísticos, mientras que la segunda se refiere a observaciones, encuestas, cuestionarios, entrevistas, procedimientos que no se pueden materializar a través de datos numéricos. De estas dos formas de evaluación, se suele dar mayor credibilidad a la cuantitativa por su supuesto valor científico. No obstante, Pastor (2003) señala que no se puede subestimar la evaluación cualitativa, dado que frecuentemente resulta clarificadora, sobre todo cuando se evalúa el proceso de aprendizaje del estudiante.

Como se observa, la evaluación de una segunda lengua o lengua extranjera (LE) se puede llevar a cabo a través de distintas maneras. Bordón (2006) se refiere a los tipos de evaluación considerando tres criterios fundamentales: 1) los objetivos que se persiguen a través de ella, 2) los procedimientos adoptados para llevarla a cabo y 3) la forma en que se valorarán los resultados. De acuerdo con los objetivos, la evaluación puede enfocarse en recoger información que permita influir en la formación del estudiante, mejorando su aprendizaje. O bien utilizar procedimientos que permitan la obtención de datos suficientes para entregar una calificación al estudiante que refleje si ha alcanzado o no los logros. El primer caso se refiere a la evaluación formativa, y el segundo a la sumativa. Según el procedimiento, hay que distinguir entre la evaluación del proceso y la evaluación de los resultados. La primera tiene como objetivo obtener información del proceso de aprendizaje del aprendiente y la segunda, averiguar si el estudiante ha alcanzado un determinado nivel o superado unos objetivos, al margen de cómo haya adquirido/aprendido para hacerlo. En cuanto a la valoración de los resultados, hay tanto una evaluación subjetiva que exige la presencia de un evaluador y otra objetiva que permite que los resultados sean revisados mediante procedimientos mecánicos (ej. máquinas de lectura óptica, análisis automático).

En contraste con la tipología de Bordón (2006), el MCER (2002) distingue una amplia variedad de tipos de evaluación, que incluye trece pares, como se muestra en la tabla 1.

1	Evaluación del aprovechamiento	Evaluación del dominio
2	Con referencia a la norma (RN)	Con referencia a un criterio (RC)
3	Maestría RC	<i>Continuum</i> RC
4	Evaluación continua	Evaluación en un momento concreto
5	Evaluación formativa	Evaluación sumativa
6	Evaluación directa	Evaluación indirecta
7	Evaluación de la actuación	Evaluación de los conocimientos
8	Evaluación subjetiva	Evaluación objetiva
9	Valoración mediante lista de control	Valoración mediante escala
10	Impresión	Valoración guiada
11	Evaluación global	Evaluación analítica
12	Evaluación en serie	Evaluación por categorías
13	Evaluación realizada por otras personas	Autoevaluación

Tabla 1: Distintos tipos de evaluación según el MCER (2002)

La amplia tipología presentada en la tabla 1 muestra las distintas posibilidades de llevar a cabo la evaluación, desde una evaluación de aprovechamiento y dominio hasta llegar a la autoevaluación, donde el propio estudiante valora sus logros y deficiencias en el transcurso del proceso de aprendizaje.

La evaluación de aprovechamiento/evaluación del dominio o proficiencia hace referencia a los conocimientos del estudiante. El aprovechamiento se refiere a si el estudiante ha sacado “provecho” de la enseñanza, mientras que la evaluación de dominio corresponde a lo que el candidato sabe hacer con la lengua, independiente de cómo lo haya aprendido. El segundo par se refiere a cómo se realiza la calificación de los estudiantes, si es con referencia a la norma o bien en función de sus propios resultados, con referencia a un criterio. La tercera clasificación responde a si la referencia al criterio (RC) se realiza en función de alcanzar un cierto nivel de maestría o bien la RC se decide de acuerdo a una línea continua que considera de forma gradual todos los niveles pertinentes del dominio. El cuarto par, opone la evaluación continua que se puede llevar a cabo a lo largo de un periodo utilizando procedimientos cualitativos, a la evaluación en un momento concreto, que supone la aplicación de un examen. En la quinta clasificación, se contrasta la evaluación formativa y la sumativa, la primera se refiere a la evaluación como un proceso continuo (se relaciona con la evaluación continua); mientras la sumativa resume en una calificación el aprovechamiento de los estudiantes al final de un curso o de una unidad. La sexta clasificación corresponde a la evaluación directa/indirecta que se refiere a la forma de acceder a evaluar las diferentes

destrezas, la primera corresponde a la actuación del estudiante en un momento específico (es observable, por ejemplo, la expresión oral), y la segunda se refiere a la evaluación de las destrezas comprensivas a las cuales no se puede acceder de manera directa (comprensión lectora). El séptimo par corresponde a la evaluación de la actuación/conocimientos. La actuación se evalúa de manera directa con una muestra de la lengua oral o escrita del estudiante y los conocimientos se evalúan de manera indirecta, por ejemplo, mediante la completación de un texto con determinadas formas lingüísticas. En el octavo lugar, la evaluación subjetiva/objetiva hace referencia a la forma en que se valora lo que realiza el estudiante. En la evaluación subjetiva el evaluador, basado en una rúbrica o pauta, puntúa el desempeño del estudiante, mientras que en la evaluación objetiva la corrección está definida de antemano, se trata de pruebas con ítems que tienen definidas las respuestas correctas. El noveno par, la valoración mediante escala consiste en asignar a un estudiante a un nivel o banda de la escala, acorde con su desempeño; mientras que la valoración a través de lista de control evalúa al estudiante en función de una lista de aspectos considerados pertinentes para un nivel específico. En el décimo lugar se encuentra la impresión versus la valoración guiada. La primera corresponde a la apreciación subjetiva del evaluador a partir de su experiencia sobre la actuación del estudiante en la clase, mientras que en la segunda se reduce esta subjetividad mediante el uso de un procedimiento evaluativo basado en unos criterios definidos sobre el desempeño del alumno. En el puesto número once se ubica el par evaluación global/evaluación analítica. La evaluación global consiste en la valoración general del evaluador sobre el desempeño del estudiante; mientras que la evaluación analítica consiste en la valoración de distintos aspectos del dominio lingüístico de manera separada. La clasificación número doce corresponde a la evaluación en serie y por categorías. La evaluación en serie involucra una sucesión de tareas aisladas que se evalúan con una calificación global según una escala de puntaje definida; mientras que la evaluación por categorías corresponde a una sola actividad donde se valora la actuación del estudiante en función de distintas categorías con puntuaciones. Finalmente, el par número trece corresponde a la evaluación realizada por otras personas y a la autoevaluación. En la primera, el profesor o examinador evalúa el dominio del estudiante; mientras que en la segunda es el propio alumno quien evalúa su desempeño en cuanto a su dominio lingüístico.

Para determinar el tipo de evaluación a utilizar, Rodríguez (2000), citado y adaptado por Bordón (2006), enuncia cinco preguntas fundamentales que ayudan a definir la evaluación pertinente. Estas preguntas son:

- 1) *¿Qué se evalúa?*: El aprendizaje de los alumnos u otra cosa.
- 2) *¿Quién evalúa?*: Los profesores, los alumnos o agentes externos.
- 3) *¿Cómo se evalúa?*: Definir qué instrumentos se usarán, cualitativos o exámenes, qué formato tendrán. La interpretación de los resultados será en relación a la norma o a un criterio.
- 4) *¿Cuándo se va a evaluar?*: En qué momento del proceso, al inicio, en el medio o al final.
- 5) *¿Para qué se evalúa?*: Controlar el proceso de enseñanza, distribuir a los estudiantes, conseguir una certificación, etc.

Teniendo claridad respecto de los objetivos, los procedimientos y la valoración de los resultados de la evaluación, se puede seleccionar el tipo de prueba que corresponda.

1.3. Tipos de pruebas

La evaluación de una segunda lengua se puede llevar a cabo a través de distintas pruebas. Pastor (2003) distingue cuatro tipos:

- *Pruebas de aptitud lingüística*: Miden la capacidad de una persona para adquirir una segunda lengua. Evalúan la capacidad del sujeto en cuanto a la memorización, inducción de significados, capacidad para codificar la gramática y fonética de la lengua que está aprendiendo.
- *Pruebas de proficiencia*: Evalúan los conocimientos que el estudiante tiene de una segunda lengua, de modo general, sin vincularlo a un programa concreto de enseñanza. No interesa el proceso de aprendizaje que ha llevado el alumno, sino lo que sabe hacer en el momento de enfrentarse a la prueba (Ej. pruebas DELE).

- *Pruebas de rendimiento:* Miden cuánto ha aprendido el estudiante de acuerdo con los objetivos del curso o programa. Se suelen aplicar al término de una unidad, un curso o de un trimestre.
- *Pruebas de nivel:* Permiten distribuir a los estudiantes en el curso que les corresponda con el objetivo de formar grupos equilibrados en cuanto a su nivel de lengua.

Bordón (2006), refiriéndose también a la tipología de pruebas las organiza en función de dos objetivos evaluativos: el aprovechamiento y el dominio.

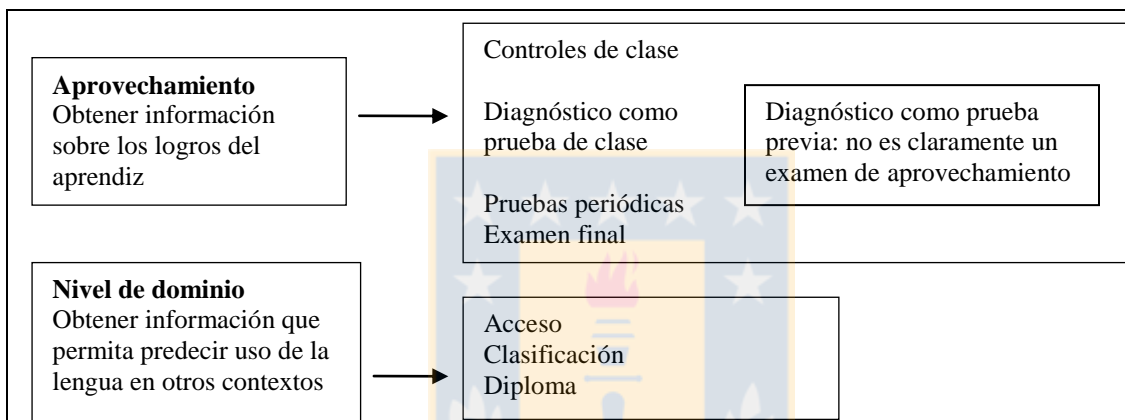


Figura 1: Clasificación de los tipos de exámenes en función del uso de los resultados (Bordón, 2006)

Esta clasificación se diferencia de la presentada por Pastor (2003), ya que organiza los tipos de pruebas atendiendo al objetivo que se persigue. Bordón (2006) divide los tipos de pruebas en torno a dos grandes objetivos, medir el dominio o el aprovechamiento. Dentro del nivel de aprovechamiento están los controles de clase, el diagnóstico, las pruebas periódicas y el examen final, y en el dominio se encuentran los exámenes de acceso, clasificación y diploma. Dicha clasificación está hecha en función de la finalidad con la que se interpretarán los resultados de estos instrumentos.

1.4. Fases en la elaboración de pruebas

La forma más efectiva (por economía, rapidez, equidad) para obtener información sobre lo que las personas saben sobre algo es hacerlo a través de una prueba. Para construir una prueba efectiva se deben tener en cuenta las distintas etapas que este proceso comprende, de modo que se logre construir un instrumento válido, confiable y viable. A partir de las propuestas recopiladas por Bordón (1991; 2006) se define las distintas fases consideradas en la confección de una prueba.

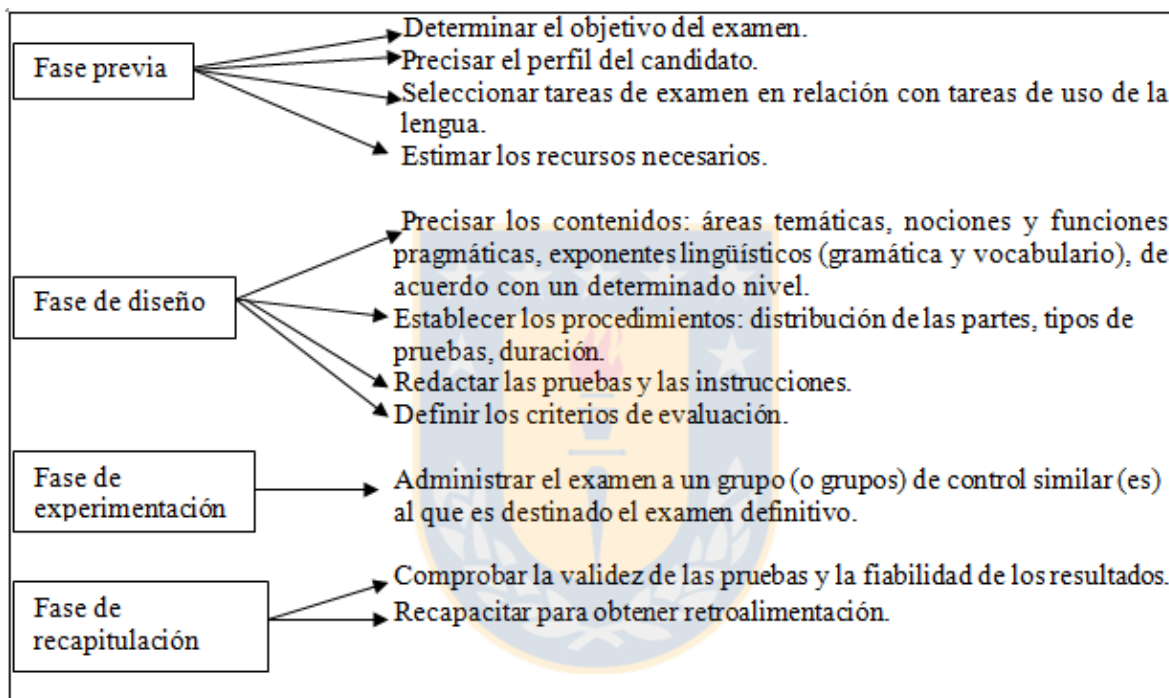


Figura 2: Fases en la elaboración de una prueba

De acuerdo con Bordón (2006), cuatro son las fases que se debe considerar en la confección de una prueba. En la fase previa lo primero que se debe determinar es el objetivo, ya que a partir de este se seleccionará el tipo de instrumento que corresponda. La elección del tipo de prueba a utilizar se realiza teniendo en cuenta el objetivo hacia el cual se orienta el instrumento. En cuanto a los tipos de pruebas, las más comunes se confeccionan e implementan para medir tanto el progreso en el desarrollo del uso de la lengua meta como el nivel de proficiencia de esta (Pastor, 2003). Con respecto a las pruebas de proficiencia, estas permiten medir los conocimientos que tiene el aprendiz de una segunda lengua, de modo general, sin un requerimiento sistemático de inscripción a un programa de enseñanza

determinado. Por otra parte, las pruebas de nivel son las que permiten distribuir a los estudiantes en un curso determinado. Finalmente, las pruebas de rendimiento evalúan los conocimientos y destrezas lingüísticas aprendidas tras el desarrollo de un curso. Este tipo de pruebas suele realizarse al final del curso y, a su vez, “permiten comprobar el nivel real de adquisición de los alumnos, así como el consiguiente éxito o fracaso de la práctica docente” (Pastor, 2003: 510). Además, en la fase previa se define el perfil de los estudiantes a quienes va dirigida la prueba, se selecciona las tareas a utilizar y se estima los recursos que se emplearán en el diseño del instrumento (recursos humanos, materiales, aulas, etc.).

En la fase de diseño se delimita los contenidos temáticos y lingüísticos acordes con el nivel de proficiencia hacia el cual se orienta la prueba, se distribuye las secciones de la prueba, la duración de ellas, además de redactar los distintos ítems y las instrucciones correspondientes. Dependiendo del tipo de prueba elegida se seleccionarán los contenidos que corresponda, si es una prueba de aprovechamiento entonces el instrumento tendrá que responder a los contenidos revisados en el curso. Si se trata de una prueba que evalúe proficiencia, entonces se debe definir todos los contenidos atendiendo a la complejidad de los distintos niveles de competencia hacia los cuales se dirige la prueba. En una prueba de nivel se busca corroborar los conocimientos que tiene el estudiante, por lo que los contenidos que abarca no son tan exhaustivos como en una prueba de proficiencia, ya que el objetivo es proporcionar una muestra de lo que el estudiante puede hacer con la lengua para poder asignarlo al curso adecuado.

La fase de experimentación corresponde a la aplicación de la prueba al grupo meta. Los resultados obtenidos en esta etapa permiten avanzar a la fase de recapitulación donde se comprueba la validez, la confiabilidad y la viabilidad del instrumento, aportando datos que permitan mejorar la prueba. Un instrumento de evaluación será eficaz mientras proporcione los datos que permitan responder a los requisitos esenciales que toda buena prueba debe tener: validez, confiabilidad y viabilidad (Bordón 2006; MCER, 2002).

El esquema presentado por Bordón (2006) es un lineamiento general para la evaluación de segundas lenguas, que sistematiza y delimita cada uno de los pasos a considerar en la elaboración de instrumentos de evaluación de lenguas. No obstante, hay aspectos en los que

este esquema no ahonda como son los criterios éticos, el proceso de elaboración y de validación, asuntos de gran relevancia para construir instrumentos que midan de manera efectiva la proficiencia en la lengua (Martínez, 2011).

1.5. La evaluación de lenguas a través del computador (CALT)

La evaluación de lenguas a través del computador, del inglés, *Computer Assisted Language Testing* (CALT) es una rama de la interdisciplina de CALL (del inglés, *Computer Assisted Language Learning*). Su constitución se debe a que los investigadores de CALL prestaban poca atención a la evaluación de la lengua, solo un grupo pequeño de investigadores hacía uso de los computadores para elaborar y validar pruebas de idiomas (Noijons, 1994). CALT se define como un procedimiento integrado en el cual se elicit y evalúa el uso de la lengua (Noijons, 1994). Sin embargo, la evaluación de la lengua a través de la tecnología no ha estado exenta de controversia y debate respecto de si son mayores las ventajas que las desventajas. Los defensores (Noijons, 1994; Chapelle y Douglas, 2006) señalan que CALT ofrece un gran número de potencialidades frente a las pruebas de aplicación tradicional (de lápiz y papel), entre ellas consideran:

- La posibilidad de medir el tiempo que el estudiante invierte en responder la prueba, así como también en sus diferentes partes.
- El registro de la información sobre las rutas de navegación que realiza el estudiante durante la prueba.
- El almacenamiento de la información en relación con la aplicación de la prueba. El candidato puede tener acceso a ella cuando lo requiera.
- Rapidez en los cálculos psicométricos de la prueba, por ejemplo, índices de dificultad o de discriminación.
- El computador permite el acceso a una variedad de información y tipos de ítems.

- La facilitación para la estandarización de una prueba, no solo en cuanto a los ítems sino también en relación con el formato. Una prueba aplicada en circunstancias similares permite obtener resultados más confiables, independiente del examinador.

Algunas de las problemáticas que se observa en el ámbito de CALT se refieren al dominio que debe mostrar el equipo que elabora una prueba. Este debe tener conocimiento sobre lo que se está evaluando y sobre plataformas para realizar pruebas. Dado que muchos profesores y desarrolladores tienen una experiencia limitada en ambas áreas no es casual que haya pocos trabajos de buena calidad en CALT (Noijons, 1994).

De acuerdo con Chapelle y Douglas (2006), la metodología de CALT considera dos etapas para la elaboración de pruebas válidas mediatizadas por computador. La primera etapa es de implementación de la prueba en la plataforma computacional, y la segunda de evaluación de la prueba a partir de los resultados obtenidos en la aplicación del instrumento. A continuación se describe cada una de estas etapas.

1.5.1. La implementación en CALT

La implementación hace referencia a los recursos necesarios para elaborar una prueba, lo que dependerá de su propósito, por ejemplo, una evaluación que hace el profesor en la clase no necesita de una mayor infraestructura. No obstante, las pruebas de alto impacto requieren de un conjunto de especificaciones y procedimientos para la selección de tareas adecuadas acorde con estas especificaciones, así como también de métodos para recopilar, almacenar y analizar las respuestas de los candidatos.

Chapelle y Douglas (2006) señalan que, anteriormente, en el contexto de CALT se trabajaba con sistemas más sofisticados, que podrían hacer del uso del computador algo costoso y poco alcanzable. Sin embargo, en la actualidad se aborda propuestas sencillas, que permitan mayor accesibilidad para trabajar con esta metodología. Chapelle y Douglas (2006) describen y analizan el proceso de implementación de una prueba en la herramienta Respondus, utilizada en la Universidad de Iowa. Para ello, proponen los siguientes pasos a nivel de implementación de una prueba en computador:

1. Creación de la prueba en la herramienta: En esta etapa corresponde señalar los pasos que involucra crear o subir el instrumento, tales como: la creación del espacio de la prueba, los ajustes, subir las preguntas, número de alternativas, guardar las preguntas, etc.
2. Publicación de la prueba en la web: Corresponde a la etapa de observación de la vista previa de la prueba y su visualización final en la plataforma donde se aplicará.
3. Dar a los estudiantes el acceso a la prueba: Corresponde poner la prueba a disposición de los estudiantes, para ello se requiere programar el tiempo que durará, la fecha, el registro del alumno y el usuario para el ingreso.
4. Aplicación de la prueba: Corresponde a la etapa en que el estudiante rinde la prueba frente al computador. Para ello, debe interactuar con la herramienta, mediante la lectura de instrucciones y selección de botones.
5. Visualización de los resultados: El examinador accede a los resultados y puntajes obtenidos por los estudiantes para realizar los análisis pertinentes.

Una vez que la prueba ha sido implementada bajo las directrices de la metodología CALT corresponde evaluarla desde una perspectiva de lo que se ha sugerido en el ámbito de CALT (Chapelle y Douglas, 2006).

1.5.2. La evaluación en CALT

La evaluación en este contexto considera dos aspectos: 1) evaluación de la prueba en su modalidad computacional por parte de expertos en CALT y 2) mediante la respuesta de quienes rinden la prueba implementada en esta modalidad. En este sentido, es útil considerar la propuesta de Fulcher (2003), quien desde la perspectiva del desarrollador de la prueba, define las siguientes fases: 1) Planificación inicial, 2) Usabilidad de la prueba y 3) Ensayos y ajustes.

Cuestiones de diseño	Consideraciones y procesos generales	Puntos de interés y sugerencias para los diseñadores
1. Diseño de prototipos	Consideraciones de hardware y de software	Especificaciones del computador, resolución de la pantalla, tiempo de descarga, compatibilidad del navegador, software de terceros, software de autoría.
2. Buen diseño de interfaz	Navegación	Botones e íconos de navegación, operación de botones del sistema y las instrucciones; facilidad y velocidad de la navegación; claridad de los títulos de página; medidas de seguridad.
	Diseño de la página	Textos breves, reducir el desplazamiento (scrolling) en la página, evitar letras mayúsculas.
	Texto	Evitar complejidades, las animaciones distraen, usar tamaño de letra sobre 10 puntos, tomar decisiones respecto de si los examinandos pueden alterar el tamaño de letra, evitar la mezcla de tipos de letras.
	Color de texto	Maximizar el contraste entre los colores si estos son significativos (instrucciones y pasajes de textos)
	Controles y barras de herramientas	Presentar el menor número de opciones posibles y poner en primer lugar la información más importante
	Íconos y gráficos	Minimizar el número de íconos y optimizar el tamaño para el aspecto visual y el tiempo de descarga; evitar imágenes animadas o con destellos
	Facilidades de ayuda	Tomar decisiones respecto de la disponibilidad de ayuda
	Aspectos externos de la prueba	Considerar el uso de fuentes externas para evaluar lo que se ha construido
	Tipos de ítems	Optimizar la cantidad de ítems necesarios en función de lo que puede realizar el desarrollador
	Multimedia	Asegurar la eficiencia de la presentación y el tiempo de descarga
Formato y espacio para la escritura y mensajes de feedback para las tareas	Dejar espacio suficiente para las respuestas; alinear y justificar los textos; momento y lugar del feedback	
3. Concurrente/Simultáneo	Desarrollo de sistemas de	

Actividades de la fase 1	administración; investigación sobre la entrega y almacenamiento del puntaje; distribución de las secciones; puntuación por humanos; rúbricas, familiarización con estudios; elaborar estudios a pequeña escala para medir la validez de constructo	
--------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

Tabla 2: Fase I de diseño de un interfaz para una prueba de lengua basada en computador

Fase	Categoría de diseño
II. Usabilidad de la prueba	<ol style="list-style-type: none"> 1. Buscar problemas y solucionarlos 2. Seleccionar a un grupo de interesados en tomar la prueba para estudios de usabilidad 3. Actividades simultáneas: redacción de ítems, pilotajes, ensayar la asignación de puntajes mediante rúbricas
III. Ensayos y ajustes	<ol style="list-style-type: none"> 1. Verificar que la interfaz funciona en los sitios y plataformas 2. Actividades simultáneas: el desarrollo de tutoriales, muestras de ejemplos de la prueba para la práctica, desarrollo de programas de capacitación y la formación de evaluadores, valoración de los estudios de escala y de entrega de puntajes, planificación de nuevos estudios de validación

Tabla 3: Fases II y III de usabilidad y de ensayos de pruebas en CALT

Estas directrices dan cuenta de las múltiples consideraciones para realizar pruebas mediatizadas por computador, teniendo en cuenta que actualmente hay un gran interés por automatizar los exámenes de lenguas extranjeras o SL.

1.5.3. Ventajas de la implementación de pruebas de EFA a través del computador

En la actualidad, hay un gran interés por automatizar los exámenes de lenguas extranjeras o segundas lenguas. A partir de 1997 algunas instituciones internacionales de evaluación de lenguas extranjeras, como el *English Testing Service* (ETS), encargada de administrar el examen TOEFL, ha implementado sistemas automáticos en la evaluación del inglés como lengua extranjera. Es así como en una primera etapa se utilizó el formato *Computer Based TOEFL* y más recientemente el *Internet Based TOEFL*. No obstante, en el caso del español como lengua extranjera se ha realizado pocos intentos de automatización de pruebas de proficiencia, de hecho las pruebas de mayor reconocimiento internacional (DELE y CELU) continúan aplicándose en formato de lápiz y papel. En este contexto, conviene preguntarse ¿Cuál es el aporte o la contribución de las pruebas automatizadas? Para García y Navarro (2008), la aplicación de pruebas en lápiz y papel tiene un costo elevado tanto para los estudiantes como para las instituciones. Este costo se observa en aspectos como: el desplazamiento del candidato al centro donde se aplica la prueba, el tiempo de entrega de los resultados y la necesidad de contar con salas, examinadores, correctores y cuidadores. En este sentido, una prueba en línea representa una ventaja porque es menos costosa. Además, mediante este formato de implementación se tiene flexibilidad en cuanto a los diferentes tipos de ítems, se puede evaluar globalmente las destrezas comunicativas, dar una corrección rápida y eficaz y las pruebas orales se pueden realizar sin la presencia de los evaluadores, etc.

1.5.4. El uso de Moodle en la elaboración de pruebas de segundas lenguas

En la actualidad existe una serie de herramientas informáticas de código libre que permiten elaborar cuestionarios y pruebas para evaluar distintos ámbitos del conocimiento. La plataforma Moodle (*Module Object-Oriented Dynamic Learning Enviroment*) se perfila como una de las herramientas de mayor funcionalidad operativa en cuanto a la realización de pruebas. Jerez *et al* (2012) compararon el uso de tres plataformas para desarrollar pruebas que permitieran evaluar de manera objetiva a distancia: Acro Tex, Lime Survey y Moodle. En dicha comparación, Moodle presentó mayores ventajas en cuanto a la funcionalidad y la seguridad que involucra la aplicación de una prueba.

De acuerdo con Jerez *et al* (2012), las principales características de Moodle en el contexto de elaboración de pruebas se refieren a:

- La administración de los usuarios que se realiza desde afuera. El administrador registra a los usuarios, les asigna un usuario y una clave.
- La creación de usuarios con distintos privilegios (estudiante, profesor).
- Un alto nivel de seguridad, dado que solo se puede acceder a una prueba si se cuenta con un usuario y clave. Además, la plataforma permite activar y desactivar la prueba por fecha y hora.
- La flexibilidad en cuanto a cambiar la configuración de la prueba, incluso cuando se está aplicando (en marcha).
- La flexibilidad en cuanto a combinar distintos tipos de ítems (verdadero-falso, emparejamiento, selección múltiple, respuesta breve, ensayo, etc.)
- La presentación de las preguntas que se puede configurar de distintas maneras. Por ejemplo, se puede mostrar todas las preguntas en una sola página.
- La posibilidad de importar documentos externos como audios o textos.
- La automatización de las calificaciones en las secciones de corrección objetiva.

Todas las funcionalidades de esta plataforma permiten visualizar los beneficios de plantear una prueba de EFA en línea.

En consecuencia, una prueba implementada en Moodle permitirá evaluar a los estudiantes en dos modalidades: a distancia o presencial, con la finalidad de obtener de manera rápida los resultados y conocer su nivel de proficiencia de español en contextos académicos.

1.6. Pruebas de proficiencia en el Español como Lengua Extranjera (ELE)

1.6.1. Pruebas de ELE con fines generales

En la actualidad, las pruebas que tienen mayor reconocimiento a nivel internacional en la certificación del dominio del ELE general son los DELE del Instituto Cervantes y el CELU del Consorcio Interuniversitario de Español como Lengua Segunda o Lengua Extranjera (ELSE) de Argentina. A nivel universitario destacan las pruebas EPLE y CELA elaboradas por el Centro de Español para Extranjeros (CEPE) de la Universidad Nacional Autónoma de México (UNAM) y la prueba CELE desarrollada por Ferreira (2010).

El siguiente esquema proporciona una panorámica general de las pruebas que certifican proficiencia en ELE a nivel internacional en un contexto del español general.



Figura 3: Exámenes de proficiencia en ELE

Acorde con lo señalado en la figura 3, en la evaluación certificativa del ELE se cuenta con instrumentos que permiten la obtención del nivel de proficiencia en español a partir del puntaje (*score*) que obtiene el candidato, por lo que no hay un nivel definido de antemano, este es el caso de los exámenes CELU y EPLE, y también está la posibilidad de rendir una prueba de un nivel específico. Así, el DELE cuenta con exámenes desde el nivel A1 hasta el nivel C2, el CELA tiene pruebas de nivel B1, B2 y C1, y la prueba CELE se orienta a un nivel B1. En el siguiente apartado se describe cada una de estas pruebas.

- **Diploma de Español como Lengua Extranjera (DELE)**

Los Diplomas de Español como Lengua Extranjera (DELE) fueron creados en 1989 por el Ministerio de Educación de España con la finalidad de difundir el español y aumentar la presencia de la cultura en español en el mundo. Los dos primeros años los exámenes fueron organizados desde el Ministerio en Madrid. A partir de 1991, el Instituto Cervantes asumió esta tarea y la Universidad de Salamanca la responsabilidad de elaborar y evaluar los exámenes para la obtención de los DELE. El Instituto Cervantes, en noviembre de 2002, asumió la dirección académica, administrativa y económica de los DELE, por lo que estos serán expedidos por el director de este organismo público, en nombre del Ministerio de Educación, Cultura y Deporte.

Los DELE son títulos oficiales que acreditan el grado de competencia y proficiencia del español, certifican el nivel logrado por un candidato, independiente de las circunstancias en que haya aprendido el español. Estos títulos tienen una validez indefinida. En la obtención del DELE se consideran válidas todas las variantes del español.

El Instituto Cervantes organiza sus diplomas desde el nivel A1 hasta el nivel C2, acorde con los parámetros del MCER (2002), como se muestra en la tabla 4:

Diploma de español DELE	MCER	Denominación
Diploma de Español Nivel A1	A1	Acceso
Diploma de Español Nivel A2	A2	Plataforma
Diploma de Español Nivel B1	B1	Umbral
Diploma de Español Nivel B2	B2	Avanzado
Diploma de Español Nivel C1	C1	Dominio operativo eficaz
Diploma de Español Nivel C2	C2	Maestría

Tabla 4: Equivalencia entre diplomas y niveles del MCER

Los DELE se componen de 4 pruebas o secciones: comprensión de lectura, comprensión auditiva, expresión en interacción escrita y expresión e interacción oral. En la comprensión de lectura se evalúa también el dominio gramatical del candidato acorde con el nivel que evalúa la prueba. A modo de ejemplo, en la tabla 5 se describe el DELE de nivel B2 (versión septiembre, 2012).

Sección	Descripción de las tareas	Duración
Comprensión de lectura	Esta sección tiene 4 tareas que suman un total de 36 preguntas: 1) Ítems de selección múltiple con tres opciones, 2) ítems de relación, respuestas en base a la lectura de 4 textos breves, 3) Ítems de completación de fragmentos en un texto, indicando la letra del fragmento que debería ir en cada espacio, 4) Ítems de selección múltiple para completar los espacios del texto con las formas gramaticales adecuadas.	70 min.
Comprensión auditiva	Esta sección se compone de 5 tareas que suman un total de 30 preguntas: 1) Ítems de selección múltiple a partir de la escucha de seis conversaciones, 2) Ítems de relación a partir de la escucha de una conversación entre amigos, 3) Ítems de selección múltiple a partir de la escucha de una entrevista, 4) Selección de enunciado a partir de la escucha de seis personas que dan consejos para formar una empresa, 5) ítems de selección múltiple a partir de un audio donde una mujer habla de su empresa.	40 min.
Expresión e interacción escritas	Esta sección se compone de dos tareas: 1) Elaborar una carta al director sobre un tema que escucha, y 2) elegir una de las dos opciones que se presentan: escribir un artículo para una revista basado en un gráfico o escribir una crítica sobre una obra de teatro.	80 min.
Expresión e interacción orales	Esta sección consta de tres tareas: 1) Hablar sobre ventajas e inconvenientes sobre diferentes soluciones propuestas (6-7 minutos), 2) Describir una situación imaginaria a partir de una fotografía y conversar sobre ella (5-6 minutos) y 3) Opinar sobre los datos de una encuesta (3-4 minutos).	17 min.+ 20 min. de preparación para las tareas 1 y 2

Tabla 5: Descripción del DELE B2 (versión septiembre, 2012)

- **Certificado de Español: Lengua y Uso (CELU)**

En el año 2004, un grupo de universidades nacionales argentinas unió sus esfuerzos para crear un consorcio interuniversitario orientado a la enseñanza, evaluación y certificación del Español como Lengua Segunda y Extranjera (ELSE). El objetivo de este consorcio es contribuir a una política lingüística y educativa regional que promueva la valoración de la diversidad y tome en cuenta la relevancia de los códigos interculturales. El primer resultado

de dicha iniciativa fue el Certificado de Español: Lengua y Uso (CELU), avalado por el Ministerio de Educación y la Cancillería Argentina.

En la actualidad, el Consorcio se encuentra integrado por casi dos tercios de las universidades nacionales del país, y sus propósitos centrales son el desarrollo de la enseñanza y la evaluación del español como lengua segunda y extranjera. Aun cuando también tienen otros objetivos, como: promover la investigación, formación de docentes y especialistas, producción y publicación de material didáctico y otras actividades relacionadas con el español como lengua segunda y extranjera.

El CELU es un examen de proficiencia que evalúa la capacidad lingüística actual del estudiante para leer, escribir y hablar en contextos de la vida real. Por lo tanto, no se presentan preguntas específicas sobre la lengua y su gramática, sino que con una sola prueba, escrita y oral se evalúa en qué grado el individuo puede cumplir con las tareas lingüísticas que se le proponen. De acuerdo con su desempeño, el candidato alcanza los niveles: Básico, Intermedio o Avanzado. Solo se certifican estos dos últimos. Es un examen reconocido oficialmente y avalado por el Ministerio de Educación y el Ministerio de Relaciones Exteriores y Culto de la República Argentina. Se trata de una certificación que tiene validez internacional. Este certificado lo pueden obtener todos los extranjeros (cuya L1 no sea el español) que deseen validar su capacidad de uso del idioma, ya sea para trabajo o estudio. El certificado de aprobación incluye el nombre, documento y nacionalidad del candidato, la fecha del examen, el nivel alcanzado (Intermedio o Avanzado) y una mención de distinción en cada nivel, (bueno, muy bueno, excelente). Además, en el mismo certificado se agrega una descripción de los dos niveles de uso de la lengua y su correlación con los niveles reconocidos en otros países. Este certificado no tiene fecha de vencimiento

El CELU se toma en varias sedes de Argentina y de Brasil. Desde 2008 también se toma en Europa. Actualmente, están preparando la administración del examen en países de Asia y de América del Norte.

El examen CELU se compone de dos partes: una escrita y otra oral como se describe en la tabla 6.

Sección	Descripción de los ítems	Duración
Escrita	-Tres fragmentos de programas radiales que el candidato tiene que escuchar y a partir de ellos escribir distintos tipos de textos: cartas, correos electrónicos y folletos. -Tres textos que el candidato tiene que leer para orientar la escritura de distintos tipos de textos: artículo de opinión, carta, correo electrónico.	3 hrs.
Oral	Se divide en tres partes: 1) Presentación del candidato (información personal, intereses, estudios/trabajo, etc.); 2) Elección de una lámina y realizar un comentario (describir, explicar y opinar) sobre lo que observa y 3) Representación de una situación de habla (la información sobre la propuesta a representar se encuentra al reverso de la lámina elegida en el punto 2).	15 min.

Tabla 6: Descripción del CELU (versión n°3)

En este examen se observa la operacionalización de los métodos comunicativos, específicamente, el enfoque basado en tareas, utilizado en la enseñanza de segundas lenguas. Esto se refleja a través de las tareas integradas que se proponen. Sin embargo, este tipo de tareas tiene sus complejidades, puesto que si el candidato tiene dificultades en comprender algún tipo de texto, oral o escrito, eso afectará su desempeño en la tarea de escritura o de expresión oral (Figueras y Puig, 2013; Mendoza, 2015).

- **Prueba EPLE**

El Centro de Enseñanza para Extranjeros (CEPE) de la Universidad Nacional Autónoma de México (UNAM) tiene una trayectoria en la enseñanza de ELE que data desde 1921. Con la finalidad de internacionalizar el español y la cultura mexicana e hispanoamericana, el CEPE ha desarrollado una serie de instrumentos que permiten certificar, a nivel nacional e internacional, el grado de dominio del español.

En el contexto de la evaluación de proficiencia con fines generales, el CEPE cuenta con la prueba EPLE y el CELA. El EPLE fue la primera prueba que desarrolló el CEPE a partir del año 1992. Este instrumento tiene como objetivo la acreditación del dominio del español de los extranjeros que cada año llegan a México. Los candidatos rinden el EPLE para desempeñarse en actividades académicas, laborales y profesionales. Esta prueba considera la variante mexicana del español. La UNAM entrega un diploma oficial con validez

institucional que indica los resultados obtenidos en cada una de las destrezas comunicativas evaluadas. La vigencia de los resultados del examen es de dos años. Este examen se rinde en línea, por lo que cada candidato tiene que asistir a uno de los centros del CEPE para realizarlo.

Sección	Descripción	Duración	Puntaje
Comprensión auditiva	El candidato escuchará audios con conversaciones cortas, noticias, anuncios radiofónicos, entrevistas, refranes y monólogos sobre algún tema especializado. Deberá responder preguntas de opción múltiple con 5 posibilidades de respuesta.	50 minutos	275
Comprensión de lectura y vocabulario	El candidato leerá fragmentos de textos sobre todo tipo de temas: recados, notas periodísticas, históricas, de divulgación científica, etc., cuya extensión va desde 45 hasta 250 palabras. Deberá responder preguntas de opción múltiple con 5 posibilidades de respuesta.	55 minutos	335
Estructuras y expresión escrita	El candidato deberá llenar espacios con la palabra adecuada; reconocer errores dentro de un párrafo; discriminar entre oraciones bien formadas o no.	25 minutos	220
Expresión oral	Se evalúa con base en 3 actividades: 1. Entrevista: el candidato deberá contestar tres preguntas sobre su persona. 2. Monólogo: El candidato dispondrá de 4 minutos para desarrollar un monólogo. Para llevar a cabo esta tarea deberá organizar sus ideas, jerarquizar, opinar y argumentar sobre uno de los temas de una lista. 3. Preguntas sobre el monólogo: El participante deberá contestar tres preguntas sobre el tema desarrollado en el monólogo. El candidato dispondrá de dos minutos para contestar cada pregunta. Esta sección se graba y se guarda en un archivo para posteriormente ser evaluada.	15 minutos	170
	TOTAL	145 minutos	1000

Tabla 7: Descripción del EPLE

- **Certificado de Español como Lengua Adicional (CELA)**

El proyecto del CELA aparece como respuesta a un convenio de la UNAM con la Universidad de Salamanca el año 2002. En este contexto se desarrollan tres nuevos exámenes teniendo, en principio, como modelo los DELE elaborados por la misma Universidad de Salamanca y aplicados por el Instituto Cervantes. Luego, este Certificado se fue actualizando y ajustando a los lineamientos del Nuevo Plan de Estudios del CEPE. El CELA evalúa el dominio del español por niveles, estos corresponden a: inicial, intermedio y superior. A diferencia del EPLE, este instrumento se ajusta a las necesidades de los estudiantes, quienes pueden decidir en qué nivel desean certificarse. Llorián (2012) clasifica al CELA como una prueba con fines específicos académicos, no obstante, en la descripción que proporciona el CEPE de la UNAM se dice que este examen evalúa el dominio del conocimiento y uso del español para interactuar en los ámbitos personal, público y profesional. Por lo tanto, se trata de una prueba de ELE con fines generales, puesto que se orienta a aquellas personas que necesitan acreditar su dominio en español para desarrollarse en contextos laborales o sociales. A modo de ejemplo, en la tabla 8 se describe el CELA de nivel B2.

Sección	Descripción	Duración	Puntaje
Comprensión lectora	El candidato lee textos de 250 a 500 palabras y contesta 20 preguntas de opción múltiple con cuatro opciones.	45 min.	15%
Producción escrita	El candidato escribe dos textos de entre 150 y 200 palabras: una carta y una redacción (ej. un reporte). Estos textos abordan temas generales o de su especialidad.	1 hora	25%
Comprensión auditiva	El candidato escucha (dos veces) tres textos auditivos que tienen una duración entre 2 y 3 minutos, y responden preguntas de selección múltiple con cuatro alternativas. Estos textos corresponden a diálogos, monólogos o a programas de radio y televisión.	20 min.	15%
Competencia lingüística	Esta sección consta de 60 reactivos, donde el candidato debe completar diálogos, llenar espacios, identificar errores, etc.	1 hora y 15 min.	20%

Expresión oral	Esta sección consta de una entrevista personal y la exposición de dos temas (seleccionados por el candidato a partir de varias opciones), en los que el candidato debe interactuar en situaciones de la vida cotidiana: narrar, argumentar, quejarse y expresar opiniones sobre temas generales o de su interés personal o profesional.	15 min.	25%
		3 hrs. 35 min.	100%

Tabla 8: Descripción del CELA Intermedio (B2)

- **Prueba de Nivel de Competencia en Español como Lengua Extranjera (CELE)**

La Prueba de Nivel de Competencia en ELE, CELE, en línea comenzó a desarrollarse a partir del año 2009 en el contexto de los proyectos de investigación Fondecyt 1080165 y 1110812, dirigidos por la Dra. Anita Ferreira. Esta prueba surge como una necesidad de evaluar y precisar el nivel de competencia lingüística en ELE de los estudiantes extranjeros participantes de procesos de intervención lingüística.

El año 2009 se comenzó a diseñar y elaborar este instrumento en su primera versión. El año 2010 se realizó una primera aplicación del instrumento a estudiantes de ELE. Entre los años 2010 y 2011 se continuó piloteando el instrumento y se realizó una nueva versión. En la actualidad, se cuenta con dos versiones equivalentes de la prueba CELE en línea, se trata de las formas A y B (Ferreira, 2010). Esta prueba se orienta a un nivel de proficiencia B1 de ELE y tiene como objetivo medir el nivel de competencia en ELE de los estudiantes extranjeros en un ámbito general, como se observa en la figura 4.

Prueba de Proficiencia Nivel B1

El Programa de Español como lengua Extranjera de la Universidad de Concepción acredita el grado de competencia y dominio del idioma español acorde con el Marco Común Europeo de referencia para las lenguas. Los certificados se otorgan en los niveles principiante (A1 y A2) , intermedio (B1 y B2) y avanzado (C1 y C2).

La evaluación se hace en dos modalidades: a distancia y presencial. La modalidad a distancia (por internet) contempla la comprensión oral y escrita, producción oral y escrita. La evaluación presencial se hace cuando el estudiante se incorpora al Programa de Español como Lengua Extranjera de la Universidad (ELE-UdeC).



[Ingresar Prueba de Proficiencia Nivel B1](#)

Figura 4: Pantalla de acceso a la prueba CELE

La prueba determina qué estudiantes tienen un nivel de dominio B1, quienes están por sobre este nivel y quienes están bajo este nivel de competencia. La estructura de la prueba CELE, en sus dos versiones, es la siguiente:

Sección	N° de textos e ítems	Textos y temas	Total de ítems	Porcentaje
Comprensión auditiva	4 audios y 3 preguntas por cada uno	1) económico-cultural: presentación de un grupo musical en un concierto. 2)cultural: Biografía de un poeta chileno 3)científico: adjudicación de un proyecto científico para Chile 4)entrevista: proyecto para superar la pobreza	12 ítems	20%
Comprensión lectora	4 textos y 3 preguntas por cada uno	1)texto expositivo: exportaciones de Chile 2)texto narrativo: biografía de un escritor chileno 3)texto descriptivo: turismo 4)noticia: actualidad	12 ítems	20%
Gramática	Un texto con la técnica cloze con 20 opciones	Texto expositivo	20 ítems	10%

Vocabulario	18 ejercicios de diálogos	Diálogos de acciones cotidianas	18 ítems	10%
Producción escrita	Escritura de un texto con 2 opciones de temas	Opción 1: Propuesta de viaje Opción 2: Dar una opinión sobre un tema (prohibición de fumar)	1 ítem (20 puntos)	20%
Producción oral	Una entrevista con 8 preguntas	-Presentación del estudiante -Conversación sobre temas culturales: qué le parece la gente de Chile, la comida, la universidad, etc.	8 ítems (20 puntos)	20%
TOTAL				100%

Tabla 9: Estructura de la prueba CELE


La prueba CELE en sus formas A y B se encuentra implementada en la plataforma *Moodle* y está anclada en el Aula Virtual desarrollada por Ferreira (2010) en el contexto del proyecto Fondecyt 1080165, como se observa en la figura 5. En este entorno virtual se alojan los cursos de español como lengua extranjera y la prueba CELE en sus dos versiones.



Figura 5: Acceso al Aula Virtual de ELE

Las secciones de comprensión auditiva, lectora, gramática, léxico y producción escrita se rinden en computador, como se muestra en la figura 6. De estas, la sección de producción escrita es corregida por dos evaluadores, las otras secciones se corrigen de automática por la misma plataforma. La sección de producción oral se realiza de manera presencial con dos evaluadores.

Comprensión auditiva



1) Según el audio, la entrada de mayor costo para escuchar a "Los Jaivas" en el Teatro de Concepción es de 45 mil pesos.

2) De acuerdo al audio anterior, uno de los mayores obstáculos fue la subida de los instrumentos hacia el cerro

3) El audio indica que "Los Jaivas" cumplen 30 años en Machu Picchu.

Figura 6: Muestra de la sección de comprensión auditiva prueba CELE

La prueba CELE se puede rendir en dos modalidades: presencial o a distancia. En la modalidad a distancia, la sección de producción oral se evalúa a través de la herramienta Skype.

Con la finalidad de observar el comportamiento de la prueba en sus dos versiones, entre los años 2011 y 2013 se llevó a cabo un estudio con distintas muestras de estudiantes de ELE. Los resultados evidenciaron que las dos versiones de la prueba son equivalentes, puesto que permiten determinar con precisión el nivel de competencia de ELE de los estudiantes (Ferreira, Vine y Elejalde, 2013).

Actualmente, este instrumento permite que el Programa de Español como Lengua Extranjera de la Universidad de Concepción (ELE-UdeC) pueda evaluar y precisar el nivel de competencia de ELE de los estudiantes extranjeros e indicarles el curso de ELE más adecuado acorde con su nivel de español.

1.6.2. Pruebas de proficiencia en Español con Fines Específicos Académicos

En cuanto a las pruebas de EFA, actualmente, solo hay dos pruebas en este ámbito, se trata del examen EXELEAA desarrollado por el CEPE y el examen eLADE (Examen en línea de Acreditación de Dominio de Español B1/B2) del Centro de Lenguas Modernas de la Universidad de Granada.

El EXELEAA es reconocido por dos universidades mexicanas y una de Puerto Rico, sin embargo, no cuenta con una difusión ni reconocimiento nacional o internacional. El eLADE es reconocido por la Universidad de Granada, por ACLES (Asociación de Centros de Lenguas en la Enseñanza Superior), CERCLES (European Confederation of Language Centres in Higher Education) y CRUE (Conferencia de Rectores de las Universidades Españolas). Ambos instrumentos tienen una data reciente, por lo que aún no son conocidos a nivel internacional. Recientemente, el eLADE ha publicado algunos documentos que muestran cómo han abordado el proceso de validación de la prueba.

- **Examen de Español como Lengua Extranjera para el Ámbito Académico (EXELEAA)**

El EXELEAA es un examen tipo *score* que tiene como objetivo evaluar el nivel de proficiencia en español de los candidatos que desean cursar estudios superiores, tanto a nivel de pregrado como de posgrado. A través de la rendición de este examen los candidatos obtienen un documento institucional que acredita su nivel de dominio en español para desenvolverse en un ámbito académico. Sin embargo, las publicaciones que hay sobre este examen son de tipo descriptivas o relacionadas con las tareas de escritura (Mendoza, 2015). Por lo tanto, se requiere de estudios que muestren los procedimientos adoptados para el diseño y la validación de esta prueba.

Este examen evalúa las cuatro habilidades comunicativas de la lengua y la competencia lingüística, como se describe en la tabla 10.

Pruebas	N° de reactivos	Duración en minutos
<p>Competencia lingüística: Consta de 4 secciones:</p> <p><i>Sección I:</i> 10 preguntas de selección múltiple</p> <p><i>Sección II:</i> Un texto con 20 ítems (huecos) para completar con selección múltiple.</p> <p><i>Sección III:</i> 5 ítems de completación de un texto (se dan 6 opciones, pero solo se debe elegir 5).</p> <p><i>Sección IV:</i> 5 ítems de identificación de errores en un texto con selección múltiple.</p>	20	15
<p>Comprensión lectora: Consta de 3 secciones:</p> <p><i>Sección I:</i> 6 textos breves con una pregunta de selección múltiple por cada texto.</p> <p><i>Sección II:</i> 2 textos largos: 1) científico con un total de 19 preguntas, 7 de selección múltiple, 5 de relación y 7 de V o F, 2) humanista con un total de 15 preguntas, 9 de selección múltiple y 6 de V o F.</p>	40	50
<p>Expresión oral: Consta de 2 tareas: 1) presentación del candidato y hablar sobre su trayectoria académica y profesional, 2) elección de un tema, ya sea del área científica o humanista, para preparar una exposición sobre él.</p>	No aplica	10
Descanso		
<p>Expresión escrita: Consta de 2 tareas: 1) Descripción de gráficos, 2) Ensayo argumentativo (250 palabras) sobre la importancia del deporte.</p>	No aplica	70
<p>Comprensión auditiva: Consta de 4 secciones:</p> <p><i>Sección I:</i> 10 diálogos breves (10 y 30 seg.) con una pregunta de selección múltiple por cada diálogo.</p> <p><i>Sección II:</i> Exposición de tema científico (3-5 minutos) con 8 preguntas de selección múltiple.</p> <p><i>Sección III:</i> Diálogo largo sobre un tema universitario, cómo postular a una beca (2-3 minutos), con 6 preguntas de selección múltiple.</p> <p><i>Sección IV:</i> Exposición de una clase (3-5 minutos) con 6 preguntas de selección múltiple.</p>	30	45
TOTAL	110	3 hrs. 15 minutos

Tabla 10: Descripción del EXELEAA

- **eLADE (Examen en línea de Acreditación de Dominio de Español B1/B2)**

El eLADE es un examen binivel (B1/B2) en línea que está dirigido a todos aquellos que desean obtener una acreditación de los niveles B1 o B2 de español por motivos profesionales o académicos. Dado que se trata de una prueba en línea, la habilidad de expresión oral se realiza a través de Skype o de manera presencial.

El examen tiene cuatro partes: comprensión auditiva, lectora, expresión e interacción escrita y oral, como se muestra en la tabla 11.

Comprensión auditiva	Comprensión lectora	Expresión e interacción escrita	Expresión e interacción orales
5 tareas auditivas y/o audiovisuales	5 tareas de lectura	2 tareas de escritura	3 tareas de expresión e interacción orales
45 minutos	75 minutos	60 minutos	10-15 minutos
En línea	En línea	En línea	Skype o presencial en el CLM

Tabla 11: Descripción del eLADE

En total, este examen tiene una duración de 3 horas y 15 minutos.

1.7. Formatos de aplicación de las pruebas de ELE y de EFA

Las pruebas de español general de mayor reconocimiento a nivel internacional (DELE, CELU) se aplican en un formato tradicional de lápiz y papel. No obstante, el EPLE y el CELA se rinden en computador en las secciones de corrección objetiva y la expresión oral se evalúa de manera presencial.

El EXELEAA se aplica en computador en las secciones de comprensión lectora, comprensión auditiva, competencia lingüística y expresión escrita. La expresión oral se rinde de manera presencial con un entrevistador. Esta prueba evalúa la variante mexicana del español. Este examen tiene una data reciente, puesto que ha finalizado su etapa de pilotaje el año 2014, y este año 2015 está disponible como una prueba más de las que ofrece el CEPE. En este sentido, aún no se encuentran publicados los resultados de los pilotajes y

del proceso de validación del instrumento, así como tampoco se conoce en qué tipo de plataforma se encuentra implementado.

2. Directrices para la elaboración de pruebas en Segundas Lenguas

Alderson *et al.*, (1998) se refieren a los criterios en la evaluación de lenguas con el término del inglés “standards”. Este término puede precisar dos significados distintos relacionados con: 1) los niveles de los resultados de los candidatos y 2) el concepto de “principios”. Pollitt (1990) hace la distinción entre “criterios” y “principios”. Los criterios medirían la adhesión de una institución a unos determinados principios, los cuales serían más generales. Por su parte, Alderson *et al.* (1998) indican que para la mayor parte de las personas del ámbito de la evaluación, los criterios y principios hacen referencia a lo mismo. En este sentido, definen el término “criterios” como “las directrices acordadas que deberían consultarse y, en la medida de lo posible, ser tenidas en cuenta durante la elaboración y evaluación de una prueba” (Alderson *et al.* 1998: 226). En otras palabras, los criterios vendrían a ser la base sobre los cuales se orienta y decide la presencia de ciertas cualidades o propiedades en el diseño de una prueba.

Acorde con la noción de criterios, los expertos en evaluación de segundas lenguas señalan que en la elaboración de una prueba de lengua se debe tener claridad respecto de: las consideraciones éticas y de equidad, las especificaciones del instrumento, la selección de materiales, el diseño de tareas e ítems, el pilotaje y la validación (Alderson *et al.* 1998; Martínez, 2011; Figueras y Puig, 2013).

2.1. Consideraciones éticas

A partir de los años noventa ha habido una preocupación por los aspectos éticos en la evaluación de las lenguas extranjeras (Bordón, 2006). Esto ha llevado a reflexionar sobre las consecuencias y el impacto que tienen las pruebas de lenguas para los candidatos, para las instituciones y para la sociedad en general. De acuerdo con esto, los temas tratados en las distintas secciones deben ser equitativos para los candidatos, así se tendrá la certeza de que el desempeño del candidato es producto de su competencia en la lengua meta y no de factores ajenos a esta.

En cuanto a la equidad en los exámenes, el equipo CaMLA (Cambridge Michigan Language Assessments) ha llevado a cabo un extenso proceso que busca resguardar este principio a lo largo del desarrollo de las pruebas. Fortus *et al.*, (2014) señalan que el concepto de equidad (del inglés *fairness*) es confuso porque no solo se puede abordar desde un punto de vista de lo que es o no equitativo sino también en el contexto de qué es equitativo para los candidatos, para los profesores, etc.

CaMLA aborda el principio de equidad como aquel que evita los constructos irrelevantes en el proceso de desarrollo de la prueba. En este sentido se preocupa de asegurar estos seis principios:

1. Demostrar respeto por la gente.
2. Demostrar respeto por las convicciones y creencias personales.
3. Demostrar sensibilidad hacia las diferencias de la población y su conocimiento de mundo.
4. Evitar la negatividad excesiva.
5. Evitar temas excesivamente controversiales o inquietantes.
6. Evitar constructos de conocimiento irrelevante en la evaluación de un segundo idioma, por ejemplo, de matemáticas o ciencias.

Estos principios se operacionalizan en la elaboración de pruebas a través de la selección de los temas de las tareas y de los materiales a utilizar. De acuerdo con esto, no se puede incluir imágenes alusivas a alcohol, si se dibuja a personas, estas deben estar usando alguna polera con manga y falda o pantalón. Si se usa imágenes de personas, estas deben reflejar las distintas nacionalidades para que ningún candidato se sienta excluido. Si se utiliza imágenes de personal de seguridad, se debe mostrar realizando su trabajo y con su equipamiento apropiado. En cuanto a la administración del instrumento, las instituciones deben adaptar los exámenes y facilitar el acceso al examen a aquellos candidatos que, por sufrir algún tipo de discapacidad, no puedan realizarlo.

2.2. Especificaciones de las pruebas

Las especificaciones de un examen corresponden al plan en el cual se basan quienes redactan una prueba y sus ítems (Alderson *et al.*, 1998). Para Martínez (2011) las especificaciones del examen son algo así como el “prospecto”, constituido de un conjunto de documentos donde se establece: el propósito de la prueba, el perfil de los estudiantes o candidatos, los usuarios de la prueba, el o los niveles que evalúa, el marco de referencia teórico en el cual se sustenta, las habilidades que evalúa, los tipos de tareas e ítems, las ponderaciones, la duración, entre otros. Contar con la definición de cada una de estas especificaciones es relevante para el profesor o el equipo de evaluación durante todas las etapas de elaboración y validación de una prueba.

- **Propósito de la prueba**

De acuerdo con la finalidad, la prueba puede tener diferentes propósitos: determinar el nivel de lengua de un estudiante, evaluar su progreso, evaluar su proficiencia o dominio, etc. Teniendo claro el objetivo de la prueba, el candidato seleccionará la más adecuada acorde con sus necesidades.

- **Perfil de los estudiantes o candidatos**

Corresponde a una descripción lo más detallada y exhaustiva posible respecto del grupo meta o a quienes se orienta el instrumento diseñado. Esta descripción incluye la edad, la procedencia, los saberes que poseen o deben tener para rendirla, etc.

- **Usuarios de la prueba**

Los usuarios de la prueba corresponden a los estudiantes o candidatos, al equipo de redactores y de validación y a las instituciones que harán uso o tendrán que entender los resultados y conclusiones de la prueba.

- **Niveles de la prueba**

Se establece el nivel de lengua hacia el cual se orienta la prueba o bien los niveles, si se trata de un instrumento multinivel. Todo esto acorde con las directrices del MCER (2002) y del Plan Curricular del Instituto Cervantes. En el contexto de esta investigación se considera el nivel B2, dado que es a partir de este nivel donde se considera los temas y habilidades lingüísticas propios de un contexto académico (MCER, 2002).

- **Marco de referencia teórico**

La prueba debe estar vinculada a un sistema de referencia externo, en este caso, el MCER (2002) que concibe la lengua como competencia comunicativa. El MCER (2002) es el resultado de más de diez años de investigación exhaustiva realizada por un grupo de especialistas del campo de la Lingüística Aplicada. Es un documento que forma parte esencial del proyecto general de política lingüística del Consejo de Europa que tiene como objetivo la unificación de directrices para el aprendizaje, la enseñanza y la evaluación de lenguas. Al ofrecer una base común para la descripción explícita de los objetivos, los contenidos y la metodología, el MCER favorece la transparencia de los cursos, los programas y las titulaciones. Esto facilitará el reconocimiento mutuo de las titulaciones obtenidas en distintos contextos de aprendizaje y ayudará también a la movilidad internacional. El MCER define seis niveles de referencia para evaluar proficiencia en una lengua, como se muestra en la tabla 12.

USUARIO COMPETENTE	C2	Maestría
	C1	Dominio operativo eficaz
USUARIO INDEPENDIENTE	B2	Avanzado
	B1	Umbral
USUARIO BÁSICO	A2	Plataforma
	A1	Acceso

Tabla 12: Niveles comunes de referencia definidos por el MCER (2002)

La división de estos seis niveles da cuenta de una interpretación más amplia de la delimitación tradicional de básico, intermedio y avanzado. No obstante, el establecimiento de estos niveles no limita la forma en que distintos sectores de diversas culturas pedagógicas puedan organizar su sistema de niveles y módulos.

- **Modelo de lengua a evaluar según el enfoque comunicativo**

Una de las primeras consideraciones que hay que hacer al plantearse la evaluación del uso de la lengua es definir el modelo de lengua que se va a evaluar (Bordón, 2006). Teniendo en cuenta las propuestas de Canale y Swain (1980) y del MCER (2002), la lengua en el contexto del enfoque comunicativo se concibe como competencia comunicativa. Por lo

tanto, la lengua se entiende como la capacidad de una persona para comportarse adecuadamente en una determinada comunidad de habla. En el ámbito de las segundas lenguas, Bordón (2008) define competencia comunicativa como la capacidad de un estudiante de ejecutar determinados conocimientos, por lo que no solo es necesario contar con un determinado conocimiento sino ser capaz de usarlo de acuerdo a los diversos contextos. El MCER (2002) se refiere a la competencia comunicativa de la lengua, incluyendo las competencias lingüísticas, sociolingüísticas y pragmáticas; las que a su vez se integran en las competencias generales del individuo referidas al saber, saber hacer, saber ser y saber aprender.



Figura 7: La competencia comunicativa (MCER,2002)

La competencia lingüística se refiere al dominio de las reglas gramaticales, léxicas, morfológicas, semánticas, fonológicas y ortográficas de la lengua, independiente de su valor sociolingüístico o pragmático, puesto que en esta competencia interesa la precisión que muestra el candidato en relación con los componentes mencionados. La competencia sociolingüística considera las condiciones socioculturales de uso de la lengua, por lo que son relevantes la adecuación del discurso, tanto mediante la forma como de su significado al contexto en que se produce, en este sentido interesan, por ejemplo, los marcadores lingüísticos de relaciones sociales (formas de saludo y de tratamiento, convenciones respecto a los turnos de habla); las normas de cortesía; las expresiones de sabiduría popular (dichos y expresiones); diferencias de registro; y el dialecto y acento. Y la competencia pragmática tiene que ver con el uso funcional que se hace de los recursos lingüísticos en

distintos contextos comunicativos, a esto se suma el dominio del discurso, la coherencia y la cohesión, la identificación de tipos y formas de texto, la ironía y la parodia (MCER, 2002).

Según el MCER (2002), los usuarios de la lengua para dar cumplimiento a sus intenciones comunicativas se valen de las competencias generales y de esta competencia comunicativa relacionada de manera específica con la lengua.

- **Especificación de la variedad de lengua**

Se debe especificar cuál es la variedad de lengua que se va a evaluar en un contexto comunicativo, si se trata de la evaluación de la lengua con una orientación general o con fines específicos. En el contexto de este trabajo se considera la evaluación del español como lengua extranjera con fines específicos académicos (EFA). En este sentido es relevante conocer el manejo del discurso académico en la lengua meta de los estudiantes que realizan intercambios universitarios, es decir, cómo enfrentan las lecturas de artículos científicos, las exposiciones y argumentaciones, la escritura de textos académicos, etc. (Pastor, 2006).

En el contexto del discurso académico escrito, Vázquez (2010) investigó las características de este tipo de discurso en español. Al respecto señala que este tipo de discurso se caracteriza por ir a lo concreto y evitar las digresiones, ser impersonal (evitar el uso de la 1ª persona), el uso del parafraseo y de abundantes recursos léxicos. En cuanto al análisis del discurso académico oral, la autora se centró en un objetivo didáctico, respecto de cómo enfrentar la comprensión auditiva de discursos altamente complejos y estructurados como la clase magistral.

Todo lo señalado anteriormente son aspectos que se necesita enseñar a los estudiantes de intercambio para que puedan tener una buena inserción académica y un buen desempeño.

- **Habilidades a evaluar**

De acuerdo con el objetivo de la prueba y el requerimiento de los usuarios, los instrumentos evalúan distintas habilidades. Comúnmente, las pruebas en segundas lenguas evalúan las cuatro habilidades lingüísticas: comprensión lectora y auditiva y expresión escrita y oral.

Comprensión de lectura

De acuerdo con Figueras y Puig (2013), la comprensión de textos escritos conlleva “un proceso mental, individual y complejo por el que interaccionan los textos con los conocimientos previos, lingüísticos y no lingüísticos, de quien lee o escucha” (Figueras y Puig, 2013: 63). En otras palabras, esto significa que el lector elabora su propia representación mental del texto a partir de sus conocimientos previos, tanto a nivel lingüístico como temático.

La comprensión es una habilidad receptiva, producto del diálogo entre un texto y un estudiante o candidato a un examen y “no proporciona de forma directa y observable muestras empíricas que den cuenta de qué se ha comprendido y qué no se ha comprendido” (Figueras y Puig, 2013:61). Por lo tanto, la evaluación tiene un papel relevante en el sentido que los instrumentos elaborados deben ser capaces de medir de manera indirecta el nivel de dominio de los candidatos en cuanto a la comprensión de textos.

En la evaluación de la comprensión, hay que tener en cuenta el objetivo de la lectura: para qué se lee. Estos objetivos tienen que estar en concordancia con las tareas cotidianas que realizan quienes rinden la prueba, por ejemplo, se lee para informarse de algo, para captar la idea general del texto, para conocer las opiniones de los expertos sobre un tema, etc.

El diseño de un instrumento que evalúe la comprensión lectora debe medir las distintas microhabilidades que la componen. Los especialistas en evaluación (Alderson, *et al.*, 1998; Martínez 2011; Figueras y Puig, 2013) determinan que, a partir de los textos escritos, se puede evaluar las siguientes microhabilidades, dependiendo del nivel de los candidatos:

- Identificar los detalles o información específica presente en el texto
- Comprender o captar la idea general de lo que se dice en el texto
- Comprender las ideas principales de los párrafos
- Distinguir hechos de opiniones
- Deducir las ideas implícitas en el texto
- Deducir el significado de palabras desconocidas mediante el contexto
- Comprender las relaciones entre frases y párrafos
- Determinar la función y tipología del texto
- Reconocer las función comunicativa o la intención del autor al escribir el texto

En este ámbito encontrar o descubrir el significado literal es la forma más elemental de comprensión lectora, mientras que las inferencias constituyen la esencia misma de este proceso (Jouini, 2005). De acuerdo con Sabaj y Ferrari (2005), en las preguntas literales se identifica las proposiciones específicas del texto, sin una reformulación o interpretación de la información que se presenta. Se busca que el estudiante pueda identificar y extraer información específica de los textos. Para Ainciburu (2008), la comprensión literal de un texto es esencial para realizar las operaciones que implican niveles más altos, como la lectura inferencial, interpretativa o crítica.

En las preguntas inferenciales se elicitaba un tipo de relación o información sobre el texto que no está de manera explícita en él, sino que el estudiante tiene que inferirla o deducirla a partir de la información explícita que se presenta (Sabaj y Ferrari, 2005). Hay que considerar que los textos proporcionan mayor información que la que se presenta explícitamente (Pérez, 2005). La inferencia consiste en realizar hipótesis respecto del significado del texto, se trata de hipótesis que el lector compara constantemente con los datos del texto (Alexopoulou, 2010).

Domínguez (2010) señala que el lector de una SL/LE se basa en los procesos adquiridos en relación con su L1, por lo tanto, no comienza de cero. No obstante, hay estudios empíricos que demuestran que los conocimientos previamente adquiridos en relación con aspectos textuales y de conocimiento de mundo no siempre son una garantía de una comprensión eficaz (Alexopoulou, 2010).

Respecto de las tareas e ítems mediante los cuales se evalúa las microhabilidades mencionadas anteriormente, estos pueden ser de distinta forma: selección múltiple, verdadero/falso, completación de espacios. En ocasiones, no siempre se logra contar con pruebas que midan las microhabilidades que se espera, por lo que resulta crucial realizar ensayos y análisis que ayuden a determinar qué es lo que efectivamente se está midiendo con los ítems elaborados.

De acuerdo con los descriptores del MCER (2002), un estudiante o hablante de nivel B2 en la comprensión lectora es capaz de:

Comprensión de lectura nivel B2

- Identificar información específica en textos especializados extensos.
- Comprender las ideas principales de artículos especializados e informes relativos a problemas actuales en los que los autores adoptan posturas o puntos de vista concretos.
- Comprender las ideas globales de artículos especializados, aunque no sean de su especialidad.

Como se observa, a partir del nivel B2 los estudiantes pueden comprender diferentes tipos de textos especializados, por lo que es pertinente señalar que es a partir de este nivel que se puede abordar el español con fines específicos académicos (Pastor, 2006). El MCER (2002), en los descriptores para la comprensión lectora y auditiva hace una distinción entre el tratamiento de temas concretos y abstractos en los niveles B1 y B2. Esto sin duda, orienta la búsqueda y selección de textos.

Comprensión auditiva

La principal diferencia entre la comprensión lectora y auditiva es el canal o medio por el cual se transmite la información y la forma en que se presenta (Buck, 2009). Aun cuando ambas habilidades tienen muchas características en común se evalúan por separado porque la forma en que interactúa el candidato con el texto oral y escrito es distinta. El discurso hablado es menos formal y coherente (Caplan, 1998), dado que la información no se presenta de manera completa y organizada, como sí se encuentra en los textos escritos.

Al igual que la comprensión lectora, la comprensión auditiva se evalúa de manera indirecta, dado que se ve interferida por otras destrezas (por ejemplo, la expresión oral). En la comprensión auditiva la información se procesa de manera simultánea a la emisión del input (Martín, 2010).

En las pruebas de comprensión auditiva se busca conocer cuánto de lo que capta el candidato en una escucha es capaz de entender. De acuerdo con esto, entonces en un texto oral podemos evaluar la capacidad del candidato para reconocer (Martínez, 2011):

- La idea general
- Las ideas principales
- La información y ejemplos que sustentan las ideas
- La información específica y los detalles en que se basa
- La actitud o propósito del hablante
- Las inferencias y deducciones lógicas implícitas en el mensaje
- El significado de palabras desconocidas mediante el contexto

Un estudiante de nivel B2 (MCER, 2002) en cuanto a la comprensión auditiva es capaz de:

Comprensión auditiva nivel B2

<p>-Identificar detalles e información específica en discursos orales especializados siempre que sean pronunciados en un nivel de lengua comprensible.</p> <p>-Comprender las ideas principales de un discurso complejo lingüísticamente que trate tanto temas concretos como abstractos pronunciados en un nivel de lengua estándar, incluyendo debates técnicos dentro de su especialidad.</p> <p>-Comprender las ideas globales de conferencias, charlas e informes, y otras formas de presentación académica y profesional lingüísticamente complejas.</p>

De acuerdo con los descriptores esperables para un nivel B2, el estudiante puede comprender textos orales especializados en el contexto de los diferentes fines específicos en la enseñanza y evaluación de segundas lenguas.

Competencia lingüística

La competencia lingüística se define como “el conocimiento de los recursos formales y la capacidad para utilizarlos” (MCER, 2002). Por lo tanto, la capacidad del estudiante para elaborar mensajes bien formados dependerá de su nivel de conocimiento y uso de la lengua. En este sentido, un estudiante de nivel B2 debe ser capaz de:

Competencia lingüística nivel B2

<p>-Mostrar un grado de control relativamente alto, evitando errores que produzcan malentendidos.</p>

La competencia lingüística considera una serie de subcompetencias: léxica, gramatical, semántica, fonológica, ortográfica y ortoépica.

Expresión e interacción escritas

La escritura es un proceso complejo donde intervienen distintos factores ajenos a la tarea misma de escribir. La expresión e interacción escrita se entienden como “el conjunto de actuaciones que se caracterizan por un uso eficaz y adecuado de la lengua en las situaciones comunicativas en las que se producen” (Figueras y Puig, 2013: 107). No se trata solo que el estudiante se exprese a través de un discurso monológico, sino que también interactúe con otros, mediante el proceso de escritura, por ejemplo, respondiendo a un anuncio o solicitud.

Por lo tanto, diseñar una tarea adecuada de escritura implica no solo tener en cuenta el tema sobre el que van a escribir los candidatos, sino también otras consideraciones, como: el manejo de los distintos tipos de textos y sus formatos, la habilidad para seleccionar información y contenido relevantes, la habilidad de escribir frases correctas desde una perspectiva gramatical, léxica, puntuación y ortografía; la habilidad para dar coherencia a las frases y párrafos y cohesión al texto, y la habilidad de manipular frases y palabras adecuadas al contexto y sus destinatarios (Martínez, 2011). A estas consideraciones internas, Martínez (2011) agrega otras externas relacionadas con:

- 1) Tipo y forma de las instrucciones: Deben ser simples y claras de modo que la actuación del candidato no se vea afectada por la comprensión de las instrucciones. Además, deben motivar la realización de la tarea de escritura.
- 2) Mecanismo o medio de escritura: Tener en cuenta cuál será el medio en que el candidato escribirá el texto, si será a mano o en el computador, ya que la forma de generar, procesar y reformular el texto es distinto en cada caso.

A lo mencionado anteriormente, se agrega que toda prueba de expresión e interacción escritas debe responder a tareas que eliciten la producción e interacción del candidato.

Respecto de la selección de tareas, Figueras y Puig (2013) mencionan dos tipos:

- 1) Dirigidas: Se trata de tareas cerradas donde hay instrucciones claras de lo que se espera del candidato, por ejemplo, del tipo de texto que tiene que escribir, el motivo, el formato, etc. Es posible que los candidatos se vean enfrentados con mayor frecuencia a este tipo de tareas en el mundo real.
- 2) Escritura creativa: Se trata de tareas abiertas donde solo se entregan lineamientos generales de lo que tiene que hacer el candidato. Este tipo de tareas es poco

probable que se den en el mundo real, sin embargo, tienen el potencial de elicitar al máximo lo que el candidato puede hacer con la lengua.

En el proceso de escritura se busca evitar que la evaluación de esta habilidad parezca algo artificial. Para ello, Martínez (2011) recomienda tener presente en el diseño de tareas de escritura lo siguiente: 1) los propósitos de una prueba de expresión escrita y las motivaciones reales de los candidatos, 2) el diseño de las tareas, la selección de temas, los tipos de textos y sus objetivos, 3) el diseño y la selección de escalas para calificar los escritos y 4) cómo restar subjetividad en la revisión de los textos. Además, en el diseño de tareas de escritura se debe cumplir tres requisitos básicos: validez, practicidad e impacto (Figueras y Puig, 2013). Por validez en este contexto se entiende que las tareas sean una muestra representativa de lo que puede producir por escrito el candidato, acorde con el objetivo de la prueba y el nivel de proficiencia al que se orienta. La practicidad de una tarea se refiere a que las exigencias de tiempo y de formato sean adecuadas para el nivel del candidato. Finalmente, una tarea que sea válida y practicable tendrá un impacto positivo en el candidato porque le permitirá evidenciar lo que sabe y puede hacer con la lengua.

Para dar cumplimiento a los requisitos mencionados se necesita hacer ensayos y pilotajes que permitan validar las distintas tareas, tanto a nivel de comprensión de las instrucciones, del tiempo estimado como de la redacción del tipo de texto esperado acorde con el nivel al cual se dirige la prueba.

El MCER (2002) describe que un estudiante de nivel B2 en la expresión escrita es capaz de:

Expresión e interacción escritas nivel B2
- Escribir redacciones e informes que desarrollan un argumento, razonando a favor o en contra de un punto de vista concreto y explicando las ventajas y las desventajas de varias opciones. -Escribir textos claros y detallados sobre diversos temas, así como defender un punto de vista sobre temas generales, indicando los pros y los contras de las distintas opciones. -Sintetizar información y argumentos procedentes de varias fuentes.

De acuerdo con esta descripción, el estudiante de nivel B2 tiene las competencias para elaborar diferentes géneros y tipologías textuales sobre temas generales y especializados.

Expresión e interacción orales

Desde una perspectiva comunicativa la expresión e interacción orales se concibe en función de la comunicación con otros hablantes, de forma interactiva a través de la conversación (Bordón, 2006). Por esta razón, en la actualidad se considera la expresión unida a la interacción. La expresión e interacción orales corresponde a las habilidades productivas de la lengua. En la expresión se ponen en relación los conocimientos que tiene el candidato sobre la LE, la competencia lingüística, y un conjunto de estrategias que el hablante utiliza con el objeto de mostrar una actuación adecuada a una situación comunicativa determinada. En la interacción se ponen en juego las reglas de cooperación, los turnos de habla, de modo que se construye una conversación sobre la base de la negociación de significados (Figueras y Puig, 2013).

Hay personas que, pese a dedicar mucho tiempo al estudio de una LE, tienen dificultades para hablarla o mantener una conversación (Martínez, 2011). Parte del problema radica en que no comprenden lo que escuchan y esto les impide conversar o hablar. La capacidad de hablar una LE es la más importante y la más difícil de medir porque depende de diversos factores: afectivos, sociales, personales, contextuales, etc. (Martínez, 2011).

Hablar una LE es complejo y evaluarla es todavía más. Martínez (2011) menciona que hay autores que dudan que se pueda medir efectivamente esta habilidad porque está sometida a muchas variaciones. No obstante, Widdowson (2001) señala que si bien no se puede evaluar la competencia comunicativa en su totalidad sí se puede evaluar algunos aspectos de ella. Se trata de una posición que intenta delimitar la lengua en unidades discretas, tanto desde un punto de vista estructural (gramática, uso de vocabulario, pronunciación, etc.) como pragmático (iniciar las intervenciones, turnos de habla, etc.). Otra perspectiva de evaluación de la lengua hablada sería de tipo holística donde se califica el uso global que hace el candidato en términos de los aspectos funcionales de la lengua, no penalizando, por ejemplo, si comete errores de precisión léxica o gramatical.

En relación con lo anterior, el MCER (2002) define como actividades de expresión e interacción orales las siguientes:

Expresión oral	Interacción oral
<ul style="list-style-type: none"> -Expresión oral en general -Monólogo sostenido: descripción de experiencias -Monólogo sostenido: argumentación -Declaraciones públicas -Hablar en público 	<ul style="list-style-type: none"> -Interacción oral en general -Comprender a un interlocutor nativo -Conversación -Conversación informal (con amigos) -Conversación formal y reuniones de trabajo -Colaborar para alcanzar un objetivo -Interactuar para obtener bienes y servicios -Intercambiar información -Entrevistar y ser entrevistado

Para cada una de las actividades mencionadas, el hablante de una LE necesita hacer uso de distintas estrategias que le permitan lograr una comunicación e interacción efectivas.

En relación con las tareas de producción e interacción orales, Pinilla (2004) propone las técnicas de diálogos, entrevistas, dramatizaciones, debates, conversaciones telefónicas y actividades de carácter lúdicas. De estas técnicas, la entrevista ha sido uno de los procedimientos más usados para evaluar la expresión e interacción orales. La entrevista corresponde al intercambio cara a cara que se da entre el examinador y el candidato (Bordón, 2006).

Aun cuando en algunos exámenes presentan variantes, por lo general, la entrevista se organiza en función de cuatro partes: 1) presentación del candidato, 2) exposición de un tema a elección, 3) simulación y 4) juego de roles. La tarea de presentación del candidato constituye la primera tarea que se realiza en el proceso de examinación oral y consiste en que el candidato cuente algunos aspectos sobre su vida para dar inicio a la conversación (romper el hielo), por ejemplo, en el contexto académico es importante que hable de su trayectoria en sus estudios, su carrera, etc.

En los descriptores del MCER (2002), se espera que un candidato de nivel B2 sea capaz de:

Expresión e interacción orales nivel B2

-Realizar con claridad presentaciones preparadas previamente, razonando a favor o en contra de un punto de vista concreto, y mostrando las ventajas y desventajas de varias opciones.

-Desarrollar argumentos con claridad, ampliando con cierta extensión y defendiendo sus puntos de vista con ideas complementarias y ejemplos adecuados.

-Responder a una serie de preguntas complementarias con un grado de fluidez y espontaneidad que no supone ninguna tensión para sí mismo ni para el público.

En cuanto al orden en que se debe evaluar las diferentes habilidades en una prueba de lengua, los expertos recomiendan comenzar con la comprensión lectora, pues de esta manera el candidato entra en contacto con las tareas de la prueba y rebaja su nivel de ansiedad.

Bordón (2006) sugiere el siguiente orden en una prueba que evalúe proficiencia en ELE:

1. Comprensión lectora
2. Comprensión auditiva
3. Competencia lingüística
4. Expresión e interacción escrita
5. Expresión e interacción oral

En términos prácticos, se suele dejar la parte oral para el final de todo proceso evaluativo porque facilita el proceso de aplicación.

• Tipos de ítems y tareas

Según el Diccionario de términos clave del ELE del Centro Virtual Cervantes el término ítem se utiliza en el ámbito de la evaluación para referirse a una pregunta abierta o cerrada. Además, en español, el término *ítem* puede utilizarse indistintamente como sinónimo de pregunta. Así, una tarea o ejercicio de evaluación puede incluir uno o más ítems. Para Alderson *et al.* (1998) un ítem de examen “constituye un método para provocar una actuación o un enunciado, junto con un sistema por el que tal actuación o enunciado pueda ser juzgado” (1998: 44). Esto significa que el ítem se concibe como una unidad de puntuación. Los ítems pueden ser de respuesta abierta si el candidato tiene que producir o elaborar su respuesta o de respuesta cerrada si tiene que marcar o seleccionar la opción correcta.

Para la selección de los tipos de ítems y su redacción es necesario tener en cuenta el nivel de lengua hacia el cual se orienta la prueba. Siempre se debe procurar que las instrucciones no sean el impedimento en el desempeño del candidato. De esta forma, tanto las instrucciones como las preguntas deben elaborarse con un nivel de lengua inferior al que se está evaluando, un lenguaje simple, sencillo; de modo que esto no dificulte la tarea. Además, en cada uno de los ítems se debe cautelar que la respuesta sea producto de la comprensión del texto y no del conocimiento de mundo del candidato.

En cuanto a los tipos de ítems, Figueras y Puig (2013) enumeran los siguientes:

1. Preguntas de corrección objetiva

- **Ítems de respuesta cerrada:**

- 1) **Verdadero/falso:** Se considera que este tipo de ítem es insatisfactorio, porque el candidato tiene un 50% de acertar a la respuesta por azar. Alderson *et al.*, (1998) señalan que para tener certeza respecto de la habilidad del candidato y reducir el efecto del azar se debe contar con un gran número de estos ítems, aun cuando algunos redactores reducen el azar a través de una tercera opción, por ejemplo, no se dice, no se menciona, etc. Sin embargo, indican los especialistas que esta tercera categoría solo podría ser útil en textos de comprensión lectora, porque en comprensión auditiva, y si solo escuchan una vez el audio, podría considerarse como demasiado exigente e incluso llevar a confusión.
- 2) **Selección múltiple (*Multiple choice*):** Este tipo de ítem es uno de los más usados para evaluar en las distintas disciplinas. En el caso de la evaluación de una segunda lengua su uso está dado en las habilidades de comprensión lectora, auditiva y en la gramática y el léxico. Los ítems de ese tipo pueden tener distintas formas: 1) elaborar una pregunta cerrada con opciones donde solo una es correcta, 2) elaborar una proposición incompleta (tronco) cuya respuesta para completar la oración se encuentra en las opciones donde solo una es correcta, y 3) a partir de un fragmento de texto se propone una inferencia que el candidato tiene que seleccionar de las opciones disponibles. Los especialistas recomiendan que se considere distintos tipos de ítems para medir una habilidad, pues de este modo se tendrá mayor seguridad respecto del desempeño del candidato.

3) Relacionar o emparejar (*Matching*): Este tipo de ítem se utiliza tanto para la comprensión lectora como auditiva. El estudiante tiene que leer o escuchar textos breves y unirlos con una breve afirmación que recoge la idea principal del texto. Para elevar el grado de complejidad de estos ítems se necesita dar más opciones para relacionar, de modo que no haya una relación unívoca, que permita resolver el último ítem por descarte.

4) Ordenamiento: Es un tipo de ítem utilizado en la comprensión lectora y consiste en reconstruir un texto con algunos de sus párrafos que se han desordenado. Tiene como objetivo evaluar la capacidad del estudiante para dar coherencia y cohesión al texto.

- **Ítems de respuesta dirigida:**

1) Cloze: Según Figueras y Puig (2013) es un tipo de ítem muy utilizado por su facilidad de elaboración. Estos ítems se diseñan suprimiendo de un texto una palabra, cada cinco, seis o siete, independiente de la palabra que se trate. Se suele mantener la frase inicial y final del texto para dar contexto. No obstante, Alderson *et. al* (1998) hacen algunas recomendaciones respecto a algunos de estos tipos de ítems. En este sentido, señalan que se ha demostrado que cuando los estudiantes responden pruebas del tipo cloze leen de manera distinta a la habitual, solo leen un poco del texto que está antes del espacio y no lo que sigue después de este espacio; como tampoco logran una comprensión a cabalidad del texto, porque solo leen las partes donde están los huecos y no todo el texto.

2) Tipo C (*C-Test*): Corresponde a una variante del cloze, pero en este tipo de ítem solo se omite la segunda mitad de cada segunda palabra de un texto.

3) Completar huecos o espacios en blanco: Este tipo de ítem ofrece una mayor versatilidad. Se suprime algunas palabras relevantes del texto de acuerdo con lo que se pretende evaluar. Se utiliza este ítem para evaluar comprensión lectora, auditiva, usos gramaticales y léxicos.

2. Preguntas de respuesta abierta:

Existen dos tipos de ítems de respuesta abierta, de respuesta breve y de respuesta extensa. Generalmente, son fáciles de elaborar, pero se debe tener claridad respecto de lo que estas preguntas están evaluando. Figueras y Puig (2013) señalan que estas preguntas tienen problemas de fiabilidad en cuanto a la corrección, situación que puede mejorarse con la definición de pautas claras y con procedimientos claros de corrección que permitan un análisis válido y confiable de las respuestas.

Una de las formas más utilizadas para la revisión de tareas subjetivas es el procedimiento de la doble corrección, que consiste en que cada texto escrito u oral es corregido por dos examinadores que trabajan de manera independiente. La puntuación final que recibe el estudiante es el promedio del puntaje asignado por los dos evaluadores, esto siempre que no exista más de 1 punto de diferencia entre los examinadores. Si las diferencias entre los evaluadores es de 2 o más puntos en una escala de 5 puntos, los evaluadores tienen que volver a revisar la tarea en contraste con la pauta o rúbrica de evaluación, y si no logran aproximarse en los puntajes asignados se le entregará el texto o audio a un tercer evaluador. Finalmente, la institución decide qué punto de vista o decisión adoptará en caso de que haya desacuerdo, o bien indicará si las dos puntuaciones más cercanas o todas las puntuaciones deberían tener la media (Figueras y Puig, 2013).

- **Ponderaciones y calificación en las pruebas de ELE**

La ponderación corresponde a la valoración que se le asigna a las distintas secciones o partes de la prueba, las cuales reflejan el uso de las diferentes habilidades lingüísticas. La definición de los porcentajes o formas de distribuir las ponderaciones dependerá del tipo de prueba, del objetivo, del perfil de los estudiantes, etc. A continuación se muestra las ponderaciones y formas de calificación de pruebas de ELE generales.

Los DELE, que certifican el ELE en un contexto general, se califican de dos formas: automática y mediante un examinador. Estas pruebas se dividen en dos grupos:

- 1) Pruebas de comprensión: Se componen de ítems de selección múltiple con una sola respuesta correcta. Las respuestas correctas reciben 1 punto y las incorrectas 0 puntos y no se penalizan. La puntuación máxima que se puede obtener en las pruebas

de comprensión es igual al número de ítems que contienen. Para realizar el cálculo, la puntuación directa (el número de ítems correctos) se transforma a su equivalencia en una escala de 25 puntos mediante la siguiente fórmula:

$$\frac{\text{puntuación directa} \times 25}{\text{puntuación directa máxima posible}} = \text{puntuación final}$$

- 2) Pruebas de expresión e interacción: Se componen de tareas de respuesta abierta. Los calificadores asignan una calificación de entre 0 y 3 para cada categoría de las escalas. La puntuación máxima que se puede obtener en cada prueba de expresión e interacción es 3 puntos. La puntuación directa se transforma a su equivalencia en una escala de 25 puntos, al igual que en las pruebas de comprensión.

GRUPOS	GRUPO 1		GRUPO 2	
	Comprensión de lectura	Expresión e interacción escritas	Comprensión auditiva	Expresión e interacción orales
Puntuación máxima	25 puntos	25 puntos	25 puntos	25 puntos
Puntuación mínima exigida	30 puntos		30 puntos	
Puntuación obtenida				
Calificación	Promedio de puntos del candidato		Promedio de puntos del candidato	
	Apto/No apto		Apto/No apto	
Calificación global	APTO/NO APTO			

Figura 8 Ponderación y calificación de los DELE

Como se observa en la figura 8, para que el candidato logre aprobar el examen y obtener el diploma del nivel al cual se ha presentado debe obtener un mínimo de 30 puntos en el grupo 1 de pruebas (comprensión de lectura y expresión e interacción escritas) y en el grupo 2 (expresión e interacción orales).

El examen CELA se califica teniendo en cuenta las siguientes ponderaciones: 15% en comprensión lectora, 25% en expresión escrita, 15% en comprensión auditiva, 20% en

competencia lingüística y un 25% en la expresión oral. Del EXELEAA se desconoce cuál es la ponderación que se realiza en las distintas secciones, así como tampoco es posible acceder a las pautas o escalas que utilizan para calificar las tareas orales y de escritura.

- **Duración de las pruebas**

En este apartado corresponde señalar cuál será la duración total de la prueba, como por cada una de sus secciones. Existe una similitud en cuanto a la duración total de las pruebas de proficiencia en ELE destinadas al nivel B2, como se observa en la tabla 12, tanto la prueba de EFA como las pruebas generales DELE y CELA tienen una duración similar. No obstante, al revisar la duración por secciones hay algunas variaciones. Por ejemplo, en el EXELEAA y en el CELA se evalúa la competencia lingüística, pero con distintas duraciones, 35 minutos y 1 hora 15 minutos, respectivamente.

Prueba	Duración total	Duración por secciones
EXELEAA	3 horas y 40 minutos	Competencia lingüística: 35 minutos
		Comprensión lectora: 50 minutos
		Expresión oral: 10 minutos
		Expresión escrita: 70 minutos
		Comprensión auditiva: 45 minutos
CELA (B2)	3 horas y 35 minutos	Comprensión lectora: 45 minutos
		Expresión escrita: 1 hora
		Comprensión auditiva: 20 minutos
		Competencia lingüística: 1 hora y 15 minutos
		Expresión oral: 15 minutos
DELE	3 horas y 40 minutos	Comprensión lectora: 70 minutos
		Comprensión auditiva: 40 minutos
		Expresión e interacción escrita: 80 minutos
		Expresión e interacción oral: 20 minutos+20 min. preparación

Tabla 13: Duración de las pruebas de ELE y de EFA

Una diferencia que se observa entre las pruebas generales de ELE es que el DELE no considera la competencia lingüística como una sección aparte, aunque sí la evalúa en el contexto de la comprensión lectora. Otra diferencia es el tiempo que se asigna a la sección de expresión e interacción oral, mientras el EXELEAA da 10 minutos y el CELA 15, el

DELE B2 considera 20 minutos de exposición más 20 minutos de preparación de las tareas a exponer.

En relación con la duración de los exámenes, Yagüe (2010) señala que el tiempo que involucra cada prueba es un problema importante para lograr mediciones pertinentes. Dado que, es difícil que una persona logre mantenerse concentrado durante 3 o 4 horas, porque no es una situación habitual; de hecho señala el autor que incluso asistir a clases esta cantidad de horas resulta agotador. Entre los factores que influyen en las pruebas de larga duración señala el agotamiento del estudiante y la pérdida de concentración al tener que responder distintos tipos de tareas y sobre temas diferentes. Sin embargo, se desconoce estudios que investiguen cómo influye el tiempo o duración en los resultados que obtienen los estudiantes y candidatos en las pruebas de ELE (Yagüe, 2010).

2.3. Selección de materiales

Teniendo en cuenta las recomendaciones de los especialistas, en este apartado se presenta algunas consideraciones respecto del proceso de selección de textos orales y escritos.

En relación con la selección del material para las pruebas de comprensión auditiva existe discusión respecto del criterio de autenticidad de los textos elegidos. Mientras algunos se inclinan por mantener el texto sin alteración alguna del medio de donde se obtuvo, otros como Behiels (2010), Martínez (2011) y Figueras y Puig (2013) señalan que lo auténtico no se relaciona solo con usar los textos sin adaptaciones, de hecho en ocasiones son necesarias, sino mas bien tiene que ver con la exigencia de la tarea que se le pedirá al candidato y su relación con las funciones que éste desempeñará en el mundo real. En esta misma línea, Martín (2010) da cuenta de lo difícil que es seleccionar o encontrar material auditivo que se ajuste a las condiciones y necesidades de la enseñanza y evaluación de la LE. Al respecto recomienda que dada la dificultad de acceder a buen material para trabajar esta destreza es conveniente que estos textos se construyan para los fines que se necesita. La elaboración de material auditivo no implica que se quite la noción de “material auténtico”, lo relevante no es tanto la autenticidad de los textos, en cuanto a que sean reales, sino que consideren las características lingüísticas propias del tipo de texto del que se trata, del contexto dado, y que se requiera del candidato la puesta en práctica de los mismos procesos y estrategias de comprensión que se

activarían en una situación similar fuera del contexto de aprendizaje o evaluación de una L2 o LE (Buck, 2009; Bordón, 2006; Martín, 2010).

Debido a la complejidad en la búsqueda de material auditivo adecuado para los propósitos esperados, los expertos recomiendan que para los niveles iniciales (A1, A2) en ocasiones es necesario elaborar materiales *ad hoc*. No obstante, en los textos auditivos para niveles superiores se sugiere conservar los ruidos de fondo, las reformulaciones, las repeticiones, etc. ya que se relacionan con el tipo de discurso que el candidato encuentra fuera del contexto del examen. Además, se debe cautelar que a lo largo de la prueba de comprensión auditiva haya diferentes voces (hombres, mujeres y niños), número de participantes (monólogos y diálogos), de diferentes zonas y en distintas situaciones comunicativas. En dichos textos se debe cuidar la velocidad de la locución, la calidad del audio y la duración, de modo que los fragmentos no sean muy largos para así evitar una sobrecarga de la capacidad de memoria del candidato (Behiels, 2010). En todo este proceso de selección o elaboración de materiales siempre hay que tener en cuenta que se escucha para algo, por lo que debe haber una interrelación entre texto, contexto y tarea, de modo que el contenido sea interesante para el público al que va dirigido (Behiels, 2010).

En la selección de textos para evaluar la comprensión lectora se debe tener claridad respecto del propósito de la lectura, el perfil de los alumnos y el nivel al cual se orienta. En este sentido, Figueras y Puig (2013) apuntan seis aspectos claves para la selección de textos escritos:

- 1) *Tipo de texto*: Se trata del grado de autenticidad del texto. Se recomienda que los textos sean auténticos, aun cuando requieran de algunas adaptaciones, puesto que hay textos en Internet que requieren de algunas modificaciones para hacerlos coherentes y cohesivos.
- 2) *Tema*: En la selección de temas de los textos se debe considerar qué conocimiento de mundo implica un determinado tema, para así evitar una dificultad innecesaria para la evaluación de la habilidad de comprensión lectora. Sin embargo, se debe cuidar que las preguntas sobre el texto no puedan responderse solo por el conocimiento de mundo. Los textos seleccionados no deben tratar temas localistas y deben evaluar las microhabilidades de

comprensión lectora acordes con el nivel de la prueba y no del dominio del tema. A partir de un nivel B2 se considera el uso de textos que traten temas abstractos o de mayor complejidad.

- 3) *Fuente:* Hay que señalar al final del texto la fuente del cual procede. De esta manera se le da autenticidad al material. En el caso de utilizar algún texto elaborado por el redactor de pruebas no debe figurar la autoría.
- 4) *Tipología textual:* En los niveles iniciales es adecuado seleccionar textos con una tipología narrativa o descriptiva. A partir de un nivel B2 se recomienda seleccionar textos de mayor complejidad, donde se evidencie las modalidades expositivas y argumentativas.
- 5) *Longitud:* Es importante que la longitud del texto seleccionado sea adecuado para el nivel y permita plantear un número adecuado de preguntas acorde con su longitud. Por ejemplo, se sugiere que de un texto de 300 se puede realiza cinco o más preguntas. Pese a que no hay una exigencia en cuanto a cantidad de palabras, de acuerdo a lo observado en las distintas pruebas de ELE, un número adecuado para un nivel B2 oscila entre 300 y 500 palabras.
- 6) *Tipología de preguntas:* Hay que señalar mediante qué tipos de preguntas se evaluará la comprensión lectora, por ejemplo, verdadero/falso, selección múltiple, respuesta breve, etc.

Es necesario que el redactor se asegure que lo que se entiende de la lectura es evidente y no solo la interpretación de una persona, para lo cual es relevante llevar a cabo procesos de ensayos y pilotajes.

El MCER define el texto como “cualquier secuencia de discurso (hablado o escrito) relativo a un ámbito específico y que durante la realización de una tarea constituye el eje central de una actividad de lengua” (MCER, 2002: 10).

2.4. Diseño de tareas e ítems

En relación con el diseño de tareas que eliciten la expresión e interacción escritas y orales se debe tener claridad sobre el tema que van a escribir o hablar los candidatos, el cual debe estar acorde con el nivel de lengua definido en la prueba. En las tareas de escritura se debe evaluar el manejo de los distintos tipos de textos y sus formatos, la habilidad de escribir frases adecuadas desde una perspectiva gramatical, léxica, puntuación y ortografía; la habilidad para dar coherencia a las frases, párrafos y cohesión al texto, y la habilidad de manipular frases y palabras adecuadas al contexto y sus destinatarios (Martínez, 2011).

2.4.1. Directrices para la elaboración de tareas académicas

La selección de las tareas de escritura en los exámenes de lengua para el contexto académico ha sido criticada por carecer de autenticidad en cuanto al dominio que se evalúa, esto porque los estudiantes universitarios no escriben textos en 30 minutos y sin el manejo de conocimientos previos (Mendoza, 2015). En este contexto, surge la interrogante respecto de ¿qué tareas de escritura permiten evaluar de manera generalizada a estudiantes universitarios de diferentes carreras? En este sentido, Bachman (2002) señala que mientras más auténtica es la tarea más difícil se hace su corrección y su estandarización, porque es más específica y no se puede generalizar. En este sentido, es complejo establecer un balance entre autenticidad y estandarización por lo que se recomienda que las tareas sean generalizables más que reales. Es decir, que se relacionen con las habilidades que todo estudiante universitario debe mostrar. En este contexto resulta interesante la aportación de Lee (2008), quien concluye que no hay diferencias significativas en los resultados obtenidos por los estudiantes que escribieron un texto sobre un tema general y otro de su especialidad. Acorde con esto y considerando el universo de estudiantes de diversas áreas del saber, se sugiere abordar temáticas académicas generales en el planteamiento de tareas de escritura académica.

En relación con el tipo de tareas, Mendoza (2015) explica que es necesario elaborar tareas de escritura discretas, porque permiten conocer la competencia real del estudiante en cuanto a la habilidad que se está midiendo. Las tareas discretas son aquellas que evalúan una habilidad lingüística de manera aislada, mientras que las tareas integradas ponen en relación las distintas habilidades. La discusión que se da entre estos dos tipos de tareas obedece a la dificultad para evaluar de manera válida y confiable las tareas integradas, dado que si el estudiante tiene que

escuchar un audio y a partir de este escribir un texto, cuando se evalúa no se tiene claridad respecto de la habilidad que está demostrando, si la comprensión lectora o la producción escrita. Aun cuando las tareas integradas reflejan mejor las actividades que se desarrollan en el proceso de enseñanza-aprendizaje, en la evaluación han sido poco utilizadas por la complejidad que representan a la hora medir el dominio del candidato (Figueras y Puig, 2013).

En el planteamiento de las tareas de escritura académica se debe ser lo más justo posible, para ello se debe tener claridad respecto de lo que se está evaluando con la tarea, elegir temas generales y cuidar que las instrucciones estén bien redactadas y sean claras para los estudiantes. Para ello, se sugiere entregar lineamientos respecto de la longitud y del tipo de texto que se espera del candidato. Lee (2008) menciona que los estudiantes prefieren elegir entre varias opciones de temas en una tarea, esto les da la libertad de seleccionar un tema que les parezca más familiar.

En cuanto al formato de aplicación de estas tareas, Mendoza (2015) señala que una de las interrogantes en cuanto a la aplicación de exámenes cronometrados es por qué se continúan haciendo a mano cuando los estudiantes en contextos universitarios escriben en computador. Al respecto menciona que los correctores suelen evaluar de manera distinta los textos escritos a mano y en computador. Los textos escritos en computador se perciben más cortos, no evidencian la competencia estratégica (correcciones, borrones, cambio de orden en las estructuras, etc.) puesto que presentan un producto, aparentemente, acabado. Los textos escritos a mano, dependiendo del tamaño de letra del estudiante, en ocasiones, dan la impresión de que el candidato escribió mucho, lo cual puede distorsionar la percepción del evaluador; la caligrafía en ocasiones dificulta la lectura y no se logra leer el texto completo porque hay partes donde no se entiende lo que se ha escrito (Weigle, 2000).

Dada la complejidad en el diseño y evaluación de tareas abiertas, se necesita pilotear las tareas de escritura de las pruebas de EFA con el objeto de conocer su pertinencia para ser aplicada a diferentes grupos de estudiantes. Hay que considerar que no hay pruebas de EFA que tengan un reconocimiento nacional o internacional (Mendoza, 2015), por lo que este proceso de pilotaje es esencial para ir avanzando en propuestas de tareas válidas que evalúen proficiencia en EFA.

Mendoza (2015) a partir de los resultados obtenidos en el pilotaje de las tareas de escritura del EXELEAA da cuenta de los cambios que tuvieron que realizar en cuanto a las instrucciones y los temas. Menciona dos ejemplos: 1) las modificaciones a unos gráficos donde no era posible elaborar la argumentación en el texto porque faltaban datos que eran necesarios para complementar los datos del gráfico y 2) la redacción de una instrucción relacionada con la educación de hombres y mujeres que llevó a un problema de género que impidió que los estudiantes presentaran posturas a favor y en contra. En este caso, se observa la falta de consideración de un criterio ético en la selección de la tarea (evitar temas polémicos).

En las tareas de expresión e interacción orales es importante dar opciones de tareas para que el estudiante seleccione de acuerdo con sus intereses. Teniendo en cuenta que esta habilidad se evalúa frecuentemente mediante la técnica de entrevista, Martínez (2011) otorga relevancia al papel que tiene el entrevistador, ya que en gran parte de él depende el desempeño del candidato. En ese sentido, Martínez (2011) sugiere considerar algunos de los protocolos de la entrevista de ACTFL, denominada OPI (*Oral Proficiency Interview*), en relación con el comportamiento del entrevistador. Con esto se pretende que, independiente de quién realice este procedimiento, el resultado del candidato sea el mismo, es decir, que la prueba sea confiable. El comportamiento del entrevistador debe:

- Lograr que la entrevista sea lo más parecida a una conversación natural, sin estrés y de acuerdo a los temas y gustos del candidato.
- Evitar el nerviosismo del candidato, de modo que esto no influya en el resultado final.
- No corregir, enseñar o terminar las frases
- No ayudar con vocabulario
- No cortar la conversación
- No expresar su opinión sobre algún tema tratado
- No cansar al candidato con preguntas
- No variar abruptamente en relación con los temas de los que están hablando
- No hacer juicios de la lengua durante la entrevista
- Ser paciente, esperar la respuesta del candidato y aceptar el silencio como parte de la conversación (es posible que el candidato esté pensando, organizando sus ideas)

- Mantener el control de la entrevista, evitando que el candidato ejerza el dominio
- Motivar a los candidatos que son tímidos
- Tranquilizar a los candidatos que estén nerviosos
- Centrar a los candidatos que responden con evasivas

En un contexto de enseñanza, probablemente, el profesor puede aplicar todas estas estrategias, y son válidas, pero en el contexto de la evaluación de proficiencia esto no es posible. Todas estas exigencias hacia el entrevistador permiten lograr una buena muestra respecto al dominio de la lengua por parte del candidato, independiente de su inteligencia, timidez o azar.

2.4.2. Directrices para la redacción de los ítems

Para la elaboración de los ítems es necesario tener en cuenta las sugerencias de Alderson *et. al.* (1998) y Figueras y Puig (2013) respecto de los tipos de preguntas a realizar y luego la forma de redactarlos. En el contexto de este estudio solo referenciaremos las sugerencias para la elaboración de ítems de selección múltiple y de verdadero y falso.

En los ítems de selección múltiple 1) se recomienda dar al menos cuatro opciones de respuesta, de esta forma se cautea que solo haya un 25% de probabilidad que el candidato responda al azar (Alderson *et. al.*, 1998), 2) evitar opciones que contengan un significado similar, 3) evitar que la respuesta correcta sea notablemente más larga o más corta que los distractores, 4) solo una sea la opción correcta, para lo cual se recomienda que otros, principalmente, hablantes nativos revisen o respondan la prueba para comprobar la decisión del redactor, 5) cada distractor debe atraer al menos a una parte de los candidatos, si esto no ocurre es señal de que esta opción no está funcionando, Además, cada una de las opciones debe encajar perfectamente al contexto del enunciado de la pregunta.

En relación con los ítems de verdadero y falso se recomienda evitar enunciados negativos.

En cuanto al funcionamiento de los distractores, no existe acuerdo respecto del porcentaje de respuestas que debe atraer cada distractor, sin embargo, los redactores consideran deficiente si un distractor no es elegido dentro de una muestra de 20 o 30 exámenes. Normalmente, solo uno o dos distractores llaman la atención en la selección que realizan los estudiantes. Es lógico pensar que aquellos distractores que no son seleccionados por ninguno de los candidatos

deberían eliminarse. Se considera como un buen distractor aquel que atrae al menos un 10% de las respuestas (Alderson *et al.*, 2002).

Todos los aspectos mencionados son solo una muestra de las múltiples perspectivas de la evaluación y de las consideraciones y responsabilidades que asumen quienes elaboran pruebas de lengua. Así, un instrumento bien diseñado evitará la controversia de por qué estudiantes competentes fracasan (por ejemplo, de corrección objetiva) y otros con menos dominio aprueban, sin tener que atribuir como respuesta que esto se debe a factores individuales del candidato, como el nerviosismo o la tensión que generan las pruebas. En términos concretos, se trata de eliminar el efecto del azar o la influencia de elementos externos al instrumento mismo y proceder con profesionalismo y rigurosidad, aplicando las técnicas y procedimientos necesarios para lograr la construcción de instrumentos válidos y confiables.

2.5. Cualidades de los instrumentos de evaluación

Para que un instrumento de evaluación sea eficaz y proporcione la información requerida debe cumplir con tres requisitos fundamentales: validez, confiabilidad y viabilidad (Bordón, 2006; MCER, 2002).

En términos generales, la validez se refiere a si el examen en su totalidad mide lo que supone debe medir, lo cual se relaciona con los propósitos para los cuales fue creado el instrumento, por lo tanto, debe ser válido para esos objetivos. Si no se cumple este requisito entonces la prueba no es útil, ya que no se podrá hacer inferencias adecuadas de ella. La validez es uno de los temas que merece más atención de parte de los expertos en evaluación. Existen distintos tipos de validez, mientras Bordón (2006) se inclina por la validez interna dado que no requiere de análisis estadísticos, Alderson *et al.*, (1998) recomiendan que hay que llevar a cabo la validación de una prueba lo más que se pueda, entre más tipos de validez se pueda establecer es mejor. Los diferentes tipos de validez son en realidad distintos métodos para evaluarla. Entre los distintos tipos de validez se menciona la validez aparente, de contenido, de respuesta, predictiva, concurrente y de constructo.

La validez de apariencia se refiere a la credibilidad aparente de una prueba y es considerada como poco científica por los expertos en evaluación. Este tipo de validez se obtiene pidiendo a hablantes nativos (no expertos) que rindan la prueba y luego den una opinión intuitiva respecto del instrumento, por ejemplo, una apreciación sobre las instrucciones, los ítems, la

duración, la complejidad de las preguntas etc. Los resultados obtenidos permiten una primera apreciación en relación con la puesta en práctica de la prueba para hacer los ajustes necesarios antes del pilotaje con el grupo meta. En la actualidad, está adquiriendo importancia este tipo de validez, debido a la perspectiva comunicativa que está primando en la evaluación de la lengua, lo cual supone que un examen comunicativo debería parecerse a lo que se hace en el mundo real con la lengua, y en este sentido es importante lo que piensan sobre el instrumento los hablantes nativos de la lengua (Alderson *et al.*, 1998; Martínez, 2011).

La validez de contenido se refiere a la representatividad o adecuación del contenido de un instrumento evaluativo de acuerdo a los objetivos y el nivel que se dirige. Este tipo de validez requiere de la evaluación de expertos en el ámbito. Se trata de recoger opiniones de especialistas, por ejemplo, a través de escalas de valoración construidas para dicho fin. El juicio de expertos se define como “una opinión informada de personas con trayectoria en el tema, que son reconocidas por otros como expertos cualificados en éste, y que pueden dar información, evidencia, juicios y valoraciones” (Escobar y Cuervo, 2008). En cuanto al número de expertos que se requiere hay diversidad de opiniones entre los especialistas, de hecho Cabrero y Llorente (2013) señalan que el número de expertos depende del acceso a ellos o la facilidad de contar con especialistas en el tema objeto de estudio. En relación con las técnicas de recolección de datos existen varios métodos, que se clasifican en individuales o grupales. En los primeros se encuentra el método de agregados individuales y el método Delphi, en los dos el experto evalúa de manera individual, pero en el Delphi luego que el investigador analiza las respuestas envía a cada juez la media obtenida para que éste reconsidere el juicio realizado hasta lograr un consenso. En cuanto a los técnicas grupales está la nominal y la de consenso, en las dos se reúne a los expertos, pero en la de consenso se exige un mayor nivel de acuerdo entre los jueces (Escobar y Cuervo, 2008). Para realizar un juicio de expertos se debe elaborar cuestionarios o planillas donde calificará el experto. En el momento de la aplicación de este juicio se debe explicar a cada juez el objetivo de la evaluación y en qué consiste cada una de las dimensiones a evaluar.

La validez de respuesta se refiere a la recogida de información respecto del razonamiento que hacen los participantes cuando responden los ítems de una prueba. Es un proceso introspectivo, por lo que no es fácil de obtener datos de este tipo. Una forma de recoger datos

es hacerlo retrospectivamente, una vez que el estudiante ha terminado la prueba se le entrevista para conocer los motivos por los que seleccionó tales respuestas.

La validez concurrente corresponde a la comparación de los resultados de la prueba con otra medida de los candidatos, obtenida durante el mismo periodo de aplicación del instrumento. Puede ser una prueba similar o bien otra distinta. Lo que se busca es mostrar un coeficiente de correlación entre ambas mediciones. La validez predictiva se diferencia de la validez concurrente porque recoge las medidas externas después de la aplicación de la prueba. Es bastante utilizada en las pruebas de proficiencia, donde se intenta predecir cómo actuará el candidato en el futuro.

La validez de constructo se refiere a una forma de validación que busca determinar en qué medida el examen responde a la teoría en la que se ha basado. En otras palabras, si la teoría se ve reflejada en la puesta en práctica de la prueba.

Otra cualidad relevante de los instrumentos de evaluación es la confiabilidad. Se refiere a la consistencia de la medición. Bordón (2006) afirma que un examen es confiable si proporciona resultados similares con el mismo grupo de estudiantes en situaciones distintas y en la misma situación con grupos de candidatos semejantes. Aclara la autora que en ocasiones puede darse que los resultados de un examen sean confiables (consistentes), no obstante, el instrumento puede no ser válido, en el sentido que no mida lo que debe medir.

El requisito de viabilidad se refiere a la factibilidad de administrar la prueba bajo las condiciones definidas en el diseño de la prueba. Esto implica que el instrumento debe tener un formato aceptable para los estudiantes (candidatos), si se utilizan pruebas impresas deben leerse sin problemas y si están disponibles a través de la red deben funcionar correctamente (acústica y visibilidad debe ser igual para todos los participantes en el examen). En definitiva, la actuación del estudiante no se debe ver afectada por factores externos como el calor, el frío, los ruidos extremos, etc. También es necesario considerar la viabilidad en términos de los recursos humanos, técnicos y económicos que permitan llevar a cabo el examen (Bordón, 2006).

2.6. Análisis de los ítems

Una de las maneras de aproximarse al análisis de los ítems en una prueba es la Teoría Clásica de los Tests (TCT).

2.6.1. La Teoría Clásica de los Tests (TCT)

La TCT corresponde al conjunto de principios teóricos y métodos cuantitativos que se han usado a partir de los años 50 para guiar a los elaboradores, administradores y usuarios de las pruebas en la comprensión de las propiedades estadísticas de los resultados, y cuando sea necesario mejorar los instrumentos (Figueras y Puig, 2013).

En la TCT se considera los siguientes conceptos:

- El ítem o pregunta que, dentro de un conjunto, constituye un instrumento de evaluación.
- La nota observada que corresponde a la nota resultante del cálculo de aciertos en una prueba.
- La nota real que corresponde al alumno considerando que existe una prueba perfecta.
- La media que corresponde a la suma de las notas o puntajes de una prueba dividida por el número de estudiantes que la han rendido.
- La mediana corresponde al valor que divide las puntuaciones del grupo en la mitad una vez que los datos han sido ordenados. Esta medida indica que la mitad de las puntuaciones del grupo se encuentran por debajo de este valor y la otra mitad por encima.
- La moda corresponde al valor, puntaje o nota que más se repite en el grupo.
- La desviación típica muestra la distancia entre las notas de los diferentes examinandos con respecto a la media de la prueba. Figueras y Puig (2013) ejemplifican esta medida señalando que si una prueba tiene 10 puntos y la media

de esta prueba es de 7 puntos, habrá alumnos que se desvíen 1 punto sobre esta media, pero habrá otros que se ubiquen por debajo de este promedio. Por lo tanto, la desviación típica considera la media de todas las diferencias que presentan todos los estudiantes en relación con la media de la prueba.

- La varianza indica la variabilidad de una prueba en cuanto a sus resultados, puesto que no se puede esperar que todos los estudiantes tengan las mismas notas. La varianza corresponde al cuadrado de la desviación típica.

La TCT define dos medidas que se calculan para cada ítem en una prueba objetiva (Alderson *et al.*, 1998): el índice de dificultad y el índice de discriminación.

2.6.1.1. Índice de dificultad

Para conocer el grado de dificultad de una pregunta se calcula el coeficiente de dificultad representado mediante el porcentaje de alumnos que lo contestan correctamente en relación con el número total de examinandos que rindió la prueba. La fórmula para calcularlo es la siguiente:

$$\text{IDif} = \frac{\text{N}^\circ \text{ de alumnos que contesta bien al ítem}}{\text{N}^\circ \text{ total de alumnos que realizan la prueba}}$$

El valor del coeficiente de dificultad dependerá del tipo de prueba que se trate, puesto que el índice de dificultad esperable para una prueba de nivel no será el mismo que para una prueba de dominio. Alderson *et al.*, (1998) recomiendan que si se quiere tener una amplia distribución en las notas de los estudiantes se debe seleccionar ítems con un índice de dificultad cercano al 50%, puesto que estas preguntas facilitan una mayor variación entre los estudiantes. No obstante, si se quiere conocer el nivel concreto de dificultad que tiene una prueba se debe seleccionar ítems más fáciles para lograr que el grupo obtenga la media aritmética esperada.

En relación con lo anterior, resulta interesante la propuesta de Cervantes (1989), quien presenta una tabla donde clasifica los distintos índices de dificultad de los ítems en una prueba.

- Valores comprendidos entre 0,86 y 1,00 , ítems muy fáciles
- Valores comprendidos entre 0,71 y 0,85 , ítems fáciles
- Valores comprendidos entre 0,40 y 0,70 , ítems deseables
- Valores comprendidos entre 0,15 y 0,39 , ítems difíciles
- Valores comprendidos entre 0,01 y 0,14 , ítems muy difíciles

Acorde con esta tabla, los ítems que se tengan un índice de dificultad de 0,86 hacia arriba se consideran muy fáciles, de 0,71 a 0,85 fáciles. Mientras que los índices de dificultad deseables para seleccionar los ítems de una prueba se encuentran entre 0,40 y 0,70. Los índices de dificultad de 0,39 hacia abajo se consideran ítems difíciles y a partir del 0,14 muy difíciles.

Si bien se pretende que una prueba cuente con ítems de índices de dificultad similares para que sea más precisa, López-Mezquita (2005) señala que es adecuado que la prueba considere un número de ítems con mayor y menor dificultad para así posibilitar la discriminación entre alumnos que tienen una mayor o menor proficiencia en la lengua.

2.6.1.2. Índice de discriminación

En el análisis de los ítems no solo es necesario conocer el índice de dificultad de cada ítem, sino que también se necesita averiguar si los estudiantes que respondieron bien al ítem son los que más saben o tienen mayores habilidades en la segunda lengua. El índice de discriminación (ID) se refiere a si un ítem es capaz de distinguir entre alumnos con diferentes niveles de lengua. Si un ítem funciona bien, lo esperable es que los estudiantes con mejores notas respondan correctamente y los con bajas notas no lo hagan (Figueras y Puig, 2013), de no presentarse esta situación entonces se debe revisar la pregunta. Para calcular la discriminación se ordena al grupo de estudiantes según la nota total que hayan obtenido, se divide al grupo en tres partes iguales (alto, medio y bajo), siempre el grupo más alto y más bajo deben tener el mismo número de sujetos. Enseguida, se contabiliza cuántos estudiantes del grupo alto y del bajo contestaron bien la pregunta para luego calcular las diferencias entre ambos grupos y finalmente dividir este resultado por el número total de alumnos del grupo.

En cuanto a los valores aceptables para un adecuado ID, Ebel y Frisbie (1991) y Cervantes (1989) señalan los siguientes:

- Valores por encima de 0,40 → ítems muy buenos (alta discriminación).
- Valores comprendidos entre 0,30 y 0,39 → ítems razonablemente buenos
- Valores comprendidos entre 0,20 y 0,29 → ítems regulares
- Valores por debajo de 0,19 → ítems malos (baja discriminación)

El mayor índice de discriminación posible es 1 y el valor mínimo es -1. El mayor índice de discriminación se logra siempre que todos los alumnos del grupo más alto logren responder correctamente el ítem y ninguno del grupo bajo lo haga, sin embargo, Alderson *et al.*, (1998) señala que es difícil que se de esta situación. Los valores superiores a 0,2 o 0,3 muestran que los examinandos que responden correctamente a ese ítem tienden asimismo a responder correctamente a más ítems del resto de la prueba y que, por lo tanto, se trata de un ítem discriminativo. Por el contrario, los valores negativos indican que aquellos candidatos que contestan correctamente a ese ítem obtienen peores puntuaciones en la prueba en conjunto. El valor 0 indica ausencia de relación entre contestar correctamente al ítem y la puntuación obtenida en la totalidad del test, por lo que corresponde a un ítem deficiente (López-Mezquita, 2005)

De acuerdo con López-Mezquita (2005), el porcentaje ideal de ítems en una prueba en cuanto a su índice de discriminación sería el siguiente:

- Valores por encima de 0,40 , más de un 25%
- Valores comprendidos entre 0,20 y 0,39 , más de un 25%
- Valores por debajo de 0,19 , menos de un 20%

2.7. Sociedades internacionales de evaluación de lenguas

La evaluación y las pruebas de segundas lenguas y de lenguas extranjeras son temas de larga data. No obstante, la evaluación como profesión y área científica es relativamente nueva, ya que las primeras asociaciones internacionales de pruebas de lengua dignas de ser consideradas se han establecido hace unos pocos años (años 90), aun cuando la evaluación de lenguas formara parte de las organizaciones dedicadas a la enseñanza de la lengua y a la lingüística aplicada desde mucho antes (Huhta, 2008).

Las principales organizaciones de evaluación de lenguas son: ALTE (Asociación de Evaluadores de Lengua en Europa), ILTA (Asociación Internacional de Pruebas de Lengua,) EALTA (Asociación Europea de Evaluación y Pruebas de Lengua) y en el ámbito del ELE, el SICELE (Sistema Internacional de Certificación de Español como Lengua Extranjera). Dichas entidades han tenido como objetivo común la elaboración de códigos de práctica que orientan respecto de las consideraciones éticas, de diseño y de impacto que hay que tener en cuenta al momento de evaluar lenguas.

ALTE es la asociación más antigua de estas cuatro. Se estableció en 1989 por las universidades de Cambridge y de Salamanca. El objetivo inicial de ALTE era establecer normas comunes para las pruebas de lengua en toda Europa, apoyando de esta manera el multilingüismo y ayudando a preservar la riqueza del patrimonio lingüístico europeo. A esto se suma que es imprescindible que los candidatos puedan rendir pruebas de idiomas que sean equitativas y precisas respecto de su competencia en la lengua, y que además gocen de reconocimiento a nivel mundial, de modo que los resultados puedan ser comparados con los niveles asignados en otras lenguas. En este contexto, ALTE ha fijado un conjunto de normas comunes para garantizar estándares mínimos para los exámenes de sus miembros, que cubren todas las etapas del proceso de evaluación de lenguas: 1) desarrollo de la prueba, 2) redacción de tareas e ítems, 3) administración de la prueba, 4) puntuación y calificación, 5) informe con los resultados de la prueba, 6) análisis de la prueba y 7) reporte de resultados.

ALTE, en la actualidad cuenta con 34 miembros, entre ellos algunos de los organismos de evaluación de idiomas líderes a nivel mundial, además tiene 40 instituciones afiliadas, así como también varios cientos de afiliados individuales.

En el contexto americano se encuentra ILTA que es la principal organización internacional de evaluación y pruebas de lengua. Se estableció en el año 1994 y fue creciendo en el ámbito de la principal conferencia internacional de pruebas de lengua, el Coloquio de Investigación de Pruebas de Lengua (LTRC, *Language Testing Research Colloquium*, en sus siglas en inglés), que corresponde a la conferencia oficial de esta asociación.

El propósito de ILTA es promover el mejoramiento de las pruebas de lengua en el mundo.

En este sentido, sus objetivos son:

- Estimular el desarrollo profesional mediante la celebración de talleres y conferencias.
- Promover la publicación y difusión de información relacionada con el área de las pruebas de lengua.
- Desarrollar el área de las pruebas de lengua y proporcionar personas capaces de liderar esa área.
- Proporcionar servicios profesionales a sus miembros.
- Incrementar en el público el reconocimiento y apoyo a la evaluación de lenguas como profesión.
- Estimular el orgullo profesional entre sus miembros.
- Reconocer la relevancia de los logros profesionales de sus miembros.
- Cooperar con otros grupos interesados en las pruebas de lengua.
- Cooperar con otros grupos interesados en estudios estadísticos y en lingüística aplicada.

EALTA es la Asociación Europea de Evaluación y Pruebas de Lengua. Fue creada en el año 2004 con apoyo financiero de la Comunidad Europea. La finalidad de EALTA es promover la comprensión de los principios teóricos de las pruebas de lengua y de la evaluación, así como la mejora y el uso compartido de las prácticas evaluativas en toda Europa.

Los principales objetivos de la EALTA son:

- Compartir la experiencia profesional en el campo de la evaluación y pruebas de lengua.
- Incrementar la comprensión de la evaluación y de las pruebas de lengua.

- Promover el reconocimiento de la evaluación y pruebas de lengua como disciplina profesional en Europa.
- Mejorar la práctica de los sistemas de evaluación y pruebas de lengua en Europa.
- Proveer formación en pruebas de lengua y evaluación.
- Formar expertos en pruebas de lengua y evaluación para la resolución de los problemas en evaluación.
- Establecer vínculos con otros grupos interesados en la evaluación y las pruebas de lengua.
- Participar en actividades para la mejora de las pruebas de lengua y la evaluación en Europa.

Específicamente, para el caso del español, recientemente, se ha creado el Sistema de Certificación del Español como Lengua Extranjera (SICELE). El nacimiento del SICELE tuvo su origen en el III Congreso Internacional de la Lengua Española, realizado en Rosario (Argentina) en noviembre de 2004, donde se emplazó a las instituciones que difunden el español a buscar un mecanismo que permitiera la creación de un sistema de certificación universal dirigido a los estudiantes de español. Fue así, como en octubre de 2005, en la Reunión de Rectores en Salamanca (España), con motivo de la Cumbre Iberoamericana de Jefes de Estado y de Gobierno, se acuerda la creación de un sistema de certificación internacional del español como lengua extranjera y se nombra una Comisión Académica, formada por un representante por país, para su desarrollo. En marzo de 2007, en Cartagena de Indias, Colombia con motivo del IV Congreso Internacional de la Lengua Española diferentes rectores y representantes de diferentes universidades del mundo ratificaron la constitución del SICELE.

2.7.1. Criterios evaluativos de las sociedades de evaluación de lenguas

Las distintas sociedades de evaluación han tenido como objetivo común la elaboración de códigos de práctica que orientan respecto de las consideraciones éticas, de diseño y de impacto que hay que tener en cuenta al momento de evaluar lenguas.

Con respecto a los principios éticos, es importante señalar que a partir de los años noventa ha habido una preocupación por estos aspectos en la evaluación de LE (Bordón, 2006). En este sentido, las instituciones han trabajado para establecer o elaborar sus códigos de práctica haciendo alusión a estas consideraciones.

- **ALTE**

En el ámbito del establecimiento de normas en la elaboración de pruebas, ALTE, ha publicado dos documentos de relevancia: El Código de Práctica y Los estándares mínimos.

El Código de Práctica identifica y define tres grupos principales de interesados en lo que respecta al proceso de elaboración de una prueba: *los elaboradores del examen, los candidatos y los usuarios del examen*. Este Código se divide en dos partes: 1) Responsabilidades de los miembros de ALTE y 2) Responsabilidades de los usuarios de exámenes. De esta forma, las responsabilidades de los miembros de ALTE se definen en cuatro ámbitos: 1) Confección de exámenes, 2) Interpretación de los resultados del examen, 3) Esfuerzo por conseguir la imparcialidad e 4) Información para los candidatos; y las responsabilidades de los usuarios de exámenes en: 1) Selección de exámenes adecuados, 2) Interpretación de los resultados, 3) Esfuerzo por conseguir imparcialidad e 4) Información para los candidatos. Cada uno de los participantes del proceso evaluativo y miembro de ALTE tiene la obligación de cumplir con estas exigencias desde las responsabilidades que les corresponda.

En relación con los principios éticos, ALTE, en las secciones de las responsabilidades de sus miembros y usuarios de exámenes establece un apartado denominado “Esfuerzo por conseguir imparcialidad” que apunta a los aspectos éticos a considerar tanto cuando se elabora una prueba como cuando los usuarios interpretan sus resultados. En este sentido, los miembros de ALTE se comprometen a elaborar exámenes con la mayor imparcialidad posible teniendo en cuenta que los candidatos tendrán diferentes procedencias (origen étnico, sexo, creencias, discapacidad, etc.). Concretamente, se comprometen a: 1) Revisar y

modificar las pruebas y materiales de los exámenes con el fin de evitar contenidos potencialmente ofensivos o con un lenguaje inadecuado, 2) Establecer procedimientos que ayuden a asegurar que las diferencias en las actuaciones de los candidatos se deben principalmente a las habilidades que se evalúan y no a factores como la raza, el género o el origen étnico y 3) Adaptar los exámenes y facilitar el acceso al examen a aquellos candidatos que, por sufrir algún tipo de discapacidad, no puedan realizarlo. Respecto a los usuarios del examen, estos se comprometen a seleccionar los exámenes que se hayan elaborado buscando la mayor imparcialidad posible para candidatos de diferente raza, género, origen étnico, o para candidatos que sufran de algún tipo de discapacidad.

En Los Estándares Mínimos se establece los 17 principios que ALTE considera para lograr perfiles de calidad en los exámenes. Dichos estándares se distribuyen en cinco apartados: 1) Confección de exámenes, 2) Administración y logística, 3) Puntuación y calificación, 4). Análisis de exámenes y 5). Comunicación con los partícipes.

En la actualidad, ALTE cuenta con un nuevo indicador de calidad, ALTE Q-mark, que las organizaciones miembros pueden usar para demostrar que sus exámenes han pasado una rigurosa auditoría y cumplen con los 17 estándares de calidad de ALTE. Este indicador otorga seguridad a los usuarios de las pruebas. De esta forma se cautela que cada una de las organizaciones afiliadas cumpla con las normas establecidas y que los usuarios reciban un servicio de calidad.

- **ILTA**

En el contexto de la elaboración de pruebas, ILTA, ha publicado dos documentos de interés: 1) La Guía de práctica y 2) El Código de Ética.

La Guía de Práctica se divide en dos partes, las consideraciones básicas para la buena práctica de las pruebas en todas las situaciones (diseño de pruebas, obligaciones de las instituciones y de quienes administran pruebas de alto impacto, responsabilidades de los usuarios de las pruebas, etc.) y los derechos y responsabilidades de los examinadores.

El Código de Ética de ILTA tiene como objetivo ofrecer un punto de referencia respecto de un comportamiento ético de parte de los evaluadores de lengua. Este Código se compone de

nueve principios en los cuales los evaluadores se comprometen a: mostrar respeto por la dignidad de cada uno de los candidatos (género, raza, origen étnico, orientación sexual, lengua materna, religión, afiliación política), almacenar la información obtenida en el ejercicio de la profesión teniendo en cuenta la confidencialidad de los datos, restringirse a los códigos éticos nacionales e internacionales al realizar investigación (ensayos, experimentos) en este ámbito, no permitir un mal uso de los conocimientos profesionales referidos a la evaluación, continuar el desarrollo profesional en el área y compartir este conocimiento con colegas y profesionales de lengua. Además, compartir la responsabilidad de mantener la integridad de la profesión, esforzarse por mejorar la calidad de las pruebas de lenguas, los servicios de evaluación y enseñanza con el objetivo de contribuir a la sociedad en estas temáticas, ser conscientes de sus obligaciones con la sociedad en la que trabajan, y considerar los efectos potenciales de la evaluación a corto y largo plazo sobre los proyectos de los candidatos.

- **EALTA**

En el contexto del establecimiento de normas para una buena práctica evaluativa, EALTA, ha elaborado una serie de directrices orientadas hacia tres posibles miembros:

- 1) La formación de profesores en evaluación: Este Código proporciona directrices evaluativas para la formación de profesores.
- 2) La evaluación en el aula: Orienta respecto a los objetivos y especificaciones de la evaluación en la clase, sus procedimientos y consecuencias.
- 3) El desarrollo de pruebas en organismos o centros evaluadores públicos o privados: Proporciona directrices respecto a los objetivos y especificaciones de las pruebas, la elaboración de las pruebas y la redacción de ítems, el control de calidad y análisis de las pruebas, el proceso de administración de la prueba, la revisión, la repercusión sobre la docencia y el anclaje al MCER (2002).

En estas directrices de EALTA subyacen algunos principios éticos relacionados con el respeto y equidad del profesor/evaluador hacia los estudiantes o candidatos, la

responsabilidad de quienes participan de la elaboración de pruebas, las consecuencias que tiene la evaluación en la enseñanza y para los participantes, y la colaboración entre las tres partes implicadas: formación de profesores, evaluación en el aula y desarrollo de pruebas en centros evaluadores.

- **SICELE**

El objetivo de SICELE es garantizar el cumplimiento de una serie de parámetros que permitan asegurar la calidad y transparencia en la certificación del dominio del español. De modo, que las certificaciones que tengan el sello del SICELE tengan un reconocimiento transnacional. Para ello, ha establecido sus estándares de calidad distribuidos en cuatro grandes apartados: 1) Planificación y desarrollo del examen, 2) Dispositivo de gestión y administración de las pruebas de examen, 3) Calificación y análisis de resultados e 4) Información a los usuarios. En cada uno de estos ámbitos se incluye una serie de consideraciones que se debe tener en cuenta para dar cumplimiento a cada una de estas etapas.

Los aspectos éticos considerados en los estándares del SICELE dicen relación con la preocupación por garantizar la protección y confidencialidad de los datos de los candidatos, la adaptación de las pruebas a candidatos con necesidades especiales, la aplicación de procedimientos que eviten sesgos socioculturales o de otra índole en las calificaciones, la entrega de información clara a los usuarios de las pruebas y la fijación de procedimientos para realizar las reclamaciones y apelaciones.

Los parámetros abordados por las distintas sociedades son indicaciones generales que luego cada institución o profesor aplican según sus necesidades y disponibilidad de recursos. Por lo tanto, se requiere de modelos que, a partir de la operacionalización de los principios teóricos, provean de criterios que ayuden a resolver la problemática sobre cómo elaborar una prueba de proficiencia en el ámbito del español con fines académicos.

CAPÍTULO II: METODOLÓGICO

3. Metodología

En este apartado se presenta las preguntas de investigación, los objetivos, el tipo de estudio y los procedimientos considerados para dar respuesta a las preguntas planteadas.

3.1 Preguntas de investigación

Considerando la problemática planteada en cuanto a que la evaluación del EFA es una necesidad, tanto a nivel nacional como internacional, dado que no existen pruebas reconocidas ni válidas en este ámbito, esta investigación pretende dar respuesta a las siguientes preguntas:

1. ¿Qué criterios se debe considerar en el diseño de un instrumento de evaluación válido, confiable y viable en el Español con Fines Específicos Académicos (EFA)?
2. ¿Cómo implementar estos criterios en un prototipo de prueba de proficiencia de EFA en línea para un nivel B2?
3. ¿Cuál es el comportamiento del prototipo de prueba diseñado en una aplicación piloto con una muestra de estudiantes de intercambio de Español como Lengua Extranjera?

3.2. Objetivos generales

Para dar respuesta a las preguntas de investigación mencionadas, se definió los siguientes objetivos generales:

1. Determinar un modelo de criterios evaluativos para el diseño de un prototipo de prueba de EFA en línea de nivel B2
2. Implementar los criterios evaluativos del modelo, en la elaboración y validación del prototipo de prueba de EFA en línea de nivel B2
3. Evaluar el prototipo de prueba de EFA en línea de nivel B2 a partir de una aplicación piloto

3.2.1. Objetivos específicos:

1.1. Definir un modelo de criterios evaluativos necesarios para el diseño de un prototipo de prueba de EFA en línea para un nivel B2

2.1. Diseñar un prototipo de prueba de EFA en línea para un nivel B2 que considere el modelo de criterios evaluativos

3.1. Determinar la validez de contenido y la confiabilidad del prototipo de prueba de EFA a través de un juicio de expertos

3.2. Determinar la validez de los principios tecnológicos de CALT en la implementación del prototipo de prueba de EFA en línea de nivel B2 por un grupo de expertos

3.2. Determinar la validez de apariencia del prototipo de prueba de EFA en línea a través de un grupo no experto

3.3. Determinar el índice de dificultad y de discriminación de los ítems de corrección objetiva del prototipo de prueba de EFA en línea de nivel B2

3.3. Conocer las apreciaciones de los estudiantes extranjeros sobre el prototipo de prueba de EFA en línea de nivel B2

3.4. Evidenciar la confiabilidad de las secciones de expresión escrita y oral del prototipo de prueba de EFA en línea de nivel B2

3.3. Tipo de estudio

Para responder a las preguntas de investigación planteadas se llevó a cabo un estudio con un diseño mixto de tipo exploratorio secuencial (Hernández, Fernández y Baptista, 2010). Este tipo de diseño exploratorio secuencial (DEXPLOS) implica una fase inicial de recolección y análisis de datos cualitativos seguida de otra fase donde se recaba y analiza datos cuantitativos. El DEXPLOS es utilizado cuando el investigador necesita desarrollar un instrumento nuevo porque los existentes son inadecuados o no se puede acceder a ellos. Para ello se sugiere tener en cuenta estas tres etapas: a) Recabar los datos cualitativos y analizarlos, b) Usar los resultados para construir el instrumento cuantitativo y c) Administrar

el instrumento a una muestra probabilística de la población para validarlo.

3.4. Procedimiento

De acuerdo con el tipo de estudio, en este trabajo de tesis se consideró dos etapas:

1) Cualitativa: Consistió en la revisión de los criterios evaluativos propuestos por los expertos en elaboración de pruebas de segundas lenguas y de los códigos de práctica de las distintas sociedades evaluativas. Estos datos cualitativos permitieron definir el modelo de criterios para diseñar un prototipo de prueba de EFA en línea de nivel B2, como se observa en la figura 9.

2) Cuantitativa: Basado en el modelo de criterios se diseñó y desarrolló el prototipo de prueba de EFA en línea de nivel B2. Luego, se consideró a un grupo de hablantes nativos de español para realizar la validez de apariencia y a un grupo de expertos para validar el instrumento a nivel de contenido y de implementación en línea. Finalmente, se administró el prototipo de prueba a un grupo de estudiantes extranjeros de intercambio en la Universidad de Concepción, Chile. Mediante esta aplicación piloto se pudo conocer el índice de dificultad y de discriminación de los ítems, la confiabilidad del prototipo de prueba, el desempeño de los estudiantes según sus áreas de estudio y la apreciación de los estudiantes sobre este prototipo de prueba de EFA en línea de nivel B2.

3.4.1. Etapa 1: Modelo de criterios evaluativos para el diseño del prototipo de prueba de EFA en línea

En relación con la problemática planteada y acorde con el primer objetivo general de esta investigación se definió el modelo de criterios, basado en los hallazgos investigativos de los especialistas en elaboración de pruebas de proficiencia en ELE y en EFA, y en las directrices de las sociedades internacionales de evaluación de segundas lenguas. Como se observa en la figura 9, este modelo se compone de cuatro criterios:

1. Éticos
2. Diseño
3. Elaboración
4. Validación y pilotaje

Estos cuatro criterios se consideraron fundamentales para diseñar, elaborar y validar un prototipo de prueba de EFA en línea dirigida, en este caso, a los estudiantes extranjeros de intercambio en la Universidad de Concepción. Es importante señalar que este modelo es interactivo, puesto que los criterios no se abordan solo de una manera lineal, sino que se retroalimentan entre sí. Por ejemplo, los criterios de validación pueden incidir en los criterios de diseño o de elaboración.

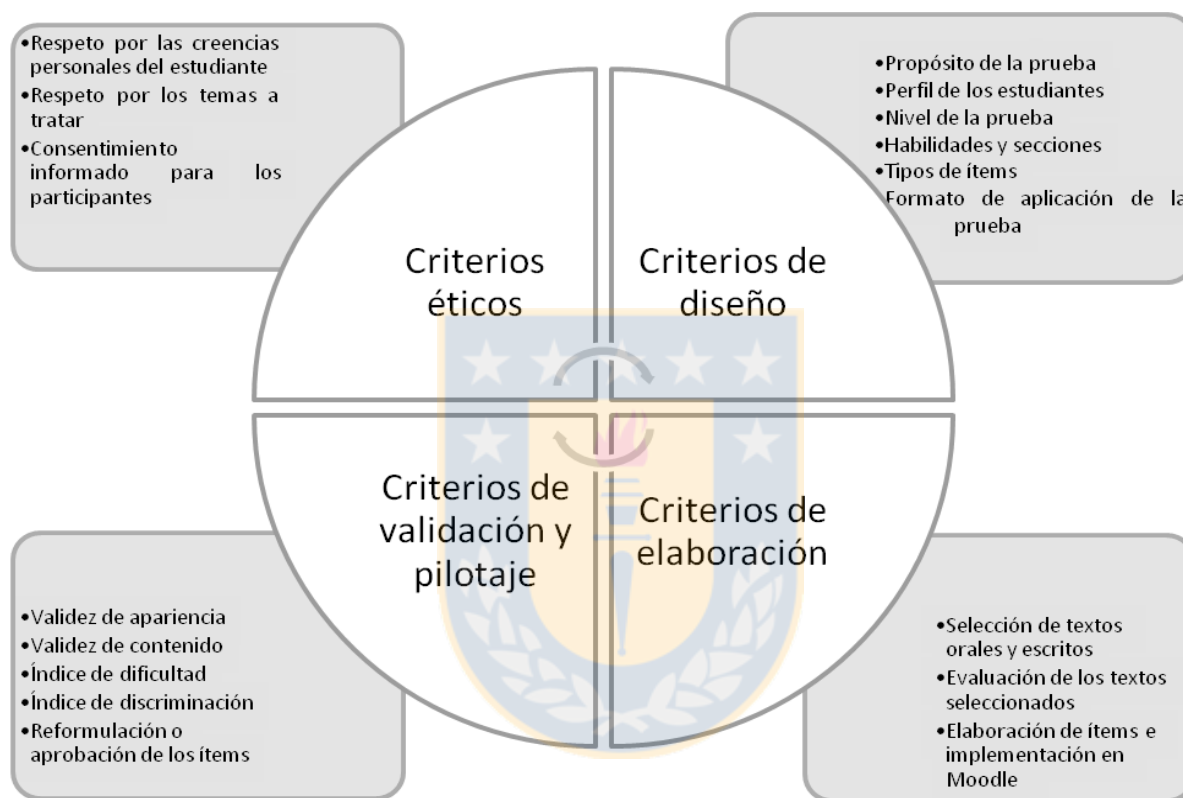


Figura 9: Modelo metodológico para la elaboración de una prueba de EFA

Los criterios éticos de este modelo son fundamentales, ya que se consideran desde el diseño hasta la validación del instrumento. Esto se refleja en la selección temática de los textos orales y escritos, y de las tareas abiertas que eliciten el dominio del español en contextos académicos; en todos ellos se muestra respeto por las creencias, la procedencia y la formación del estudiante.

Los criterios de diseño corresponden a las directrices que orientan la elaboración del instrumento. En el diseño se definió que el prototipo de prueba de EFA en línea se orienta a

los estudiantes extranjeros de intercambio en la Universidad de Concepción y que mediante ítems de corrección objetiva y de tareas abiertas se evalúa las cuatro habilidades lingüísticas y la competencia lingüística en un nivel B2. En los criterios de elaboración se considera la búsqueda de los textos orales y escritos, la evaluación de estos por un grupo de expertos, para luego avanzar a la redacción de los ítems y su implementación en línea. Finalmente, el criterio de validación y pilotaje contempla la validez de apariencia, de contenido y la aplicación de la prueba al grupo de estudiantes extranjeros para conocer el índice de dificultad y de discriminación de los ítems y el grado de confiabilidad del instrumento. Los resultados obtenidos, tanto en el proceso de validación como de pilotaje, permitirán contar con un instrumento válido y confiable para evaluar el nivel B2 de EFA.

De acuerdo con el tipo de estudio, en esta primera etapa se recopiló y definió los criterios necesarios recomendados por los expertos para elaborar el prototipo de prueba de EFA. En la etapa siguiente se aplicó los criterios de este modelo en la elaboración y validación de apariencia y de contenido del instrumento.

3.4.2. Etapa 2: Aplicación del modelo de criterios evaluativos en la elaboración del prototipo de prueba de EFA en línea de nivel B2

En esta segunda etapa se presenta la aplicación de los criterios recopilados en la etapa cualitativa, en la elaboración y validación del prototipo de prueba de EFA en línea de nivel B2.

3.4.2.1. Criterios éticos y de equidad

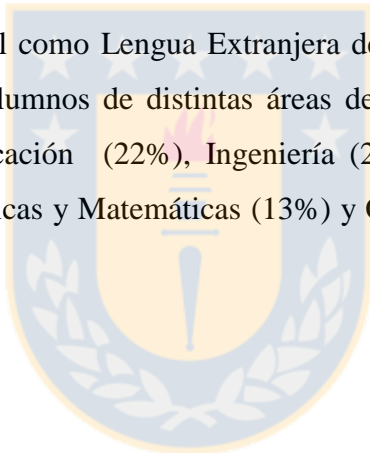
La operacionalización de los criterios éticos referidos a mostrar respeto por las convicciones y creencias personales y evitar el tratamiento de temas controversiales se consideró en la selección de los textos orales y escritos y en los temas de las tareas abiertas de expresión e interacción escrita y oral, donde se cauteló que las temáticas fueran propias de un ámbito académico. De esta manera, en el prototipo de prueba no se presenta temas polémicos, discriminatorios o que privilegien determinados contenidos disciplinares, pues lo que se evalúa está en el contexto de las habilidades lingüísticas y no del contenido temático.

El tercer criterio ético está relacionado con los principios que se debe tener en cuenta cuando los estudios implican la participación de personas. En este modelo se consideró la participación de hablantes nativos de español y de estudiantes extranjeros que se encuentran

realizando un intercambio universitario, quienes firmaron un consentimiento informado, donde se estipulaba el tipo de participación solicitada. Considerar estos criterios permite construir una prueba éticamente adecuada. Por supuesto, existen otros aspectos relevantes como la adaptación de estas pruebas a candidatos que presenten algún tipo de discapacidad, sin embargo, esto se escapa a la viabilidad de este trabajo y solo es posible en instituciones que administran pruebas a gran escala y que cuentan con los recursos para hacer estas adaptaciones.

3.4.2.2. Criterios de diseño

De acuerdo con los criterios de diseño de la figura 9, el prototipo de prueba EFA es una prueba de proficiencia que evalúa lo que es capaz de hacer (habilidades) el estudiante en cuanto al uso del español en un contexto académico y está orientada a un nivel B2. El perfil de los estudiantes extranjeros de intercambio en la Universidad de Concepción, según los datos del Programa de Español como Lengua Extranjera de la Universidad de Concepción (ELE-UdeC), corresponde a alumnos de distintas áreas del conocimiento, siendo las más representativas: Letras y Educación (22%), Ingeniería (21%), Ciencias Políticas (16%), Economía (14%), Ciencias Físicas y Matemáticas (13%) y Ciencias Sociales (6%), como se observa en la tabla 14.



Área	Carreras
Letras y Educación (22%)	Traducción (14%)
	Pedagogía en LE (8%)
Ingeniería (21%)	Mecánica (10%)
	Civil (8%)
	Materiales (3%)
Ciencias Políticas (16%)	Ciencias Políticas (16%)
Economía (14%)	Ingeniería Comercial (8%)
	Contabilidad (3%)
	Administración de Empresas (3%)
Ciencias Físicas y Matemáticas (13%)	Matemática (8%)
	Astronomía (5%)
Ciencias Sociales (6%)	Psicología (3%)
	Antropología (3%)

Tabla 14: Áreas de estudios de los estudiantes del Programa ELE-UdeC

Un estudiante de nivel B2 en un ámbito académico es capaz de comprender las ideas principales de textos complejos que tratan temas concretos y abstractos; interactuar con hablantes nativos con un grado suficiente de fluidez y naturalidad; producir textos claros y detallados sobre diversos temas, y defender un punto de vista sobre temas específicos, indicando los pros y los contras de las distintas opciones (MCER, 2002). Los contextos de uso de la lengua corresponden a los ámbitos propios de un nivel B2 de EFA, esto es, participar de conversaciones sobre temas especializados, escuchar y comprender programas científicos, charlas, conferencias; leer y escribir sobre temas especializados (Vázquez, 2006; Pastor, 2014).

En relación con las habilidades lingüísticas evaluadas, se elaboró una tabla de especificaciones (ver anexo 1) donde se definió de manera operativa lo que cada una comprendía. A modo de ejemplo, se observa en la figura 10 que la expresión e interacción escrita se evalúa en función de la adecuación al género discursivo (en este caso, el ensayo), la coherencia, la corrección y el dominio de vocabulario que presenta el estudiante.

Expresión e interacción escrita			
Escribe redacciones claras y detalladas sobre una variedad de temas en las que desarrolla argumentos a favor o en contra de un punto de vista.			
<i>Adecuación al género discursivo</i>	<i>Coherencia</i>	<i>Corrección</i>	<i>Vocabulario</i>
El texto cumple con la estructura del ensayo. Se redacta en un lenguaje formal. Escribe un título, introduce el tema, establece claramente una postura respecto del tema. Menciona las ventajas y desventajas sobre el tema. Presenta una conclusión sobre el tema.	El texto es claro y coherente. Muestra un uso controlado, pero adecuado de los mecanismos de cohesión (conectores, estructura de párrafos y puntuación) para sintetizar y relacionar las ideas.	Demuestra un elevado control gramatical, aunque puede cometer algunos errores no sistemáticos que no producen malentendidos. Muestra un buen dominio ortográfico, si comete algunos errores estos no afectan la comprensión del escrito.	Tiene un amplio repertorio léxico que le permite comunicar adecuadamente sus ideas sobre temas especializados. Si comete algún error de selección léxica, este no afecta la comprensión.

Figura 10: Ejemplo de tabla de especificaciones del prototipo de prueba EFA en línea B2

Cada una de las habilidades se evalúa de manera separada, por lo que el prototipo de prueba se compone de cinco secciones, como se observa en la tabla 15.

Sección	N° textos	Tipo y n° de ítems	Tiempo
Comprensión lectora (25%)	3 textos	V o F (1 texto con 5 preguntas)	30 minutos
		Selección múltiple con 4 opciones (2 textos con 5 preguntas cada uno)	
Comprensión auditiva (20%)	3 audios	V o F (1 audio con 5 preguntas)	30 minutos
		Selección múltiple con 4 opciones (2 audios con 5 preguntas cada uno)	
Competencia lingüística (10%)	1 texto	Selección múltiple con 4 opciones	10 minutos
Expresión e interacción escrita (25%)	---	1 tarea abierta con dos opciones (el estudiante elige una)	20 minutos
Expresión e interacción oral (20%)	---	Entrevista: Consta de 3 partes: -Presentación -Elección del tema a exponer -Conversación sobre el tema	15 minutos
TOTAL			1 hora, 45 minutos

Tabla 15: Secciones del prototipo de prueba de EFA en línea B2

La tabla 15 muestra la distribución de las secciones del prototipo de prueba, sus porcentajes, el número y tipo de ítems y el tiempo asignado para cada una de las partes. Este prototipo de prueba en línea tiene una duración total de 1 hora, 45 minutos. Las secciones de comprensión lectora, auditiva, competencia lingüística y de expresión e interacción escrita se encuentran implementadas en un formato en línea y la expresión e interacción oral se

rinde de manera presencial, una vez que el estudiante ha terminado el prototipo de prueba en línea. Cabe señalar que esta habilidad también podría evaluarse a través de Skype si el estudiante se encontrara en su país de origen o no pudiera asistir a rendir la prueba de manera presencial.

Como se observa en la tabla 15, se contó con tres tipos de ítems distintos para evitar que el instrumento se orientara hacia un método o a un tipo de estudiante (Alderson *et al.*, 1998). Los ítems de corrección objetiva para las secciones de comprensión lectora y auditiva fueron de verdadero y falso y de selección múltiple con cuatro opciones. La sección de comprensión lectora se compone de tres textos y la sección auditiva de tres audios. En el primer texto de cada una de estas secciones se plantea cinco preguntas de verdadero y falso, y en los textos 2 y 3 se plantea cinco preguntas de selección múltiple por cada texto, por lo que en total cada sección cuenta con 15 ítems de corrección objetiva. La sección de competencia lingüística, se restringe a la evaluación de la competencia gramatical y comprende un texto con veinte ítems de selección múltiple donde se evalúa los diferentes usos gramaticales definidos para el nivel B2 en el contexto del EFA. Los ítems de respuesta abierta se consideraron para evaluar la expresión e interacción escrita y oral. En la sección de expresión e interacción escrita, el estudiante debe escribir un ensayo. Para ello, se propone dos temas distintos, el alumno elige uno y escribe el texto según el formato entregado. En la expresión e interacción oral, el estudiante tiene que cumplir con tres tareas de la entrevista: 1) presentarse, 2) exponer sobre un tema a elección y argumentar un punto de vista sobre dicho tema y 3) conversar con el entrevistador sobre el tema expuesto. Cada una de estas tareas abiertas tiene 16 puntos distribuidos en cuatro criterios en una escala analítica (ver anexo 2).

En relación con la distribución de los porcentajes, Alderson *et al.*,(1998) y Martínez (2011) señalan que no existen porcentajes definidos para otorgar a cada sección de una prueba, sino que esto dependerá de los objetivos que se persiga a través de la evaluación. Martínez (2011) indica que, en el ámbito de la evaluación de una lengua extranjera, no tiene sentido dar una nota general al estudiante sino que hay que mostrar en qué habilidades presenta un mejor y un peor desempeño. Considerando que las habilidades comprensivas de la lengua se evalúan de manera indirecta, en este prototipo de prueba se asignó un 25% a la comprensión lectora y un 20% a la comprensión auditiva. Esta determinación obedece a que los

estudiantes en su periodo de intercambio tienen que realizar lecturas de diversos géneros académicos, que en ocasiones, pueden resultar difíciles si no se tiene un dominio avanzado en la lengua, del mismo modo resulta también complejo comprender el discurso académico oral. En la evaluación de las habilidades comprensivas no se busca que el estudiante comprenda todo, sino conocer qué es lo que entiende de aquello que logra captar ya sea a nivel escrito u oral (Martínez, 2011).

El examen EXELEAA, que evalúa el español con fines académicos, no muestra una distribución de porcentajes por sección, pero considera un mayor número de ítems para evaluar la comprensión lectora.

Con la finalidad de evaluar el dominio de los estudiantes extranjeros en cuanto al uso de las estructuras gramaticales del español en un contexto académico, se propuso una sección de competencia lingüística, a la que se le asignó un 10%. Sin embargo, en las tareas de expresión escrita y oral se considera el criterio de corrección, que permite evaluar si el estudiante muestra un control gramatical que le permita darse cuenta y corregirse cuando comete errores (ver anexo 2). El examen EXELEAA también considera una sección de competencia lingüística, pues es relevante conocer el nivel de dominio que presentan los estudiantes en este ámbito, puesto que se enfrentarán a tareas orales y escritas que demandarán un uso preciso desde la perspectiva de las formas gramaticales.

A la tarea de expresión e interacción escrita se le asignó un 25%, atendiendo a lo que señala Alderson *et al.*, (1998) y Mendoza (2015), que una prueba de lengua en un ámbito académico debe dar relevancia a la habilidad escrita por la complejidad que representa para el estudiante, en un tiempo limitado y en una situación controlada de evaluación, organizar sus ideas, estructurarlas y luego producir el texto solicitado sin poder recurrir a otras fuentes de información. A la tarea de expresión e interacción oral se le asignó un porcentaje de 20%, teniendo en cuenta que es otra parte importante de las actividades académicas que deben realizar los estudiantes, por ejemplo, a través de exposiciones, debates, preguntas, etc. Por lo tanto, el 20% de la tarea oral no debe interpretarse como menos relevante en este prototipo de prueba.

Como se ha señalado a lo largo de este trabajo, la prueba que se presenta está a nivel de prototipo, por lo que a partir de los resultados obtenidos en el pilotaje se observará el comportamiento de las distintas secciones y en función de esto se determinará si en el diseño

futuro de una prueba de EFA se mantendrá el número de secciones propuestos y los porcentajes asignados en este prototipo.

3. 4.2.3. Criterios de elaboración

Teniendo definido el diseño del prototipo de prueba EFA en línea se procedió a la búsqueda y selección de los textos orales y escritos, acorde con el perfil de los estudiantes. En esta selección se consideró que los textos fueran éticamente adecuados, es decir, trataran de diversos temas acordes con un nivel B2 de EFA y con los datos del Programa ELE-UdeC, por lo que se recopiló textos de diversas fuentes: revistas de divulgación académica, programas de radio, televisión con entrevistas a expertos. Los textos corresponden a fragmentos de artículos científicos, charlas y ponencias, por lo tanto son representativas las modalidades expositiva y argumentativa. La longitud de los textos escritos seleccionados fue entre 265 y 550 palabras y la de los orales, entre 1 y 3 minutos. En estos textos se utiliza un lenguaje y estilo académico, sin expresiones idiomáticas que dificulten la comprensión. En los audios se cauteló que hubiera una articulación y velocidad adecuadas para el nivel B2. En algunos casos, se hizo adaptaciones de los textos orales y escritos manteniendo su organización discursiva (coherencia y cohesión). De acuerdo con los criterios de diseño, se recopiló un conjunto de textos, de los cuales finalmente se seleccionó diez textos orales y diez escritos que se consideraron más adecuados en términos de pertinencia, contenido y formato para el nivel B2 de EFA. Luego, estos diez textos fueron evaluados por un grupo de diez expertos (5 profesores de segundas lenguas con grado de Magíster en Lingüística, 2 estudiantes de Magíster en Lingüística Aplicada y 3 Doctores en Lingüística) a través de una escala Likert (ver anexo 3) con el objeto de validar en qué medida estos materiales eran aptos para evaluar el nivel de proficiencia B2 de EFA. La evaluación de los textos consideró tres criterios: pertinencia, contenido y formato, como se observa en la figura 11.

CRITERIOS	INDICADORES
PERTINENCIA: El tipo de texto y el tema son adecuados para un nivel B2 de español con fines académicos.	1.Muy en desacuerdo 2.En desacuerdo 3.De acuerdo 4.Muy de acuerdo
CONTENIDO: El contenido gramatical y léxico es adecuado para un nivel B2 de español con fines académicos.	1.Muy en desacuerdo 2.En desacuerdo 3.De acuerdo 4.Muy de acuerdo
FORMATO DEL TEXTO: La redacción, el tamaño de letra, la longitud del texto/ la locución, el ruido de los audios son apropiados para el nivel que se orienta la prueba.	1.Muy en desacuerdo 2.En desacuerdo 3.De acuerdo 4.Muy de acuerdo

Figura 11: Criterios para evaluar los textos

Acorde con el juicio de expertos, se seleccionó los tres textos que presentaban un 80% (Escobar y Cuervo, 2008) de la categoría 4, muy de acuerdo, de la escala Likert en los tres criterios definidos: pertinencia, contenido y formato. A partir de estos resultados se seleccionó tres textos para la comprensión lectora y tres para la comprensión auditiva.

CRITERIOS	PERTINENCIA	CONTENIDO	FORMATO	
N° textos y temas	Muy de acuerdo	Muy de acuerdo	Muy de acuerdo	Promedio
1 La adolescencia	30%	20%	50%	33%
2 Aprendizaje de segundas lenguas	80%	80%	80%	80%
3 Lengua mapuche	80%	80%	80%	80%
4 El cobre	50%	30%	50%	43%
5 Planta de energía solar	80%	80%	50%	70%
6 Adquisición y aprendizaje	60%	50%	80%	63%
7 Educación y pueblos indígenas	60%	40%	30%	43%
8 Biodiversidad	90%	90%	80%	87%
9 Competencias sustentables ingeniería	80%	80%	80%	80%
10 El agua	60%	80%	60%	67%

Tabla 16: Evaluación de los textos de comprensión lectora

Para evaluar la habilidad de comprensión lectora, se seleccionó el texto 2, el texto 8 y el 9, referidos respectivamente, al aprendizaje de un segundo idioma, la biodiversidad y las competencias sustentables en ingeniería. Estos textos corresponden a un *abstract*, un fragmento de artículo científico y una reseña, que en la secuencia de aparición del prototipo de prueba se organizaron en el siguiente orden: 1) El sistema de competencias sustentables

para el desempeño profesional en ingeniería, 2) El aprendizaje de una segunda lengua y 3) La biodiversidad.

De acuerdo con la tabla 16, se observa que tanto el texto 2 como el 3, ambos del área de letras, obtuvieron un 80% de muy de acuerdo en la evaluación de los expertos; no obstante, se decidió seleccionar el texto 2 porque trata de un tema más cercano a cada estudiante de intercambio, el aprendizaje de un segundo idioma, puesto que el texto 3 trataba sobre la situación de la lengua mapuche en una zona de Chile.

Del mismo modo, en la comprensión auditiva se seleccionó los textos orales que presentaban un mínimo de 80% de muy de acuerdo en los tres criterios definidos para la evaluación, como se aprecia en la tabla 17.

CRITERIOS	PERTINENCIA	CONTENIDO	FORMATO	
N° audios y temas	Muy de acuerdo	Muy de acuerdo	Muy de acuerdo	Promedio
1 Energía geotérmica	70%	100%	70%	80%
2 Planta termosolar	90%	70%	60%	73%
3 Lengua materna y segunda lengua	90%	70%	50%	70%
4 Energía eólica	80%	100%	90%	90%
5 Proyecto Agua Copiapó	100%	80%	90%	90%
6 El aporte del folclor a Chile	90%	70%	70%	77%
7 Desafíos de las lenguas indígenas	70%	70%	20%	53%
8 Biocombustibles	80%	90%	80%	83%
9 Métodos para estudiar mejor	70%	80%	50%	67%
10 Transducción de señales en células	80%	80%	80%	80%

Tabla 17: Evaluación de textos orales de comprensión auditiva

Según la evaluación del juicio de expertos, los audios mejor evaluados en los tres criterios fueron el audio 4, 5 y 8, referidos respectivamente a la energía eólica, un proyecto para evitar la escasez del agua en el norte de Chile y los biocombustibles. Estos textos corresponden a dos exposiciones de trabajos científicos-tecnológicos y a un fragmento de un

programa televisivo sobre ciencia, que en la secuencia de aparición del prototipo de prueba se organizaron de la siguiente manera: 1) La energía eólica, 2) Un proyecto para evitar la escasez de agua en la zona norte de Chile y 3) Los biocombustibles.

La tabla 18 sintetiza las características de los textos y audios seleccionados para el prototipo de prueba. Se observa que los textos y audios fueron aumentando gradualmente en longitud y duración. Para evaluar la competencia lingüística en un nivel B2 se utilizó un ensayo sobre el cobre y su impacto en la economía chilena, de 227 palabras.

Sección	Géneros académicos	Tipo de ítem	Longitud/duración
Comprensión lectora	Texto 1: <i>Abstract</i> (Sistema de competencias sustentables para el desempeño profesional en ingeniería)	5 ítems verdadero/falso	265 palabras
	Texto 2: Fragmento de artículo científico (El aprendizaje de una segunda lengua)	5 ítems selección múltiple	338 palabras
	Texto 3: Reseña (La biodiversidad)	5 ítems selección múltiple	517 palabras
Comprensión auditiva	Audio 1: Exposición (Energía eólica)	5 ítems verdadero/falso	1:22 minutos
	Audio 2: Exposición (Proyecto Agua Copiapó)	5 ítems selección múltiple	1:36 minutos
	Audio 3: Exposición (Los biocombustibles)	5 ítems selección múltiple	1:57 minutos
Competencia lingüística	Ensayo (El cobre y el futuro de Chile)	20 ítems de selección múltiple	227 palabras

Tabla 18: Textos y audios que componen el prototipo de prueba

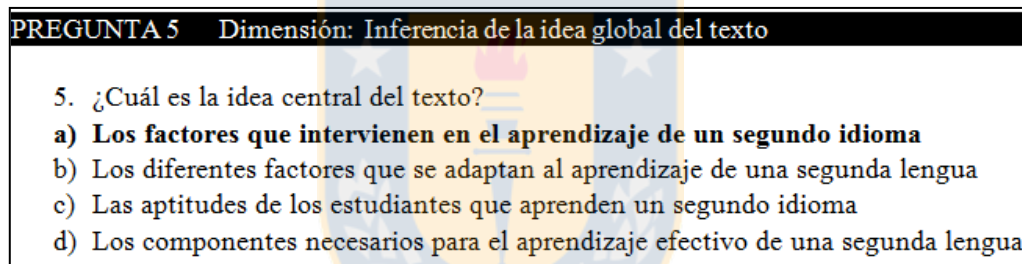
Una vez que se contó con la selección de textos definitiva para el prototipo de prueba se procedió a la elaboración y redacción de los ítems e instrucciones acorde con los criterios de diseño del modelo.

Los ítems de las secciones de comprensión lectora y auditiva se elaboraron teniendo en cuenta tres dimensiones o microhabilidades: 1) la identificación de información específica en textos especializados, 2) la inferencia de las ideas principales de los párrafos y 3) la

inferencia de la idea global del texto (MCER, 2002). Los ítems de la sección de competencia lingüística miden la dimensión del dominio gramatical correspondiente a un nivel B2 de EFA. Las tareas de expresión escrita y oral evalúan dos dimensiones: 1) las modalidades expositiva y argumentativa y 2) el género académico ensayo y exposición oral. Se decidió abordar la modalidad expositiva y argumentativa en las tareas de escritura y oralidad porque son las más utilizadas en el ámbito académico (Vázquez, 2004; Mendoza, 2015), y en este contexto se prestaba bien considerar el género ensayo y la exposición oral.

En consecuencia, el prototipo de prueba en línea evalúa seis dimensiones en el dominio del EFA para el nivel B2.

En el proceso de elaboración de los ítems de comprensión lectora y auditiva se registró, junto a la pregunta planteada, la dimensión que cada una evaluaba, como se observa en la figura 12. Esto, sin duda, facilitó posteriormente la validación del instrumento por parte del grupo de expertos.



PREGUNTA 5 Dimensión: Inferencia de la idea global del texto

5. ¿Cuál es la idea central del texto? —

- a) **Los factores que intervienen en el aprendizaje de un segundo idioma**
- b) Los diferentes factores que se adaptan al aprendizaje de una segunda lengua
- c) Las aptitudes de los estudiantes que aprenden un segundo idioma
- d) Los componentes necesarios para el aprendizaje efectivo de una segunda lengua

Figura 12: Ejemplo de elaboración de preguntas

En relación con la elaboración de tareas abiertas se consideró dos temas para la escritura y dos para la expresión oral. En la expresión e interacción escrita se consideró la formación académica a distancia y El día Mundial sin compras. De estos, el estudiante selecciona uno y realiza la tarea en el tiempo asignado, como se observa en la figura 13.

Sección IV: Expresión e interacción escrita

La sección de expresión escrita consta de dos opciones. Solo tienes que elegir UNA opción y escribir un ensayo de 250 palabras (20-25 líneas)

Opción 1

Una alternativa a la formación académica tradicional es tomar cursos a distancia (por Internet), modalidad que permite un acceso real al conocimiento de personas que antes no podían hacerlo, ya sea por razones económicas o de tiempo. La modalidad de aprendizaje a distancia, cada vez va tomando más fuerza, pero aún queda el escepticismo de algunos que dudan si es efectiva o no.

En relación con este tema, escribe un ensayo sobre la formación académica a través de cursos a distancia por Internet.

En el ensayo tienes que incluir:

- Escribir un título
- Introducir el tema
- Establecer una postura respecto del tema
- Mencionar las ventajas y desventajas de la formación a través de internet
- Concluir

Figura 13: Tarea de escritura

La evaluación de la expresión e interacción oral se llevó a cabo mediante la técnica de entrevista, compuesta de tres partes: 1) presentación del estudiante, 2) exposición y argumentación sobre un tema a elección (opción 1: el uso de pesticidas en la agricultura, opción 2: La inmersión o el aprendizaje formal de un segundo idioma) y 3) conversación con el entrevistador sobre el tema expuesto. Esta sección del prototipo de prueba tiene una duración máxima de 15 minutos.

El prototipo de prueba EFA de nivel B2 se implementó en la plataforma Moodle (del inglés *Object-Oriented Dynamic Learning Environment*) con cuatro secciones (ver anexo 4). La sección cinco de expresión e interacción oral se realizó de manera presencial. En el anexo 5 se puede visualizar una muestra del prototipo de prueba, respondido por uno de los estudiantes extranjeros (ver anexo 5).

Las opciones avanzadas de Moodle permiten el registro de los estudiantes y la programación de la prueba en cuanto a la fecha, la hora y la duración. El formato de aplicación de este prototipo de prueba representa un aporte a la evaluación del EFA, dado que tanto las pruebas generales de ELE como las pruebas de EFA, generalmente, se aplican en formato de lápiz y papel. En este sentido, el formato de este prototipo de prueba resulta beneficioso, puesto que permite obtener de manera rápida y eficaz los resultados sobre el nivel de proficiencia de los estudiantes. Además, Moodle otorga seguridad en diversos aspectos: restringe el número de intentos del estudiante, por lo que una vez que ha comenzado la prueba no puede detenerla; muestra el tiempo del que dispone el alumno;

facilita la navegación, ya que para avanzar a la sección siguiente solo se presiona el botón “siguiente”; en la misma pantalla el alumno dispone del texto y de las preguntas (a diferencia del papel donde tiene el texto y la hoja de respuestas). Del mismo modo, en la sección de competencia lingüística el alumno va leyendo el texto y va seleccionando la forma gramatical adecuada sin tener que salirse del texto para marcar la alternativa correcta. En cuanto a la escritura, la plataforma da la opción de desactivar el procesador de texto, lo cual permite acceder a muestras de lengua auténticas. Sin duda, este formato promueve una interacción entre el estudiante y la prueba (Figueras y Puig, 2013) y permite dos formas de aplicación: presencial y a distancia.

3. 4.2.4. Criterios de validación y pilotaje

En este modelo el criterio de validación y pilotaje se considera en tres niveles: no experta, experta y mediante el pilotaje con una muestra de estudiantes. Como se observa en la figura 14, el criterio de validación y pilotaje del prototipo de prueba EFA en línea considera la validez de apariencia con la participación de un grupo de hablantes nativos; la validez de contenido a través del juicio de expertos, y el pilotaje con una muestra de estudiantes extranjeros de intercambio universitario. Los resultados del pilotaje darán a conocer el índice de dificultad y de discriminación de los ítems y el grado de confiabilidad del instrumento. Los datos e información obtenida en las distintas etapas de validación y pilotaje permitirán retroalimentar el prototipo de prueba EFA en línea, considerando que la validación es un proceso continuo, que no termina con la aplicación de la prueba al grupo objeto, sino que siempre se necesita estar probando el funcionamiento de los ítems (Bordón y Liskin-Gasparro, 2007).

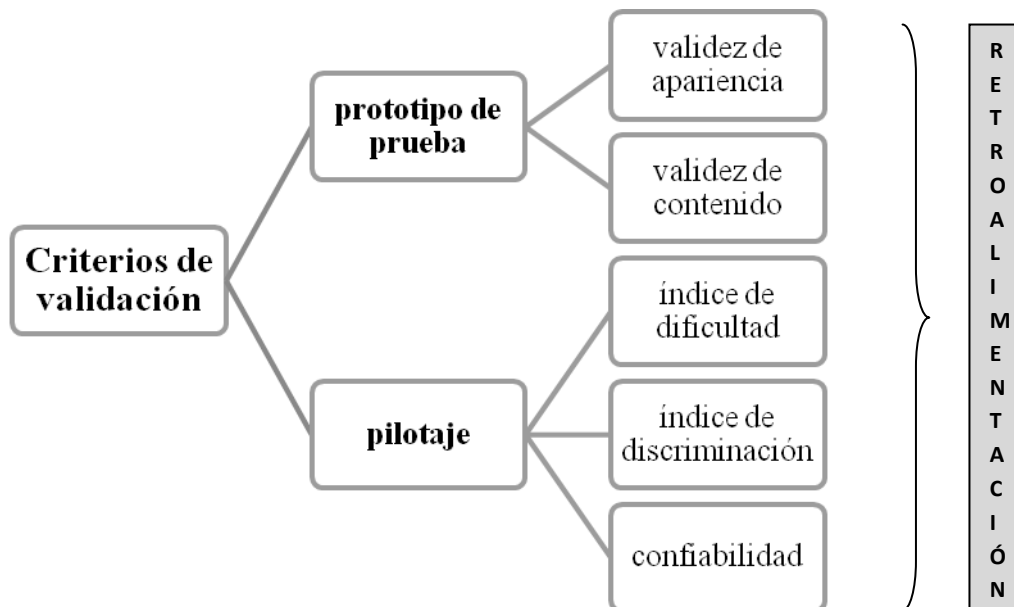


Figura 14: Criterios de validación y pilotaje

- **Validez de apariencia**

La validez de apariencia se realizó a través del juicio de un grupo no experto, en el que participaron 9 estudiantes de primer año (en su tercer mes de universidad) de la carrera de Traducción e Interpretación de la Universidad de Concepción, 5 mujeres y 4 hombres, con un promedio de edades de 18 años. Ellos rindieron el prototipo de prueba EFA en línea de nivel B2, para ello, se les creó una cuenta a cada uno en la plataforma Moodle y se definió la fecha y hora de aplicación de la prueba en el Laboratorio de Aplicaciones Tecnológicas en Lingüística Aplicada (LATLA) de la Universidad de Concepción. Al finalizar el prototipo de prueba, se aplicó a cada estudiante un cuestionario de apreciación (ver anexo 6) sobre la prueba en su conjunto (claridad en las instrucciones, comprensibilidad de los textos, duración de la prueba, implementación de la prueba en línea, etc.).

- **Validez de contenido**

La validez de contenido del prototipo de la prueba EFA se llevó a cabo mediante la técnica de juicio de expertos. El objetivo de este tipo de validación es evaluar la pertinencia y claridad de los ítems en relación con la dimensión y el nivel de lengua que evalúan, y evaluar el funcionamiento del prototipo de prueba en su modalidad computacional.

En el proceso de validación de los ítems del prototipo de prueba participaron ocho expertos, profesionales de lenguas (2 estudiantes de Magíster en Lingüística Aplicada, 1 Magíster en Lingüística, 3 candidatos al grado de Doctor en Lingüística y 2 Doctores en Lingüística). De manera individual, cada experto, evaluó a través de una escala Likert de 5 puntos (1: Muy en desacuerdo, 2: En desacuerdo, 3: Indiferente, 4: De acuerdo, 5: Muy de acuerdo) en qué medida los ítems de las diferentes secciones del prototipo de prueba en línea medían las dimensiones definidas en los criterios de elaboración. Para ello, se les entregó una pauta (ver anexo 7) donde se les explicaba en qué consistía la evaluación y luego se proporcionaba la lista de ítems en las distintas dimensiones. Los evaluadores leían la pauta y luego accedían al prototipo de prueba en línea, con una clave que se les entregó. Enseguida, leían los textos y las preguntas y evaluaban si los ítems medían las distintas dimensiones, y así sucesivamente con cada sección del prototipo de prueba.

En la figura 15, se observa parte del instrumento que respondieron los evaluadores, específicamente, las preguntas 3 y 4 correspondientes al primer texto de comprensión lectora (Sistemas de competencias sustentables para el desempeño profesional en Ingeniería) que evalúan la dimensión II del prototipo de prueba.

DIMENSIÓN II: INFIERE LAS IDEAS PRINCIPALES DE LOS PÁRRAFOS

Secciones: Comprensión de lectura	EVALÚA DIMENSIÓN	INDICADORES				
Preguntas	SI/NO (observación)	1	2	3	4	5
<i>Texto I</i>						
3. Las competencias de egreso se reflejan en el dominio disciplinar que debe mostrar todo profesional de ingeniería.						
4. Se puede inferir que las escuelas de Ingeniería han intentado implementar cambios en la formación académica de sus estudiantes.						

Figura 15: Pauta de evaluación de contenido para los jueces

En la evaluación del prototipo de prueba EFA en línea participaron 10 expertos (3 estudiantes de Magíster en Lingüística Aplicada, 2 Magíster en Lingüística Aplicada, 3 candidatos al grado de Doctor en Lingüística y 2 Doctores en Lingüística) que tenían

experiencia en la enseñanza y evaluación de lenguas en un contexto mediatizado por computador. Acorde con los principios de la metodología de CALT el prototipo de prueba fue evaluado en función de dos aspectos: 1) la dimensión de evaluación de segundas lenguas y 2) la dimensión técnica.

Una vez que se completó la evaluación de los expertos, se tabularon los datos en Excel atendiendo a las calificaciones de los evaluadores. Los resultados obtenidos muestran que el prototipo de prueba presenta validez de contenido, en tanto que los ítems miden las dimensiones definidas y el instrumento cumple con los principios metodológicos de lo que significa plantear una prueba de EFA en línea.

- **Pilotaje del prototipo de prueba EFA en línea de nivel B2**

En este proceso de pilotaje del prototipo de prueba EFA en línea de nivel B2 participó un grupo de 28 estudiantes de intercambio en la Universidad de Concepción, 14 hombres y 14 mujeres, cuyas edades fluctúan entre los 19 y 27 años. En las tablas 19 y 20 se describe esta muestra en función de su lengua materna y área de estudios.

Lengua materna	N° de estudiantes
Francés	9
Alemán	8
Inglés	4
Sueco	2
Portugués	2
Italiano	1
Checo	1
Ruso	1
TOTAL	28 estudiantes

Tabla 19: Descripción de la muestra

De acuerdo con estos datos, las lenguas maternas más representadas en la muestra son el francés con 9 estudiantes, el alemán con 8 estudiantes y el inglés con 4.

En la tabla 20 se muestra las áreas de estudios del grupo de estudiantes. Se observa que 12 son estudiantes del área de Ingeniería, 7 de Letras, 3 de Ciencias Políticas, 2 de Administración y Economía, 2 de Ciencias Químicas, 1 de Ciencias Biológicas y 1 de Ciencias de la salud.

Área de estudios	N° de estudiantes
Ingeniería	12
Letras	7
Ciencias Políticas	3
Administración y Economía	2
Ciencias Químicas	2
Ciencias Biológicas	1
Ciencias de la salud	1

Tabla 20: Área de estudios de la muestra piloto

En la aplicación del prototipo de prueba de EFA en línea, los estudiantes rindieron la prueba por un tiempo de 1 hora y 30 minutos y luego se les administró la tarea de entrevista para evaluar la expresión e interacción oral. Hay que señalar que la prueba no se aplicó de una sola vez a los 28 estudiantes, sino que se citó por grupos acorde con la disponibilidad horaria de los alumnos. Esto, además, facilitó la aplicación de la tarea oral.

Una vez que los estudiantes finalizaron este prototipo de prueba, se les aplicó un cuestionario tipo Likert y dos preguntas abiertas, con el objeto de conocer sus percepciones respecto del instrumento, tanto en su estructura general, como también de las secciones y textos.

Los resultados obtenidos en el pilotaje del prototipo de la prueba EFA en línea han permitido conocer el comportamiento de los diferentes ítems de la prueba en relación con el índice de dificultad y de discriminación. Con esta información, en un futuro, se debiera revisar el instrumento, ampliar la muestra y decidir qué ítems aprobar, reformular o rechazar. Del mismo modo, son interesantes las respuestas de los estudiantes en función de sus percepciones sobre las instrucciones, los textos, la duración de la prueba, etc.

Hay que tener en cuenta que puede haber ambigüedades en el planteamiento de las preguntas, en la redacción de las opciones de respuestas, en las tareas, etc. y esto solo se hace evidente una vez que se aplica a una muestra real de estudiantes. Este criterio permite enmendar errores que en el futuro podrían ser graves, si nos atenemos al impacto que cada prueba genera en el estudiante y en su entorno.

CAPÍTULO III: RESULTADOS Y DISCUSIÓN

4. Resultados

En este apartado se presenta los resultados obtenidos durante el proceso de evaluación del prototipo de la prueba EFA en línea de nivel B2. De acuerdo con el procedimiento descrito en la metodología, primero, se muestra los resultados obtenidos en la validez de apariencia, luego los de la validez de contenido en cuanto a los ítems y a la implementación en línea según los principios de CALT, y finalmente los resultados que se obtuvo mediante la aplicación piloto del prototipo de prueba al grupo de 28 estudiantes extranjeros de intercambio en la Universidad de Concepción.

4.1. Resultados de la validez de apariencia

A partir de los resultados obtenidos por los 9 estudiantes de Traducción que participaron de la validez de apariencia del prototipo de prueba EFA en línea de nivel B2, se construyó una matriz con el número y porcentaje de aciertos obtenidos en las respuestas en las distintas secciones de corrección objetiva del prototipo de prueba: comprensión lectora, comprensión auditiva y competencia lingüística, como se observa en las siguientes tablas.

- **Comprensión lectora**

Sección I: Comprensión lectora															
Pregunta	Texto 1 (V/F)					Texto 2 (S/M)					Texto 3 (S/M)				
	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
N° sujeto/clave	b	a	a	b	b	d	a	b	c	a	b	c	a	c	b
S1	b	a	b	b	b	d	a	d	d	a	b	a	a	d	a
S2	b	a	b	a	b	d	a	c	a	a	b	d	a	d	c
S3	b	a	b	b	b	b	a	d	a	b	a	d	a	b	c
S4	b	a	b	b	b	d	a	b	c	d	b	c	a	c	c
S5	b	a	b	b	b	d	a	c	c	a	b	d	a	c	c
S6	b	a	b	b	b	d	a	a	a	a	b	c	a	d	c
S7	b	a	a	b	b	d	a	b	c	a	b	c	a	b	a
S8	b	a	b	b	a	b	a	c	a	b	b	c	b	d	c
S9	b	a	b	b	b	d	a	b	a	c	a	c	a	c	b
N° de aciertos	9	9	1	8	8	7	9	3	3	5	7	5	8	3	1
% de aciertos	100%	100%	11%	89%	89%	78%	100%	33%	33%	56%	78%	56%	89%	33%	11%

Tabla 21: Número y porcentaje de aciertos en comprensión lectora

De acuerdo con la tabla 21, en el texto 1, de comprensión lectora, la mayoría de los estudiantes acierta con la respuesta correcta, excepto en la pregunta 3, donde solo hubo un 11% de acierto. En el texto 2, hay tres preguntas que presentan bajos porcentajes de aciertos,

se trata de las preguntas 3, 4 y 5 con un 33% de acierto las dos primeras y la última con un 56%. En el texto 3, las preguntas con bajo porcentaje de acierto son las preguntas 2, 4 y 5 con un 56%, 33% y 11% respectivamente.

A partir de las respuestas dadas por este grupo no experto se revisó las preguntas con un bajo porcentaje de aciertos. En este contexto y dado el reducido número de sujetos que participó en la validez de apariencia, se consideró pertinente examinar los ítems que tuvieran un porcentaje de aciertos igual o inferior al 56% , es decir, 5 aciertos; teniendo en cuenta que los estudiantes son hablantes nativos de español y el prototipo de prueba se orienta a hablantes de español como lengua extranjera.

Al revisar los ítems seleccionados se comprobó la ambigüedad en las opciones de respuesta, lo cual confundió a los estudiantes, especialmente, aquellas preguntas donde hubo entre 1 y 3 aciertos, como es el caso de la pregunta 3 del texto 1:

PREGUNTA 3	Estrategia: Inferencia de la idea principal del párrafo
3. Las competencias de egreso se reflejan en el dominio disciplinar que debe mostrar todo profesional de ingeniería.	
a) Falso	
b) Verdadero	

Figura 16: Pregunta 3 del texto 1 de comprensión lectora

Esta pregunta confundió a los estudiantes, porque no era tan evidente la falsedad de ella, puesto que las competencias de egreso incluyen el dominio disciplinar, además de otros atributos personales, por lo tanto, la forma en que estaba planteada la pregunta la hacía verdadera. Este ítem se modificó de la siguiente manera: *De acuerdo con el texto, las competencias de egreso corresponden al plan de estudios de una institución educativa.* De esta forma quedaba claro que la respuesta era falsa.

En el texto 2, las preguntas 3 y 4 fueron ambiguas para los estudiantes por la forma en que estaban planteadas, los distractores seleccionados fueron la alternativa c y d. La pregunta 3, que solo tuvo un 33% de acierto, estaba planteada de la siguiente forma:

PREGUNTA 3 Dimensión: Inferencia de la idea principal del párrafo

3. Según el texto, el estudiante de un segundo idioma aprende mejor la habilidad auditiva...
- a) en las clases de la lengua que está aprendiendo
 - b) en un entorno donde se habla la segunda lengua**
 - c) dependiendo de su forma de aprendizaje de otras lenguas
 - d) según la capacidad para procesar otras lenguas

Figura 17: Pregunta 3 del texto 2 de comprensión lectora

De acuerdo con el texto, la opción c solo era plausible en una lectura rápida del texto, puesto que en el texto se dice “*Normalmente se acepta que la habilidad auditiva del alumno está claramente relacionada con la manera en la que ha aprendido esta segunda lengua. La mejor forma de aprender es de forma natural, interactuando,...*”. Considerando las respuestas de los estudiantes se decidió modificar, tanto el enunciado de la pregunta como las opciones b y c, quedando como se muestra en la figura 18.

PREGUNTA 3 Dimensión: Inferencia de la idea principal del párrafo

3. De acuerdo con el texto, la mejor forma de aprender la habilidad auditiva es...
- a) en las clases de la lengua que está aprendiendo
 - b) en un lugar donde se habla la segunda lengua**
 - c) como se ha aprendido otras lenguas
 - d) según la capacidad para procesar otras lenguas

Figura 18: Pregunta 3 corregida

La pregunta 4 del texto 2 presenta una situación similar, puesto que 5 estudiantes marcaron como opción correcta la a. En esta pregunta se modificó la alternativa a, por: a) *permite el control del sistema muscular humano*, de esta manera se eliminó la ambigüedad del ítem.

PREGUNTA 4 Dimensión: Inferencia de la idea principal del párrafo

4. En el aprendizaje de una segunda lengua, la habilidad motora...
- a) permite la producción del habla
 - b) presenta dificultades en el habla de algunas lenguas
 - c) debe practicarse de manera constante**
 - d) da resultados garantizados a una edad temprana

Figura 19: Pregunta 4 corregida

La pregunta 5 del texto 3 con la opción b como correcta fue acertada solo por 1 estudiante, 6 alumnos marcaron la c como alternativa correcta, por lo que se revisó las alternativas en función del texto y se modificó la opción b) *Los retos futuros para conservar la biodiversidad* por b) *Las investigaciones científicas sobre la biodiversidad* y se dejó la c como correcta.

PREGUNTA 5 Dimensión: Inferencia de la idea global del texto

5. ¿Cuál es la idea central del texto?

a) La situación de la biodiversidad a nivel mundial

b) Las investigaciones científicas sobre la biodiversidad

c) La relevancia de la biodiversidad para el planeta

d) Las consecuencias de la disminución de la biodiversidad

Figura 20: pregunta 5 corregida

- **Comprensión auditiva**

En relación con la comprensión auditiva, se consideró la revisión de las preguntas 2 y 4 del audio 1, las preguntas 1, 2, 3 y 4 del audio 2 y las preguntas 1, 2, 4 y 5 del audio 3, como se muestra en la tabla 22.

Sección II: Comprensión auditiva															
Pregunta	Audio 1 (V/F)					Audio 2 (S/M)					Audio 3 (S/M)				
	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
N° sujeto/clave	b	a	b	a	b	b	d	a	b	c	b	d	c	a	c
S1	b	b	b	a	b	a	c	d	a	c	a	c	b	b	c
S2	b	b	b	a	b	a	d	a	b	c	a	c	c	d	c
S3	b	b	b	a	b	c	b	a	c	c	b	c	c	d	b
S4	b	b	b	b	b	a	a	c	b	c	b	c	c	a	c
S5	b	a	b	b	b	a	a	a	a	c	c	c	c	a	c
S6	b	a	b	b	b	c	c	a	d	c	b	c	c	d	a
S7	b	a	b	b	b	c	a	c	d	c	c	c	c	c	d
S8	b	b	b	a	b	b	a	d	d	c	c	b	b	b	b
S9	b	b	b	b	b	a	a	a	b	c	a	c	b	a	a
N° de aciertos	9	3	9	4	9	1	1	5	3	9	3	0	6	3	4
% de aciertos	100%	33%	100%	44%	100%	11%	11%	56%	33%	100%	33%	0%	67%	33%	44%

Tabla 22: Número y porcentaje de aciertos en comprensión auditiva

En las preguntas 2 y 4 del audio 1, el problema estaba en una parte del encabezado del ítem que causaba confusión, estas se modificaron como se muestra en la figura 21, en la pregunta 2 la parte marcada se reemplazó por *aerodinámica* y en la pregunta 4 por *carece de recursos económicos*.

PREGUNTA 2 Dimensión: Identificación de información específica
2. Según el audio, el objetivo de la instalación de la turbina es transformar la energía eléctrica en una energía limpia .
a) Falso
b) Verdadero
PREGUNTA 4 Dimensión: Inferencia de ideas principales del audio
4. De acuerdo con el audio, la gente que vive en zonas rurales presenta bajos niveles de calidad de vida .
a) Falso
b) Verdadero

Figura 21: Preguntas 2 y 4 corregidas

En las preguntas 1 y 2 del audio 2 la ambigüedad estaba en las opciones, principalmente, en la opción *a*, por lo que tuvieron que modificarse, como se muestra en la figura 22.

PREGUNTA 1 Dimensión: Identificación de información específica
1. De acuerdo con el audio ¿Cuál es la situación del río Copiapó?
a) La cuenca del río recibe poco agua (Está rodeada de un valle seco)
b) Tiene poca disponibilidad de agua
c) Presenta una sequía permanente
d) Permanece sin agua
PREGUNTA 2 Dimensión: Identificación de información específica
2. El objetivo del proyecto Agua Copiapó es...
a) tener un control sobre el uso del agua en el Valle de Copiapó (consumo, la zona)
b) introducir tecnología para mejorar la calidad del agua
c) mejorar la distribución del agua en la zona
d) optimizar el uso del agua para el riego

Figura 22: Preguntas 1 y 2 del audio 2 modificadas

En el audio 3 se modificó la pregunta 1 solo cambiando la palabra elevado por competitivo (sugerencia de la evaluación de expertos) a) *de precio elevado*, puesto que las demás opciones estaban bien planteadas en función del audio, según el juicio de expertos. La pregunta 2 de este audio resultó problemática, dado que hubo 0 aciertos, 8 seleccionaron la alternativa c como correcta, por lo que se decidió modificarla para así evitar la ambigüedad del ítem. De hecho, esta alternativa era plausible, puesto que en el audio se dice que el objetivo de este centro es *desarrollar tecnologías y formar capacidades humanas que permitan la implementación en Chile de una industria de biocombustible...* La alternativa c quedó c) *implementar un centro de investigación sobre biocombustibles*.

PREGUNTA 2 Dimensión: Identificación de información específica
<p>2. El Centro de Biotecnología de la Universidad de Concepción está interesado en...</p> <p>a) invertir en la formación de capital humano avanzado</p> <p>b) desarrollar tecnología para la obtención de petróleo</p> <p>c) implementar una industria de biocombustible</p> <p>d) encontrar fuentes de combustible renovables</p>

Figura 23: Pregunta 2 del audio 3 corregida

La pregunta 5 del audio 3 se modificó levemente en cuanto a una de las opciones de respuesta, que presentaba ambigüedad. Se trata de la opción a) *El aporte de un centro de investigación a los problemas energéticos del país*, que se cambió por a) *El aporte de la investigación a los problemas energéticos del país*.

- **Competencia lingüística**

En cuanto a la tercera sección de competencia lingüística se observó que las preguntas 3, 9 y 17 tenían un bajo porcentaje de aciertos (11% cada una), como se muestra en la tabla 23.

Sección III: Competencia lingüística																				
Pregunta	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N° sujeto/clave	b	a	d	c	d	a	c	a	b	d	c	a	c	b	c	a	b	c	a	d
S1	b	a	c	c	d	d	d	a	a	d	c	a	c	b	c	a	d	c	a	d
S2	b	a	c	c	d	d	c	b	c	d	c	a	c	b	a	a	d	c	d	a
S3	b	a	c	c	d	d	d	a	d	d	c	a	c	b	c	d	d	c	a	b
S4	b	a	c	c	d	a	c	a	d	d	c	a	c	b	c	d	d	c	a	d
S5	b	a	c	c	d	a	c	a	c	d	c	a	c	b	c	a	d	c	d	d
S6	b	a	d	c	d	a	c	a	d	d	c	a	c	b	c	a	d	c	c	a
S7	d	a	c	c	d	a	c	a	d	d	b	a	c	a	a	a	b	c	a	d
S8	b	a	c	d	d	a	c	a	d	d	c	a	c	b	c	a	c	c	c	b
S9	b	a	c	c	a	a	c	a	b	d	c	a	c	b	b	a	d	c	a	d
N° de aciertos	8	9	1	8	8	6	7	8	1	9	8	9	9	8	6	7	1	9	5	5
% de aciertos	89%	100%	11%	89%	89%	67%	78%	89%	11%	100%	89%	100%	100%	89%	67%	78%	11%	100%	56%	56%

Tabla 23: Número y porcentaje de aciertos en la competencia lingüística

La pregunta 3 presentó solo un acierto porque la alternativa *d) pues* referida a un conector causal se confundía con la opción *c) dado que* de un conector de consecuencia; por lo que se modificó la alternativa, quedando la opción correcta como conector de consecuencia *b) por lo tanto*. Esta confusión se puede deber a la cercanía y ambigüedad para separar en el uso los conectores causales de los de consecuencia, por esta razón se optó por dejar la idea de consecuencia que sería más comprensibles para estudiantes de ELE de nivel B2 “*El cobre siempre ha sido importante en Chile. Pero después de 1990 ha pasado a ser doblemente más importante. Ha aumentado notablemente su relevancia en la economía, por lo tanto gracias a él Chile cambió.*”

La pregunta 9 era una pregunta sobre el uso de adverbios, la correcta era la *b) Aún más*, pero hubo distintas selecciones, por lo que se optó por eliminar ese ítem y elaborar uno nuevo sobre el uso del pronombre personal *se*.

El ítem 17 solo tuvo 1 acierto, por lo que se cambió la alternativa correcta, que era: *es decir* por el conector de consecuencia *d) por lo tanto* como opción correcta, dado que se presenta una idea de consecuencia en el texto.

En la pregunta 19 sobre el uso de preposiciones se cambió las opciones *c) hacia* y *d) por*, por las preposiciones *c) de* y *d) entre*, que habían confundido a 4 estudiantes de Traducción. No obstante, se mantuvo como correcta la opción *a) en*.

- **Tareas de expresión e interacción oral y escrita**

Este grupo de 9 estudiantes de Traducción también rindió la parte de expresión e interacción escrita y oral para observar el funcionamiento de las instrucciones. En la escritura, las instrucciones de la tarea funcionaron como estaba previsto, los estudiantes seleccionaron el tema y escribieron el texto de la longitud solicitada. Sin embargo, en la sección de expresión e interacción oral hubo problema con la opción 2: El aprendizaje de un segundo idioma, que en la forma como se había planteado no elicitaba las desventajas, pues la instrucción de esta opción solo daba para hablar sobre el tema de manera positiva.

Opción 2: El aprendizaje de una segunda lengua a través de un proceso de inmersión. Se considera más efectivo el aprendizaje de una segunda lengua o lengua extranjera cuando se tiene la posibilidad de vivir en el país donde se habla esa lengua. ~~No obstante, hay estudios que muestran que quienes solo han tenido contacto con la lengua extranjera a través de la enseñanza también logran un buen dominio en la lengua.~~

A partir de esta observación se modificó parte de la instrucción de la opción 2. La parte marcada fue reemplazada por *Sin embargo, no todas las personas tienen la posibilidad de viajar y vivir esta experiencia.* Esta modificación fue efectiva luego en el proceso de pilotaje con los estudiantes extranjeros, puesto que elicitó las ventajas y desventajas sobre el tema.

Una vez que los estudiantes finalizaron el prototipo de prueba respondieron un cuestionario tipo Likert y de respuesta abierta para conocer sus actitudes y percepciones sobre la prueba EFA en su conjunto. Ellos juzgaron aspectos como la calidad de las instrucciones, la pertinencia de los textos y de la implementación del prototipo de prueba en línea, la dificultad de los ítems, la duración, etc. (ver anexo 6), como se muestra en la tabla 24.

- **Actitudes de los estudiantes frente al prototipo de prueba de EFA**

La tabla 24 presenta las respuestas que dieron los estudiantes luego de responder el prototipo de prueba de EFA en línea.

Ítems a evaluar	INDICADORES	
	De acuerdo %	Muy de acuerdo %
<i>Apreciación global sobre la prueba</i>		
¿Le parece interesante que la prueba esté implementada en computador?	22	78
El entorno de la prueba es fácil de usar	11	89
¿Considera aceptable esta prueba?	22	78
<i>Formato de la prueba</i>		
¿Le parece adecuado el número de secciones de la prueba: comprensión lectora, comprensión auditiva, competencia lingüística, expresión e interacción escrita y oral?	22	78
Las instrucciones de la prueba son claras	11	89
¿Le parece adecuada la duración total de la prueba (1 hora y 45 minutos)?	33	67
¿Le parece adecuada la duración de las distintas secciones de la prueba: comprensión lectora (30min.), comprensión auditiva (30 min.), expresión escrita (20 min.), competencia lingüística (10 min.) y expresión oral (15 min.)?	33	67
¿Le parecieron comprensibles los textos de la sección de comprensión lectora?	44	56
¿Le parecieron comprensibles los audios de la sección de comprensión auditiva?	22	78

Tabla 24: Actitudes de los estudiantes sobre el prototipo de prueba de EFA

De acuerdo con la tabla 24, las respuestas de los estudiantes se situaron en las categorías 4 y 5, de acuerdo y muy de acuerdo, lo cual da cuenta de una buena apreciación en cuanto a la validez aparente del prototipo de prueba de EFA en línea de nivel B2. Los ítems más destacados corresponden al formato de aplicación de la prueba, puesto que un 78% del grupo está muy de acuerdo con el hecho de que la prueba esté implementada en línea, lo que se reafirma con el 89% que está muy de acuerdo en que el entorno donde se encuentra la

prueba es fácil de utilizar. Esto da cuenta que a los estudiantes les parece interesante contar con instrumentos evaluativos automatizados.

En relación con la aceptabilidad de la prueba, a un 78% le parece muy aceptable esta prueba. Sin duda, este es un dato relevante en la validez de apariencia, puesto que si a los hablantes nativos no les parece aceptable una prueba que mide el uso de la lengua, difícilmente podría utilizarse en el contexto de una segunda lengua (Alderson *et al.*, 1998; Martínez, 2001).

En cuanto a los ítems de esta escala referidos al formato del prototipo de prueba, existe un alto nivel de acuerdo entre los estudiantes. Un 78% del grupo considera muy adecuado el número de secciones del prototipo de prueba y un 89% está muy de acuerdo en que las instrucciones de este prototipo de prueba son claras. Esta información es relevante para la aplicación de este prototipo de prueba a estudiantes de español como lengua extranjera, puesto que si las instrucciones son comprensibles para hablantes nativos no expertos, esto asegura que se está usando un lenguaje aceptado y usado socialmente y no restrictivo, por lo que no debería representar un problema para un estudiante de ELE de nivel B2.

En relación con la duración del prototipo de prueba, un 67% de los estudiantes está muy de acuerdo con la duración de 1 hora y 45 minutos que tiene la prueba y un 33% está de acuerdo con esta duración. Se observa los mismos valores en la duración de las distintas secciones del prototipo de la prueba, por lo que se considera que el tiempo asignado sería adecuado para evaluar el nivel de proficiencia en EFA de los estudiantes extranjeros de intercambio, cuestión que tendrá que corroborarse con la percepción del grupo meta.

En cuanto a la comprensión de los textos escritos, un 56% señaló estar muy de acuerdo y un 44% de acuerdo, es decir, la comprensión de los textos escritos no representó un problema. Del mismo modo, en la comprensión de los audios, un 78% indicó estar muy de acuerdo en que pudo comprender estos textos orales sin problemas.

Las respuestas de los estudiantes indican que los textos escritos y orales utilizados en la prueba, los ítems y las tareas les parecen representativos de las actividades académicas a las que se ven enfrentados, puesto que no las consideraron difíciles.

- **Respuestas abiertas del cuestionario de validez de apariencia**

La parte del cuestionario de preguntas abiertas consta de 10 ítems que tienen como objetivo conocer algunos aspectos no considerados en la escala Likert, por ejemplo, qué parte del prototipo de la prueba les pareció más fácil o difícil, cómo resolvieron los ítems de las distintas secciones, etc. De acuerdo con Hernández, Fernández y Baptista (2010), se codificó las preguntas abiertas agrupando las respuestas de los estudiantes en categorías, puesto que siempre hay patrones o ideas que se repiten en las respuestas y estas se pueden codificar para conocer su frecuencia. En este proceso de codificación de respuestas abiertas se debe tener en cuenta que una misma idea puede expresarse de distintas formas, por lo que corresponden a una misma categoría. Atendiendo a este procedimiento, se codificó las 10 preguntas de los estudiantes en la tabla 25.

Preguntas	Categorías	Frecuencia de mención
1. ¿Qué tal le pareció la prueba que acaba de realizar? (adecuada/inadecuada)	Es una prueba muy adecuada para evaluar a estudiantes extranjeros	8
	Es una buena prueba para extranjeros que tienen un dominio avanzado del español	1
	Es una buena prueba porque el contenido es adecuado, no aburre y las instrucciones son claras	1
2. ¿Qué parte de la prueba le pareció más fácil o más difícil? ¿Por qué?	La parte más fácil fue la de competencia lingüística	3
	La parte más fácil fue la de comprensión lectora porque podía volver a leer el texto	2
	La parte más difícil fue la comprensión lectora porque me cuesta esta habilidad	2
	La parte más difícil fue la escritura porque me cuesta organizar mis ideas	2
	Cada parte de la prueba tenía preguntas complejas y otras más fáciles	2
3. ¿Hubo algún texto oral o escrito que le haya parecido más fácil de comprender? ¿Cuál?	El texto sobre aprendizaje de una segunda lengua fue fácil de comprender	4
	El texto sobre competencias en ingeniería fue fácil de comprender	2
	Cada texto escrito y oral tenía sus complicaciones porque las respuestas no estaban explícitas	2
	El texto sobre biodiversidad fue fácil de comprender	1
	El audio de la escasez de agua en Copiapó fue fácil de entender	1
	El audio de la energía eólica fue fácil de entender	1
	Todos los textos tenían el mismo nivel de dificultad	1

4. ¿Hubo algún texto oral o escrito que le haya parecido más difícil de comprender? ¿Cuál?	El texto más complejo fue el de competencias en ingeniería	4
	El audio más difícil es el de los biocombustibles	2
	Ningún texto o audio fue difícil de comprender	2
	Fue difícil comprender los textos y audios sobre los que no tenía conocimientos previos	1
5. En la sección de comprensión lectora ¿las preguntas que respondió se relacionaban con el tema que cada texto trataba?	Todas las preguntas tenían relación con los temas tratados en los textos	9
6. En la sección de comprensión auditiva ¿las preguntas que respondió se relacionaban con el tema que cada audio trataba?	Todas las preguntas tenían relación con los temas tratados en los textos	9
7. ¿Qué hizo para resolver las preguntas en la sección de comprensión lectora?	Leer el texto completo por primera vez, luego leer las preguntas y sus opciones y luego una relectura para seleccionar la respuesta	3
	Leer atentamente el texto, releer e inferir	3
	Releer el texto para encontrar las respuestas literales	2
	Primero leer las preguntas del texto	2
	Las respuestas explícitas las buscaba en el texto, y para las implícitas releía donde creía que estaba la respuesta	1
8. ¿Qué hizo para resolver las preguntas de comprensión auditiva?	Primero leer las preguntas y sus opciones, luego escuchar el audio e ir respondiendo a medida que avanzaba	4
	Escuchar atentamente el audio	3
	Escuchar el audio en su totalidad y analizar globalmente sobre lo que trataba	2
	Poner atención a las palabras claves de los audios	1
9. ¿Cómo resolvió la tarea de la sección de competencia lingüística?	Analizando la coherencia de cada oración y párrafo del texto	8
	Pensando en el uso cotidiano de las formas gramaticales a seleccionar	1
10. ¿Cómo realizó las tareas de escritura y de oralidad? ¿Tuvo alguna dificultad? ¿Cuál?	No hubo dificultad para realizar las tareas de escritura y de oralidad	9
	Seleccionando el tema, mostrando una postura clara sobre él y desarrollando las ideas al respecto	3
	Me ayudó que se diera la estructura del texto que se debe escribir porque orienta	2
	En la expresión oral fue difícil mantener un registro formal	1
	En la expresión oral fue difícil enlazar las ideas para hablar fluidamente sobre el tema	1
	Realicé las tareas de escritura y oralidad con	1

	tranquilidad	
--	--------------	--

Tabla 25: Codificación de las preguntas abiertas

Como se observa en la tabla 25, en la pregunta 1 se menciona frecuentemente que el prototipo de prueba es adecuada para evaluar a estudiantes extranjeros (8 menciones). Otro aspecto interesante de la respuesta a esta pregunta, aunque solo la mencionó 1 estudiante es que la prueba no es tediosa, como suele ocurrir en las pruebas de segundas lenguas (Yagüe, 2010), sino que es adecuada en cuanto a duración y contenido.

En la pregunta 2, tres estudiantes mencionaron que la parte más fácil del prototipo de prueba fue la sección de competencia lingüística. Para dos estudiantes la parte más fácil fue la comprensión lectora porque tenían la posibilidad de volver a leer el texto. Sin embargo, otros dos participantes consideraron esta parte como la más difícil de la prueba porque tienen dificultades con la habilidad de comprensión lectora. La sección de escritura fue difícil para dos estudiantes porque les cuesta organizar sus ideas. Dentro del grupo también se hace mención 2 veces que cada parte de la prueba tenía preguntas simples y otras más complejas. Esto muestra que se dieron cuenta que la prueba planteaba preguntas de distintos niveles de complejidad.

En las respuestas a la pregunta 3, se menciona 4 veces que el texto más fácil de comprender en la sección de comprensión lectora fue el de aprendizaje de segundas lenguas. Podría pensarse que por la carrera que estudian los alumnos este texto le resultó más familiar, pero no más fácil, puesto que los estudiantes tuvieron más aciertos en el texto de competencias en ingeniería que en el de aprendizaje de segundas lenguas. Tampoco se observa diferencias entre sus respuestas en este texto de aprendizaje y el de biodiversidad.

Dos estudiantes mencionan que el texto de competencias en ingeniería resultó fácil de comprender. No obstante, hay también 2 estudiantes que señalaron que cada texto oral y escrito tenía sus complicaciones porque las respuestas no estaban explícitas.

Las respuestas a la pregunta 4 indican que el texto más difícil de comprender fue el texto de competencias en ingeniería y el audio más difícil el de los biocombustibles con dos menciones. Hay dos menciones que señalan que no hubo algún texto o audio que fuera difícil de comprender. Solo un estudiante expone que le fue difícil comprender los textos y audios sobre los que no tenía conocimientos previos, lo cual no implica que sobre los textos que tenía conocimientos previos haya tenido mejores resultados.

En cuanto a las respuestas a las preguntas 5 y 6, existe mención unánime entre los estudiantes en señalar que las preguntas de las secciones de comprensión lectora y auditiva tienen una relación directa con los temas tratados en los textos escritos y orales.

Respecto de la pregunta 7 y de las estrategias utilizadas por los estudiantes para resolver las preguntas de comprensión lectora, 3 participantes señalan haber leído el texto completo por primera vez para luego leer las preguntas y sus opciones y finalmente hacer una relectura del texto para seleccionar la respuesta; 3 estudiantes mencionan que leyeron atentamente el texto, lo releeron e infirieron las respuestas; 2 releeron el texto para buscar las respuestas literales y solo dos mencionan haber leído primero las preguntas antes de leer el texto.

Las respuestas a la pregunta 8, relacionadas con las estrategias utilizadas para responder en la sección de comprensión auditiva, dan cuenta que lo primero que hicieron cuatro estudiantes fue leer las preguntas sobre el audio y sus opciones para luego escucharlo e ir respondiendo en la medida que este avanzaba; tres mencionaron que escucharon atentamente el audio y solo dos señalan haber escuchado el audio en su totalidad para analizar globalmente lo que trataba.

En relación con las respuestas de la pregunta 9 sobre la sección de competencia lingüística se puede establecer que hay 8 menciones que reafirman que la estrategia que utilizaron los estudiantes fue la de analizar la coherencia de cada oración, párrafo y del texto en general para seleccionar la forma gramatical adecuada.

Finalmente, en cuanto a los resultados de la pregunta 10 sobre las tareas de expresión e interacción escrita y oral, se observa que todos los estudiantes señalaron no tener dificultad para realizar estas tareas, y dieron a conocer algunas de las estrategias utilizadas. En la sección escrita y oral hay 3 estudiantes que abordaron este tipo de tarea seleccionando el tema sobre el cual iban a escribir o a hablar, elaboraron una postura sobre él y desarrollaron sus ideas al respecto; dos estudiantes indicaron que la estructura para la elaboración del texto escrito fue de gran ayuda, puesto que orientó el proceso de escritura. Solo se hizo una mención a que fue difícil en la tarea de expresión e interacción oral mantener un registro formal en el uso del español. De acuerdo con Mendoza (2015), se aprecia en es acertado proporcionar la estructura del texto que se espera que el estudiante elabore.

4.2. Resultados de la validez de contenido de los ítems del prototipo de prueba EFA

En este apartado se presenta los resultados correspondientes a la validez de contenido llevada a cabo mediante la técnica de juicio de expertos. Como se describió anteriormente, cada juez evaluó de manera individual los ítems del prototipo de prueba. A continuación se presenta el resumen de las calificaciones dadas en las distintas dimensiones que miden los ítems y el grado de acuerdo que presentan los 8 evaluadores (2 estudiantes de Magíster en Lingüística Aplicada, 1 Magíster en Lingüística, 3 candidatos al grado de Doctor en Lingüística y 2 Doctores en Lingüística). Dicho acuerdo fue medido mediante el estadístico W de Kendall.

Como se observa en la tabla 26, en la dimensión I de la comprensión lectora y auditiva, que corresponde a las preguntas 1 y 2 de cada texto, todos los jueces calificaron los ítems entre 4 y 5 en la escala Likert, lo que significa que hay acuerdo entre el grupo de expertos en señalar que los ítems están midiendo la dimensión de identificar información específica en los textos orales y escritos. De hecho, calificaron 7 preguntas de esta dimensión con un 100 de muy de acuerdo, lo que otorga seguridad en cuanto a que el instrumento está evaluando lo que dice evaluar.

SECCIONES	INDICADORES	
	De acuerdo	Muy de acuerdo
Comprensión lectora y auditiva		
<i>Texto I</i>		
1. De acuerdo con el texto, la ingeniería realiza aportes relevantes al desarrollo de un país.		100
2. Según el texto, la formación que reciben los ingenieros les impide encontrar trabajo.	75	25
<i>Texto II</i>		
1. De acuerdo con el texto...		100
2. Según el texto, la aptitud lingüística es...		100
<i>Texto III</i>		
1. Según el texto, el aumento de la población a nivel mundial ha...		100
2. De acuerdo con el texto, la biodiversidad...	12,5	87,5
<i>Audio I</i>		
1. La energía eólica corresponde a la energía del viento.		100
2. Según el audio, el objetivo de la instalación de la turbina es transformar la energía eléctrica en una energía limpia.	25	75
<i>Audio II</i>		
1. De acuerdo con el audio ¿Cuál es la situación del río		

Copiapó?	25	75
2. El objetivo del proyecto Agua Copiapó es...		100
<i>Audio III</i>		
1. Según el audio, el petróleo es un combustible...	12,5	87,5
2. El Centro de Biotecnología de la Universidad de Concepción está interesado en...	25	75

Tabla 26: Dimensión I. Identificación de información específica en textos especializados

En la dimensión II sobre la inferencia de las ideas principales de los párrafos, que corresponde a las preguntas 3 y 4 de comprensión lectora y auditiva, los ítems también fueron calificados entre 4 y 5 en la escala Likert. En este caso, un evaluador calificó con 3 las preguntas 3 y 4 del texto 2, argumentando que si bien la pregunta evaluaba la dimensión la forma en que estaba redactada no la elicataba, se hizo las modificaciones y se le pidió al experto que evaluara estos cambios, en esta ocasión calificó los ítems con la categoría 5.

SECCIONES	INDICADORES	
	De acuerdo	Muy de acuerdo
Comprensión lectora y auditiva		
<i>Texto I</i>		
3. Las competencias de egreso se reflejan en el dominio disciplinar que debe mostrar todo profesional de ingeniería.	12,5	87,5
4. Se puede inferir que las escuelas de Ingeniería han intentado implementar cambios en la formación académica de sus estudiantes.	37,5	62,5
<i>Texto II</i>		
3. Según el texto, el estudiante de un segundo idioma aprende mejor la habilidad auditiva...	37,5	62,5
4. En el aprendizaje de una segunda lengua, la habilidad motora...	25	75
<i>Texto III</i>		
3. La situación actual de la biodiversidad indica que	25	75
4. En el futuro se espera que haya...		100
<i>Audio I</i>		
3. Según el audio, la instalación de turbinas mejorará la calidad energética en las zonas alejadas de la ciudad.	12,5	87,5
4. De acuerdo con el audio, la gente que vive en zonas rurales presenta bajos niveles de calidad de vida.	25	75
<i>Audio II</i>		
3. La tecnología utilizada en el proyecto Agua Copiapó...		100
4. Según el audio...	25	75
<i>Audio III</i>		

3. El Centro de Biotecnología puede llevar a cabo su investigación porque la región donde se encuentra cuenta con...	25	75
4. De acuerdo con el audio, las microalgas...	25	75

Tabla 27: Dimensión II. Inferencia de las ideas principales de los párrafos

Como se aprecia en la tabla 27, todos los ítems de la dimensión II fueron calificados por la mayoría de los evaluadores por sobre el 60% en la categoría de muy de acuerdo, lo que indica que son aptos para ser considerados en el prototipo de prueba. No obstante, se consideró la sugerencia de los evaluadores en relación con las preguntas 3 y 4 del texto 2, en lo referido a mejorar la redacción para elicitación de la inferencia, quedando de la siguiente manera: *De acuerdo con el texto, la mejor forma de aprender la habilidad auditiva es...* (pregunta 3, texto 2) y *Según el texto, la habilidad motora de una segunda lengua...* (pregunta 4, texto 2).

La tabla 28 presenta los resultados de la dimensión III, correspondiente a las preguntas 5 de cada texto escrito y oral. Estas preguntas evalúan la inferencia de la idea central del texto/audio.

SECCIONES	INDICADORES	
	De acuerdo	Muy de acuerdo
<i>Comprensión lectora y auditiva</i>		
<i>Texto I</i>		
5. A partir de la lectura del texto se puede inferir que la sociedad requiere de egresados vinculados al ámbito laboral.	12,5	87,5
<i>Texto II</i>		
5. ¿Cuál es la idea central del texto?	37,5	62,5
<i>Texto III</i>		
5. ¿Cuál es la idea central del texto?	37,5	62,5
<i>Audio I</i>		
5. La idea central del audio es dar cuenta del aporte de las energías renovables al suministro eléctrico.		100
<i>Audio II</i>		
5. ¿Cuál es la idea central del audio?	12,5	87,5
<i>Audio III</i>		
5. ¿Cuál es el tema principal del audio?	25	75

Tabla 28: Dimensión III: inferencia de la idea central del texto

Los resultados del juicio de expertos en la dimensión III del prototipo de prueba evidencian que las preguntas planteadas evalúan la dimensión que dicen evaluar, puesto que todos los expertos situaron sus respuestas en las categorías de acuerdo y muy de acuerdo. Aún más, sobre el 60% de los evaluadores está muy de acuerdo con que estos ítems evalúan la inferencia de la idea central de cada texto oral y escrito.

La dimensión IV de esta pauta, que corresponde al dominio de la competencia lingüística para el nivel B2 de EFA, presenta un alto nivel de acuerdo entre los evaluadores, 6 muy de acuerdo y 2 de acuerdo, lo que representa un 75% y un 25% respectivamente. En la dimensión V, correspondiente a si las tareas abiertas de expresión e interacción escrita y oral evalúan las modalidades expositivas y argumentativas, los evaluadores las calificaron como muy de acuerdo en un 87,5% para la escritura y un 75% muy de acuerdo para la oral. La dimensión VI que evalúa si las tareas de expresión e interacción escrita y oral miden el género académico ensayo y exposición oral, respectivamente, la primera fue calificada como muy de acuerdo por un 87,5% de los evaluadores y la segunda como muy de acuerdo en un 62,5% y un 37,5% de acuerdo. En la tarea de expresión escrita, dos expertos sugirieron que se abordara otros temas más académicos en la opción 2 de esta tarea (El Día Mundial sin compras).

Finalmente, se consideró cada una de las sugerencias de los expertos y se realizó las modificaciones al prototipo de prueba EFA con la finalidad de mejorar la calidad y validez del instrumento. La opción 2 de la tarea de escritura fue reemplazada por el tema de la crisis del agua a nivel mundial, de la misma forma se cambió la redacción de las preguntas 3 y 4 del texto 2, de modo que elicitaran la inferencia a nivel de párrafo.

Las sugerencias indicadas por los expertos se relacionaban con aspectos de presentación de las opciones y de redacción de los ítems.

A partir de los resultados de este juicio de expertos, se evidencia la efectividad del modelo de criterios, dado que para la elaboración de este prototipo de prueba y sus ítems fue esencial contar con los criterios definidos en el diseño, lo cual permitió organizar de manera secuencial y rigurosa el proceso de desarrollo del instrumento. Este juicio de expertos ha servido, además, para corroborar que los ítems y tareas son apropiados para medir el nivel de proficiencia B2 en EFA en un contexto mediatizado por computador.

En efecto, se sugiere el uso de instrumentos que permitan recoger datos de manera ordenada y que faciliten el trabajo de quienes participen como jueces expertos. Como señala Alderson *et al.*, (1998), en la validez de contenido se debe elaborar documentos claros que permitan interpretar de manera adecuada las opiniones de los expertos en función de los objetivos de la investigación.

4.2.1. Análisis estadístico de concordancia entre los jueces (W de Kendall)

Para conocer el grado de acuerdo entre los jueces expertos se aplicó el estadístico W de Kendall, que se utiliza para conocer el grado de asociación entre k conjuntos de rangos. El mínimo valor asumido por el coeficiente es 0 y el máximo es 1. Se utiliza este coeficiente cuando se trata de variables ordinales, en este caso, la escala ordinal es de 5 puntos, de esta manera se estima la confiabilidad de un juicio de expertos que considera elementos subjetivos. Cuando el valor de acuerdo es alto significa que hay consenso en la calificación que han realizado los evaluadores. De acuerdo con Escobar y Cuervo (2008), el programa SPSS indica el nivel de significancia, y cuando es inferior al 0.05, se rechaza la H0 y se acepta que hay concordancia significativa entre los rangos asignados por los jueces.

En este caso la escala que evaluaron los expertos contiene 35 ítems, distribuidos en las seis dimensiones que evalúa el prototipo de prueba: dimensión 1: identificación de información literal o específica en textos especializados, dimensión 2: inferencia de las ideas principales de los párrafos, dimensión 3: inferencia de la idea global del texto, dimensión 4: dominio gramatical nivel B2 de EFA, dimensión 5: modalidades expositiva y argumentativa, y dimensión 6: género académico.

De acuerdo con la tabla 29, el valor del coeficiente de acuerdo de Kendall es de .427 con un valor p inferior a 0,05, por lo que se acepta la hipótesis de que hay un nivel significativo de concordancia entre los evaluadores.

N	35
W de Kendall	,427
Gl	8
Sig. asintót.	,000

Tabla 29: W de Kendall

Esto significa que hubo acuerdo significativo entre los ocho jueces en cuanto a que el prototipo de prueba está evaluando 6 dimensiones distintas en los 35 ítems. Esto da cuenta que el conjunto de evaluadores aplicó los mismos criterios a la hora de asignar los puntajes a los diferentes ítems. Este resultado confirma que los ítems están midiendo lo que dicen medir.

A partir de los datos cualitativos y cuantitativos obtenidos en la validez de apariencia y de contenido se ha podido retroalimentar el prototipo de prueba y así aumentar su grado de validez estadística. La información obtenida en ambos procesos de validación ha permitido hacer las modificaciones y/o ajustes necesarios antes de pilotear el instrumento con el grupo de estudiantes extranjeros de intercambio en la Universidad de Concepción.

Hay que señalar que una perspectiva válida es la opinión de los expertos, en cuyas opiniones se está dispuesto a confiar, sin embargo, también se debe considerar en la elaboración y evaluación de un prototipo de prueba la etapa de pilotaje.

4.3. Resultados de la validación del prototipo de prueba EFA según los principios de CALT

Otra parte de la validación de contenido de la prueba se refiere a la evaluación del instrumento en su modalidad de aplicación en línea. Para ello, se aplicó una escala de Likert de 5 puntos (5: Muy de acuerdo, 4: De acuerdo, 3: Indiferente, 2: En desacuerdo, 1: Muy en desacuerdo), que se elaboró a partir de las directrices metodológicas de CALT (Chapelle y Douglas, 2006; Noijons, 1994). El objetivo de esta evaluación fue evidenciar en qué medida el prototipo de prueba en línea cumple con los principios metodológicos de CALT en lo referido a la implementación de una prueba de lengua.

El instrumento para la validación del prototipo de prueba en línea se dividió en dos dimensiones, una primera dimensión sobre el enfoque evaluativo bajo el cual se ha elaborado la prueba, y una segunda dimensión sobre los aspectos técnicos de lo que implica plantear una prueba de segundas lenguas en línea. Cada evaluador recibió un usuario y clave para poder ingresar al prototipo de prueba en un periodo de tiempo determinado. Las tablas 30 y 31 muestran los resultados de esta validación.

Ítems	4	5
1. La prueba cumple con el objetivo de evaluar el nivel de proficiencia B2 en Español como Lengua Extranjera con fines académicos (EFA)		100%
2. Los materiales (textos, audios, tareas) de la prueba evalúan el español en contextos auténticos, acorde con el enfoque comunicativo de enseñanza y evaluación de segundas lenguas	10%	90%
3. Los diferentes tipos de ítems de la prueba son adecuados para evaluar las distintas habilidades lingüísticas acordes con el nivel B2 de EFA	10%	90%
4. Considera adecuada la distribución de las secciones de la prueba: comprensión lectora, auditiva, competencia lingüística, expresión e interacción escrita y oral	20%	80%
5. Considera adecuada la duración de la prueba	20%	80%

Tabla 30: Resultados dimensión evaluación de segundas lenguas

En términos generales se observa que la primera dimensión del prototipo de prueba tuvo una evaluación positiva de parte de los expertos, puesto que sus respuestas se situaron en los niveles de acuerdo y muy de acuerdo (4 y 5 de la escala). De hecho, la evaluación mayoritaria está centrada en el valor máximo de la escala. Por ejemplo, en el primer ítem se observa que el prototipo de prueba en línea está cumpliendo en un 100% con los propósitos de evaluar el español con fines específicos académicos en un nivel B2. El segundo ítem muestra que un 90% de los jueces señala estar muy de acuerdo en que los materiales evalúan el español en un contexto comunicativo auténtico. Del mismo modo, un 90% marcó la opción muy de acuerdo cuando se les consultó si los diferentes tipos de ítems eran los adecuados para evaluar las diferentes habilidades lingüísticas; mientras que un 80% de los jueces está muy de acuerdo con la distribución de las secciones del prototipo de prueba y con la duración.

Ítems	4	5
1. Es fácil la navegación en la página (botones visibles y simples)	20%	80%
2. Es simple acceder y finalizar la sesión de la prueba	10%	90%
3. La prueba tiene una conectividad segura (requiere de usuario y clave, de fecha y hora de aplicación)		100%
4. La prueba en línea evita elementos distractores (animaciones, figuras, exceso de botones, etc.)	10%	90%
5. La prueba tiene una velocidad de carga adecuada		100%
6. Las instrucciones de la prueba se presentan con claridad	10%	90%
7. El volumen de los audios es adecuado		100%
8. Las preguntas se presentan de manera ordenada y con un tamaño de letra adecuado	20%	80%
9. Le parece interesante que se presenten todas las preguntas de una sección en una sola página	10%	90%
10. En general, considera que esta prueba en línea está implementada en un formato simple	10%	90%

Tabla 31: Resultados de la validación en la dimensión técnica del prototipo de prueba

Los resultados de la segunda dimensión del prototipo de prueba indican que, efectivamente, se trata de un instrumento que cumple con lo exigido por la metodología de CALT. En el primer ítem, un 80% de los jueces indicó estar muy de acuerdo en que el ambiente donde se encuentra la prueba es de fácil navegación, lo que es esencial en una prueba de segundas lenguas, dado que el estudiante puede desplazarse sin problemas en la página y no requiere ayuda del profesor. En el segundo ítem, un 90% de los expertos señaló que el acceso y el cierre de la sesión en el prototipo de prueba es una tarea simple.

En cuanto a la seguridad de la prueba, requisito fundamental en la evaluación de una segunda lengua, los 10 jueces indicaron estar muy de acuerdo en que el prototipo de prueba es seguro, puesto que se necesita que un profesor/administrador registre a los estudiantes y les genere un usuario, una clave y les asigne la fecha y la duración.

Los resultados del ítem 4 indican que el 90% de los jueces está muy de acuerdo con que el prototipo de prueba evita elementos externos que distraigan a los estudiantes. En el ítem 5 el 100% de los jueces señaló que la prueba tiene una velocidad de carga adecuada. En el ítem 6, un 90% de los jueces está muy de acuerdo en que las instrucciones del prototipo de prueba son adecuadas, es decir, son comprensibles y suficientes para quienes lo rinden. Los

resultados del ítem 7 muestran que el volumen de los audios de la prueba es adecuado. Asimismo, en el ítem 8, un 80% está muy de acuerdo con que las preguntas de la prueba se presentan de manera ordenada y con un tamaño de letra adecuado. Para el 90% de los jueces es un aporte relevante el hecho de que las preguntas de cada sección de este prototipo de prueba estén en una sola página, lo que beneficia al estudiante en cuanto a la disminución de tiempo y de navegación.

En términos generales, el 90% de los evaluadores consideró que el prototipo de prueba EFA en línea de nivel B2 está implementada en un formato de fácil navegación y uso, lo que se condice con lo señalado por Chapelle y Douglas (2006) en cuanto a los aportes de CALT a la evaluación en una segunda lengua.

4.4. Resultados del pilotaje

En este apartado se presenta los resultados obtenidos por la muestra de 28 estudiantes extranjeros de intercambio universitario.

La tabla 32 muestra los porcentajes generales obtenidos por los estudiantes en las diferentes secciones del prototipo de prueba de EFA en línea de nivel B2, acorde con las exigencias definidas en el criterio de diseño.

N° sujeto	C.lectora 25%	C.auditiva 20%	C.lingüística 10%	E.I escrita 25%	E.I oral 20%	Total (100%)
1	25%	16%	8%	14%	12%	74%
2	22%	15%	9%	14%	16%	75%
3	18%	9%	7%	16%	14%	64%
4	15%	12%	9%	14%	8%	57%
5	17%	12%	9%	16%	16%	68%
6	18%	13%	8%	13%	15%	67%
7	10%	16%	7%	13%	14%	60%
8	17%	15%	8%	10%	13%	62%
9	18%	11%	9%	19%	16%	72%
10	20%	9%	6%	9%	4%	49%
11	12%	16%	8%	16%	11%	63%
12	17%	9%	7%	11%	11%	54%
13	13%	9%	8%	7%	13%	50%
14	20%	13%	7%	9%	12%	62%

15	17%	11%	7%	6%	10%	51%
16	15%	12%	9%	20%	15%	71%
17	18%	12%	6%	9%	7%	53%
18	18%	9%	8%	13%	11%	58%
19	15%	9%	8%	13%	11%	56%
20	20%	15%	8%	13%	12%	67%
21	18%	12%	7%	13%	12%	62%
22	18%	15%	8%	16%	13%	69%
23	15%	7%	8%	15%	13%	57%
24	15%	9%	8%	18%	12%	62%
25	18%	15%	7%	16%	14%	70%
26	22%	15%	7%	11%	12%	66%
27	20%	15%	8%	13%	13%	68%
28	17%	8%	7%	13%	13%	58%
PROMEDIO	17%	12%	7%	13%	12%	62%

Tabla 32: Porcentajes de logro de la muestra en el prototipo de la prueba EFA de nivel B2

El prototipo de prueba EFA en línea consta de un total de 82 puntos, 15 puntos la sección de comprensión lectora, 15 puntos la sección de comprensión auditiva, 20 puntos la sección de competencia lingüística, 16 puntos la sección de expresión e interacción escrita y 16 puntos la sección de expresión e interacción oral. Sin embargo, para conocer el rendimiento real de cada sujeto se calculó el porcentaje logrado en cada sección, como se especificó en el criterio de diseño del modelo evaluativo. Así, la comprensión lectora equivale a un 25% del total de la prueba, la comprensión auditiva a un 20%, la competencia lingüística a un 10%, la expresión e interacción oral a un 20% y la expresión e interacción escrita a un 25%.

En términos generales se aprecia en la tabla 32 que el mayor logro obtenido por el grupo fue en la competencia lingüística con un 7% del 10% exigido, seguido de la comprensión lectora con un promedio de un 17% y en la comprensión auditiva con un 12%, del 20% exigido. Por el contrario, las habilidades donde mostraron un bajo resultado fueron la expresión e interacción escrita con un 13% y la expresión e interacción oral con un 12%. Si bien la interacción oral no representa un problema en la vida cotidiana de un estudiante de nivel intermedio sí tiene implicancias en un contexto académico. No es lo mismo conversar sobre temas cotidianos que abordar un género académico como la exposición, donde se debe argumentar sobre un tema, mantener un registro formal, sin la ayuda del profesor, mostrar un dominio léxico, etc.

A partir de estos resultados, habrá que revisar si en un futuro, cuando se plantee una prueba de EFA a gran escala, se debería mantener el número de secciones de este prototipo y la distribución de porcentajes por sección, puesto que esta muestra de 28 estudiantes está evidenciando que la comprensión auditiva fue más difícil para los sujetos.

4.4.1. Porcentaje de aprobación para determinar el nivel de EFA de la muestra

Teniendo en cuenta que las universidades chilenas exigen un 60% como mínimo para la aprobación de las asignaturas y que los estudiantes se insertan en este sistema educativo y evaluativo, en esta prueba se aplicó un 60% de exigencia. Es decir, aquellos estudiantes que muestren un 60% o más de logro en este prototipo de prueba de EFA en línea de nivel B2 podrían considerarse como estudiantes que presentan un nivel B2 de español, y por lo tanto, están en mejores condiciones para enfrentar las diversas tareas que demanda el ambiente académico, lo cual no implica que no requieran de algún tipo de ayuda. Contar con un nivel B2 garantiza que el alumno puede atender adecuadamente a las clases y otras tareas académicas en la lengua meta; no obstante, los estudiantes necesitan interiorizarse respecto de cómo funciona el discurso académico en el país, la universidad, la facultad o la carrera donde realizan su estadía (Pastor, 2006).

Acorde con la exigencia de un 60% para aprobar como estudiante de nivel B2, se observa en la tabla 32 que 18 estudiantes estarían en condiciones de cumplir con este requisito. Es decir, de los 28 sujetos, 18 están mejor preparados para insertarse a las distintas actividades que exige el ambiente académico, por ejemplo, comprender textos orales y escritos, atender y participar de las clases, hacer exposiciones, etc.

Estos resultados están en concordancia con lo señalado por Pastor (2006, 2010), en cuanto a que los estudiantes que llegan a las universidades españolas no siempre presentan este nivel, y que sería el ideal que las universidades exigieran algún tipo de certificación, pues así se podría garantizar que el estudiante pudiera comprender y participar en las clases adecuadamente.

Los resultados obtenidos por los estudiantes en el prototipo de prueba EFA en línea de nivel B2 fueron comparados y luego correlacionados con los datos obtenidos por la muestra en la

prueba CELE en su forma A (Ferreira, 2010). Como se explicó en el apartado 1.6.1, la prueba CELE está orientada a un nivel B1 de ELE en un contexto general. Esta prueba se aplica de manera regular, cada semestre, a los estudiantes extranjeros que llegan a la Universidad de Concepción a realizar un periodo de intercambio.

El objetivo de esta comparación es mostrar que tanto el prototipo de prueba EFA en línea como la prueba CELE tienen algún grado de correlación, dado que evalúan las mismas habilidades lingüísticas, aun cuando son pruebas diferentes. Para realizar esta correlación se re-escaló los puntajes de ambas pruebas (a 100 puntos) para que pudieran ser comparables en sus distintas secciones, excepto en la sección de léxico que no se consideró como una sección aparte en el prototipo de prueba EFA, sino incluida en tarea de escritura y de expresión oral. Las tablas 33 y 34 muestran este re-escalamiento y los resultados que obtuvo la muestra, tanto en el prototipo de prueba EFA como en la prueba general CELE.

Sujeto	C.lectora	C.auditiva	C.linguística	E.escrita	E. oral	Total
1	100,00	80,00	75,00	56,25	59,38	73,78
2	86,67	73,33	85,00	56,25	81,25	76,83
3	73,33	46,67	70,00	62,50	68,75	64,63
4	60,00	60,00	85,00	56,25	37,50	60,98
5	66,67	60,00	85,00	62,50	78,13	71,34
6	73,33	66,67	80,00	50,00	75,00	69,51
7	40,00	80,00	65,00	53,13	68,75	61,59
8	66,67	73,33	75,00	40,63	65,63	64,63
9	73,33	53,33	85,00	75,00	78,13	73,78
10	80,00	46,67	60,00	37,50	21,88	49,39
11	46,67	80,00	80,00	65,63	53,13	65,85
12	66,67	46,67	65,00	43,75	53,13	55,49
13	53,33	46,67	75,00	28,13	62,50	54,27
14	80,00	66,67	70,00	37,50	59,38	62,80
15	66,67	53,33	70,00	25,00	50,00	53,66
16	60,00	60,00	85,00	81,25	75,00	73,17
17	73,33	60,00	60,00	37,50	34,38	53,05
18	73,33	46,67	75,00	50,00	53,13	60,37
19	60,00	46,67	80,00	53,13	53,13	59,76
20	80,00	73,33	75,00	50,00	59,38	67,68
21	73,33	60,00	65,00	53,13	59,38	62,20
22	73,33	73,33	75,00	65,63	62,50	70,12

23	60,00	33,33	75,00	59,38	62,50	59,15
24	60,00	46,67	75,00	71,88	59,38	63,41
25	73,33	73,33	70,00	65,63	68,75	70,12
26	86,67	73,33	65,00	43,75	59,38	65,24
27	80,00	73,33	80,00	50,00	62,50	69,51
28	66,67	40,00	70,00	53,13	65,63	59,76
Promedio	69,76	60,48	74,11	53,01	60,27	64,00

Tabla 33: Puntajes re-escalados prototipo de prueba EFA en línea nivel B2

Sujeto	C.lectora	C.auditiva	C.lingüística	E.escrita	E. oral	Total
1	100,00	100,00	90,00	95,00	95,00	95,24
2	100,00	100,00	100,00	95,00	100,00	98,81
3	66,67	100,00	80,00	95,00	95,00	88,10
4	75,00	83,33	90,00	80,00	85,00	83,33
5	66,67	100,00	90,00	100,00	95,00	91,67
6	75,00	100,00	100,00	90,00	90,00	91,67
7	91,67	100,00	100,00	90,00	90,00	94,05
8	75,00	100,00	100,00	90,00	90,00	91,67
9	91,67	91,67	100,00	95,00	95,00	95,24
10	66,67	91,67	65,00	80,00	85,00	77,38
11	66,67	66,67	80,00	100,00	95,00	84,52
12	83,33	91,67	45,00	80,00	85,00	75,00
13	66,67	75,00	80,00	75,00	80,00	76,19
14	83,33	66,67	55,00	85,00	90,00	76,19
15	83,33	83,33	70,00	70,00	80,00	76,19
16	75,00	83,33	95,00	72,50	77,50	80,95
17	75,00	75,00	80,00	50,00	50,00	64,29
18	83,33	83,33	75,00	65,00	57,50	70,83
19	50,00	75,00	80,00	70,00	75,00	71,43
20	58,33	83,33	90,00	60,00	65,00	71,43
21	75,00	58,33	90,00	60,00	55,00	67,86
22	75,00	41,67	95,00	75,00	65,00	72,62
23	58,33	75,00	85,00	82,50	62,50	73,81
24	83,33	75,00	70,00	72,50	80,00	75,60
25	58,33	100,00	95,00	75,00	62,50	77,98
26	66,67	75,00	95,00	57,50	45,00	67,26
27	58,33	83,33	75,00	60,00	70,00	69,05
28	58,33	66,67	100,00	55,00	60,00	69,05
Promedio	73,81	83,04	84,64	77,68	77,68	79,55

Tabla 34: Puntajes re-escalados prueba CELE nivel B1

En términos generales, se observa en las tablas 33 y 34 que la mayoría de los estudiantes obtuvo mejores resultados en la prueba general CELE que en el prototipo de prueba EFA B2. El promedio general obtenido en el prototipo de prueba EFA fue de 64 puntos y en la prueba CELE fue de 79,55 puntos. Además, los resultados confirman la tendencia general en cuanto a que los sujetos que tuvieron puntajes más altos en el prototipo de prueba EFA también presentaron un buen desempeño en la prueba CELE.

El gráfico 1 muestra la distribución de los porcentajes obtenidos por los estudiantes en las dos pruebas.

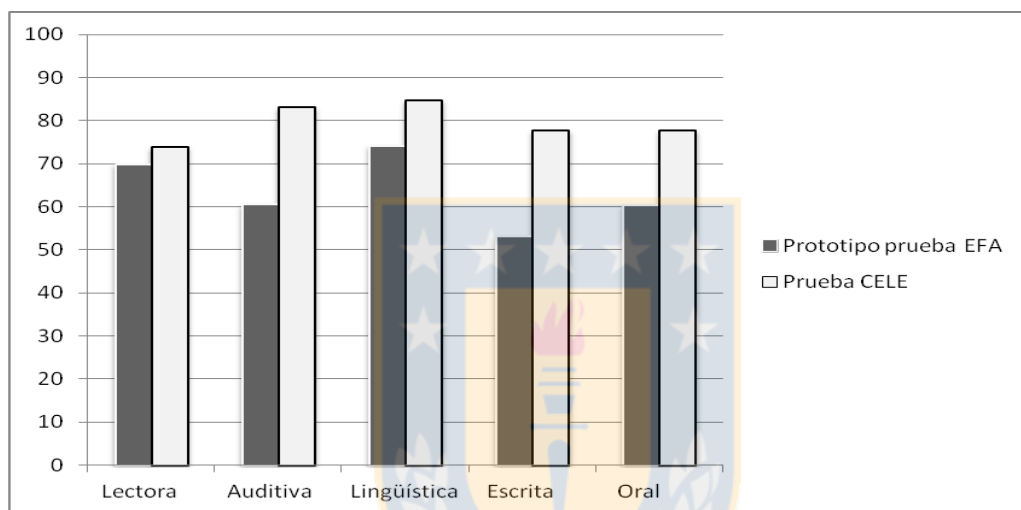


Gráfico 1: Resultados generales prototipo de prueba EFA y prueba CELE

Como se observa en el gráfico 1, la comprensión lectora presenta resultados similares en ambas pruebas. Según los promedios presentados en las tablas 33 y 34, los resultados del grupo en cuanto a la comprensión lectora en las dos pruebas difiere en 4,05 puntos. Si bien se observa diferencias, estas no son significativas. Esto se puede deber a que los estudiantes han estado más expuestos, en un contexto académico, a la lengua escrita, por lo tanto, tienen un mayor dominio de esta destreza. Además, en la comprensión lectora del prototipo de prueba, las preguntas se plantearon de manera ordenada, según la aparición de la información en el texto, lo cual agiliza el tiempo y facilita la búsqueda de las respuestas; a esto se suma que los estudiantes disponen todo el tiempo del texto, por lo que pueden hacer relecturas.

En cambio, se observa diferencias notorias en los resultados en la comprensión auditiva con 22,6 puntos de diferencia, esto se explica porque en los textos orales es complejo retener toda la información de los textos, aun cuando también las preguntas van planteadas en orden de aparición. Se trata de dos instrumentos orientados a niveles y fines distintos.

Estos resultados podrían sugerir que en un futuro, en una prueba de EFA, se reconsiderara la distribución de los porcentajes que se ha hecho en este prototipo de prueba.

Las diferencias de la expresión e interacción escrita corresponden a 24,7 puntos, lo que se explica por el tipo de tarea que debe realizar el estudiante (escribir un ensayo sobre: a) formación a distancia por internet, o b) crisis mundial del agua). Se trata de un género académico más complejo que el exigido por la prueba CELE, donde tiene que escribir o una propuesta de viaje o una opinión sobre un tema.

La competencia lingüística tiene un mayor logro en ambas pruebas, aunque en la prueba CELE se obtuvo un promedio más alto, la diferencia entre las dos pruebas en esta sección es de 10 puntos. Esto se condice con el hecho de que esta parte del prototipo de prueba EFA resultó más fácil para el grupo. No obstante, se observa bajos puntajes en las habilidades de expresión escrita y oral, lo que indica que fue difícil para la muestra producir un ensayo y realizar una exposición oral sobre uno de los temas propuestos (a) aprendizaje de una segunda lengua a través del proceso de inmersión o b) el uso de pesticidas o elementos químicos en la agricultura).

4.4.2. Resultados aplicación de la prueba T- Student

Con el objeto de conocer si ambas pruebas presentaban diferencias estadísticamente significativas, es decir, se trataba de pruebas distintas, puesto que la prueba CELE se orienta a un nivel B1 de ELE general y el prototipo de prueba EFA en línea se orienta a un nivel B2 con fines específicos académicos, se aplicó la prueba T-Student para muestras pareadas.

Secciones	EFA-UdeC	ELE-UdeC	valor p
Lectora	69,8 (12,5)	73,8 (13,0)	0,1947
Auditiva	60,5 (13,6)	83,0 (15,0)	<0,0001
Lingüística	74,1 (7,6)	84,6 (14,3)	0,0003
Escrita	53,0 (13,4)	77,7 (14,6)	<0,0001
Oral	60,3 (13,3)	77,7 (15,8)	<0,0001
Total	64,0 (7,1)	79,6 (10,1)	<0,0001

Tabla 35: Valor p en las distintas secciones de las pruebas

De acuerdo con estos datos es posible afirmar que hay diferencias estadísticamente significativas entre las distintas secciones de la prueba, excepto en la comprensión lectora donde solo hay 4 puntos de diferencia entre los resultados. En las demás secciones de la prueba se observa que los valores p están por debajo de 0,05. Por lo tanto, es posible afirmar que se trata de pruebas diferentes (con distintos objetivos y niveles de complejidad) aun cuando evalúan las mismas habilidades lingüísticas.

Sin duda, estos datos son relevantes y justifican el planteamiento de pruebas que midan el uso del español en un contexto específico, académico, dado que permiten conocer con mayor precisión cuáles son las habilidades lingüísticas de los estudiantes en este contexto. Si bien estos estudiantes logran sus metas comunicativas fuera del ambiente académico, con la enseñanza y evaluación del EFA se busca que ese éxito pueda reflejarse también en sus estudios. Como señala Pastor (2006) es indudable que los estudiantes que hacen intercambio obtienen beneficios en cuanto a lo lingüístico, lo social y lo cultural. Sin embargo, no está claro qué tan enriquecedor es el aprovechamiento en el plano académico.

4.4.3. Resultados de las correlaciones

Para observar el grado de correlación entre ambas pruebas se aplicó el coeficiente de correlación de Pearson, como se muestra en la tabla 36.

Secciones	Correlación
Lectora	0,20
Auditiva	0,13
Lingüística	0,36
Escrita	0,30
Oral	0,27
Total	0,53

Tabla 36: Coeficiente de correlación de Pearson

De acuerdo con estos resultados, se observa un coeficiente de correlación moderado entre ambas pruebas. Los expertos señalan que un nivel adecuado de correlación oscila entre 0,4 y 0,6 (Alderson, *et al* 1998 y Martínez, 2011). Se observa en la tabla que la correlación más baja se presenta en la habilidad de comprensión auditiva, puesto que en esta sección los resultados que obtuvieron los estudiantes en el prototipo de prueba EFA fueron notoriamente más bajos que los de la prueba CELE. Sin embargo todas las secciones están correlacionadas con el total con un valor de 0,53.

Pese a las bajas correlaciones se puede decir que ambas pruebas tienen algún grado de similitud puesto que las dos evalúan dominio en español. De hecho, Alderson *et al.*, (1998) señalan que es difícil que una prueba de lenguas no presente algún nivel de correlación.

En cuanto a la correlación interna del prototipo de prueba EFA en línea de nivel B2, esta se calculó relacionando las diferentes partes de la prueba con el puntaje total.

	<i>Lectora</i>	<i>Auditiva</i>	<i>C. Lingüística</i>	<i>Escrita</i>	<i>Oral</i>
Total	0,32614917	0,56201313	0,65470978	0,67238286	0,75409624

Tabla 37: Correlaciones entre las diferentes partes del prototipo de la prueba EFA y el total

Acorde con estos coeficientes, se observa un buen nivel de correlación entre las distintas secciones y el total del prototipo de la prueba, siendo la correlación más baja la que se da entre la comprensión lectora y el total con un 0,32, puesto que fue en esta sección donde los estudiantes tuvieron mejores resultados. Según Alderson *et al.*, (1998) las correlaciones entre cada parte de la prueba y toda la prueba estará alrededor de 0,7 o más, por lo que, en este caso, solo sería problemática la comprensión lectora.

4.4.4. Análisis y validación de los ítems del prototipo de prueba

En este apartado se presenta los resultados y el análisis de las respuestas de los estudiantes en las diferentes secciones del prototipo de la prueba EFA en línea nivel B2. Para ello, se construyó una matriz con cada una de las respuestas de los estudiantes en las diferentes secciones, como se ejemplifica en la tabla 38 con la comprensión lectora.

N° sujeto	Texto 1 (V/F)					Texto 2					Texto 3				
	b	a	a	b	b	d	a	b	c	a	b	c	a	c	c
	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
1	b	a	a	b	b	d	a	b	c	a	b	c	a	c	c
2	b	a	a	b	b	d	a	b	a	a	b	c	a	c	a
3	b	a	a	b	b	d	a	b	d	d	b	c	b	c	a
4	b	b	b	b	b	d	a	b	d	c	b	c	b	c	a
5	b	b	a	b	b	d	a	b	a	b	b	c	b	c	a
6	b	a	b	b	b	d	a	b	a	a	b	c	a	b	a
7	a	b	b	b	a	d	a	b	a	a	b	d	b	b	a
8	b	a	b	b	b	d	a	b	c	b	b	c	b	b	a
9	b	b	a	a	b	d	a	b	a	a	b	c	b	c	c
10	b	a	b	b	b	d	a	b	c	d	b	c	a	a	c
11	b	b	b	b	a	c	a	b	a	a	c	c	d	c	a
12	b	a	b	a	b	d	a	b	b	a	b	c	a	b	a
13	b	a	a	b	a	d	a	b	a	d	b	d	b	b	a
14	b	b	a	b	b	d	a	b	c	a	b	c	b	b	c
15	b	a	b	b	b	d	a	b	a	b	b	c	a	d	b
16	a	a	b	b	a	d	a	b	c	a	b	d	b	c	a
17	b	a	a	b	b	d	a	b	a	a	b	c	b	b	a
18	b	b	a	b	b	d	a	b	a	b	b	c	a	c	a
19	b	a	b	b	a	d	a	b	d	a	c	a	a	a	c
20	b	a	b	b	b	d	a	b	c	a	b	c	a	b	a
21	b	b	a	b	b	d	a	b	a	a	b	c	b	d	c
22	b	a	a	b	b	d	a	b	c	b	a	c	b	c	a
23	b	a	a	b	b	b	a	b	b	d	b	c	b	b	a
24	b	b	b	b	a	d	a	b	d	a	a	c	a	b	c
25	b	a	a	b	b	d	a	b	c	b	a	c	a	b	a
26	b	a	a	b	b	d	a	b	c	a	b	c	b	c	a
27	b	a	a	b	b	d	a	b	c	d	b	c	d	c	a
28	b	a	a	b	b	d	a	c	a	a	b	c	b	d	a
N° aciertos	26	19	16	26	22	26	28	27	10	16	23	24	11	12	7

Tabla 38: Muestra del número de aciertos por ítem en comprensión lectora

- **Comprensión lectora**

En términos generales, se observa que en el texto 1 de comprensión lectora la mayoría de los estudiantes acertó en las respuestas a las preguntas 1 y 4. Del mismo modo, en el texto 2 las preguntas con mayor número de aciertos fueron la 1, 2 y 3. Estos resultados, sin duda,

repercutirán en el índice de dificultad y de discriminación de dichos ítems, puesto que no presentan un comportamiento diferenciador frente al grupo. No obstante, hay ítems que solo tuvieron 7 y 10 aciertos, se trata de las preguntas 5 del texto 3 y de la pregunta 4 del texto 2, solo mediante el análisis estadístico se podrá conocer qué nivel de dificultad y de discriminación presentan estos ítems. En el anexo 4 se puede visualizar el prototipo de prueba EFA en línea B2 con todas las preguntas.

Acorde con la Teoría Clásica de los Test (TCT), dos índices comúnmente utilizados para determinar la calidad de una prueba son el coeficiente o índice de dificultad y el índice de discriminación (Ebel y Frisbie, 1991). En la tabla 39 se observa el índice de dificultad y de discriminación de los ítems que evalúan la habilidad de comprensión lectora.

	Texto 1 (V/F)					Texto 2					Texto 3				
Preguntas	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
Índice de dificultad	0,93	0,68	0,57	0,93	0,79	0,93	1,00	0,96	0,36	0,57	0,82	0,86	0,39	0,43	0,25
Índice de discriminación	0,25	0,36	0,46	0,00	0,75	0,25	0,00	0,00	0,73	0,09	0,38	0,50	0,32	0,20	0,18

Tabla 39: Dificultad y discriminación comprensión lectora

En la primera fila de la tabla 39 se muestra todas las preguntas de los textos de comprensión lectora, en la segunda fila se presenta el índice de dificultad del ítem, y en la última fila se muestra los valores del índice de discriminación. Para calcular el índice de discriminación se ordenó a los sujetos acorde con los puntajes obtenidos en cada sección y se dividió el grupo en 3, los de mejor rendimiento, los de rendimiento medio y los de rendimiento bajo.

Acorde con los índices recomendados por Ebel y Frisbie (1991) y Cervantes (1989, citado en López-Mezquita, 2005), los rangos para la clasificación de los ítems en función del nivel de dificultad son muy fáciles, fáciles, deseables, difíciles y muy difíciles. Los ítems cuyos valores se acercan a 1 resultan más fáciles.

Nivel dificultad	Ítems texto1	Ítems texto2	Ítems texto3
Muy fáciles (0,86-1)	2	3	1
Fáciles (0,71-0,85)	1	0	1
Deseables (0,40-0,70)	2	1	1
Difíciles (0,15-0,39)	0	1	2
Muy difíciles (0,01-0,14)	0	0	0

Tabla 40: Índice de dificultad ítems de comprensión lectora

En general, en la sección de comprensión lectora, de los 15 ítems 6 resultaron muy fáciles para los estudiantes, 2 fáciles, 4 deseables y 3 difíciles. Si observamos la tabla 39, las preguntas que tuvieron mayores aciertos resultaron fáciles o muy fáciles para esta muestra, sin embargo, dado que el número de sujetos evaluados es bajo no se puede desechar parte de los ítems muy fáciles, puesto que se necesita seguir incrementando la muestra para tener claridad respecto de las tendencias observadas hasta el momento en cuanto al comportamiento de los ítems del prototipo de prueba.

El texto 2 y 3, con ítems de selección múltiple, muestra que las preguntas 4 y 5 presentan mejores niveles de dificultad, puesto que son aquellas que miden la inferencia de la idea principal a nivel de párrafo y la inferencia de la idea global del texto. Por otro lado, se observa que las preguntas 1 y 2 de los tres textos resultaron fáciles puesto que miden información específica o literal de los textos.

Se considera que el índice de dificultad promedio de una prueba debe oscilar entre 0,5 y 0,6 (Backhoff, Larrazolo y Rosas, 2000). De acuerdo con esto, la sección de comprensión lectora presenta una media de 0,7, esto indica que esta parte de la prueba resultó fácil para la mayoría de los sujetos de la muestra.

Otro indicador dentro de la TCT es el índice de discriminación, acorde con estos valores la discriminación de la sección de comprensión lectora del prototipo de la prueba EFA en línea nivel B2 es la siguiente.

Nivel discriminación	Ítems texto 1	Ítems texto 2	Ítems texto 3
Muy buenos (>0,39)	2	1	1
Buenos (0,30-0,39)	1	0	2
Regulares (0,20-0,29)	1	1	1
Malos (0,00-0,19)	1	3	1
Pésimos (< -0,01)	0	0	0

Tabla 41: Índice de discriminación ítems comprensión lectora

La sección de comprensión lectora en total presenta 4 ítems que discriminaron muy bien y 3 ítems que discriminaron bien entre la muestra. Además, se observa 3 ítems regulares y 5 ítems malos, se trata de la pregunta 4 del texto 1, de las preguntas 2, 3 y 5 del texto 2 y de la pregunta 5 del texto 3 que no discriminaron entre los sujetos que presentaban un rendimiento más alto y los que tuvieron un resultado más bajo. Ebel y Frisbie (1991) recomiendan descartar o revisar aquellos ítems cuyo ID esté entre 0,00 y 0,29, sin embargo, en este caso se espera aumentar el número de la muestra para observar la variabilidad de estos ítems acorde con los diferentes niveles de proficiencia de los estudiantes.

- **Comprensión auditiva**

En cuanto a la comprensión auditiva, en la tabla 42 se muestra el número de aciertos, el índice de dificultad y de discriminación de los ítems de esta sección.

Preguntas	Audio 1 (V/F)					Audio 2					Audio 3				
	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
Índice de dificultad	0,89	0,93	0,93	0,46	0,64	0,46	0,39	0,50	0,39	0,75	0,57	0,50	0,68	0,75	0,21
Índice de discriminación	0,22	0,11	0,22	0,58	0,04	0,27	0,59	0,48	-0,04	0,33	0,47	0,48	0,67	-0,08	0,30

Tabla 42: Dificultad y discriminación comprensión auditiva

Los resultados muestran que las preguntas 2 y 3 del audio 1 presentaron un mayor número de aciertos, lo cual se corresponde con los resultados de la tabla 43 que muestra que estos ítems resultaron muy fáciles para los estudiantes.

Nivel dificultad/ texto	Ítems audio 1	Ítems audio 2	Ítems audio 3
Muy fáciles (0,86-1)	3	0	0
Fáciles (0,71-0,85)	0	1	1
Deseables (0,40-0,70)	2	2	3
Difíciles (0,15-0,39)	0	2	1
Muy difíciles (0,01-0,14)	0	0	0

Tabla 43: Índice de dificultad ítems comprensión auditiva

La media del índice de dificultad de la comprensión auditiva es de 0,60 por lo que cumple con la media de dificultad señalada por Ebel y Frisbie (1991). En general, la comprensión auditiva presenta 3 ítems muy fáciles, 2 fáciles, 7 deseables y 3 difíciles. Acorde con estos resultados, las tres primeras preguntas del audio 1 sobre la energía eólica fueron más fáciles para los estudiantes. Mientras que las preguntas de los audios 2 y 3 generalmente presentan mejores índices de dificultad.

En cuanto a la discriminación de los ítems de comprensión auditiva, en la tabla 44 se muestra que en general en esta sección hay 6 ítems que discriminan muy bien y 2 que discriminan bien, lo cual es un resultado interesante para mostrar validez a nivel de los ítems del prototipo de la prueba EFA en línea de nivel B2.

Nivel discriminación	Ítems audio 1	Ítems audio 2	Ítems audio 3
Muy buenos (>0,39)	1	2	3
Buenos (0,30-0,39)	0	1	1
Regulares (0,20-0,29)	2	1	0
Malos (0,00-0,19)	2	0	0
Pésimos (< -0,01)	0	1	1

Tabla 44: Índice de discriminación ítems comprensión auditiva

Además se puede observar que hay 3 ítems regulares, 2 malos y 2 pésimos. Los dos ítems con discriminación negativa corresponden a las preguntas 4 del audio 2 y 3, que tratan respectivamente sobre el proyecto de Agua Copiapó y del biocombustible. Ambas preguntas las contestó correctamente el mismo número de sujetos del grupo alto y del grupo bajo, por lo que se trataría de ítems poco adecuados que habría que revisar y corroborar en futuras aplicaciones para observar si esta tendencia se mantiene o varía.

- **Competencia lingüística**

La última parte de corrección objetiva del prototipo de prueba corresponde a la sección de competencia lingüística. La tabla 45 muestra los resultados obtenidos por la muestra.

Preguntas	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Índice de dificultad	0,75	0,86	0,43	0,54	0,18	0,75	0,89	0,89	0,89	0,89	0,86	0,89	0,86	0,75	0,89	0,39	0,93	0,82	0,68	0,68
Índice de discriminación	0,34	0,18	0,19	0,10	-0,07	0,34	0,07	0,07	0,18	0,18	-0,02	-0,02	0,16	0,25	0,07	0,19	0,09	0,18	0,45	0,32

Tabla 45: Índice de dificultad y discriminación competencia lingüística

Como se puede observar esta parte de la prueba fue la más fácil para los estudiantes, dado que presenta bajos índices de dificultad y de discriminación, es decir, la mayoría respondió correctamente y no hubo grandes diferencias entre los estudiantes del grupo alto y del grupo bajo.

De acuerdo con la tabla 46, los índices de dificultad de esta sección indican que 10 ítems resultaron muy fáciles, 4 fáciles, 4 deseables y 2 difíciles. Los ítems deseables corresponden a los 3, 4, 19 y 20. La pregunta 3 evalúa el uso de conectores consecutivos, la 4 el uso de preposiciones de finalidad, la 19 el uso de la preposición *en* y la 20 el uso de demostrativos.

Nivel dificultad	N° de ítems
Muy fáciles (0,86-1)	10
Fáciles (0,71-0,85)	4
Deseables (0,40-0,70)	4
Difíciles (0,15-0,39)	2
Muy difíciles (0,01-0,14)	0

Tabla 46: Índice de dificultad ítems competencia lingüística

De acuerdo con estos resultados, dos ítems resultaron difíciles para los estudiantes, se trata de los ítems 5 y 16, que tratan, respectivamente, del uso del pretérito imperfecto de subjuntivo y del uso de conectores. En la pregunta 5 solo acertaron 5 estudiantes y en la 16 solo 11 de los 28 participantes.

Respecto de la discriminación de los ítems, en esta sección solo 1 ítem discriminó muy bien, 3 discriminaron bien, 1 regular y 12 no discriminaron.

Nivel discriminación	Nº de ítems
Muy buenos (>0,39)	1
Buenos (0,30-0,39)	3
Regulares (0,20-0,29)	1
Malos (0,00-0,19)	12
Pésimos (< -0,01)	3

Tabla 47: Índice de discriminación ítems competencia lingüística

Los 4 ítems que mejor discriminaron fueron el 19 que evalúa el uso de la preposición *en*, el ítem 1 que evalúa el uso del pretérito perfecto compuesto de indicativo, el 6 que evalúa el uso del presente de subjuntivo, y el 20 el uso de los demostrativos.

El hecho de que 12 ítems no hayan discriminado, es decir, tanto el grupo con mejores resultados como el grupo más bajo obtuvieran los mismos resultados, no significa que los estudiantes muestren un buen nivel de competencia en este ámbito. Así lo confirman los puntajes de las secciones de expresión oral y escrita, cuya rúbrica considera el criterio de corrección como uno de los aspectos para evaluar estas tareas. Frente a estos resultados es posible señalar que el tipo de ítem mediante el cual se está evaluando pudo influir en la selección de las respuestas. Alderson *et al.*, (1998) señalan que se ha demostrado que cuando los estudiantes responden pruebas del tipo cloze leen de manera distinta a la habitual, solo leen un poco del texto que está antes del espacio y no lo que sigue después de este espacio; como tampoco logran una comprensión a cabalidad del texto, porque solo leen las partes donde están los huecos y no todo el texto.

4.4.4.1. Análisis de los ítems de competencia lingüística

De acuerdo con los resultados de la validez de contenido, los ítems de la sección de competencia lingüística están midiendo el nivel B2 de EFA. Si se observa los resultados generales del grupo (ver tabla 32) en este prototipo de prueba, se aprecia que 18 estudiantes presentan un nivel de competencia B2 en EFA; no obstante, en la sección de competencia lingüística la mayoría de los estudiantes acertó en las respuestas, lo cual no es evidencia de que dominen los usos gramaticales planteados en la prueba. De hecho, si se compara estos resultados con los obtenidos en el criterio de corrección de la expresión escrita y oral se observa que los estudiantes cometen errores en el uso de las estructuras gramaticales.

Los resultados observados en esta sección de competencia lingüística evidencian que se tiene que avanzar hacia tareas e ítems más acordes con el enfoque comunicativo de la lengua para evaluar el uso real, en contexto, que hacen los estudiantes de las formas gramaticales (Rocha y Cuadros, 2008). Esta competencia de la lengua, tradicionalmente, se ha evaluado con ítems de selección múltiple, que dan pistas al estudiante, por lo que, sus respuestas, en ocasiones, muestran el uso de diversas estrategias aprendidas por sobre el conocimiento del idioma.

En las secciones de comprensión oral y escrita, comúnmente, se evalúa con preguntas de selección múltiple, sin embargo, la naturaleza de la competencia lingüística es distinta, ya que cada opción referida a un mismo tema gramatical facilita la selección de la respuesta por parte del estudiante. Por lo tanto, a partir de los resultados obtenidos y de lo señalado en este apartado, se propone que la evaluación, en cuanto al dominio de las formas gramaticales se considere en las secciones de expresión escrita y oral, que evalúan de manera directa el uso de la lengua en contexto.

Rea (1991) critica que la gramática sea la parte de la evaluación que menos ha evolucionado hacia el modelo comunicativo, ya que solo se puede observar el uso real de la lengua cuando el estudiante la utiliza para lograr metas comunicativas, por lo que proponer un nuevo modelo de evaluación de la gramática no obedece solo al ajuste con el enfoque comunicativo que prima en la enseñanza, sino con lograr mejores observaciones y evidencias del uso real que hacen los estudiantes de los aspectos formales de la lengua. De esta forma, se podrá evaluar de manera válida y confiable lo que realmente puede hacer el estudiante con los contenidos lingüísticos que maneja y no la memorización de estos.

4.4.5. Confiabilidad de las secciones de expresión e interacción oral y escrita

La confiabilidad de las tareas de expresión e interacción escrita y oral se realizó mediante el sistema de doble corrección, y en aquellos casos donde hubo más de 1 punto de diferencia entre los evaluadores intervino un tercer evaluador. En la tabla 48 se observa los promedios de los puntajes que otorgaron las evaluadoras acorde con cada uno de los criterios de la rúbrica analítica.

Sección IV: Criterios de corrección expresión e interacción escrita					
Sujetos	Adecuación	Coherencia	Corrección	Vocabulario	TOTAL
1	2	2	3	2	9
2	2	2	3	2	9
3	3	3	2	2	10
4	2,5	2,5	2	2	9
5	2	2	3	3	10
6	1	2	3	2	8
7	1,5	2	3	2	8,5
8	1	1,5	2	2	6,5
9	3	3	3	3	12
10	1	1	2	2	6
11	2,5	3	2	3	10,5
12	1	2	2	2	7
13	0,5	1,5	1	1,5	4,5
14	2	2	1	1	6
15	1	1	1	1	4
16	3,5	3	3	3	12,5
17	1	1,5	1,5	2	6
18	2	2	2	2	8
19	2	2	2,5	2	8,5
20	2	2	2	2	8
21	2,5	2	2	2	8,5
22	3	2	2,5	3	10,5
23	2	2,5	2	3	9,5
24	3	3	2,5	3	11,5
25	3	2	2,5	3	10,5
26	1	2	2	2	7
27	2	2	2	2	8
28	2,5	2	2	2	8,5
Promedio	2,0	2,1	2,2	2,2	8,5

Tabla 48: Puntajes de la expresión e interacción escrita

Como se observa en las tablas 48 y 49, se mantiene una tendencia a nivel de grupo en cuanto que son más bajos los resultados en la expresión e interacción escrita que en la oral, puesto que la mayoría de los sujetos obtiene en la tarea escrita un puntaje más bajo que en la tarea oral. Se aprecia una similitud en los promedios en los distintos criterios de la rúbrica. El puntaje más bajo es el de adecuación al género discursivo ensayo, aun cuando en la instrucción de la tarea se les entregaba el formato del texto que tenían que escribir. En

general, para los estudiantes les fue más fácil introducir el tema que habían seleccionado, pero luego les fue difícil establecer una postura clara sobre el tema, así como también referirse a las ventajas y desventajas, y elaborar una conclusión de lo escrito. En la figura 24 se presenta un ejemplo de texto escrito por el sujeto 13 de la muestra, de lengua materna alemán y del área de ingeniería.

Falta agua o disponibilidad?

Agua es el alimento más importante del vida. Sin agua no hubiera vida en la tierra y muchos investigadores piensan que este regla esta válido también para el universo completo. Pero en el mundo industrial agua no solamente permite a beberlo. Es una resursa muy importante para la industria de ropa, agricultura, metales o quimica y también en las casas utilizan agua para mucho más que beber. Al lado de un consumo creciendo otro problema es que por el cambio climatico los extremos de tiempo aumentan. Por eso hay menos lluvia en regiones que tienen normalmente poco lluvia y los desiertos crecen.

Para mejorar la situación se tienen que seperar de regiones secas y no humidas. En las regiones humidas muchas veces no la cantidad de agua es el problema sino la calidad dado que industrias de ropa y quimica y el agua de los baños de los habitantes dejan en los rios (en los paises muy pobres sin filtración). Para mejorar la calidad de agua es necesario, utilizar filtros para limpiar el agua de tóxicos y bacterias. Es posible utilizar filtros biologicos que comen las substancias cuales son tóxicos para humanos o artificiales cuales utilizan un principio físico, quimico y biológico (enzymos).

En las regiones secas nescestiamos una transporte de agua, para mejorar la situaion dado que el cambio climático no se va.

Figura 24: Ejemplo de texto escrito por el sujeto 13 de la muestra

En este texto se observa que el estudiante no tuvo en cuenta el formato del ensayo solicitado. Acorde con los criterios de la rúbrica, este texto fue calificado por las evaluadoras de la siguiente forma:

1. Adecuación al género ensayo: El texto no cumple con la estructura del ensayo. Solo escribe el título y presenta el tema, luego propone soluciones aunque esto no se solicitó; no escribe cuál es su punto de vista sobre el tema. Así como tampoco se refiere de manera clara a las ventajas y desventajas del uso industrial del agua. En consecuencia se calificó con 0,5 puntos de un máximo de 4 que tiene la rúbrica para este criterio.

2. Coherencia: El texto plantea ideas poco claras, desconectadas. Se debe hacer varias relecturas para comprender el sentido de lo que se intenta comunicar, por ejemplo, en *“Al lado del consumo creciendo...”*. Se calificó con 1,5 puntos.
3. Corrección: El estudiante muestra un dominio básico en cuanto al uso de estructuras gramaticales. Comete errores de concordancia, conjugación de tiempos verbales, uso de preposiciones. Se calificó con 1 punto.
4. Vocabulario: Utiliza un conjunto de palabras sencillas que le permiten comunicar de manera limitada sus ideas. En ocasiones, comete errores de selección léxica *“...los extremos de tiempo aumentan”*, *“... el cambio climático no se va”*. Se calificó con 1,5 puntos.

En el anexo 8 se presenta una muestra más amplia de los textos escritos por la muestra de estudiantes (ver anexo 8).

En la tabla 49 se presenta los puntajes que obtuvo la muestra en la tarea de expresión e interacción oral.

Sección V: Expresión e interacción oral					
Sujetos/criterios	Coherencia	Fluidez	Corrección	Vocabulario	TOTAL
1	2	2	2,5	3	9,5
2	4	3	3	3	13
3	3	3	2	3	11
4	1	2	1	2	6
5	3	3	3	3,5	12,5
6	3	3	3	3	12
7	3	3	2	3	11
8	3	1,5	3	3	10,5
9	3	3	3	3,5	12,5
10	1	1	0,5	1	3,5
11	2,5	2	2	2	8,5
12	2	2	2	2,5	8,5
13	3	3	2	2	10
14	2,5	2	2	3	9,5
15	2	2	2,5	1,5	8
16	3	3	3	3	12
17	1	2	1	1,5	5,5

18	2	2	2,5	2	8,5
19	2	2	2	2,5	8,5
20	2	2	2,5	3	9,5
21	2	2,5	2	3	9,5
22	2	2	3	3	10
23	3	2	2	3	10
24	2	3	2	2,5	9,5
25	3	3	2	3	11
26	2	2,5	2	3	9,5
27	2	3	2	3	10
28	3	3	2,5	2	10,5
Promedio	2,4	2,4	2,2	2,6	9,6

Tabla 49: Puntajes de la expresión e interacción oral

La tabla 49 muestra que en la tarea de expresión e interacción oral el criterio mejor logrado fue el dominio de vocabulario con 2,6 puntos de un total de 4. El resultado más bajo fue el de corrección con 2,2 puntos. En la primera parte de la entrevista, los estudiantes se presentan y hablan de sus estudios, esta parte no resultó problemática, solo con algunos errores de precisión gramatical y léxica. Sin embargo, en la segunda parte tienen que seleccionar uno de los temas que se presenta (aprendizaje de una segunda lengua a través de un proceso de inmersión o el uso de pesticidas en la agricultura) y elaborar una exposición donde deben presentar el tema y su relevancia, dar a conocer su punto de vista sobre el tema, indicar las ventajas y desventajas sobre el tema y concluir. Esta segunda parte evidenció varias dificultades, por ejemplo, los estudiantes hacían pausas largas, titubeaban, cometían errores de imprecisión léxica y gramatical, poca cohesión en las ideas. Hay que señalar que realizar una exposición oral resulta también una tarea compleja en la lengua materna, por lo tanto, se necesita preparar a los estudiantes de intercambio respecto de cómo deben abordar una actividad de esta naturaleza en el contexto donde realizan su intercambio.

Para abordar la segunda parte de esta entrevista, se le entregó a cada estudiante una hoja con las dos opciones de temas y con la estructura de la exposición que debía elaborar, lo que fue muy útil porque guió a los alumnos en cuanto a la organización de su discurso. A modo de ejemplo, se sugiere revisar el anexo 9 con una muestra de entrevista calificada.

El gráfico 2 muestra las leves diferencias en cuanto al promedio de los sujetos en los resultados de ambas habilidades lingüísticas. Se trata de una diferencia de 1,73 puntos.

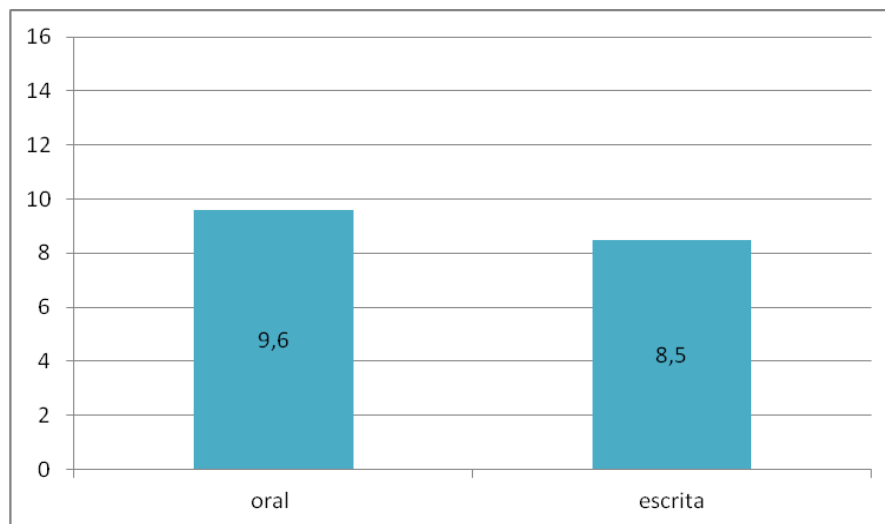


Gráfico 2: Resultados de la muestra en las secciones de interacción oral y escrita

Acorde con lo que señala Pastor (2006) no basta con que el estudiante de intercambio maneje el discurso académico en su lengua materna, es necesario también que aprenda las peculiaridades del discurso en la lengua meta, atendiendo a cuestiones relacionadas con el uso de formas impersonales, el grado de concreción de las ideas, el manejo de un léxico adecuado, etc. (Vázquez, 2010). Al respecto, en el proceso de corrección de los audios se percibió que fue difícil para los estudiantes mantener un registro formal de la lengua, incluso en los textos escritos usaron palabras o frases coloquiales. Pese a que se comunican de manera fluida en la lengua en cuanto a asuntos cotidianos, dicha fluidez en el prototipo de la prueba EFA disminuyó porque la tarea exige la producción de un género académico determinado, como es la exposición, por lo tanto, el estudiante debe abordarla de una manera distinta a como resuelve sus actividades diarias.

4.4.6. Análisis de las distintas dimensiones de la prueba

Como se ha mencionado anteriormente, el prototipo de la prueba EFA en línea nivel B2 evalúa seis dimensiones de la competencia comunicativa en el español con fines específicos académicos. De estas seis dimensiones, cuatro se miden a través de ítems de corrección objetiva. Las dimensiones 1, 2 y 3 corresponden a las secciones de comprensión lectora y

auditiva; y la 4 evalúa la competencia lingüística de los estudiantes. El gráfico 3, elaborado a partir del promedio logrado por el grupo en cada dimensión, muestra el nivel de rendimiento en cada una.

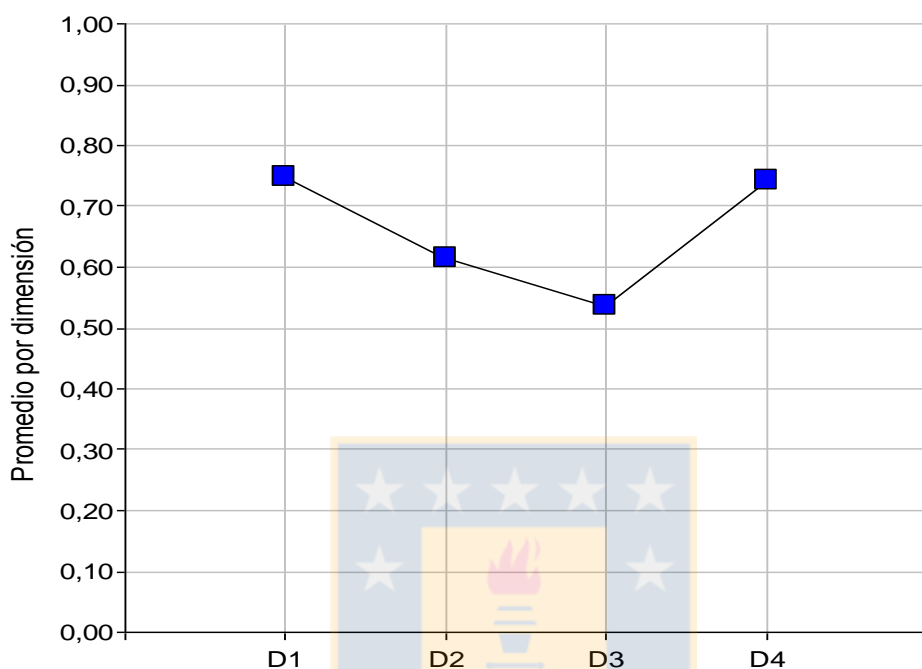


Gráfico 3: Comportamiento dimensiones de corrección objetiva del prototipo de prueba EFA

Este gráfico muestra la distribución del nivel de complejidad en las diferentes dimensiones del prototipo de la prueba. La dimensión 1 de la sección de comprensión lectora y auditiva, que corresponde a las preguntas 1 y 2 de cada texto /audio que miden información específica o literal se presenta con mejores resultados entre los estudiantes, lo cual evidencia que el nivel de complejidad es menor, dado que se trata de buscar o seleccionar las ideas que aparecen de manera explícita en los textos/audios. Sin embargo, los resultados de la dimensión 2 que corresponde a las preguntas 3 y 4 de los textos/audios que miden la capacidad de hacer inferencias a nivel de párrafo, se presenta con resultados más bajos que la dimensión 1.

Los resultados de la dimensión 3 de las secciones de comprensión lectora y auditiva, medidas con las preguntas 5 de cada texto/audio, son los más bajos dentro de las tres dimensiones comprensivas. Dicho comportamiento era esperable, puesto que las

dimensiones inferenciales a nivel de la idea global y de las ideas centrales a nivel párrafo se consideran más complejas en relación con la información literal (Martínez, 2011). En consecuencia, los resultados confirman las propuestas teóricas y de diseño que sustentan este prototipo de prueba en relación con los niveles de complejidad de las dimensiones comprensivas, puesto que los sujetos presentan mayores logros en la dimensión más fácil y menos en las más difíciles, y esta disminución es de manera gradual, puesto que la D1 mide la identificación de información específica en los textos, la D2 evalúa la idea principal a nivel de párrafo y la D3 la inferencia de la idea global de los textos, por lo que se corrobora la validez de contenido del prototipo de prueba, dado que se observa que los ítems están midiendo lo que dicen medir (las dimensiones que dicen medir).

La D4 que evalúa la competencia lingüística en un nivel B2 de EFA se presenta con mejores resultados, esto porque la mayoría de los estudiantes acertó a los ítems de esta sección, lo que no implica, como se ha mostrado, que los estudiantes tengan un buen dominio a nivel gramatical.

El gráfico 4 muestra el comportamiento de todas las dimensiones de este prototipo de prueba, incluyendo las dimensiones 5 y 6 que corresponden a las secciones de expresión e interacción escrita y oral. Acorde con esto, se observa que la comprensión lectora y la competencia lingüística fueron las que presentaron un mayor rendimiento debido a los factores ya explicados en el apartado 4.4.4.1. Por el contrario, el menor logro se ubicó en la dimensión 5 de expresión e interacción escrita con un promedio de 0,53.

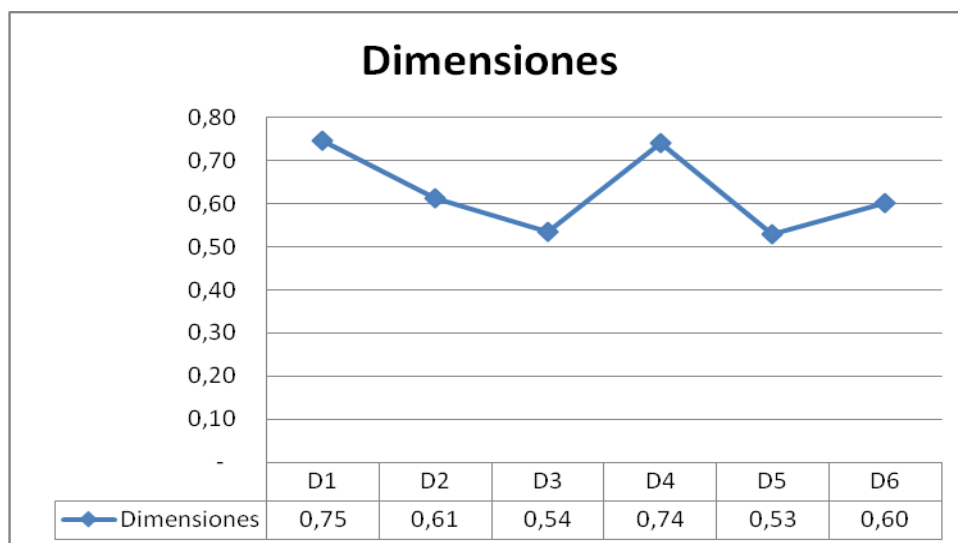


Gráfico 4: Dimensiones del prototipo de la prueba EFA

El gráfico 5 muestra de manera detallada el promedio de los resultados en los distintos criterios utilizados en la corrección de la tarea de expresión e interacción escrita.

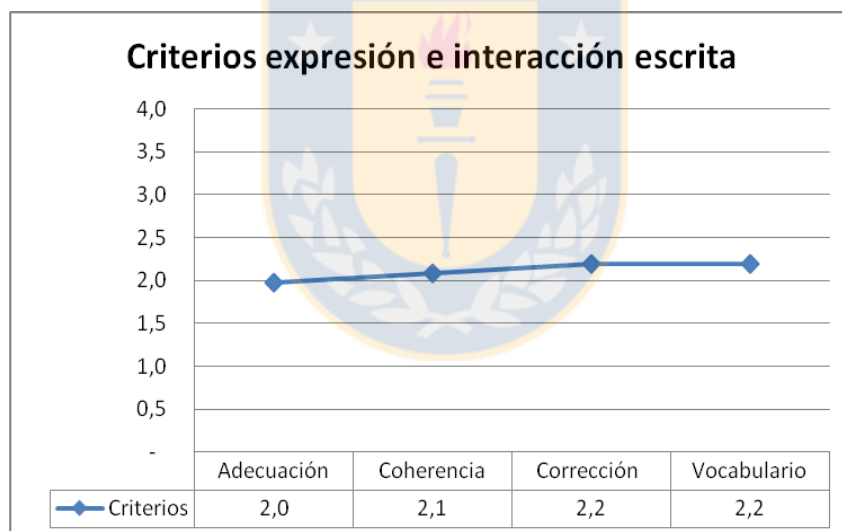


Gráfico 5: Criterios de corrección en la expresión e interacción escrita

El puntaje total de esta sección se basa en cuatro criterios de 4 puntos cada uno. De acuerdo con el gráfico, se observa que el criterio menos logrado fue el de adecuación al género discursivo ensayo con un promedio de 2 puntos, pese a que se entregó el formato del texto esperado. La coherencia presentó un promedio de 2,1 puntos y el dominio del vocabulario un promedio de 2,2 puntos, lo que indica que les fue difícil a los estudiantes expresar de

manera clara y precisa sus ideas. Finalmente, el criterio de corrección obtuvo un promedio de 2,2 puntos. Esto da cuenta que los estudiantes cometen errores gramaticales que afectan la comunicación, dado que no presentan una competencia adecuada para un nivel B2 de EFA.

El gráfico 6 presenta los resultados obtenidos por los estudiantes en la expresión e interacción oral.

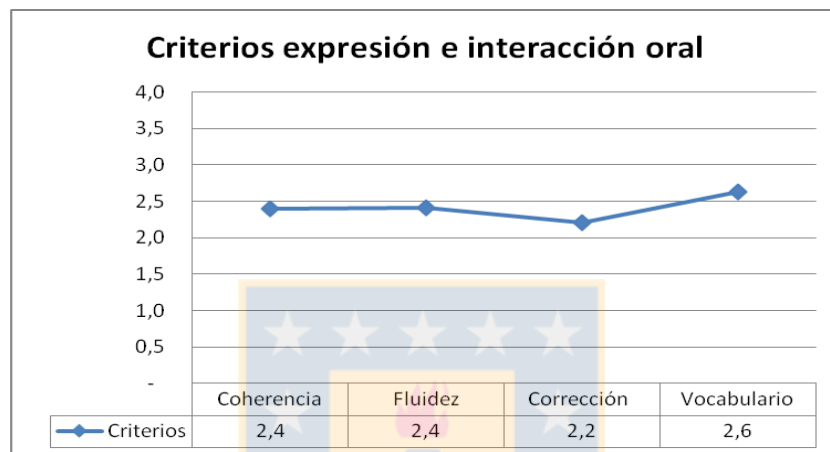


Gráfico 6: Criterios de corrección expresión e interacción oral

De acuerdo con el gráfico 6, el promedio más alto se observa en el dominio de vocabulario con 2,6 puntos, seguido de la coherencia y la fluidez con 2,4 puntos. La corrección obtuvo un promedio similar al de la tarea de escritura con 2,2 puntos, pese a que en la escritura tenían el tiempo suficiente para organizar el escrito y corregirlo, no lo hicieron. Esto evidencia que los estudiantes no muestran una precisión lingüística acorde con un nivel B2 de EFA, puesto que, independiente, del modo de expresarse (por medio escrito u oral) se mantienen las imprecisiones gramaticales que interfieren en la comprensión del texto.

La fluidez presenta un promedio de 2,4 puntos, un valor bajo si se considera que el máximo es de 4 puntos y los estudiantes son capaces de comunicarse en español. En este sentido, es interesante señalar que en la entrevista los estudiantes mostraron fluidez al inicio, donde tenían que presentarse, sin embargo, luego de seleccionar el tema sobre el cual realizarían la exposición comenzaron a titubear, a realizar pausas largas, a hacer selecciones de formas gramaticales y léxicas poco adecuadas, etc. lo que sin duda influyó en el puntaje asignado.

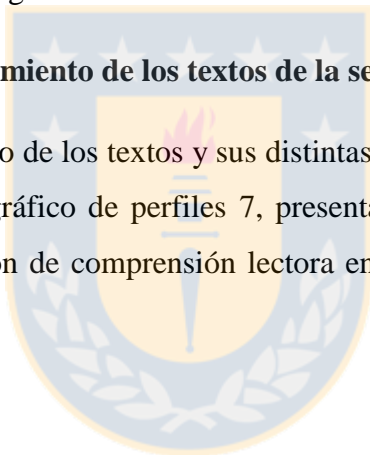
Esto indica que, si bien los estudiantes se comunican en sus actividades cotidianas les resulta complejo hacerlo en un contexto académico.

El dominio de vocabulario presenta 2,6 puntos de promedio, lo que implica que, a pesar de que les falta un mayor nivel de proficiencia, intentan seleccionar las palabras más adecuadas, acordes con el contexto temático.

En términos generales, si se analiza los gráficos 5 y 6 de expresión e interacción escrita y oral se observa que se mantuvo la tendencia en ambas habilidades y que en general, el grupo no sobrepasó los 2,6 puntos. Este promedio, nos indica que, en general, se requiere apoyar el proceso de inserción de los estudiantes extranjeros, de modo que su intercambio sea provechoso también en el plano académico, y esto solo se logrará si se expone al estudiante a este tipo de tareas antes de exigírselas como sabidas.

4.4.6.1. Análisis del comportamiento de los textos de la sección de comprensión lectora

Para conocer el comportamiento de los textos y sus distintas preguntas se elaboró en *Infostat* unos gráficos de perfiles. El gráfico de perfiles 7, presenta el comportamiento de los tres textos que componen la sección de comprensión lectora en función de las cinco preguntas que cada uno plantea.



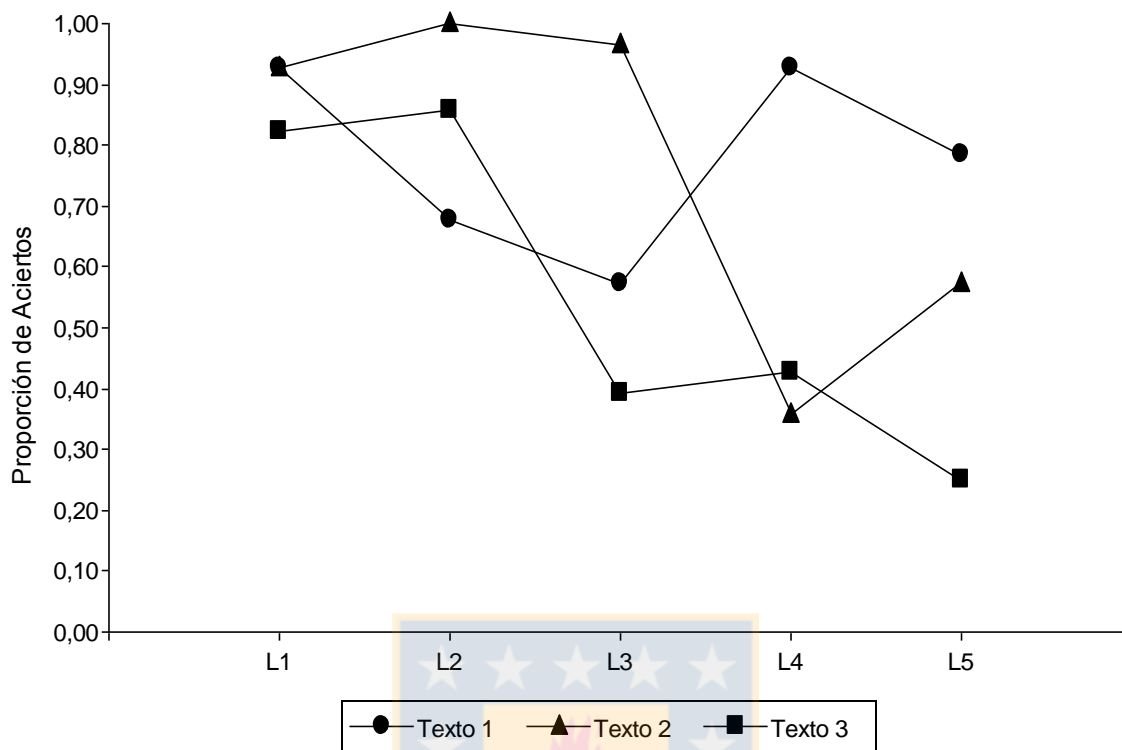


Gráfico 7: Comportamiento de los textos e ítems de comprensión lectora

Como se observa en el gráfico 7, el texto 1 referido al sistema de competencias en ingeniería muestra un comportamiento poco acorde con el planteamiento de la prueba, dado que lo esperable sería que las preguntas 3, 4 y 5 presenten resultados más bajos, dado que evalúan la capacidad de los estudiantes de captar la idea principal del párrafo y la idea global del texto, que son microhabilidades más complejas que identificar información literal. No obstante, dado el bajo número de la muestra es posible que al incrementarla, se modifique el comportamiento en la selección de las respuestas. Pese a esta situación se observa que la pregunta 3 presenta un nivel más bajo de respuestas que la 1 y la 2.

El texto 2 referido al aprendizaje de segundas lenguas muestra un comportamiento adecuado, dado que las preguntas que miden las dimensiones inferenciales presentan un rendimiento menor. Se observa que de la pregunta 3 a la 4 hay una gran diferencia, lo cual muestra el nivel de dificultad de esta pregunta, del mismo modo la pregunta 5 presenta un rendimiento bajo, pese a que se mantiene en términos generales por sobre la pregunta 4.

De acuerdo con el gráfico, el texto 3, referido a la biodiversidad, presenta el mejor comportamiento frente a los otros dos, puesto que descienden los resultados en la medida que se avanza en las preguntas. Así, las preguntas 1 y 2 resultaron más fáciles que las 3, 4 y 5, siendo esta última la que presenta un resultado más bajo, dado que dentro de las dimensiones de la prueba es la que representa una mayor complejidad para ser respondida.

En términos generales se observa que, a pesar de que todas las preguntas de los textos no muestran el comportamiento esperado, las preguntas que miden la dimensión 2 y 3 presentan resultados más bajos respecto de las preguntas de la dimensión 1. Esto indica que, efectivamente, estas preguntas están midiendo lo que se pretendía medir y que es necesario incrementar el número de sujetos para observar una mayor variabilidad en los resultados.

4.4.6.2. Análisis del comportamiento de los textos de la sección de comprensión auditiva

El gráfico 8 muestra el comportamiento que presentaron los tres audios y sus distintas preguntas.

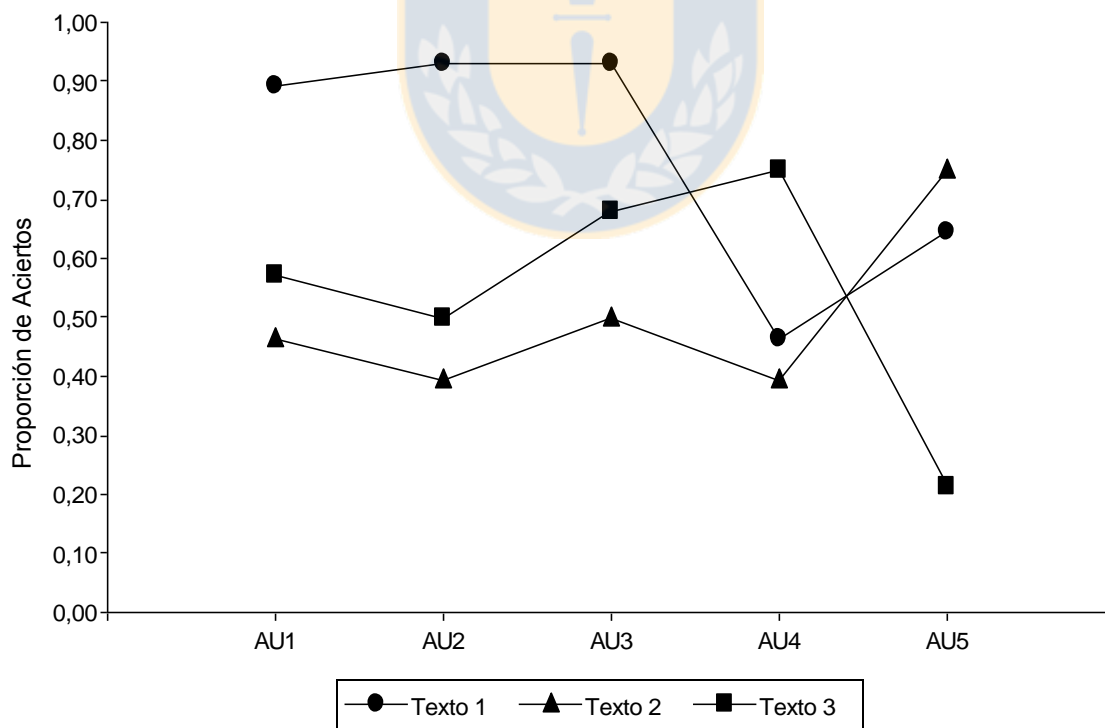


Gráfico 8: Comportamiento de los textos e ítems de comprensión auditiva

Se observa que el audio 1 del prototipo de la prueba EFA en línea B2 tiene un comportamiento adecuado, puesto que en las preguntas 4 y 5, que evalúan la inferencia de la idea central a nivel de párrafo y la idea global del texto, se obtuvo resultados más bajos respecto de las preguntas 1 y 2. Aun cuando la pregunta 3 de este audio presentó un comportamiento no diferenciador respecto de las primeras dos preguntas.

El audio 2 no presenta el comportamiento esperado, pues a medida que se avanza en las respuestas se obtiene mejores resultados. Sin duda, estas tendencias deberían ser corroboradas y de mantenerse habría que modificar las preguntas 2 y 5. Una situación similar ocurre con el audio 3 donde las preguntas 3 y 4 obtuvieron más aciertos que las 1 y 2. No obstante, en este caso la pregunta 5 muestra un resultado bajo lo que indica que fue difícil para el grupo resolver este ítem, dado que evalúa la idea principal del texto.

4.4.7. Resultados de la escala Likert aplicada a la muestra

Una vez que rindieron la prueba, se les aplicó un cuestionario de escala Likert y dos preguntas abiertas (ver anexo 10), con la finalidad de conocer la opinión de los estudiantes extranjeros frente al prototipo de prueba. En la tabla 50 se presenta estos resultados.

Preguntas	1(%)	2(%)	3(%)	4(%)	5(%)
1.Me sentí cómodo (a) respondiendo la prueba en computador			26	54	20
2. Me parece interesante que la prueba esté en computador			40	40	20
3. El ambiente de la prueba es fácil de usar			13	60	27
4. La prueba me pareció fácil de responder	7	13	46	27	7
5. Las instrucciones de la prueba son claras				53	47
6. Los textos y audios de la prueba son interesantes			20	67	13
7. Fue fácil comprender los textos de la prueba		20	53	13	13
8. Fue fácil comprender los audios de la prueba		7	27	53	13
9. Encontré fácil la sección de gramática		13	60	20	7
10. Me pareció fácil la tarea de escritura		13	60	27	
11. Me pareció fácil la tarea oral		13	20	60	7
12. El tiempo de duración de la prueba es adecuado	7	7	27	53	7

Tabla 50: Porcentajes de apreciación de la muestra sobre el prototipo de la prueba EFA B2

Como se observa en la tabla 50, los tres primeros ítems se refieren al entorno de la prueba, mientras que los siguientes hacen alusión de manera específica a las secciones y el formato del instrumento.

En relación con el primer ítem, un 54% declaró estar de acuerdo con haberse sentido cómodo rindiendo la prueba y un 20% señaló estar muy de acuerdo con este ítem. Esto se relaciona con los porcentajes obtenidos en el segundo ítem donde un 40% indicó estar de acuerdo con la afirmación de que es interesante que esta prueba esté en línea y un 20% se manifestó muy de acuerdo al respecto. No obstante, hay un 40% que se muestra indiferente frente a la modalidad de la prueba, esto se debe a que los estudiantes al estar en un proceso de inmersión valoran todo aquello que sea presencial, de interacción con el profesor, lo cual no quita el valor añadido ni las ventajas de las pruebas en líneas en el ámbito del EFA. Un 60% indicó estar de acuerdo en que el ambiente de la prueba es fácil de usar y un 27% señaló estar muy de acuerdo, esto se explica por la continua interacción de los estudiantes con plataformas sociales o de enseñanza, puesto que el uso del computador forma parte de sus actividades cotidianas académicas y no académicas.

La pregunta 4 donde se consulta por el grado de dificultad de la prueba, muestra una dispersión en la apreciación de los estudiantes, puesto que hay estudiantes que consideran que la prueba fue muy difícil (7%), mientras otros piensan que estuvo fácil (27%) y muy fácil (7%), estos datos se relacionan con el nivel de español de los estudiantes, puesto que para los que tenían un mayor dominio del español la prueba no fue tan difícil, sin embargo, para otros fue compleja (un 20% la consideró difícil).

Un aspecto importante en todo instrumento evaluativo es la claridad en las instrucciones (Alderson *et al.*, 1998; Martínez, 2011). De acuerdo con los resultados del ítem 5, los estudiantes dieron una muy buena calificación a este componente de la prueba, lo que indica que cada uno comprendió lo que debía hacer en cada sección y tarea y no tuvo inconvenientes para realizarlo.

En relación con el interés de los estudiantes en los temas que tratan los textos orales y escritos, la mayoría considera interesantes las temáticas que estos tratan, con un 67% de acuerdo y un 13% muy de acuerdo.

Desde las preguntas 8 a la 11 se busca conocer qué secciones del prototipo de prueba les parecieron fáciles o difíciles a los estudiantes. En términos generales se observa que las

secciones más fáciles para los estudiantes fueron la expresión e interacción oral y la comprensión auditiva con un 60% y un 53% de acuerdo, respectivamente. No obstante, los resultados obtenidos en la habilidad oral muestran que esta fue una de las partes más difíciles de la prueba, pues aunque los estudiantes hablan, les es difícil abordar una tarea del ámbito académico donde tienen que mantener un discurso claro, fluido, coherente y con un registro formal del español.

En cuanto a la sección más difícil de la prueba, un 20% señaló que fue la comprensión lectora, mientras un 53% marcó la opción indiferente. Un 60% indicó sentirse indiferente frente a la pregunta respecto de si la tarea de escritura había sido fácil y un 13% la consideró difícil, esta percepción se relaciona con los resultados obtenidos en la prueba, puesto que la habilidad donde mostraron menos dominio fue, precisamente, en la expresión e interacción escrita.

En cuanto a la duración del prototipo de prueba, un 53% está de acuerdo con el tiempo total otorgado de 1 hora y 45 minutos.

Para abordar las dos preguntas abiertas de este cuestionario se codificó las respuestas de los estudiantes, con el objeto de conocer en mayor detalle qué parte de la prueba le había parecido más fácil o difícil y por qué.

Preguntas	Categorías	Frecuencia de mención
1. ¿Qué parte de la prueba te pareció más fácil? ¿Por qué?	La sección de gramática porque cuando se aprende una lengua se hace este tipo de ejercicios	13
	La sección de comprensión lectora porque las palabras difíciles se entendían por contexto	8
	La sección de comprensión auditiva porque no tenía pronunciaciones difíciles	4
	La sección de escritura porque en la clase de español hemos escrito textos	3
2. ¿Qué parte de la prueba te pareció más difícil? ¿Por qué?	La sección de escritura porque es difícil escribir y se necesita de mucho vocabulario	15
	La sección de comprensión lectora porque los textos eran extensos	3
	La sección de comprensión lectora porque desconocía el vocabulario	3
	La sección de comprensión auditiva	3
	La sección de gramática porque son difíciles las conjugaciones verbales	2

	Todas las partes tenían su dificultad, por lo que no era posible saber todo	2
--	-----------------------------------------------------------------------------	---

Tabla 51: Codificación de las respuestas abiertas

Los resultados evidencian la información obtenida en el análisis del prototipo de la prueba EFA en línea de nivel B2. La idea que tuvo mayor mención dentro de la primera pregunta fue que la parte más fácil de la prueba había sido la competencia lingüística porque el tipo de ejercicio le resultó familiar, puesto que se trata de ejercicios que se realizan cuando se aprende un segundo idioma. Estas opiniones se relacionan con los datos de la prueba, puesto que esta sección fue la más fácil para todo el grupo.

En relación con la parte de la prueba que les había parecido más difícil la idea más mencionada fue la escritura del género académico ensayo, porque escribir es una habilidad compleja que requiere del dominio de otras competencias como la léxica, la gramatical, la discursiva, etc.

4.4.8. Análisis de los resultados de la muestra por área de estudio

En el ámbito de la evaluación de una segunda lengua con fines específicos académicos existe una controversia respecto de la influencia que pueden tener los temas que tratan los textos y tareas en los resultados que obtienen los estudiantes según sus áreas de estudio (Mendoza, 2015). Acorde con esto, se analizó los resultados obtenidos por la muestra en las secciones habilidades de comprensión lectora y auditiva en función de sus áreas de estudio. Hay que recordar que la distribución por área es la siguiente: Ingeniería 12 estudiantes, Letras 7, Ciencias Políticas 3, Administración y Economía 2, Ciencias Químicas 2, Ciencias Biológicas 1 y Ciencias de la Salud 1. Por lo tanto, los dos grupos mayoritarios son ingeniería y letras.

La tabla 52 muestra los resultados generales de los sujetos por sus áreas de estudio en la habilidad de comprensión lectora.

Área	Sección I: Comprensión lectora															PROMEDIO
	Texto 1 (V/F)					Texto 2					Texto 3					
	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5	
Ingeniería	91,67	50	50	100	66,67	91,67	100	100	16,67	50	83,33	83,33	25	50	16,67	65,0
Letras	85,71	85,71	57,14	85,71	71,43	85,71	100	100	42,86	42,86	57,14	71,43	57,14	42,86	14,29	66,7
Cs Políticas	100	100	66,67	100	100	100	100	66,67	33,33	66,67	100	100	33,33	33,33	0	73,3
Economía	100	50	100	100	100	100	100	100	100	100	100	100	50	50	100	90,0
Cs Químicas	100	50	50	50	100	100	100	100	50	100	100	100	50	50	50	76,7
Cs Biológicas	100	100	0	100	100	100	100	100	100	0	100	100	100	0	100	80,0
Cs Salud	100	100	100	100	100	100	100	100	0	100	100	100	0	0	0	73,3

Tabla 52: Resultados de la comprensión lectora por área de estudio

Se puede observar en las tablas 52 y 53 que entre los dos grupos mayoritarios de la muestra no hubo diferencias significativas en los promedios obtenidos en la habilidad de comprensión lectora. En los demás grupos se presenta diferencias, no obstante, dado el bajo número de sujetos (2, 3 y 1 estudiante) que integran esas áreas no es conveniente extraer conclusiones apresuradas.

En un análisis pormenorizado del promedio obtenido por cada grupo en los distintos textos, se observa que el comportamiento de los dos grupos mayoritarios fue similar. Esto se corroboró mediante el análisis de la prueba T- Student que arrojó un valor $p=0,29$, lo que indica que no hay diferencias significativas entre los estudiantes del área de ingeniería y de letras.

Área	T1	T2	T3
Ingeniería	71,7	71,7	51,7
Letras	77,1	74,3	48,6
Cs Políticas	93,3	73,3	53,3
Economía	90	100	80
Cs Químicas	70	90	70
Cs Biológicas	80	80	80
Cs Salud	100	80	40

Tabla 53: Resultados de cada texto por áreas de estudio

Los datos encontrados son relevantes, dado que muestran que, efectivamente, en esta sección se está midiendo habilidades comprensivas y no dominio disciplinar de los

diferentes temas tratados en los textos, que es uno de los cuestionamientos que se le hace a la evaluación del EFA.

Del mismo modo, en la tabla 54 se presenta los resultados generales obtenidos por las diferentes áreas en la habilidad de comprensión auditiva.

Área	Sección II: Comprensión auditiva															PROMEDIO
	Audio 1 (V/F)					Audio 2					Audio 3					
	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5	
Ingeniería	91,7	91,7	100,0	58,3	50,0	58,3	50,0	50,0	41,7	75,0	75,0	41,7	75,0	66,7	33,3	63,9
Letras	100	100	85,71	42,86	71,43	42,86	28,57	14,29	57,14	57,14	28,57	42,86	57,14	85,71	0	54,3
Cs Políticas	100	66,67	100	66,67	66,67	33,33	33,33	66,67	0	100	33,33	66,67	33,33	66,67	0	55,6
Economía	50	100	100	50	100	50	50	100	50	100	50	50	100	100	50	73,3
Cs Químicas	100	100	100	0	100	50	0	100	0	50	100	50	100	100	0	63,3
Cs Biológicas	0	100	0	0	100	0	100	0	0	100	100	100	0	100	0	46,7
Cs Salud	100	100	100	0	0	0	0	100	100	100	0	100	100	0	100	60,0

Tabla 54: Resultados de la comprensión auditiva por área de estudio

Se observa en la tabla 54 que hubo una leve diferencia entre los dos grupos mayoritarios. Además, se aprecia que los estudiantes de las diferentes áreas tuvieron promedios similares en esta sección.

La tabla 55 muestra los promedios obtenidos por cada área en los distintos audios. Se observa que el audio 1 resultó más fácil para toda la muestra, seguido del audio 2.

Área	A1	A2	A3
Ingeniería	78,3	55,0	58,3
Letras	80,0	40,0	42,9
Cs Políticas	80,0	46,7	40,0
Economía	80	70	70
Cs Químicas	80	40	70
Cs Biológicas	40	40	60
Cs Salud	60	60	60

Tabla 55: Resultados de cada audio por áreas de estudio

Con el objetivo de conocer si la diferencia entre los estudiantes de ingeniería y letras era significativa, se aplicó la prueba T-Student que arrojó un valor $p= 0.12$. Esto indica que las diferencias entre estas dos áreas no son significativas.

Los resultados obtenidos en este análisis evidencian que se hace necesario evaluar el español en un contexto académico, puesto que lo que se está midiendo son las habilidades que los estudiantes deben mostrar en este ámbito y no el dominio sobre un tema de su especialidad. En efecto, estos resultados apoyan la idea de Lee (2008), en tanto que los temas abordados en esta prueba en la comprensión lectora y auditiva se pueden concebir dentro de su especificidad como temáticas generales, comprensibles para todos los estudiantes, independiente de sus áreas de estudio.



CAPÍTULO IV: CONCLUSIONES Y PROYECCIONES

5. Conclusiones

Este trabajo de tesis se enmarcó en la interdisciplina de la Lingüística Aplicada, en los ámbitos referidos a la adquisición, enseñanza y evaluación de segundas lenguas o lenguas extranjeras. En dichos ámbitos resulta complejo determinar con certeza los niveles de adquisición y dominio que presenta un hablante no nativo en la lengua que aprende, por lo que se requiere de propuestas investigativas desde la evaluación que permitan determinar de manera precisa las competencias lingüísticas que presenta el estudiante.

El ámbito de la evaluación ha tenido un fuerte impulso desde la adquisición y enseñanza de segundas lenguas en cuanto a los lineamientos que se debe considerar para evaluar las distintas habilidades y competencias lingüísticas, según los distintos niveles de proficiencia (MCER, 2002). Específicamente, desde la adquisición y enseñanza se sugiere qué input presentar a los estudiantes según el nivel de lengua, cómo abordar las habilidades comprensivas y productivas de la lengua (Figueras y Puig, 2013), por lo que la evaluación debe contribuir mediante la definición de criterios que, desde una perspectiva evaluativa empírica den cuenta cómo medir de manera precisa el nivel de competencia lingüística que presentan los estudiantes en determinados contextos. En este sentido, se ha puesto énfasis en los requisitos de validez y de confiabilidad que debe mostrar todo instrumento que mida competencias en una lengua extranjera; y en épocas más recientes se ha manifestado un interés por investigar en qué medida las nuevas tecnologías contribuyen a agilizar el proceso de evaluación. En este contexto, la evaluación del Español como Lengua Extranjera presenta avances notables que se reflejan en los distintos instrumentos que evalúan proficiencia en ELE en un ámbito general.

Sin embargo, en la actualidad, la evaluación del ELE exige nuevas propuestas de instrumentos acordes con las nuevas realidades que presentan los intercambios académicos, profesionales y laborales, por lo que se necesita evaluar los niveles de proficiencia del ELE en estos ámbitos. Así, en el ámbito del Español como Lengua Extranjera con Fines Específicos Académicos (EFA) está el EXELEAA y el eLADE, que como se señaló en el

marco teórico, se encuentran en etapas de pilotaje, y de los cuales se ha publicado algunos aspectos en un plano descriptivo. El eLADE, recientemente, ha puesto a disposición del público un documento donde describe el análisis estadístico que han llevado a cabo para validar el instrumento, pero aún no explicitan qué criterios se consideraron en la selección de los materiales y en la definición de las dimensiones que evalúan los ítems de la prueba. Además, esta prueba se orienta a los fines académicos y profesionales y no queda claro cómo han hecho la delimitación entre lo académico y lo profesional en dicho instrumento.

Hasta ahora, lo que existe a nivel de la evaluación del EFA no satisface las problemáticas del ámbito, puesto que no queda claro cuáles han sido los criterios considerados para elaborar y validar estas pruebas. En este sentido, se requiere de propuestas investigativas, desde la Lingüística Aplicada, que sugieran criterios y directrices que permitan, en un futuro, evaluar con validez y precisión los niveles de proficiencia de EFA que presentan los estudiantes extranjeros de intercambio.

Teniendo en cuenta las problemáticas que presenta la evaluación del español con fines específicos académicos (EFA) y que, actualmente, no hay pruebas de EFA que tengan un reconocimiento nacional o internacional (Mendoza, 2015), la principal contribución de este trabajo de tesis, al ámbito de la Lingüística Aplicada, es determinar con mayor precisión el nivel de competencia lingüística y comunicativa del EFA que presentan los estudiantes de intercambio. Dicha contribución se sustenta en el modelo de criterios evaluativos que ha permitido el diseño de un prototipo de prueba de EFA en línea para un nivel B2 con validez de apariencia y de contenido.

La primera pregunta de investigación de este trabajo, relacionada con los criterios que se necesita para diseñar un prototipo de prueba de EFA en línea para un nivel B2 válido y confiable, se ha respondido a partir de las propuestas de los especialistas y de las sociedades internacionales de evaluación de lenguas en lo que respecta a la elaboración de instrumentos válidos para medir proficiencia en EFA. Acorde con esto, se elaboró el modelo con los cuatro criterios necesarios que han permitido luego diseñar, elaborar, implementar y evaluar el prototipo de la prueba EFA en línea nivel B2 en términos de la validez de apariencia y de contenido.

Un modelo de criterios para elaborar un prototipo de prueba de proficiencia en EFA debe contar con criterios éticos, de diseño, de elaboración y de validación y pilotaje. Destaca en este modelo la consideración de principios éticos, que han permitido el diseño de una prueba equilibrada y justa para los estudiantes. Se trata de un tema poco abordado en la elaboración de pruebas de segundas lenguas (Bordón, 2008), pero que cada vez cobra mayor interés en las prácticas investigativas internacionales.

La segunda pregunta de investigación, referida a cómo aplicar los criterios del modelo en la implementación de un prototipo de prueba de proficiencia en línea, de nivel B2 de EFA, se respondió a partir de la operacionalización de cada uno de los criterios mencionados en el modelo. Así, los criterios éticos se consideraron en la selección de los materiales y temas tratados. Los criterios de diseño se operacionalizaron a partir del propósito de la prueba, el perfil de los estudiantes a los cuales se orientó el prototipo de prueba y el formato de implementación en línea.

El aporte de los criterios de elaboración en este modelo se refiere a que se debe evaluar y validar los materiales antes de decidir si pueden formar parte de una prueba. Una forma de hacer esto es mediante un juicio de expertos, quienes evalúan de manera más objetiva e imparcial, porque no están involucrados en la elaboración del instrumento. La consideración de la validación de los materiales fue relevante para esta investigación, ya que se pudo hacer uso de los textos orales y escritos, que presentaron un alto porcentaje de acuerdo entre los evaluadores.

En cuanto a los criterios de validación y pilotaje es interesante señalar que el modelo propone contar con distintos tipos de validación del instrumento. Esto, sin duda, permite tener certeza de lo que, efectivamente, el prototipo de prueba evalúa. En este sentido, los resultados de la validez de apariencia fueron de gran relevancia, puesto que aportaron datos cualitativos y cuantitativos que permitieron conocer el primer comportamiento de este prototipo de prueba y mejorarlo. Esta validez se suele obviar en la mayoría de los análisis de pruebas de segundas lenguas por considerarse poco objetiva (Alderson *et al.*, 1998), sin embargo, en esta investigación aportó información valiosa que permitió reformular algunos

ítems y tareas de la prueba que presentaban una redacción ambigua, por lo que se sugiere considerar la participación de hablantes nativos en pilotajes de futuras pruebas de EFA.

Los resultados de la validez de contenido mostraron un alto grado de acuerdo entre los evaluadores, en las seis dimensiones de la prueba, dichos datos se encuentran avalados por el estadístico W de Kendall, cuyo índice señala que el grupo de expertos ha coincidido en que, efectivamente, los ítems están midiendo las dimensiones definidas para evaluar proficiencia en un nivel B2 de EFA. Del mismo modo, la validez del prototipo de prueba en línea en cuanto a los principios tecnológicos que la sustentan, muestra que se trata de un buen instrumento, puesto que evalúa el español en un contexto académico comunicativo y en un entorno seguro y de fácil uso.

En este sentido, la validación desde una perspectiva tecnológica del prototipo de prueba EFA es un aporte interesante de esta investigación, dado que este formato de aplicación permite evaluar de manera efectiva, rápida y con altos niveles de seguridad. En efecto, ha sido relevante poder validar este prototipo en línea desde una perspectiva evaluativa y tecnológica, por lo que se sugiere la consideración de pruebas de EFA automatizadas, que se implementen en ambientes de fácil acceso y navegación. Además, en el contexto del EFA es interesante proponer instrumentos en línea, que estén acordes con el perfil de los estudiantes a los cuales se orienta, quienes realizan gran parte de sus tareas académicas frente a un computador, por lo que rendir una prueba en línea les resulta una actividad habitual.

Los resultados obtenidos en la validación del prototipo de prueba EFA en línea, desde una perspectiva tecnológica reflejan que se cumplió con los principios metodológicos de CALT y que se trata de un prototipo de prueba válido tanto desde una perspectiva tecnológica como de contenido de los ítems.

La tercera pregunta de esta tesis tuvo como objetivo observar el comportamiento del prototipo de la prueba EFA en línea de nivel B2 en una aplicación piloto con una muestra de 28 estudiantes extranjeros de intercambio en la Universidad de Concepción. Los resultados obtenidos permitieron conocer los índices de dificultad y de discriminación de los ítems de corrección objetiva de la prueba. Si bien algunos de estos coeficientes están con valores levemente por debajo de lo indicado por los especialistas es importante señalar que este es un primer aporte, y que se necesita incrementar la muestra para observar una mayor

variabilidad en el comportamiento de los ítems. Hay que señalar que la validación de una prueba no es algo que finaliza cuando esta se ha administrado y evaluado, sino que se trata de un proceso que se realiza continuamente (Bordón y Liskin-Gasparro, 2007), y es lo que se hace en pruebas de alto impacto a nivel internacional como el TOEFL o los exámenes de Cambridge, donde ha habido mucho trabajo e inversión de parte de las instituciones patrocinantes para contar con instrumentos válidos y confiables. Y es hacia allá donde se necesita avanzar en las pruebas de proficiencia de EFA en línea.

En cuanto a las apreciaciones de los estudiantes sobre el prototipo de la prueba EFA se observó que el instrumento les pareció claro en sus instrucciones e interesante en cuanto a los temas tratados en los audios y textos. Aun cuando no se observa un interés marcado por el formato de aplicación de esta prueba. Esto se debe a que los alumnos, al estar en un proceso de inmersión, prefieren las actividades presenciales que requieren de una interacción con el profesor, lo cual no resta viabilidad ni interés a esta modalidad de la prueba.

Los resultados obtenidos, a través de la comparación de la prueba CELE y el prototipo de la prueba EFA, permiten ver que estos instrumentos están correlacionados, pues en ambos se evalúan las habilidades lingüísticas en el Español como Lengua Extranjera. Sin embargo, se trata de dos pruebas que tienen orientaciones y niveles de complejidad distintos, lo que fue avalado por el estadístico T-Student. Se observa, entonces, que el prototipo de prueba EFA se ha complementado muy bien con la prueba CELE. Dicho contraste ha permitido conocer de manera más precisa el nivel de competencia de los estudiantes, sus aciertos y sus dificultades.

En términos generales, se observa que cada criterio del modelo se ha operacionalizado con la finalidad de contribuir en un futuro a tener un instrumento que permita medir de manera válida y confiable el nivel de proficiencia B2 en el EFA. Se debe considerar que no hay pruebas de EFA reconocidas a nivel nacional o internacional y que existe la necesidad en el ámbito de contar con un instrumento que permita evaluar a los estudiantes extranjeros en cuanto a sus competencias lingüísticas en un contexto académico. Por lo tanto, esta investigación contribuye con datos cualitativos y cuantitativos que evidencian desde los

procesos iniciales de planeación de la prueba en línea hasta las etapas finales de elaboración y validación, cuestión que los expertos reclaman a las instituciones que actualmente ofrecen exámenes de EFA (Llorián, 2012). No obstante, hay que enfatizar que, el principal aporte de esta investigación está a nivel de los procedimientos que se requiere para diseñar y elaborar un instrumento de EFA que permita, en el futuro, evaluar a estudiantes que requieran precisar su dominio en el Español como Lengua Extranjera con fines específicos Académicos.

En el contexto de esta investigación, la prueba CELE se constituyó en un elemento clave, puesto que se consideró como una referencia en cuanto a la medida del dominio lingüístico que presentaban los estudiantes. Una evaluación en el español general proporciona información valiosa de los niveles de proficiencia de los estudiantes; no obstante, se hace necesario contar con pruebas que permitan evaluar las habilidades de los estudiantes extranjeros de intercambio en un contexto específico, académico. Del mismo modo, como se necesita saber cómo se desenvolverá un profesional en su entorno laboral específico, se requiere conocer cómo se desempeñará un estudiante universitario, no nativo, en cuanto a sus competencias lingüísticas. En el caso del español, se necesita investigar cuáles son las concepciones que traen los estudiantes extranjeros de su cultura de origen, en relación con los géneros académicos más representativos y utilizados en el contexto universitario, y cómo se conciben estos en la cultura meta. Como explica Vázquez (2004), uno de los principales problemas que enfrentan los estudiantes que realizan intercambios universitarios se refiere al desconocimiento que tienen respecto de las tradiciones académicas de las universidades donde se insertan, por ejemplo, cómo se estructuran las clases, las formas de evaluación, la elaboración de informes, trabajos, etc. Por lo tanto, solo se tendrá un conocimiento real y preciso del desempeño de los estudiantes evaluándolos en esos términos. De esta manera, los resultados serán útiles para planificar estrategias de apoyo a la inserción académica. En efecto, no solo se trata del dominio en la lengua, sino de las habilidades y los géneros implicados en el contexto académico. Por lo tanto, se requiere contar con una prueba específica de EFA que permita conocer de manera precisa las competencias lingüísticas de los estudiantes en contextos académicos.

En relación con la controversia respecto de lo que se está evaluando en una prueba de EFA, los resultados de este prototipo de prueba avalan que, efectivamente, se midió el nivel de proficiencia de español situado en un contexto académico y no contenidos disciplinares, ya que tanto los estudiantes de ingeniería como los de letras obtuvieron promedios similares. En este sentido, se propone abordar temáticas académicas más generales que interesen a distintos grupos de estudiantes, independiente de sus áreas de formación (Mendoza, 2015).

En el planteamiento de las tareas de escritura se discute en qué medida los textos que pueden producir los estudiantes extranjeros universitarios pueden ser representativos de su competencia real de español en un ámbito específico. En efecto, los resultados obtenidos en la aplicación de este prototipo de prueba señalan que los temas tratados parecen interesantes a la muestra, por lo que se sugiere que las pruebas de EFA aborden, por ejemplo, en las secciones productivas, temas generales dentro de un ámbito académico, que eliciten las modalidades expositivas y argumentativas, representativas de un contexto académico. Como señala Llorián (2012), las instituciones deben dar un mayor impulso a la evaluación del español en contextos académicos, de modo que sea un requisito para los estudiantes que realicen intercambios o estadías en universidades de habla hispana, el acreditar su dominio en la lengua.

Este trabajo ha servido para dar a conocer los procedimientos abordados para realizar los distintos tipos de validación y pilotajes del prototipo de prueba EFA en línea de nivel B2. Aunque el pilotaje se realizó con una muestra reducida de sujetos, fue de gran utilidad para comprobar que el modelo propuesto ha sido útil para guiar la compleja tarea de elaborar una prueba de EFA, a nivel de prototipo, que sea válida y confiable. Este prototipo de prueba representa una contribución a pequeña escala de lo que significa evaluar el EFA. Los datos cualitativos y cuantitativos serán fundamentales en la elaboración futura de un instrumento que permita precisar el nivel de EFA de los estudiantes que deseen realizar un intercambio en una universidad de habla hispana.

Limitaciones y proyecciones de la investigación

Teniendo en cuenta que son múltiples las dimensiones bajo las cuales se enmarca la evaluación del EFA, en los siguientes apartados se menciona las limitaciones y las proyecciones de este trabajo de tesis.

Una de las limitaciones de esta investigación se refiere al modelo de criterios evaluativos, dado que es difícil proporcionar un modelo de elaboración y validación de un prototipo de prueba de EFA que considere todas las variables y pueda ser replicado sin modificaciones, ya que son múltiples los escenarios y realidades educativas, el modelo puede ser mejorado o adaptado en función de las necesidades del usuario.

Otro aspecto a considerar corresponde a la estructura del prototipo de prueba de EFA, en cuanto a las habilidades y competencias que evalúa. Como se observó en el diseño del prototipo de prueba, se propuso cinco secciones, una de las cuales mide de manera específica la competencia gramatical del estudiante. Sin embargo, los resultados mostraron que el tipo de ítem propuesto (técnica cloze) para evaluar esta competencia resulta inadecuado, dado que no evidenció el dominio real de los estudiantes, porque las alternativas en este tipo de ítem elicitaban la respuesta correcta. Hay que recordar que los sujetos obtuvieron bajos resultados en las habilidades productivas de la lengua (tareas orales y escritas), donde uno de los criterios de revisión de las tareas fue el de corrección, por lo que se sugiere, que en una prueba de EFA se evalúe la competencia gramatical en contextos auténticos de uso del español.

En relación con lo anterior, otra limitación de este trabajo se refiere a la distribución de los porcentajes por cada sección. Tal como se presentó en el prototipo de prueba, se observa una mayor consideración de las habilidades escritas, puesto que a la comprensión lectora y expresión escrita se les asignó un 25% respectivamente. Sin duda, esta distribución se debe mejorar en una prueba de EFA, otorgando una consideración similar a las distintas habilidades de la lengua o bien dar un mayor énfasis a aquellas donde se haya observado mayores dificultades, por ejemplo, la comprensión auditiva. Esto porque los resultados mostraron que los estudiantes tienen mejor dominio de la comprensión lectora que de la comprensión auditiva; no obstante, se trata de una decisión que debe ser analizada.

Otra limitación de este trabajo puede observarse en la selección de los textos orales, donde los audios mejor evaluados por los expertos solo consideraban temáticas científicas. Si bien los resultados evidenciaron que no hubo diferencias significativas entre los diferentes grupos, sería interesante equilibrar esta sección en cuanto la variedad temática, de modo que los estudiantes no tengan la sensación de que se privilegia ciertas áreas del conocimiento por sobre otras, lo cual contribuye a diseñar un instrumento acorde con el perfil de los estudiantes.

Por otra parte, corresponde mencionar el número reducido de sujetos que participaron del pilotaje de este prototipo de prueba de EFA. Generalmente, en materia de evaluación, se espera trabajar con grupos más numerosos, sobre 100 estudiantes, en el contexto de la Universidad de Concepción es difícil acceder a grandes muestras de estudiantes de Español como Lengua Extranjera, pues la mayoría de estudiantes que realizan intercambios universitarios en esta institución son latinos. No obstante, los nuevos análisis estadísticos permiten validar ítems independiente de la muestra que se disponga, lo cual es una proyección interesante para continuar con el mejoramiento de este prototipo de prueba, de modo que en el futuro se pueda contar con una prueba de EFA que mida con exactitud las competencias de los estudiantes en el ámbito académico. Esto además, permitiría corroborar la validación realizada por expertos en cuanto a las dimensiones que evalúan los distintos ítems de un prototipo de prueba de EFA.

Finalmente, en lo referido a los contextos de uso del EFA, es posible mejorar la consideración de los géneros académicos abordados en este prototipo de prueba de EFA, los cuales fueron elegidos en función de lo señalado en el ámbito de la enseñanza del EFA y lo descrito en la literatura y no acorde con la realidad que tienen que vivir los estudiantes extranjeros de intercambio en la Universidad de Concepción.

En relación con las limitaciones señaladas se puede proyectar desarrollar distintos aspectos de esta investigación.

Una proyección de este trabajo consiste en definir el número de secciones que debería considerar una futura prueba de EFA, teniendo en cuenta los alcances señalados con respecto a no incluir un apartado que mida de manera aislada y automática el dominio gramatical de los estudiantes. Del mismo modo, se debe decidir los porcentajes que se asignará a las distintas habilidades lingüísticas en el ámbito académico, de modo que no se dé preponderancia a una por sobre la otra, para así no sobredimensionar el dominio del estudiante. En este sentido se proyecta continuar evaluando este prototipo de prueba de EFA con estudiantes extranjeros de intercambio, para ello se tendrá que realizar las modificaciones correspondientes, en lo que respecta al número de secciones y distribución de porcentajes, revisión de los ítems con bajos índices de discriminación y propuestas de textos orales con temas variados para evaluar la comprensión auditiva.

Otra proyección de este trabajo se refiere a plantear un análisis de los ítems del prototipo de prueba EFA utilizando la Teoría de Respuesta al Ítem (TRI) para así determinar el nivel de dificultad de los ítems con independencia de la muestra a la cual se ha aplicado. De esta manera, se podrá comparar el funcionamiento de los ítems con estudiantes de diferentes niveles de proficiencia. Para ello, se necesita aumentar considerablemente el número de ítems de la prueba, con el objeto de construir un banco de ítems que incluya preguntas que evalúen dimensiones más avanzadas en las habilidades comprensivas (análisis crítico, evaluación, reflexión).

También es interesante en un largo plazo avanzar en delimitar los contextos de uso del español con fines académicos, dado que no se dispone de datos concluyentes al respecto, pues lo que existe está en un plano didáctico de la enseñanza del EFA (Vázquez, 2006), pero no en términos evaluativos. Para ello, sería interesante conocer qué leen, escuchan, escriben y exponen los estudiantes extranjeros de intercambio en sus actividades académicas en la Universidad de Concepción.

BIBLIOGRAFÍA

Ainciburu, M. (2008). ¿De qué color era el caballo blanco de Bolívar?: La evaluación de la comprensión literal de textos en las pruebas DELE INICIAL. S. Pastor Cesteros, y S. Roca

Marín (coords.), La evaluación en el aprendizaje y la enseñanza del español como lengua extranjera/segunda lengua, 107-115.

Alexopoulou, A. (2010). Tipología textual y comprensión lectora en E/LE. *Revista Nebrija de Lingüística Aplicada* (9), pp. 1-7.

Alderson, Ch.; Clapham, C. y Wall, D. (1998). *Exámenes de idiomas*. Madrid: Cambridge University Press.

Backhoff, E., Larrazolo, N. y Rosas, M. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista Electrónica de Investigación Educativa*, 2 (1). Disponible en: <http://redie.uabc.mx/vol2no1/contenido-backhoff.html> [Fecha de consulta: julio de 2015].

Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman, L. (2000). Modern Language testing at the turn of the century: assuring that what count counts". *Language Testing*, 17(1), pp.1-42.

Bachman, L. (2002). Some reflections on task-based language performance assessment. *Language testing*, 19(4), 453-476.

Behiels, L. (2010). Estrategias para la comprensión auditiva. Monográficos Marcoeel, [Revista electrónica], 11. Disponible en:

<http://www.marcoele.com/descargas/navas/08.behiels.pdf> [Fecha de consulta: noviembre 20 de 2014].

Bordón, T. (1991). El enfoque comunicativo aplicado a la enseñanza de segundas lenguas. Técnicas de evaluación. Tesis doctoral. Madrid, España: Universidad Autónoma de Madrid.

Bordón, T. (2006). *La evaluación de la lengua en el marco de E/L2: Bases y procedimientos*. Madrid. Editorial: ArcoLibros.

Bordón, T y Liskin-Gasparro, J. (2007). Evaluación. En: M. Lacorte. (ed.), *Lingüística aplicada del español*. Madrid: Arco/Libros. pp. 211-252.

Bordón, T. (2008). Panorama histórico del desarrollo de algunas de las cuestiones fundamentales en la evaluación de segundas lenguas. *MarcoELE*, 7, 4-25.

Consejo de Europa. (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: Anaya.

Buck, G. (2009). *Assessing Listening*. Cambridge: Cambridge University Press.

Cabrero, C., y Llorente, L. (2013). La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información y comunicación (TIC). *Eduweb*, 7(2), 11-22.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied linguistics*, 1(1), 1-47.

Caplan, D. (1998). *El lenguaje. Estructura, procesamiento y trastornos*. Buenos Aires: Docencia.

Cervantes, E.P. (1989). *English Teaching Forum*, 27, 1: 12-13.

Chapelle, C & Douglas, D. (2006). *Assessing language through computer technology*. United Kingdom. Edit: Cambridge University Press.

Domínguez, I. (2010). Las inferencias en la comprensión lectora: una ventana hacia los procesos cognitivos en segundas lenguas. *Revista Nebrija de Lingüística aplicada a la enseñanza de Lenguas*, (7), 1-31.

Douglas, D. (2000). *Assessing Languages for Specific Purposes*. Cambridge: Cambridge University Press.

Dudley-Evans, T., & St John, M. J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge University Press.

Ebel, R.L. y Frisbie, D.A. (1991). *Essentials of Education Measurement*. Englewood Cliffs, NJ: Prentice Hall.

Escobar, J., y Cuervo, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en medición*, 6, 27-36.

Ferreira, A. (2010). *Aula Virtual para la Enseñanza de Español como Lengua Extranjera*. Universidad de Concepción.

Ferreira, A. (2010). CELE, Prueba de nivel de competencia en ELE, 2010. Universidad de Concepción.

Ferreira, A, Vine, A y Elejalde, J. (2013). Hacia una prueba de nivel en español como lengua extranjera. *RLA*, 51 (2), 73-103.

Figueras, N. y Puig, F. (2013). *Pautas para la evaluación del español como lengua extranjera*. Madrid: Edinumen.

Fortus, R., Coriat, R., Fund, S., Khalifa, H., Weir, C. J., & Kunnan, A. (2014). Safeguarding fairness principles through the test development process: A tale of two organisations. ISSN 1756-509X, 3.

Fulcher, G. (2003). Testing second language speaking. Pearson Education.

García, J. y Navarro, C. (2008). Automatizar los exámenes del Diploma de Español como Lengua Extranjera (DELE): ¿Utopía o realidad? *Revista de Lingüística Teórica y Aplicada* 46 (2), 81-93.

Hernández, R., Fernández, C., y P. Baptista.(2010). Metodología de la investigación. Mc Graw Hill: México.

Huhta, A. (2008). Asociaciones internacionales de evaluación y pruebas de lengua.¿ Qué son y qué hacen? *MarcoELE*, 7, 42-51.

Hyland, K. (2002). Specificity revisited: how far should we go now?. *English for Specific Purposes*, 21(4), 385-395.

Jerez, M., Robles, M. D., Serrano, G., Mauricio, A., Bujosa, M., Hiernaux, A. G., & Casals, J. (2012). Sistemas de evaluación objetiva a distancia en métodos cuantitativos: valoración de plataformas alternativas. *RELADA-Revista Electrónica de ADA-Madrid*, 6(3).

Jouini, K. (2005). Estrategias inferenciales en la comprensión lectora. *Glosas didácticas* (13), 95-114.

Lee, H. (2008). The relationship between writers' perceptions and their performance on a field specific writing test. *Assessing Writing* (13), 93-110.

López-Mezquita, M. T. (2005). La evaluación de la competencia léxica: test de vocabulario. Su fiabilidad y validez (Vol. 177). Ministerio de Educación.

Llorián, S. (2012). Los retos de la certificación del español con fines específicos. *Revista de Nebrija de Lingüística Aplicada*, 12 (6), 3-34.

Martín, S. (2010). La elaboración de textos y preguntas para evaluar la comprensión oral. En *Actas del XXI Congreso Internacional ASELE* (pp.557-569).

Martínez, A. (2011). *La evaluación de las lenguas: garantías y limitaciones*. Granada: Ediciones Mágina.

Mendoza, A. (2015). La selección de tareas de escritura en los exámenes de lengua extranjera destinados al ámbito académico. *Revista de Nebrija de Lingüística Aplicada*, 18 (9).

Noijons, J. (1994). "Testing computer assisted language testing: Towards a checklist for CALT", en *CALICO Journal*, 12(1), pp. 37-58.

Parodi, G. (2012). ¿Qué se lee en los estudios doctorales?: Estudio empírico basado en géneros a través del discurso académico de seis disciplinas, *RLA*. 50 (2), pp. 89-119.

Pastor, S. (2003). La evaluación del proceso de aprendizaje de Segundas Lenguas. *Perspectivas teóricas y metodológicas: lengua de acogida, educación intercultural y contextos inclusivos*, Madrid, Consejería de Educación, 503-513.

Pastor, S. (2006). La enseñanza del español como lengua vehicular en contextos académicos", *Marcoele*, nº 2, ISSN 1885-2211.

Pastor, S. (2010). Enseñanza de español con fines profesionales y académicos y aprendizaje por contenidos en contexto universitario, *Testi e Linguaggi*, 4, pp.71-88.

Pastor, S. (2014). Hacia una certificación oficial del español académico. II Congreso Internacional del SICELE. Avances en la certificación del ELE, México D.F.

Pérez, M. (2005). Evaluación de la comprensión lectora: dificultades y limitaciones. *Revista de educación*, (1), 121-138.

Pinilla, R. (2004). “La expresión oral” en SÁNCHEZ LOBATO, J. y SANTOS GARGALLO, I. (eds.). *Vademécum para la formación de profesores*. Madrid: SGEL.

Pollit, A. (1990). Standards. Notes prepared for a meeting to discuss language testing standards. University of Cambridge Local Examinations Syndicate, Cambridge.

Regueiro, M. y D. Sáez. (2013). *El español académico*. Guía práctica para la elaboración de textos académicos. Madrid. Editorial: ArcoLibros.

Rocha, C. & R. Cuadros. (2008). ¿Es posible evaluar objetivamente contenidos gramaticales en producciones escritas?. En *La evaluación en el aprendizaje y la enseñanza del español como lengua extranjera/segunda lengua: XVIII Congreso Internacional de la Asociación para la Enseñanza del Español como lengua Extranjera (ASELE)*: Alicante, 19-22 de septiembre de 2007 (pp. 179-184). Servicio de Publicaciones.

Rodríguez, T.(coord.).(2000). *La evaluación en el aula*. Barceloa, Ediciones Nobel.

Sabaj, O., & Ferrari, S. (2005). La comprensión de textos especializados en sujetos con formación profesional diferenciada. *RLA*, 43(2), 41-51.

Vázquez, G. (2001). *Guía didáctica del discurso académico escrito:¿ cómo se escribe una monografía?* Editorial Edinumen.

Vázquez, G. (2004). *La enseñanza del español con fines académicos. Vademécum para la formación de profesores. Enseñar español como segunda lengua y lengua extranjera*. Madrid: SGEL, 1129-1147.

Vázquez, G. (2010). ¿Por qué el discurso académico? un análisis de necesidades. *MarcoELE* (10), 213-221.

Yagiie, A. (2010). Examinando exámenes. Consideraciones acerca de las pruebas de medición en ELE. *MarcoELE*, 10,1-53.

Weigle, S.C.(2000). *Assessing Writing*. Cambridge: Cambridge University Press.

Widdowson, H. G. (2001). *Communicative Language Testing: The art of the possible*. *Experimenting with uncertainty: Essays in honour of Alan Davies*, 12-21.

Sitografía

ALTE. (2007). Estándares mínimos. Disponible en http://www.alte.org/attachments/files/minimum_standards_es.pdf. [Consulta: 24/08/2014].

ALTE. (2009). Código de Práctica. Disponible en http://www.alte.org/attachments/files/code_practice_es.pdf. [Consulta: 24/08/2014].

Centro de enseñanza para extranjeros. Certificado de Español como Lengua Adicional (CELA). Disponible en <http://132.248.130.236/cela/>. [Consulta:26/08/2014].

Centro de enseñanza para extranjeros. Examen de Posesión de la Lengua Española (EPL). Disponible en <http://132.248.130.236/eple/>. [Consulta:26/08/2014].

Centro de enseñanza para extranjeros. Examen de Español como Lengua Extranjera para el Ámbito Académico (EXELEAA). [Consulta: 26/08/2014].

CONSORCIO ELSE. Certificado de Español Lengua y Uso (CELU). Disponible en <http://www.celu.edu.ar/>. [Consulta:26/08/2014].

EALTA. (2006). Guidelines for Good Practice in Language Testing and Assessment. Disponible en <http://www.ealta.eu.org/guidelines.htm>. [Consulta: 24/08/2014].

ILTA. (2000). Code of Ethics. Disponible en <http://iltaonline.com/index.php/enUS/resources/ilta-code-of-ethics>. [Consulta: 24/08/2014].

INSTITUTO CERVANTES. Diplomas de Español como Lengua Extranjera (DELE). Disponible en <http://dele.cervantes.es/>. [Consulta: 26/08/2014].

SICELE. (2015). Estándares de calidad SICELE. Disponible en <https://sites.google.com/a/sicele.org/sicele/>. [Consulta: 24/08/2014].





ANEXO 1: TABLA DE ESPECIFICACIONES NIVEL B2 (MCER, 2002)

Comprensión de lectura			
Identifica información específica en textos especializados extensos. Comprende las ideas principales y globales de textos especializados. Infiere el significado de los párrafos y textos a partir de la información presentada en textos especializados.			
Comprensión auditiva			
Identifica información específica en discursos extensos y complejos lingüísticamente (entrevistas a expertos, charlas) sobre temas académicos. Comprende las ideas principales y globales de discursos especializados. Infiere el significado de diferentes partes de los discursos especializados.			
Expresión e interacción escrita			
Escribe redacciones claras y detalladas sobre una variedad de temas en las que desarrolla argumentos a favor o en contra de un punto de vista.			
<i>Adecuación al género discursivo</i>	<i>Coherencia</i>	<i>Corrección</i>	<i>Vocabulario</i>
El texto cumple con la estructura del ensayo. Se redacta en un lenguaje formal. Escribe un título, introduce el tema, establece claramente una postura respecto del tema. Menciona las ventajas y desventajas sobre el tema. Presenta una conclusión sobre el tema.	El texto es claro y coherente. Muestra un uso controlado, pero adecuado de los mecanismos de cohesión (conectores, estructura de párrafos y puntuación) para sintetizar y relacionar las ideas.	Demuestra un elevado control gramatical, aunque puede cometer algunos errores no sistemáticos que no producen malentendidos. Muestra un buen dominio ortográfico, si comete algunos errores estos no afectan la comprensión del escrito.	Tiene un amplio repertorio léxico que le permite comunicar adecuadamente sus ideas sobre temas especializados. Si comete algún error de selección léxica, este no afecta la comprensión.
Expresión e interacción orales			
Se comunica con fluidez, precisión y eficacia con hablantes nativos sobre temas académicos. Expone con claridad temas académicos y argumenta sus opiniones y puntos de vista. Participa activamente de una entrevista y amplía y desarrolla sus ideas sobre temas académicos.			
<i>Coherencia</i>	<i>Fluidez</i>	<i>Corrección</i>	<i>Vocabulario</i>
Presenta un discurso claro, comprensible y bien estructurado, con un uso adecuado de mecanismos de cohesión. Si se extiende en su discurso puede mostrar cierta pérdida de control. Mantiene la conversación de manera adecuada, colaborando con el entrevistador.	Se expresa de manera fluida y a un ritmo bastante uniforme, aunque puede dudar mientras busca las palabras adecuadas, las pausas no son largas. Su pronunciación es claramente comprensible, aunque cometa algunos errores esporádicos debido a su acento extranjero.	Muestra un alto dominio y control gramatical. Comete errores que no afectan la comprensión y casi todos los corrige.	Tiene un amplio dominio léxico que le permite expresarse adecuadamente sobre diversos temas especializados, sin evidenciar la búsqueda de palabras y utilizando oraciones complejas. Si comete algún error de selección léxica, este no afecta la comunicación.

ANEXO 2: RÚBRICAS PARA EVALUAR LAS TAREAS DE EXPRESIÓN ESCRITA Y ORAL

Rúbrica analítica para evaluar expresión e interacción escrita

Dimensiones	Niveles de desempeño					PUNTAJE
	4	3	2	1	0	
Adecuación al género discursivo	El texto cumple con la estructura del ensayo. Se redacta en un lenguaje formal. Escribe un título, introduce el tema, establece claramente una postura respecto del tema. Menciona las ventajas y desventajas sobre el tema. Presenta una conclusión sobre el tema.	El texto cumple con la estructura del ensayo, pero no mantiene un registro formal de la lengua. Establece una postura clara sobre el tema, pero no articula claramente las ventajas de las desventajas. Presenta una conclusión sobre el tema.	El texto cumple medianamente con la estructura del ensayo, puesto que carece de algunas partes y está redactado en un estilo informal. No establece una postura clara sobre el tema. No se distingue las ventajas de las desventajas. Presenta ideas desconectadas al término del texto.	El texto presenta una o dos partes de la estructura del ensayo. Está redactado en un registro informal. No establece una postura sobre el tema. No menciona ventajas y desventajas. Carece de una conclusión	El texto no cumple con la estructura del ensayo. No introduce el tema, ni establece una postura sobre él. No menciona ventajas y desventajas	
Coherencia	El texto es claro y coherente. Muestra un uso controlado, pero adecuado de los mecanismos de cohesión (conectores, estructura de párrafos y puntuación) para sintetizar y relacionar las ideas.	El texto es claro, aunque presenta elementos básicos para organizar las ideas. La distribución de los párrafos y los signos de puntuación son limitados, lo que obliga a una relectura para su comprensión.	El texto es poco claro y coherente, las ideas están dispuestas de manera lineal y carecen de mecanismos de cohesión. Se debe hacer relecturas para comprender la idea principal de lo que se quiere comunicar.	El texto plantea ideas poco claras y desconectadas. Ausencia de elementos cohesivos. Aunque se haga relecturas es difícil comprender el texto.	No se comprende las ideas que se plantean en el texto.	
Corrección	Demuestra un elevado control	Demuestra un buen control de estructuras	Demuestra un control gramatical limitado,	Muestra un dominio básico	No demuestra un uso y	

	gramatical, aunque puede cometer algunos errores no sistemáticos que no producen malentendidos. Muestra un buen dominio ortográfico, si comete algunos errores estos no afectan la comprensión del escrito.	gramaticales habituales, aunque comete algunos errores sistemáticos que, a veces, producen malentendidos. Muestra un buen dominio ortográfico, aunque comete varios errores que, en ocasiones, dificultan la comprensión del texto.	comete un gran número de errores sistemáticos que producen malentendidos. Muestra un dominio ortográfico deficiente, comete una gran cantidad de errores que dificultan la comprensión.	de estructuras gramaticales y oracionales memorizadas. Su escrito presenta una gran cantidad de errores ortográficos sistemáticos que impiden la comprensión del texto.	control gramatical u ortográfico.	
Vocabulario	Tiene un amplio repertorio léxico que le permite comunicar adecuadamente sus ideas sobre temas especializados. Si comete algún error de selección léxica, este no afecta la comprensión	Tiene un dominio léxico que le permite comunicar casi todas las ideas sobre temas especializados. Comete errores de selección y morfología léxica que, en ocasiones, afectan la comprensión.	Usa un conjunto de palabras sencillas que le permiten comunicar algunas ideas limitadas sobre el tema. Comete una gran cantidad de errores de selección y morfológica léxica que afectan la comprensión.	Utiliza un conjunto reducido de palabras y formas memorizadas, que dan cuenta vagamente de lo que se quiere comunicar.	Utiliza un conjunto reducido de palabras y fórmulas aprendidas que no logran comunicar una idea.	
TOTAL						

Rúbrica analítica para evaluar la expresión e interacción oral

Dimensiones	Niveles de desempeño					PUNTAJE
	4 puntos	3 puntos	2 puntos	1 punto	0 punto	
Coherencia	<p>Presenta un discurso claro, comprensible y bien estructurado, con un uso adecuado de mecanismos de cohesión. Mantiene la conversación de manera adecuada, colaborando con el entrevistador.</p>	<p>Presenta secuencias de ideas (lineales) comprensibles, unidas por conectores de uso habitual (es que, además, por eso). Solo presenta algunas partes de la exposición. Mantiene la conversación con ayuda del interlocutor, y en ocasiones necesita aclaraciones o repeticiones.</p>	<p>Presenta algunas ideas lineales, poco comprensibles, unidas por grupos de palabras y conectores sencillos (y, pero, porque). Presenta dos partes de la exposición del tema. Es Necesita de la colaboración del interlocutor para continuar la conversación y responder a preguntas y afirmaciones.</p>	<p>Presenta algunas ideas enlazadas con conectores muy básicos y lineales (y, entonces). Sin embargo, es difícil comprender la idea general de lo que desea comunicar. Sin la ayuda del interlocutor no puede continuar la conversación.</p>	<p>Presenta ideas desconectadas e incomprensibles sobre el tema de la conversación</p>	
Fluidez	<p>Se expresa de manera fluida y a un ritmo bastante uniforme, aunque puede dudar mientras busca las palabras adecuadas, las pausas no son largas. Su pronunciación es claramente comprensible, aunque cometa algunos errores esporádicos debido a su acento extranjero.</p>	<p>Habla de forma continuada y comprensible, aunque son evidentes las pausas para planificar su discurso y pensar en la gramática y el léxico adecuados. Su pronunciación es comprensible, pero su acento extranjero resulta evidente y comete errores esporádicos.</p>	<p>Se hace entender con expresiones muy breves, y resultan evidentes las pausas, las dudas iniciales y la reformulación. Su pronunciación y articulación son generalmente claras y comprensibles, aunque su acento y sus errores ocasionales pueden provocar que la comprensión requiera</p>	<p>Se expresa con frases breves y poco comprensibles. Realiza pausas largas, titubeos y reformulaciones. Su pronunciación y articulación es poco clara y distinguible, por lo que es difícil comprender lo que intenta decir sobre el tema.</p>	<p>Se expresa con titubeos, pausas y dudas. No se logra distinguir o comprender lo que intenta decir.</p>	

			cierto esfuerzo.			
Corrección	Muestra un alto dominio y control gramatical. Comete errores que no afectan la comprensión y casi todos los corrige.	Muestra un control razonable de formas y estructuras gramaticales de uso habitual. Comete errores gramaticales que no afectan la comprensión.	Muestra un control limitado y adecuado de estructuras sencillas, aunque comete errores básicos de forma sistemática, por ejemplo, confusión de tiempos verbales y faltas de concordancia que afectan la comprensión	Muestra un escaso control de estructuras gramaticales sencillas. La mayor parte de su discurso la elabora con errores básicos sistemáticos.	No muestra dominio ni control gramatical en su discurso.	
Vocabulario	Tiene un amplio dominio léxico que le permite expresarse adecuadamente sobre diversos temas especializados, sin evidenciar la búsqueda de palabras y utilizando oraciones complejas. Si comete algún error de selección léxica, este no afecta la comunicación.	Tiene un dominio léxico que le permite transmitir de manera razonable información sencilla sobre el tema. Sin embargo, le cuesta hacer descripciones o expresar su punto de vista de manera clara sobre el tema. Utiliza principalmente oraciones sencillas y necesita buscar las palabras adecuadas. Comete algunas incorrecciones e imprecisiones léxicas.	Tiene un dominio léxico limitado que le permite transmitir información básica y sencilla sobre el tema de la conversación. Tiene que buscar las palabras adecuadas para elaborar su discurso. Comete errores e imprecisiones léxicas que afectan la comprensión.	Tiene un escaso dominio léxico, le cuesta transmitir ideas sobre un tema. Comete un gran número de imprecisiones y errores léxicos que dificultan la comprensión.	Carece de un dominio léxico, por lo que le es imposible hablar sobre un tema.	
					TOTAL	

ANEXO 3: ESCALA PARA LA EVALUACIÓN DE TEXTOS

Nombre evaluador: _____

Estimado evaluador:

Solicito su colaboración para evaluar un conjunto de textos orales y escritos a través de una escala Likert. Los textos seleccionados formarán parte de un prototipo de prueba de proficiencia en ELE de un nivel B2 con una orientación académica.

Usted deberá juzgar a través de esta escala Likert si el texto es adecuado para evaluar comprensión lectora y comprensión auditiva en un nivel B2 de español, esto, en función de tres criterios: **pertinencia, contenido y formato.**

Para facilitar la revisión, se adjunta una descripción de los criterios mencionados y de los textos que se espera comprenda un estudiante de nivel B2, según el MCER (2002).

Su evaluación ayudará a mejorar la calidad, rigurosidad y transparencia en la elaboración de esta prueba de ELE.

I. Descripción de los criterios para evaluar los textos

CRITERIOS	INDICADORES
PERTINENCIA: El tipo de texto y el tema son adecuados para un nivel B2 de español con fines académicos.	1.Muy en desacuerdo 2.En desacuerdo 3.De acuerdo 4.Muy de acuerdo
CONTENIDO: El contenido gramatical y léxico es adecuado para un nivel B2 de español con fines académicos.	1.Muy en desacuerdo 2.En desacuerdo 3.De acuerdo 4.Muy de acuerdo
FORMATO DEL TEXTO: La redacción, el tamaño de letra, la longitud del texto/ la locución, el ruido de los audios son apropiados para el nivel que se orienta la prueba.	1.Muy en desacuerdo 2.En desacuerdo 3.De acuerdo 4.Muy de acuerdo

II. Textos que comprende un estudiante de nivel B2 en un contexto académico

DESTREZAS LINGÜÍSTICAS	Textos que comprende un nivel B2
Comprensión lectora	Textos extensos y especializados de diversas áreas del saber: artículos, informes, ensayos.
Comprensión auditiva	Programas de los medios de comunicación, conferencias, discusiones, charlas y otras

	formas de presentación, siempre que estén articulados con claridad, aunque tengan ruidos de fondo.
--	----------------------------------------------------------------------------------------------------

ESCALA LIKERT: De acuerdo con los criterios explicados en la tabla anterior evalúe primero los textos escritos y luego los orales. Para ello, escriba el número que mejor representa su opinión para cada texto

EJEMPLO:

Si usted considera que el texto 1 es poco pertinente en cuanto al tema para un B2, y los contenidos lingüísticos y el formato son adecuados, escribirá

CRITERIOS TEXTOS	PERTINENCIA	CONTENIDO	FORMATO
Texto 1	3	4	4

TEXTOS ESCRITOS

CRITERIOS TEXTOS	PERTINENCIA	CONTENIDO	FORMATO
Texto 1			
Texto 2			
Texto 3			
Texto 4			
Texto 5			
Texto 6			
Texto 7			
Texto 8			
Texto 9			
Texto 10			

Si tiene alguna observación sobre los textos evaluados, escríbala aquí:

TEXTOS ORALES

CRITERIOS AUDIOS	PERTINENCIA	CONTENIDO	FORMATO
Audio 1			
Audio 2			
Audio 3			
Audio 4			
Audio 5			

Audio 6			
Audio 7			
Audio 8			
Audio 9			
Audio 10			

Si tiene alguna observación sobre los textos evaluados, escríbala aquí:



ANEXO 4: IMPLEMENTACIÓN DEL PROTOTIPO DE PRUEBA DE EFA EN LÍNEA NIVEL B2

1. Presentación de la prueba e instrucciones generales

Prueba ELE Académico

Prueba de Español como Lengua Extranjera con fines académicos

Esta prueba de Español como Lengua Extranjera tiene como objetivo evaluar el nivel de proficiencia de los estudiantes extranjeros de intercambio en la Universidad de Concepción.

Esta prueba se compone de cinco secciones:

- 1) Comprensión lectora:** Corresponde a la primera sección de la prueba. Se presenta tres textos escritos, de cada uno se plantean cinco preguntas. El texto 1 se responde con preguntas de verdadero y falso y los textos 2 y 3 con preguntas de selección múltiple.
- 2) Comprensión auditiva:** Corresponde a la segunda sección de la prueba. Se presenta tres audios, de cada uno se plantean cinco preguntas. El texto 1 se responde con preguntas de verdadero y falso y los textos 2 y 3 con preguntas de selección múltiple.
- 3) Competencia lingüística:** Corresponde a la tercera sección de la prueba. Se presenta un texto con 20 preguntas de selección múltiple sobre algunos usos gramaticales.
- 4) Expresión e interacción escrita:** Corresponde a la cuarta sección de la prueba y consta de una tarea donde se presentan dos opciones, tienes que elegir UNA y escribir un texto de 250 palabras (20-25 líneas).
- 5) Expresión e interacción oral:** Esta sección se realiza de manera presencial, luego de que has terminado la prueba en el computador. Se trata de una entrevista oral.

Porcentajes de la prueba

El porcentaje de la prueba es el siguiente: comprensión lectora (25%), comprensión auditiva (20%), competencia lingüística (10%), expresión e interacción escrita (25%) y expresión e interacción oral (20%).

Ingreso y desarrollo de la prueba

Para ingresar a la prueba presiona el botón "Comenzar intento". Sigue las instrucciones y responde lo que se te solicita en cada sección. Cuando termines cada sección, presiona el botón "Siguiente".

Dispones de 1 hora y 30 minutos para realizar esta prueba en línea.

Administración

- ▼ Administración del cuestionario
 - Editar ajustes
 - Anulaciones de grupo
 - Anulaciones de usuario
- ⚙ Editar cuestionario
- 🔍 Vista previa
- ▶ Resultados
- Roles asignados localmente
- Permisos
- Compruebe los permisos
- Filtros
- Registros
- Copia de seguridad
- Restaurar
- ▶ Banco de preguntas

- ▶ Administración del curso
- ▶ Cambiar rol a...
- ▶ Ajustes de mi perfil

2. Sección I: Comprensión de lectura

Texto I: Lee el texto y selecciona si los enunciados son verdaderos o falsos

Sistema de competencias sustentables para el desempeño profesional en Ingeniería

El reconocido aporte de la Ingeniería al avance socioeconómico de los países, a través del desarrollo industrial, de los servicios y de la infraestructura, entre otros factores, ha generado una fuerte presión sobre las universidades y otros centros de educación superior. Estas entidades educativas deben formar ingenieros con la mayor capacidad posible de una inserción laboral efectiva.

Recogiendo una tradición educativa de larga data, con raíces en la educación escolar y educación superior, la formación de profesionales se espera culmine, en el presente, en el egreso de ingenieros que tengan ciertas capacidades y atributos personales que los hagan aptos para insertarse al trabajo productivo en forma rápida y eficaz. Las capacidades y atributos indicados suelen denominarse, en general, "competencias de egreso". Se espera que esas competencias deben resultar de un conjunto de aprendizajes acumulativos, con fuerte énfasis en la aplicación del conocimiento a tareas propias de la Ingeniería profesional, aspecto que permite guiar la formación hacia la deseada capacidad de efectiva inserción laboral.

Traer el contexto laboral, en alguna medida, a las aulas, o acercar a los estudiantes a dicho contexto es, por lo tanto, un importante desafío educativo en la actualidad. Evaluar esos deseados resultados longitudinales de aprendizaje constituye otro de los grandes desafíos para las escuelas de Ingeniería.

Las dificultades indicadas precedentemente, así como los recursos y capacidades institucionales disponibles, han hecho muy lento el progreso en los cambios curriculares que deberían acompañar a la formación de ingenieros basada en competencias, como lo atestiguan las publicaciones nacionales e internacionales en Educación en Ingeniería.

(Fuente: *Revista Facultad de Ingeniería*, Universidad de Tarapacá)

1. De acuerdo con el texto, la ingeniería realiza aportes relevantes al desarrollo de un país.
2. Según el texto, la formación que reciben los ingenieros les impide encontrar trabajo.
3. De acuerdo con el texto, las competencias de egreso corresponden al plan de estudios de una institución educativa.
4. A partir del texto, se puede inferir que las escuelas de Ingeniería están intentando implementar cambios en la formación académica de sus estudiantes.
5. A partir de la lectura del texto se puede inferir que la sociedad requiere de egresados vinculados al ámbito laboral.

Texto II: Lee el texto y selecciona la alternativa correcta (a, b, c, d)

El aprendizaje de una segunda lengua

Todas las personas poseen un "equipamiento básico" para aprender las características propias de distintas lenguas. No obstante, hay personas que aprenden las destrezas lingüísticas con mayor facilidad que otras debido a factores personales como su capacidad intelectual o el entorno social en el que se encuentran.

Para algunos autores, inteligencia y aptitud son dos factores diferentes, y es que la inteligencia es la habilidad general para aprender, mientras que la aptitud es la habilidad específica para aprender una segunda lengua. La aptitud tiene una importancia relevante en la enseñanza dentro del aula, aunque no desempeña ningún papel decisivo en lo referente a las habilidades comunicativas que se establecen en un entorno natural.

Uno de los factores más importantes de la aptitud lingüística es la habilidad auditiva del aprendiz con respecto a la segunda lengua. Normalmente se acepta que la habilidad auditiva del alumno está claramente relacionada con la manera en la que ha aprendido esta segunda lengua. La mejor forma de aprender es de forma natural, interactuando, no obstante se desconoce si la precisión del dominio de esta habilidad oral se debe únicamente a esta característica "natural".

Relacionada con la habilidad oral del aprendiz se encuentra también la "habilidad motora". Esta habilidad está relacionada con la capacidad para desarrollar y controlar ciertos músculos del cuerpo, ya que para la producción del habla se necesita poner en movimiento una serie de órganos (mandíbula, labios, lenguas, cuerdas vocales, etc.). Los movimientos de estos órganos difieren de una lengua a otra, por lo tanto el "adiestramiento" muscular de estas partes debe ser ejecutado por el hablante en todo momento, con mejores resultados si el aprendiz comienza a dominar estos músculos desde una edad temprana.

Otro factor crucial que permite la adaptación de la segunda lengua es el tipo de personalidad del estudiante. Hoy en día todavía los investigadores no se ponen de acuerdo en definir lo que significa personalidad, puesto que dentro de esta definición destacan cientos de minúsculos factores, y muchos de ellos escapan a la razón del ser humano.

(Fuente:Revista digital *Innovación y experiencias educativas*)

1. Según el texto, en el aprendizaje de una segunda lengua...
 2. De acuerdo con el texto, la aptitud lingüística es...
 3. De acuerdo con el texto, la mejor forma de aprender la hab
 4. Según el texto, la habilidad motora de una segunda lengua
 5. ¿Cuál es la idea central del texto?
- a) hay habilidades que son más fáciles de aprender

b) el entorno social tiene un papel decisivo

c) la inteligencia ayuda a conocer mejor el idioma

d) hay personas que aprenden con mayor facilidad

Texto III: Lee el texto y selecciona la alternativa correcta (a, b, c, d)

La biodiversidad

El crecimiento continuo de la población humana y el consiguiente aumento del consumo de recursos ha llevado a una explotación insostenible de la diversidad biológica del planeta, exacerbada por el cambio climático, la acidificación oceánica y otros impactos ambientales de origen antrópico.

La biodiversidad —la variedad de genes, especies y ecosistemas que constituyen la vida en la tierra— proporciona numerosos servicios esenciales a la sociedad. Cabe mencionar los bienes materiales (alimentos, madera, medicinas y fibra), las funciones de soporte (control de inundaciones, regulación del clima y ciclos de nutrientes) y los beneficios inmateriales como el ocio. La biodiversidad puede favorecer la agricultura a través de la polinización y el control de plagas, facilitar el almacenamiento y secuestro de carbono y promover la salud física y mental de las personas. Además, también proporciona resiliencia frente a las perturbaciones y los cambios ambientales. Esas y otras ventajas socioeconómicas son fundamentales, y los cálculos recientes afirman que el valor económico que aportan los ecosistemas pueden superar de 10 a 100 veces el coste de su mantenimiento.

A pesar de algunos éxitos de conservación (en especial, a escala global) y del interés creciente de la población y de los gobiernos acerca de la sostenibilidad del planeta, la biodiversidad sigue disminuyendo.

La restauración ecológica representa una práctica cada vez más común en todo el planeta. Destaca el éxito de las acciones realizadas a favor de las especies, como la gestión de su hábitat, la erradicación de especies invasoras, la cría en cautividad y la reintroducción de especies. Entre 1994 y 2004 se ha evitado por lo menos 16 extinciones de especies de aves y, desde 1992, la red mundial de zonas protegidas se ha ido extendiendo, con un incremento anual del 2,5 % y 1,4 %, respectivamente, de la superficie total y del número de zonas protegidas. A pesar de algunos inconvenientes, las áreas protegidas siguen siendo una estrategia fundamental de la conservación de la biodiversidad.

Varias iniciativas reflejan las recomendaciones de los científicos sobre la importancia de mantener hábitats de gran extensión y conectados entre sí. También se ha dado impulso a las medidas de conservación que generan beneficios económicos, derivados tanto de recursos renovables como no renovables, como el ecoturismo. Algunas de esas estrategias permiten hacer frente a los costes de conservación y además proporcionan rendimientos, aspecto de vital importancia en las comunidades rurales pobres. Una de ellas es el pago directo por la conservación de la biodiversidad o el pago de los servicios ofrecidos por los ecosistemas.

Pero las presiones sobre la biodiversidad continúan, principalmente debido a la sobreexplotación de especies, la presencia de especies exóticas invasoras, la contaminación, el cambio climático y, en especial, la degradación, fragmentación y destrucción de hábitats.

En el futuro, las estrategias eficaces de conservación deben recibir un mayor impulso y suficiente financiación. Además, debe reconocerse la biodiversidad como un bien público general, un bien que conviene integrar en las políticas y la toma de decisiones relativas a la producción y consumo de recursos. Es necesario centrarse en los cambios institucionales y sociales generales para permitir una aplicación más eficaz de las medidas políticas.

(Fuente: <http://www.investigacionyciencia.es/noticias/retos-futuros-para-conservar-la-biodiversidad-9793>)

1. Según el texto, el aumento de la población a nivel mundial ha...
2. De acuerdo con el texto, la biodiversidad...
3. La situación actual de la biodiversidad indica que...
4. En el futuro se espera que haya...
5. ¿Cuál es la idea central del texto?

3. Sección II. Comprensión auditiva

Sección II: Comprensión auditiva

Esta sección tiene tres audios sobre distintos temas académicos. De cada audio se plantea cinco preguntas. El primer audio se responde con ítems de verdadero/falso y los siguientes con preguntas de selección múltiple



Audio 1: Escucha el audio y selecciona si los enunciados son *verdaderos* o *falsos*

1. La energía eólica corresponde a la energía del viento.
2. Según el audio, el objetivo de la instalación de la turbina es transformar la energía eléctrica en energía libre.
3. Según el audio, la instalación de turbinas mejorará la calidad energética en las zonas alejadas de la ciudad.
4. De acuerdo con el audio, la gente que vive en zonas rurales carece de recursos económicos.
5. La idea central del audio es dar cuenta del aporte de las energías renovables al suministro eléctrico.



Audio 2: Escucha el audio y selecciona la opción correcta (a, b, c, d)

1. De acuerdo con el audio ¿Cuál es la situación de la cuenca del río Copiapó?
2. El objetivo del proyecto Agua Copiapó es...
3. La tecnología utilizada en el proyecto Agua Copiapó...
4. Según el audio, el proyecto Agua Copiapó...
5. ¿Cuál es la idea central del audio?



Audio 3: Escucha el audio y selecciona la opción correcta (a, b, c, d)

- Según el audio, el petróleo es un combustible...
- El Centro de Biotecnología de la Universidad de Concepción está interesado en...
- El Centro de Biotecnología puede llevar a cabo su investigación porque la región donde se encuentra cuenta con...
- De acuerdo con el audio, las microalgas...
- ¿Cuál es el tema principal del audio?

6. Sección V: Expresión e interacción oral

Esta sección tiene una duración de 15 minutos y consta de TRES partes:

1. **Presentación del estudiante:** El estudiante da a conocer su nombre, su nacionalidad, habla de sus estudios, de su trayectoria académica/profesional en Chile y de sus proyecciones a largo plazo en su formación o trabajo.
2. **Exposición de un tema:** Se presenta dos opciones de temas para exponer. El estudiante elige una y tiene 2 minutos para preparar la exposición.

En esta exposición debes:

- Presentar el tema y su relevancia
 - Dar a conocer tu punto de vista sobre el tema
 - Indicar las ventajas y desventajas sobre el tema
 - Concluir
- Opción 1: El uso de pesticidas en la agricultura, sus efectos en la salud y el medioambiente. Actualmente no es posible una agricultura con altos rendimientos sin la utilización de medidas de protección de plantas, entre las cuales los pesticidas siguen teniendo una participación considerable.
 - Opción 2: El aprendizaje de una segunda lengua a través de un proceso de inmersión. Se considera más efectivo el aprendizaje de una segunda lengua o cuando se tiene la posibilidad de vivir en el país donde se habla esa lengua. Sin embargo, no todas las personas tienen la posibilidad de viajar y vivir esta experiencia.
3. Conversación con el entrevistador sobre el tema expuesto

ANEXO 5: MUESTRA DE UN PROTOTIPO DE PRUEBA DE EFA EN LÍNEA NIVEL B2 RESPONDIDA

Sujeto: 22

Lengua materna: inglés, área de estudios: Letras

Texto I: Lee el texto y selecciona si los enunciados son *verdaderos o falsos*

Sistema de competencias sustentables para el desempeño profesional en Ingeniería

El reconocido aporte de la Ingeniería al avance socioeconómico de los países, a través del desarrollo industrial, de los servicios y de la infraestructura, entre otros factores, ha generado una fuerte presión sobre las universidades y otros centros de educación superior. Estas entidades educativas deben formar ingenieros con la mayor capacidad posible de una inserción laboral efectiva.

Recogiendo una tradición educativa de larga data, con raíces en la educación escolar y educación superior, la formación de profesionales se espera culmine, en el presente, en el egreso de ingenieros que tengan ciertas capacidades y atributos personales que los hagan aptos para insertarse al trabajo productivo en forma rápida y eficaz. Las capacidades y atributos indicados suelen denominarse, en general, "competencias de egreso". Se espera que esas competencias deben resultar de un conjunto de aprendizajes acumulativos, con fuerte énfasis en la aplicación del conocimiento a tareas propias de la Ingeniería profesional, aspecto que permite guiar la formación hacia la deseada capacidad de efectiva inserción laboral.

Traer el contexto laboral, en alguna medida, a las aulas, o acercar a los estudiantes a dicho contexto es, por lo tanto, un importante desafío educativo en la actualidad. Evaluar esos deseados resultados longitudinales de aprendizaje constituye otro de los grandes desafíos para las escuelas de Ingeniería.

Las dificultades indicadas precedentemente, así como los recursos y capacidades institucionales disponibles, han hecho muy lento el progreso en los cambios curriculares que deberían acompañar a la formación de ingenieros basada en competencias, como lo atestiguan las publicaciones nacionales e internacionales en Educación en Ingeniería.

(Fuente: *Revista Facultad de Ingeniería, Universidad de Tarapacá*)

1. De acuerdo con el texto, la ingeniería realiza aportes relevantes al desarrollo de un país. b) Verdadero
2. Según el texto, la formación que reciben los ingenieros les impide encontrar trabajo. a) Falso
3. De acuerdo con el texto, las competencias de egreso corresponden al plan de estudios de una institución educativa. a) Falso
4. A partir del texto, se puede inferir que las escuelas de Ingeniería están intentando implementar cambios en la formación académica de sus estudiantes. b) Verdadero
5. A partir de la lectura del texto se puede inferir que la sociedad requiere de egresados vinculados al ámbito laboral. b) Verdadero

Texto II: Lee el texto y selecciona la alternativa correcta (a, b, c, d)

El aprendizaje de una segunda lengua

Todas las personas poseen un "equipamiento básico" para aprender las características propias de distintas lenguas. No obstante, hay personas que aprenden las destrezas lingüísticas con mayor facilidad que otras debido a factores personales como su capacidad intelectual o el entorno social en el que se encuentran.

Para algunos autores, inteligencia y aptitud son dos factores diferentes, y es que la inteligencia es la habilidad general para aprender, mientras que la aptitud es la habilidad específica para aprender una segunda lengua. La aptitud tiene una importancia relevante en la enseñanza dentro del aula, aunque no desempeña ningún papel decisivo en lo referente a las habilidades comunicativas que se establecen en un entorno natural.

Uno de los factores más importantes de la aptitud lingüística es la habilidad auditiva del aprendiz con respecto a la segunda lengua. Normalmente se acepta que la habilidad auditiva del alumno está claramente relacionada con la manera en la que ha aprendido esta segunda lengua. La mejor forma de aprender es de forma natural, interactuando, no obstante se desconoce si la precisión del dominio de esta habilidad oral se debe únicamente a esta característica "natural".

Relacionada con la habilidad oral del aprendiz se encuentra también la "habilidad motora". Esta habilidad está relacionada con la capacidad para desarrollar y controlar ciertos músculos del cuerpo, ya que para la producción del habla se necesita poner en movimiento una serie de órganos (mandíbula, labios, lenguas, cuerdas vocales, etc.). Los movimientos de estos órganos difieren de una lengua a otra, por lo tanto el "adiestramiento" muscular de estas partes debe ser ejecutado por el hablante en todo momento, con mejores resultados si el aprendiz comienza a dominar estos músculos desde una edad temprana.

Otro factor crucial que permite la adaptación de la segunda lengua es el tipo de personalidad del estudiante. Hoy en día todavía los investigadores no se ponen de acuerdo en definir lo que significa personalidad, puesto que dentro de esta definición destacan cientos de minúsculos factores, y muchos de ellos escapan a la razón del ser humano.

(Fuente: Revista digital *Innovación y experiencias educativas*)

1. Según el texto, en el aprendizaje de una segunda lengua... d) hay personas que aprenden con mayor facilidad
2. De acuerdo con el texto, la aptitud lingüística es... a) la habilidad específica para aprender un segundo idioma
3. De acuerdo con el texto, la mejor forma de aprender la habilidad auditiva es... b) en un lugar donde se habla la segunda lengua
4. Según el texto, la habilidad motora de una segunda lengua... c) debe practicarse de manera constante
5. ¿Cuál es la idea central del texto? b) Los diferentes aspectos que afectan la adquisición de una segunda lengua

Texto III: Lee el texto y selecciona la alternativa correcta (a, b, c, d)

La biodiversidad

El crecimiento continuo de la población humana y el consiguiente aumento del consumo de recursos ha llevado a una explotación insostenible de la diversidad biológica del planeta, exacerbada por el cambio climático, la acidificación oceánica y otros impactos ambientales de origen antropico.

La biodiversidad —la variedad de genes, especies y ecosistemas que constituyen la vida en la tierra— proporciona numerosos servicios esenciales a la sociedad. Cabe mencionar los bienes materiales (alimentos, madera, medicinas y fibra), las funciones de soporte (control de inundaciones, regulación del clima y ciclos de nutrientes) y los beneficios inmateriales como el ocio. La biodiversidad puede favorecer la agricultura a través de la polinización y el control de plagas, facilitar el almacenamiento y secuestro de carbono y promover la salud física y mental de las personas. Además, también proporciona resiliencia frente a las perturbaciones y los cambios ambientales. Esas y otras ventajas socioeconómicas son fundamentales, y los cálculos recientes afirman que el valor económico que aportan los ecosistemas pueden superar de 10 a 100 veces el coste de su mantenimiento.

A pesar de algunos éxitos de conservación (en especial, a escala global) y del interés creciente de la población y de los gobiernos acerca de la sostenibilidad del planeta, la biodiversidad sigue disminuyendo.

La restauración ecológica representa una práctica cada vez más común en todo el planeta. Destaca el éxito de las acciones realizadas a favor de las especies, como la gestión de su hábitat, la erradicación de especies invasoras, la cría en cautividad y la reintroducción de especies. Entre 1994 y 2004 se ha evitado por lo menos 16 extinciones de especies de aves y, desde 1992, la red mundial de zonas protegidas se ha ido extendiendo, con un incremento anual del 2,5 % y 1,4 %, respectivamente, de la superficie total y del número de zonas protegidas. A pesar de algunos inconvenientes, las áreas protegidas siguen siendo una estrategia fundamental de la conservación de la biodiversidad.

Varias iniciativas reflejan las recomendaciones de los científicos sobre la importancia de mantener hábitats de gran extensión y conectados entre sí. También se ha dado impulso a las medidas de conservación que generan beneficios económicos, derivados tanto de recursos renovables como no renovables, como el ecoturismo. Algunas de esas estrategias permiten hacer frente a los costes de conservación y además proporcionan rendimientos, aspecto de vital importancia en las comunidades rurales pobres. Una de ellas es el pago directo por la conservación de la biodiversidad o el pago de los servicios ofrecidos por los ecosistemas.

Pero las presiones sobre la biodiversidad continúan, principalmente debido a la sobreexplotación de especies, la presencia de especies exóticas invasoras, la contaminación, el cambio climático y, en especial, la degradación, fragmentación y destrucción de hábitats.

En el futuro, las estrategias eficaces de conservación deben recibir un mayor impulso y suficiente financiación. Además, debe reconocerse la biodiversidad como un bien público general, un bien que conviene integrar en las políticas y la toma de decisiones relativas a la producción y consumo de recursos. Es necesario centrarse en los cambios institucionales y sociales generales para permitir una aplicación más eficaz de las medidas políticas.

(Fuente: <http://www.investigacionyciencia.es/noticias/retos-futuros-para-conservar-la-biodiversidad-9793>)

- Según el texto, el aumento de la población a nivel mundial ha... a) deteriorado los recursos naturales
- De acuerdo con el texto, la biodiversidad... c) beneficia a la sociedad
- La situación actual de la biodiversidad indica que... b) es necesaria la conservación de especies para evitar su desaparición
- En el futuro se espera que haya... c) mayor consideración sobre la biodiversidad en las instituciones
- ¿Cuál es la idea central del texto? a) La situación de la biodiversidad a nivel mundial

Audio 1: Escucha el audio y selecciona si los enunciados son verdaderos o falsos

- La energía eólica corresponde a la energía del viento. b) Verdadero
 - Según el audio, el objetivo de la instalación de la turbina es transformar la energía eléctrica en energía libre. a) Falso
 - Según el audio, la instalación de turbinas mejorará la calidad energética en las zonas alejadas de la ciudad. b) Verdadero
 - De acuerdo con el audio, la gente que vive en zonas rurales carece de recursos económicos. a) Falso
- Puntúa 1,0 sobre 1,0
5. La idea central del audio es dar cuenta del aporte de las energías renovables al suministro eléctrico. a) Falso
- Puntúa 0,0 sobre 1,0

Escribir comentario o corregir la calificación

Historial de respuestas

Paso	Hora	Acción	Estado	Puntos
1	29/09/2015 15:01	Iniciado/a	Sin responder aún	
2	29/09/2015 15:17	Guardada: part 1: ; part 2: ; part 3: ; part 4: ; part 5:	Sin responder aún	
3	29/09/2015 15:32	Guardada: part 1: b) Verdadero; part 2: a) Falso; part 3: b) Verdadero; part 4: a) Falso; part 5: a) Falso	Respuesta guardada	
4	29/09/2015 16:14	Intento finalizado	Parcialmente correcta	4,0

Audio 2: Escucha el audio y selecciona la opción correcta (a, b, c, d)

1. De acuerdo con el audio ¿Cuál es la situación de la cuenca del río Copiapó? ❌
2. El objetivo del proyecto Agua Copiapó es... ✔️
3. La tecnología utilizada en el proyecto Agua Copiapó... ✔️
4. Según el audio, el proyecto Agua Copiapó... ❌
5. ¿Cuál es la idea central del audio? ✔️

Escribir comentario o corregir la calificación

Historial de respuestas

Paso	Hora	Acción	Estado	Puntos
1	29/09/2015 15:01	Iniciado/a	Sin responder aún	
2	29/09/2015 15:17	Guardada: part 1: ; part 2: ; part 3: ; part 4: ; part 5:	Sin responder aún	
3	29/09/2015 15:32	Guardada: part 1: d) Permanece sin agua; part 2: d) optimizar el uso del agua para el riego; part 3: a) evita la pérdida de agua innecesaria ; part 4: a) solucionó el problema del agua para los habitantes de la zona ; part 5: c) El mejoramiento en la gestión y administración del agua en el Valle de Copiapó	Respuesta guardada	
4	29/09/2015 16:14	Intento finalizado	Parcialmente correcta	3,0

Audio 3: Escucha el audio y selecciona la opción correcta (a, b, c, d)

- Según el audio, el petróleo es un combustible... ✔️
- El Centro de Biotecnología de la Universidad de Concepción está interesado en... ✔️
- El Centro de Biotecnología puede llevar a cabo su investigación porque la región donde se encuentra cuenta con... ✔️
- De acuerdo con el audio, las microalgas... ✔️
- ¿Cuál es el tema principal del audio? ❌

Escribir comentario o corregir la calificación

Sección III: Competencia lingüística

Instrucciones: Completa los espacios del texto, seleccionando la opción correcta (a, b, c, d)

Historial de respuestas

Paso	Hora	Acción	Estado
1	29/09/2015 15:01	Iniciado/a	
2	29/09/2015 15:37	Visto	
3	29/09/2015 16:14	Intento finalizado	

El cobre y el futuro de Chile

El cobre siempre d) había sido importante en Chile. Pero a) después de 1990 ha pasado a ser doblemente más importante. Ha aumentado notablemente su relevancia en la economía, c) por consiguiente gracias a él Chile cambió. Es un país que exporta mucho más. El cobre es una de las principales fuentes de recursos c) para el Fisco. Permite que las reservas internacionales b) aumentaron significativamente y que haya considerables ahorros para enfrentar el futuro. Además, permitió que Chile a) sea un país menos endeudado, sus ciudadanos paguen menos impuestos y c) tengan mayores ingresos.

El cobre ha aumentado el nivel de bienestar a) de la sociedad chilena. Aún más, la sociedad chilena a) la ha vuelto "adicta c) al cobre" (con precios superiores a US\$ 3 la libra); los chilenos a) consumen cada vez más bienes importados y exigen mayores gastos públicos en bienes sociales.

La gran diferencia c) entre los países desarrollados y los países en desarrollo es la existencia de un horizonte de largo plazo. b) Cómo introducir esta perspectiva en el debate chileno? La interrogante de fondo que ha estado presente en la historia chilena c) durante los últimos 100 años: ¿Cómo puede Chile a) llegar a ser un país desarrollado? b) Es decir , lo que hay que debatir no es lo c) que le pasa al cobre o a la minería; el tema no es sectorial. b) Lo que interesa es la transformación del país a) en uno avanzado. El cobre permite generar d) este debate.

Patricio Meller (*Académico Ingeniería Industrial U. Chile y director ejecutivo Cieplan*)

¿Agua o trajes? Una pregunta para el futuro

Según UNESCO hay una crisis de agua a nivel mundial. Con nuestra población en crecimiento constante, simplemente no tenemos suficiente agua para todos los seres humanos puedan beber y vivir. Sin embargo, parece que el crecimiento de la población mundial no es el problema y que son las empresas que utilizan demasiada agua y las fábricas de ropa son los que consumen lo más. Desde lo que dijo el informe de 2014 para fabricar un traje de caballero la empresa usa 5,550 litros de agua y 4,400 litros para zapatos. Entonces, necesitamos preguntarnos lo que necesitamos más, trajes o el agua.

Desafortunadamente no es tan simple y necesitamos enfocar nuestros esfuerzos en identificar los lugares donde se podría ahorrar agua y el primer punto importante es que yo, como habitante de la clase media en un país desarrollado, no me dí cuenta que había una crisis. De hecho, desde lo que veo en mi país no hacemos nada para conservar agua y aún más soy consciente de que yo podría ahorrar mucho más agua en mi día típico para ahorrar. Sin embargo, así que no me está afectando uso una cantidad de agua libre y es lo mismo para las empresas, en una escala mucho ms grande, que usan unos millones litros de agua cada mes. Necesitamos ahorrar agua, pero en primer lugar deberíamos sensibilizar la gente en que hay una necesidad de ahorrar agua para que los otros puedan beber.

Sin obstante, no es para decir que no necesitamos usar agua en las empresas y, sin duda, las empresas necesitan este recurso para funcionar, pero me sorprendería si no habría una opción más conservadora en cuanto a fabricar un traje. Es decir que el Estado necesita controlar la cantidad de agua que la industria de ropa usa y, si posible, buscar si hay tecnologías para cambiar el método de fabricar sus productos porque la ropa es imprescindible en el mundo y no podemos castigar a las empresas sin tratar de ayudarlas antes. No es un tema simple y hay muchos factores que hacen una parte del debate y estos factores crean de varios argumentos diferentes. Por ejemplo, se podría preguntar si es más la culpa de los gobiernos en los países donde hay una tasa de nacimiento muy alto y si ellos podrían hacer más para inhibir el crecimiento de la población global.

Para concluir, este tema es muy sensible y debería ser la decisión de un país sobre la cantidad de agua que cada individuo en el mundo puede recibir. El agua es un bien natural y vital para la vida y tenemos suficiente para todos los seres humanos pero necesitamos re-evaluar nuestras prioridades que, actualmente, permiten que las empresas hacen lo que quieren y usan tanta agua que quieren mientras que personas mueren de sed en el otro lado del mundo. No estoy de acuerdo que es un problema de tamaño de población y no necesitamos regular el nacimiento en otros países sino enfocarnos en nuestras empresas grandes que podrían usar menos agua y hacer un esfuerzo para modernizar su tecnología con el objetivo de resolver esta crisis.

ANEXO 6: CUESTIONARIO DE VALIDEZ DE APARIENCIA

INSTRUMENTO PARA CONOCER LA VALIDEZ DE APARIENCIA DE LA PRUEBA DE ELE PARA EL ÁMBITO ACADÉMICO

Estimado estudiante:

Usted ha sido seleccionado para evaluar esta prueba que tiene como objetivo evaluar el nivel de proficiencia B2 en Español como Lengua Extranjera de los estudiantes extranjeros de intercambio en la Universidad de Concepción. **La prueba tiene una duración total de 1 hora y 45 minutos y se compone de 5 secciones:** comprensión lectora, comprensión auditiva, competencia lingüística, expresión e interacción escrita y expresión e interacción oral.

Identificación del estudiante

Nombre	
Edad	
Carrera	
Años de estudio en la universidad	

Actividad evaluativa

Con el objetivo de asegurar la calidad de la prueba que ha rendido, solicito a usted que evalúe mediante una escala de Likert su grado de apreciación en relación con este instrumento. Para ello, se presenta una serie de ítems que usted debe juzgar con esta escala de 5 puntos, considerando los siguientes indicadores:

1	2	3	4	5
Muy en desacuerdo	en desacuerdo	Indiferente	De acuerdo	Muy de acuerdo

Una vez que complete la escala Likert responda el cuestionario que se adjunta.

Agradezco su colaboración, puesto que su evaluación ayudará a mejorar la calidad, rigurosidad y transparencia en la elaboración de esta prueba de ELE.

Muchas gracias.

ESCALA LIKERT

Ítems a evaluar	INDICADORES				
	1	2	3	4	5
<i>Apreciación global sobre la prueba</i>					
Le parece interesante que la prueba esté implementada en computador.					
El entorno de la prueba es fácil de usar.					
¿Considera aceptable esta prueba?					
<i>Formato de la prueba</i>					
Considera adecuado el número de secciones de la prueba: comprensión lectora, comprensión auditiva, competencia lingüística, expresión e interacción escrita y oral.					
Le parecieron claras todas las instrucciones de la prueba.					
Le parece adecuada la duración total de la prueba (1 hora y 45 minutos).					
Le parece adecuada la duración de las distintas secciones de la prueba: comprensión lectora (30min.), comprensión auditiva (30 min.), expresión escrita (20 min.), competencia lingüística (10 min.) y expresión oral (15 min.).					
Le parecieron comprensibles los textos de la sección de comprensión lectora.					
Le parecieron comprensibles los audios de la sección de comprensión auditiva.					

Preguntas

Responda estas preguntas a partir de su apreciación sobre la prueba que acaba de rendir

- 1) Escriba su apreciación general sobre la prueba que acaba de realizar (buena/mala) ¿Qué tal le pareció esta prueba?

- 2) ¿Qué parte de la prueba le pareció más fácil o más difícil? ¿Por qué?

- 3) ¿Hubo algún texto oral o escrito que le haya parecido más fácil de comprender? ¿Cuál?

- 4) ¿Hubo algún texto oral o escrito que le haya parecido más difícil de comprender? ¿Cuál?

- 5) En la sección de comprensión lectora ¿las preguntas que respondió se relacionaban con el tema que cada texto trataba?

- 6) En la sección de comprensión auditiva ¿las preguntas que respondió se relacionaban con el tema que cada audio trataba?

- 7) ¿Qué hizo para resolver las preguntas en la sección de comprensión lectora?

8) ¿Qué hizo para resolver las preguntas de comprensión auditiva?

9) ¿Cómo resolvió la tarea de la sección de competencia lingüística?

10) ¿Cómo realizó las tareas de escritura y de oralidad? ¿Tuvo alguna dificultad? ¿Cuál?



ANEXO 7: PAUTA PARA LA EVALUACIÓN DEL JUICIO DE EXPERTOS

Estimado evaluador:

Usted ha sido seleccionado para evaluar esta prueba que tiene como objetivo evaluar el nivel de proficiencia B2 en ELE de los estudiantes extranjeros de intercambio en la Universidad de Concepción. Esta prueba tiene una orientación académica.

Identificación del evaluador experto

Nombre	
Título profesional/ grados	
Años de experiencia en la enseñanza/evaluación de SL	
Cargo actual	
Lugar de trabajo	

Esta prueba se compone de 5 secciones: comprensión lectora, comprensión auditiva, competencia lingüística, expresión e interacción escrita y expresión e interacción oral.

- Los ítems de las secciones de comprensión lectora y auditiva evalúan tres dimensiones: 1) la identificación de información específica en textos especializados, 2) la inferencia de las ideas principales de los párrafos y 3) la inferencia de las ideas globales del texto.
- Los ítems de la sección de competencia lingüística miden la dimensión del dominio gramatical correspondiente a un nivel B2 en ELE.
- Las tareas de expresión escrita y oral evalúan dos dimensiones: 1) la modalidad expositiva y argumentativa y 2) el género académico.

Actividad evaluativa

Con el objetivo de asegurar la calidad del instrumento diseñado, solicito a usted que evalúe los ítems teniendo en cuenta los siguientes criterios:

1. Si usted considera que la pregunta evalúa la dimensión correspondiente, escriba “**SI**” en el espacio asignado y marque su opción en la escala de 5 puntos.
2. Si usted considera que la pregunta no evalúa la dimensión correspondiente, escriba “**NO**” en el espacio asignado.
3. Si usted considera que las preguntas podrían mejorarse, escriba sus observaciones o sugerencias en el mismo espacio donde evaluó SI/NO
4. Mediante una escala Likert de 5 puntos evalúe en qué medida las preguntas miden la dimensión señalada. Para ello, considere los siguientes indicadores:

1	2	3	4	5
Muy en desacuerdo	En desacuerdo	Indiferente	De acuerdo	Muy de acuerdo

DIMENSIÓN I: IDENTIFICA INFORMACIÓN ESPECÍFICA EN TEXTOS ESPECIALIZADOS

Secciones: Comprensión de lectura y comprensión auditiva	EVALÚA DIMENSIÓN	INDICADORES				
		1	2	3	4	5
Preguntas	SI/NO (observación)					
<i>Texto I</i>						
1. De acuerdo con el texto, la ingeniería realiza aportes relevantes al desarrollo de un país.						
2. Según el texto, la formación que reciben los ingenieros les impide encontrar trabajo.						
<i>Texto II</i>						
1. De acuerdo con el texto...						
2. Según el texto, la aptitud lingüística es...						
<i>Texto III</i>						
1. Según el texto, el aumento de la población a nivel mundial ha...						
2. De acuerdo con el texto, la biodiversidad...						
<i>Audio I</i>						
1. La energía eólica corresponde a la energía del viento.						
2. Según el audio, el objetivo de la instalación de la turbina es transforman la energía eléctrica en una energía limpia.						
<i>Audio II</i>						
1. De acuerdo con el audio ¿Cuál es la situación del río Copiapó?						
2. El objetivo del proyecto Agua Copiapó es...						
<i>Audio III</i>						
1. Según el audio, el petróleo es un combustible...						

2. El Centro de Biotecnología de la Universidad de Concepción está interesado en						
----------------------------------------------------------------------------------	--	--	--	--	--	--

DIMENSIÓN II: INFIERE LAS IDEAS PRINCIPALES DE LOS PÁRRAFOS

Secciones: Comprensión de lectura y comprensión auditiva	EVALÚA DIMENSIÓN	INDICADORES				
		1	2	3	4	5
Preguntas	SI/NO (observación)					
<i>Texto I</i>						
3. Las competencias de egreso se reflejan en el dominio disciplinar que debe mostrar todo profesional de ingeniería.						
4. Se puede inferir que las escuelas de Ingeniería han intentado implementar cambios en la formación académica de sus estudiantes.						
<i>Texto II</i>						
3. Según el texto, el estudiante de un segundo idioma aprende mejor la habilidad auditiva...						
4. En el aprendizaje de una segunda lengua, la habilidad motora...						
<i>Texto III</i>						
3. La situación actual de la biodiversidad indica que						
4. En el futuro se espera que haya...						
<i>Audio I</i>						
3. Según el audio, la instalación de turbinas mejorará la calidad energética en las zonas alejadas de la ciudad.						
4. De acuerdo con el audio, la gente que vive en zonas rurales presenta bajos niveles de calidad de vida.						
<i>Audio II</i>						
3. La tecnología utilizada en el proyecto Agua Copiapó...						
4. Según el audio...						

<i>Audio III</i>						
3.El Centro de Biotecnología puede llevar a cabo su investigación porque la región donde se encuentra cuenta con						
4. De acuerdo con el audio, las microalgas...						

DIMENSIÓN III: INFIERE LA IDEA CENTRAL DEL TEXTO

Secciones: Comprensión de lectura y comprensión auditiva	EVALÚA DIMENSIÓN	INDICADORES				
		1	2	3	4	5
Preguntas	SI/NO (observación)					
<i>Texto I</i>						
5. A partir de la lectura del texto se puede inferir que la sociedad requiere de egresados vinculados al ámbito laboral.						
<i>Texto II</i>						
5. ¿Cuál es la idea central del texto?						
<i>Texto III</i>						
5. ¿Cuál es la idea central del texto?						
<i>Audio I</i>						
5. La idea central del audio es dar cuenta del aporte de las energías renovables al suministro eléctrico.						
<i>Audio II</i>						
5. ¿Cuál es la idea central del audio?						
<i>Audio III</i>						
5. ¿Cuál es el tema principal del audio?						

DIMENSIÓN IV: DOMINIO GRAMATICAL

	EVALÚA DIMENSIÓN	INDICADORES				
Sección III: Competencia lingüística	SI/NO (observación)	1	2	3	4	5
Texto para completar con las formas gramaticales adecuadas.						

DIMENSIÓN V: MODALIDAD EXPOSITIVA Y ARGUMENTATIVA

	EVALÚA DIMENSIÓN	INDICADORES				
Tareas	SI/NO (observación)	1	2	3	4	5
<i>Tarea de expresión e interacción escrita:</i> -Escribe un ensayo sobre la formación académica a través de cursos a distancia por Internet/ <i>El Día Mundial sin Compras</i> . En el ensayo tienes que incluir los siguientes aspectos: Introduce tema, establece una postura respecto del tema, menciona las ventajas y desventajas sobre el tema.						
<i>Tarea de expresión e interacción oral:</i> La presentación incluye: Presentar el tema y su relevancia, dar a conocer tu punto de vista sobre el tema, indicar las ventajas y desventajas sobre el tema y concluir.						

DIMENSIÓN VI: GÉNERO ACADÉMICO

	EVALÚA DIMENSIÓN	INDICADORES				
Tareas	SI/NO (observación)	1	2	3	4	5
<i>Tarea de expresión e interacción escrita:</i> Ensayo						
<i>Tarea de expresión e interacción oral:</i> Exposición oral						

ANEXO 8: MUESTRAS DE TEXTOS ESCRITOS POR LA MUESTRA

Sujeto: 3

Lengua materna: francés

Área de estudios: Ingeniería

El agua, un futuro llena de incertumbres

Hace muchos años que estamos preocupado por el problema del petróleo, pero un problema tan grave se presenta también : el agua. Para nosotros, los países desarrollados, el agua no es un problema pero a un nivel mundial podemos encontrar muchas países donde el agua carece.

Eso cree guerras, muertos y problemas sociales. Por el momento no es un problema, pero cada años las reservas de agua potable disminuye, y en un futuro cerca eso podría provocar conflictos.

Hay muchas maneras para resolver eso, como la sensibilización de la gente a no gastar el agua. Campañas para informar la gente de los problemas que puede generar un consumo abusivo del agua, y como evitar eso.

Además la instalación de edificios o de tecnologías que permiten de controlar el agua para las grandes empresas sería una buena cosa.

Porque realmente no es la gente que utiliza más el agua sino que la industria donde no hay ningún control y pueden hacer lo que quieren.

Creo que no hay ventajas al consumo de agua en las industrias, pero es como un obligación para el mundo, hay muchos procesos donde el agua es necesario, por ejemplo para producir ropas, autos todas las cosas necesarias para la vida.

Pero hay muchas desventajas, en un primer lugar podemos ver que el consumo de agua es demasiado, por ejemplo para fabricar un traje de caballero requiere de 5.550 litros de agua.

En un segundo lugar hay problemas de contaminación, porque el agua va directamente en los ríos y los animales y la gente están directamente tocado.

Por fin, el agua no es problema tan importante por el momento pero es una cosa que la gente utiliza con abuso, como muchas cosas. Eso cree problemas ambientales, contaminación donde la biodiversidad sufre. Vamos a ver lo que políticos y científicos van a hacer para cambiar y mejorar los problemas.

Sujeto: 5

Lengua materna: alemán

Área de estudios: Ingeniería

La crisis de agua a nivel mundial

En el año 2014, la UNESCO informó que hay una crisis de agua a nivel mundial. La crisis está causada por el cambio climático, un tratamiento inadecuado de agua en algunos sectores de la industria y distintas condiciones del acceso a agua. En general, la crisis significa problemas graves para todos, porque la escasez de agua es un problema que afecta a millones de personas. Pero al otro lado están los consumidores en los países desarrollados, los cuales no quieren renunciar a sus productos que causan un gran volumen de agua.

En mi opinión, la crisis de agua es un tema muy importante que necesita la atención de la gente. En los países desarrollados consumimos una gran cantidad de agua para asegurar nuestra alta calidad de vida. Pero en muchos puntos hay un sobreconsumo que no se puede continuar. Me gusta mucho tener la posibilidad de seleccionar entre distintos productos aunque sé que las condiciones y métodos de muchas empresas no son adecuadas. Creo que es necesario cambiar su punto de vista y tener más atención cual tipo de ropa o otros bienes de consumo se compran. Eso no es solamente relacionado al problema de la escasez de agua, también es importante tomar en cuenta las condiciones de los trabajadores y la calidad de los productos, la cual está influenciada de nuestra demanda de precios bajos. Para resumir eso, no creo que el consumo de algunos países esté permitido de afectar la situación de la mayoría de la población mundial en una forma negativa. Tenemos que cuidarnos de nuestro consumo directo e indirecto de agua para evitar la existencia de gente sin acceso a agua limpia y potable.

A un lado, las ventajas del consumo de agua en la industria son evidentes: el consumo de agua permite la producción de bienes, por ejemplo, ropa, alimentos o papel. Y muchas veces no hay la opción de disminuir esta cantidad de agua utilizada. Pero este aspecto que se puede influir es el consumo innecesario de bienes. Al otro lado hay las desventajas del consumo de agua que se componen de distintos factores. Un factor es la contaminación de agua a causa del tratamiento industrial. Además se producen todos los bienes un consumo alto de agua utilizada, como ropa, en muchos países. La crisis causada por este consumo va a influir los países de los productores más grave que otros.

Finalmente, el sector de la industria afecta y apoya a la crisis de agua a nivel mundial. Pero, para resolver este problema es importante cuidarse también de su consumo y uso de agua. Además sea la tarea de todos dar apoyo a la gente que está afectada muy mal por la crisis.

Sujeto: 9

Lengua materna: sueco

Área de estudios: Ciencias Químicas

Nuestro huella de agua

La cantidad de agua que usamos en la vida cotidiana muchas veces es más grande de lo que creemos. Especialmente nuestro vestimientto está producido con procesos que consumen inmensas cantidades de agua, y aunque puede parecer que el agua potable es un recurso renovable natural todavía puede agotarse en sectores, y el agua contaminado a veces es difícil de limpiar para que sea útil de nuevo.

Las industrias de ropa usan mucho agua, por procesos de limpieza del producto o para diluir por ejemplo colorantes o otros compuestos que dan características al producto. Hay ventajas con usar grandes cantidades de agua, especialmente en limpiar el producto. Al limpiar el producto con más agua más de los tóxicos o sustancias innecesarios se van a diluir en el agua y separarse del producto.

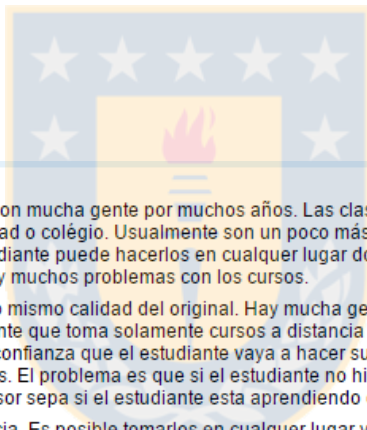
Lamentablemente esto también implica que estos compuestos, a veces muy tóxicos, se quedan en el agua que tiene que dejar la fábrica como residuos. Hay casos donde este agua no ha sido purificado o solo parcialmente purificado antes de que se deshace de ella, y que entonces contamina las aguas en las napas subteraneas y los ríos en el area al rededor de la producción.

Es difícil encontrar ropa producida en una manera ecológica en nuestras tiendas comunes y corrientes, pero sí se puede como parte de la sociedad aportar en alguna manera a disminuir el consumo, por ejemplo evitando a comprar grandes cantidades de ropa que se usa pocas veces y despues se deja en algún basurero. Se puede reciclar ropa dejandolo a una tienda de segunda mano, cambiar con un amigo cuando ya no lo usa tanto o en otras maneras disminuir su consumo de ropa nueva. Al final esto es una manera muy económica para asegurarnos que no nos quedamos sin el agua potable esencial para la vida.

Sujeto 12:

Lengua materna: inglés

Área de estudios: Letras



Cursos a Distancia

Los cursos a distancia han estado populares con mucha gente por muchos años. Las clases sirven como una manera de tomar clases académicas si no es posible ir a una universidad o colegio. Usualmente son un poco más barata de los cursos tradicionales. También los cursos son más fáciles acceder porque el estudiante puede hacerlos en cualquier lugar donde hay internet. Estos cursos parecen que una buena idea por mucha gente pero aún hay muchos problemas con los cursos.

A veces cuando algo es más barata no tiene lo mismo calidad del original. Hay mucha gente que piensa que esto es la verdad por los cursos a distancia. Es muy probable que la gente que toma solamente cursos a distancia no reciba una educación buena. Cuando no hay un profesor con el estudiante hay mucha confianza que el estudiante vaya a hacer su propio trabajo. Es muy probable que esto no sea la verdad por la mayoría de los estudiantes. El problema es que si el estudiante no hiciera su propio trabajo sin ayuda, no aprendería nada. Por lo tanto, no es posible que el profesor sepa si el estudiante esta aprendiendo el material o no.

Hay muchos beneficios de los cursos a distancia. Es posible tomarlos en cualquier lugar y cuando hay tiempo. Son más baratas y más fáciles por muchas pero no es lo mismo de los cursos normales. Es imposible obtener un educación real con solamente cursos a distancia.

Sujeto: 16

Lengua materna: francés

Área de estudios: Letras

Opción 1:

La enseñanza a distancia

Hoy en día, los cursos a distancia son más frecuentes que antes, gracias al desarrollo de internet y a su impacto sobre nuestras vidas. Así, los estudiantes pueden tomar clases sobre el internet y avanzar según su propio ritmo. Ese tipo de enseñanza presenta algunos aspectos muy positivos.

Primero, permite a cada estudiante de progresar a su ritmo. Si quiere terminar la materia de un ramo en una semana, puede hacerlo, o si prefiere tomar su tiempo, también lo puede hacer. Segundo, permite a algunas personas de poder tener un trabajo mientras estudiando, como a los mayores que quieren empezar nuevos estudios por ejemplo pero al mismo tiempo tienen que cuidar a su familia. Además, los cursos a distancia son muy cómodos para las personas que viven lejos de la universidad y que no tienen bastante recursos para pagar todo lo que concierne los cursos (como un departamento para estudiantes, los trayectos en coche o en tren cada día, la comida sobre el campus, gastos diversos, etc.)

Pero, este metodo presenta también algunos aspectos negativos. Uno de estos es la supresión del aspecto social de las clases reales. Quedarse solo en casa todos los días no es algo muy positivo para las relaciones sociales de los estudiantes, especialmente de los jovenes. Además, si el estudiante no entiende algo del curso, obviamente si puede enviar un correo electrónico a un profesor para preguntarle, pero no tendrá su respuesta inmediatamente y así deberá dejar su trabajo hasta que reciba su respuesta. Y eso es muy incómodo. También, a veces, no conocemos la autenticidad de los cursos. Y a he leído un artículo sobre este tema (los cursos presenciales y no presenciales) y se decía que había diplomas obtenidos de esta manera pero que no eran validos o que eran falsos. Así que hay que ser muy prudente.

En lo que me concierne, yo prefiero seguir clases reales que clases virtuales. Lo que me gusta más es el aspecto social que ningún clase sobre el internet puede reemplazar. Y además, creo que el contacto con el profesor es muy importante también. Las relaciones que se crean entre los profesores y sus estudiantes no pueden existir a través del internet.

Sujeto: 21

Lengua materna: francés

Área de estudios: Ingeniería

El desenvolvimiento de la formación académica a través internet

Hoy día, con la aceleración del tiempo, la mas grande movilidad de la gente y las ganas de siempre hacer mas, una nueva forma de aprender va tomando más fuerza: los cursos a distancia. Eses cursos permiten a los viajadores, a la gente que no tiene suficiente tiempo, a los que quieren aumentar sus conocimientos, ..., a todo el mundo de tomar cursos a cualquier momento y en cualquier lugar.

Este concepto me parece interesante en el sentido que permite a cualquier persona de poder tener acceso a la cultura con mucha libertad. Cualquier persona que quiere aprender, puede hacer lo. Pués corespone a enlargar el conocimiento a más gente y eso me parece fundamental.

El desventaja que puede estar encontrado es lo de la motivación. La escuela es un lugar hecho para aprender, y cualquier alumno que va a la escuela tiene que trabajar. En el caso de los cursos a distancia, el alumno puede faltar sus cursos y así pasar al lado de conocimientos necesarios a su educación. Los padres deben entonces estar atento a los niños para estar seguros que atenten a todos los cursos.

Para los cursos de idioma, un curso por Internet no es tan bueno que una verdadera discusion para aprender a expresarse.

Para concluir, pienso que este forma de formación es muy practica porqué puede servir a todo el mundo y en todo el mundo. Pero en el caso de niños, hay que cuidar a lo que hacen para estar seguro que toman bien los cursos que necesitan para la educación. Además, aprendendo una idioma, es necesario de incluir real discusiones en la formación para expresarse correctamente.

Sujeto: 28

Lengua materna: portugués

Área de estudios: Ciencias Políticas

El Uso de la Internet para la Formación Académica: Un Planteamiento Sobre la Realidad Actual

Con la Revolución industrial en el periodo pasado el mundo conoció un conjunto de cambios que contituen la llamada globalización. Además, esta globalización se caracteriza pela adopción y uso de la tecnologia para mejorar la vida del hombre. Así, en la actualidad hay lo que se llama "aldea global" dónde las personas a través de la internet, estando en diferentes cantos del mundo pueden comunicarse al mismo tiempo. Antes, la formación acdémica a distancia o por internet consistia en los alumnos se hacen presentes en la escuela para llevar materiales físicos y tenerem orientaciones para estudiar pero, hoy dia con la internet es posible que uno estudie sin necesitar ir la escuela, pode tener clases dónde habla, escribe, escucha, recibe y envia información, hace evaluaciones sin necesitar salir de su casa, país o dónde esté. Es importante adoptar las tecnologias porque ajudam na reducción de costos como tiempo y diñero y en la maximización de los gaños. No obstante, hay que señalar que hay igual desventajas en terminos de uso de la internet para la formación académica porque los alumnos no tienen el contacto con la realidad física, no tienen interacción con otros colegas para cambiar experiencias y la tecnologia crea problemas para la salud de los usuários. Así, haty que mezclar el uso de la internet con la realidad practica para permitir que los alumnos tengam conocimiento adecuado en su formación.

ANEXO 9: TRANSCRIPCIÓN DE UNA ENTREVISTA

Sujeto: 17

Lengua materna: francés

Área de estudios: Salud

Transcripción de la entrevista

Entrevistadora: (Introduce la entrevista y solicita a la estudiante que se presente y hable de sus estudios)

Sujeto 17: Me llamo XXX, estoy soy francesa y estudio en Bélgica, estudio en Bélgica. Eeee..Estoy acá en Chile para hacer un intercambio en Kinesiología mmm mmmm eeeee, como... Estoy en Chile desde un mes, y voy a quedar acá durante tres meses en el todo, todo el tiempo y voy a hacer como tres meses de prácticas un mes en una o después un otro en otro área y el tercero en otro área también yyy... vivo acá en Concepción con un pariente de dos personas mayores y con otras dos chicas de intercambio.

Entrevistadora: ¿Cómo ha sido tu experiencia en Kinesiología, en Chile?

Sujeto 17: Por el momento muy bien porque estoy en un área que es el área que conocí bien porque tuve una práctica en Francia antes de venir acá... me sentí bien, me sentí como acostumbré al ambiente y todo este... sí, las personas muy simpáticas y me explica bien todo lo que debo hacer.

[2 minutos para la preparación de la exposición]

Sujeto 17: Elige la opción dos el aprendizaje de un segunda lengua a través de un proceso de inmersión mmmm... elijo esta opción porque es mi caso en este momento so me parece elige conocer mejor mmm entonces mmm aprender una segunda idioma, lengua es, parece es muy difícil, es difícil si la aprendes, si la aprendes como si habla un idioma en Francia se va a hacer más difícil porque eeee porque cuando tú eres un inmersión tú puedes hablar las personas y entender todo el tiempo este idioma y y mmm[se ríe nerviosa] no sé mmm por ejemplo cuando aprendí, aprendí español en colegios parece más difícil porque estaba como mas mmm como decir, conjugación aprender palabras.... Todo esto y cuando tú eres un inmersión es como cosas de la vida, cosas que se pueden entender mejor que las cosas que tú vive como experiencia a.... para un punto de vista es que lo que dice que prefiero este, pero mmm... para las ventajas son que (disculpe no es bien...separado) las ventajas aa.. que como le dec, le he dicho entiende todo el tiempo el mismo idioma eee.. como decir es más fácil porque no sé [piensa...]

Entrevistadora: ¿Qué desventajas le ves tú al tema? ¿Es posible para todas las personas hacer un proceso de inmersión?

Sujeto 17: No, debe, debes tener una oportunidad de ir en otro lugar porque claro no todas las personas hacen este, las personas deben tener el *abio de hacer este, debe.... tener dinero para hacer este porque también es muy caro y ... debes, deben estar motivada porque ir lejos de su casa a de los

cosas que tú conoces diferente y difícil también porque son experiencias de vida mmm.. si este mmm... creo q...está una cosa que quería hacer ir en otro país para descubrir una otra idioma porque...pienso que para la abierta de *spread* (como se dice) que las personas de ver otras cosas que conocen, que conoces como formador o es como una buena cosa y ser inmersión no solamente por ejemplo ir en vacaciones durante una o dos semanas no es inmersión para mí, es como quedar más que un meses, dos meses si conocer las gentes y las costumbres y todo este.

Duración total de la entrevista: 9:23 minutos

Calificación de la entrevista:

Criterios	Comentarios	Puntaje asignado
Coherencia	Presenta algunas ideas enlazadas con conectores muy básicos (usa bastante el conector y). La exposición no se ciñe a la estructura indicada, solo menciona ventajas, no establece cuál es su punto de vista al respecto, no concluye.	1
Fluidez	Se hace entender con expresiones muy breves, y realiza un gran número de pausas y titubeos. Su pronunciación y articulación, a veces, dificulta la comprensión de lo que intenta decir (inmersión, Larrea, *abio).	2
Corrección	Muestra un escaso control gramatical. Comete un gran número de errores de pronunciación y de conjugación verbal, concordancia gramatical, etc.	1
Vocabulario	Tiene un dominio léxico limitado que le permite transmitir información básica y sencilla sobre el tema. En ocasiones, no se entiende lo que plantea (las ventares, abio, spread...). Comete errores de selección léxica, utiliza el pronombre “este” y el sustantivo “cosa” como comodines.	1,5

ANEXO 10: CUESTIONARIO DE EVALUACIÓN SOBRE LA PRUEBA EFA PARA LOS ESTUDIANTES EXTRANJEROS

Cuestionario de evaluación sobre la prueba de español

Nombre: _____

I. Marca el número que mejor represente tu opinión en relación con la pregunta

1: Muy en desacuerdo 2: En desacuerdo 3: Indiferente 4: De acuerdo 5: Muy de acuerdo

Preguntas	1	2	3	4	5
1. Me sentí cómodo (a) respondiendo la prueba en computador					
2. Me parece interesante que la prueba esté en computador					
3. El ambiente de la prueba es fácil de usar					
4. La prueba me pareció fácil de responder					
5. Las instrucciones de la prueba son claras					
6. Los textos y audios de la prueba son interesantes					
7. Fue fácil comprender los textos de la prueba					
8. Fue fácil comprender los audios de la prueba					
9. Encontré fácil la sección de gramática					
10. Me pareció fácil la tarea de escritura					
11. Me pareció fácil la tarea oral					
12. Considero adecuado el tiempo de duración de la prueba					

II. Responde estas preguntas

1. ¿Qué parte de la prueba te pareció más fácil? ¿Por qué?

2. ¿Qué parte de la prueba te pareció más difícil? ¿Por qué?
