



UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
INSTITUTO DE INVESTIGACIÓN Y DESARROLLO EDUCATIVO
MAESTRÍA EN CIENCIAS EDUCATIVAS



**DISEÑO DE UN SISTEMA PARA ADMINISTRAR EN LÍNEA INSTRUMENTOS DE
EVALUACIÓN DEL APRENDIZAJE.**

TESIS

**Para obtener el grado de
MAESTRA EN CIENCIAS EDUCATIVAS**

Presenta

Zulma Viridiana Sarabia Ocampo.

Ensenada, B. C. México, Noviembre de 2014



Universidad Autónoma de Baja California
Instituto de Investigación y Desarrollo Educativo
Maestría en Ciencias Educativas



**“Diseño de un sistema para administrar en línea instrumentos de
evaluación del aprendizaje”**

TESIS

Que para obtener el grado de
MAESTRA EN CIENCIAS EDUCATIVAS

Presenta

Zulma Viridiana Sarabia Ocampo

APROBADO POR:

Dr. Luis Ángel Contreras Niño
Directora de tesis

**Dr. Juan Carlos Rodríguez
Macías**
Sinodal

Dr. Joaquín Caso Niebla
Sinodal

Dr. Lewis McAnally Salas
Sinodal





Ensenada, B.C. a 07 de noviembre de 2014

ASUNTO: Voto aprobatorio sobre trabajo de tesis de grado de Maestría.

Alicia Alelí Chaparro Caso-López
Coordinadora de la Maestría en Ciencias Educativas
Presente.

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por la C. **ZULMA VIRIDIANA SARABIA OCAMPO** para poder presentar la defensa de su examen y obtener el grado de Maestría en Ciencias Educativas, me permito comunicarle que he dado mi **VOTO APROBATORIO**, sobre su trabajo intitulado:

“DISEÑO DE UN SISTEMA PARA ADMINISTRAR EN LÍNEA INSTRUMENTOS DE EVALUACIÓN DEL APRENDIZAJE”.

Esperando reciba el presente de conformidad, quedo de Usted.

Atentamente

Una firma manuscrita en tinta negra, que parece ser "L. Contreras", sobre una línea horizontal.

Dr. Luis Ángel Contreras Niño



Ensenada, B.C. a 07 de noviembre 2014

ASUNTO: Voto aprobatorio sobre trabajo de tesis de grado de Maestría.

Alicia Alelí Chaparro Caso-López
Coordinadora de la Maestría en Ciencias Educativas
Presente.

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por la **C. ZULMA VIRIDIANA SARABIA OCAMPO** para poder presentar la defensa de su examen y obtener el grado de Maestría en Ciencias Educativas, me permito comunicarle que he dado mi **VOTO APROBATORIO**, sobre su trabajo intitulado:

“DISEÑO DE UN SISTEMA PARA ADMINISTRAR EN LÍNEA INSTRUMENTOS DE EVALUACIÓN DEL APRENDIZAJE”.

Esperando reciba el presente de conformidad, quedo de Usted.

Atentamente

Dr. Juan Carlos Rodríguez Macías



Ensenada, B.C. a 07 de noviembre de 2014

ASUNTO: Voto aprobatorio sobre trabajo
de tesis de grado de Maestría.

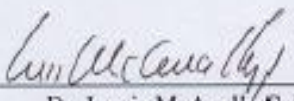
Alicia Aleli Chaparro Caso-López
Coordinadora de la Maestría en Ciencias Educativas
Presente.

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por la C.
ZULMA VIRIDIANA SARABIA OCAMPO para poder presentar la defensa de su examen y
obtener el grado de Maestría en Ciencias Educativas, me permito comunicarle que he dado mi
VOTO APROBATORIO, sobre su trabajo intitulado:

***"DISEÑO DE UN SISTEMA PARA ADMINISTRAR EN LÍNEA INSTRUMENTOS DE
EVALUACIÓN DEL APRENDIZAJE".***

Esperando reciba el presente de conformidad, quedo de Usted.

Atentamente



Dr. Lewis McAnally Salas



Ensenada, B.C. a 07 de noviembre 2014

ASUNTO: Voto aprobatorio sobre trabajo de tesis de grado de Maestría.

Alicia Aleli Chaparro Caso-López
Coordinadora de la Maestría en Ciencias Educativas
Presente.

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por la **C. ZULMA VIRIDIANA SARABIA OCAMPO** para poder presentar la defensa de su examen y obtener el grado de Maestría en Ciencias Educativas, me permito comunicarle que he dado mi **VOTO APROBATORIO**, sobre su trabajo intitulado:

"DISEÑO DE UN SISTEMA PARA ADMINISTRAR EN LÍNEA INSTRUMENTOS DE EVALUACIÓN DEL APRENDIZAJE".

Esperando reciba el presente de conformidad, quedo de Usted.

Atentamente

Una firma manuscrita en tinta negra, que parece ser "J. Caso", sobre una línea horizontal.

Dr. Joaquín Caso Niebla

Dedicatoria

*A mi papá por todo el cariño y paciencia, pero
sobre todo, por siempre confiar en mí.*

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), al Instituto de Investigación y Desarrollo Educativo (IIDE), y a la Unidad de Evaluación Educativa (UEE) por el apoyo económico brindado durante la realización de esta tesis.

Agradezco de todo corazón a mi tutor Dr. Luis Ángel Contreras, por todo el apoyo, dedicación, paciencia y orientación durante la elaboración de la tesis. Gracias por la confianza que depositó en mí para la realización de esta tesis tan controversial.

Gracias a mis sinodales, Dr. Juan Carlos Rodríguez, Dr. Joaquín Caso y Dr. Lewis McAnally, por tomarse el tiempo de hacer revisiones, observaciones y sugerencias.

Gracias a todos mis profesores de la maestría de quienes aprendí mucho. Un especial agradecimiento al M.T.R.I. José Luis Ramírez por todos los consejos dentro y fuera del salón de clases.

Muchas gracias a mis amigos de la maestría, los cuales se convirtieron en mi familia, Anita, Dalia, Fabián, Ivonne, Liliana, Lorvic y Nancy, gracias por compartir alegrías, tristezas, risas, lágrimas, estrés, abrazos, comidas, cafés y viajes, esta etapa de mi vida no hubiera sido la misma sin ustedes, gracias por estar siempre al pendiente.

Gracias a mis padres, por ser un soporte en mi vida.

Gracias Coste, por acompañarme en cada desvelo.

Gracias Víctor, por quererme tanto.

Índice general

1	Capítulo I	1
1.1	Antecedentes	1
1.1.1	Conceptualizaciones de evaluación	1
1.1.2	Instituciones de evaluación del aprendizaje a gran escala en México	2
1.1.3	Evaluación del aprendizaje mediante el uso de la computadora	5
1.1.4	Software para la evaluación del aprendizaje en línea	7
1.2	Planteamiento del problema	8
1.3	Objetivos del estudio	10
1.3.1	Objetivo general	10
1.3.2	Objetivos específicos	11
1.4	Justificación del estudio	11
2	Capítulo II Marco conceptual	15
2.1	Evaluación del aprendizaje a gran escala	15
2.2	Evaluación del aprendizaje a gran escala en línea	19
2.2.1	Antecedentes de la evaluación a gran escala en línea	19
2.2.2	Desarrollo de la evaluación del aprendizaje a gran escala en línea	21
2.2.3	Tipos de software que se emplean para desarrollar y administrar pruebas	22
2.2.4	Sistemas para administrar evaluaciones en línea	25
2.3	Estándares para la administración en línea de pruebas a gran escala.	29
2.3.1	Características de los administradores de exámenes	32
2.3.2	Derechos y responsabilidades de quienes administran y responden las pruebas estandarizadas	33
2.3.3	Derechos de los examinados	39
2.4	Estándares de seguridad para las evaluaciones en línea	40
2.5	Principios relativos al diseño universal de las evaluaciones	42
2.6	Diseño de la interfaz gráfica	48
3	Capítulo III Método	52
3.1	Apego a la normatividad para la administración estandarizada en línea de exámenes aplicados a gran escala.	53
3.1.1	Incorporación de los estándares psicométricos de APA-AERA-NCME y CENEVAL	54
3.1.2	Incorporación de principios del diseño universal de las evaluaciones	59
3.2	Análisis de requerimientos	62
4	Capítulo IV Resultados	68
4.1	Resultados del análisis de requerimientos	68
4.1.1	Referencias	68
4.1.2	Interfaces con el Software	69
4.1.3	Proveedor	70
4.1.4	Funciones del sistema	70
4.1.5	Requisitos de funcionalidad	70
4.2	Resultados del diseño de SALIDEA	81

4.2.1	Estructura general del diseño de SALIDEA	81
4.2.2	Interfaz con el administrador	84
4.2.3	Interfaz con el examinado	86
4.2.4	Interfaz con los padres de familia	87
4.2.5	Interfaz con los profesores	88
4.2.6	Interfaz con los directores	88
4.2.7	Interfaz con autoridades educativas	89
5.	Discusión	90
5.1	Limitaciones y comentarios finales	102
Referencias bibliográficas		106
Apéndices		114

Índice de tablas

2.1	Críticas y limitaciones expresadas sobre la evaluación del aprendizaje a gran escala	17
2.2	Principales ventajas de la evaluación del aprendizaje a gran escala	18
2.3	Principales programas de cómputo para administrar pruebas para evaluar el aprendizaje	23
2.4	Sistemas para administrar en línea evaluaciones del aprendizaje	25
2.5	Derechos y responsabilidades de los desarrolladores y administradores de pruebas al momento de desarrollar y seleccionar las pruebas adecuadas	33
2.6	Derechos y responsabilidades de los desarrolladores y administradores de pruebas al momento de administrar la prueba y el reporte de resultados	34
2.7	Derechos y responsabilidades de los desarrolladores y administradores de pruebas al momento de reportar e interpretar los resultados	36
2.8	Nuevos estándares relacionados con la administración, calificación, reporte e interpretación de los resultados de las pruebas	37
2.9	Elementos de la evaluación basada en el diseño universal	46
2.10	Dimensiones y características de máxima legibilidad en el párrafo	47
3.1	Estándares psicométricos relativos al desarrollo de la prueba, que fueron atendidos al diseñar SALIDEA	55
3.2	Estándares psicométricos relativos a la administración de las pruebas, que fueron atendidos al diseñar SALIDEA	56
3.3	Estándares psicométricos relativos a la calificación de la prueba, la interpretación de sus resultados y elaboración de informes de resultados de la prueba, que fueron atendidos al diseñar SALIDEA	58
3.4	Etapas del procedimiento para incorporar el diseño universal de las evaluaciones al diseñar SALIDEA	60
3.5	Características de máxima legibilidad textual consideradas al diseñar SALIDEA	62
3.6	Análisis de requerimientos para el diseño de SALIDEA	63

Índice de figuras

2.1	Rueda de libros de Ramelli	20
2.2	Máquina para aplicar pruebas de opción múltiple de Pressey	21
2.3	Pantalla para elegir opciones	28
2.4	Ejemplo de interfaz gráfica de la prueba	29
2.5	Ejemplos de interfaces graficas con organización, relaciones entre elementos e íconos, inapropiados (izquierda) y apropiados (derecha).	50
3.1	Etapas que se siguieron para diseñar SALIDEA.	53
4.1	Estructura general de SALIDEA	81

Resumen

El presente estudio tuvo como objetivo diseñar un sistema de cómputo, alineado a estándares de calidad técnica de las evaluaciones, que permita administrar en línea instrumentos estandarizados como escalas y exámenes de referencia criterial para evaluar el aprendizaje, así como analizar y elaborar reportes de resultados específicos para diferentes usuarios. El diseño del software, al que se denominó SALIDEA (Sigla de Sistema para Administrar en Línea Instrumentos De Evaluación del Aprendizaje) considera que, una vez elaborado el sistema, permitirá incorporar a él un examen de referencia criterial y la correspondiente documentación que avala su calidad técnica, de manera que sea posible llevar a cabo su aplicación a gran escala de manera estandarizada y, posteriormente al concluir la evaluación, analizar y reportar de manera inmediata, simple y significativa los resultados generales y específicos, para los diferentes usuarios de la información. Todo lo anterior apegado a estándares psicométricos internacionales propuestos por el *Joint Committee on Testing Practices* (AERA, APA y NCM, 1999, 2014), por diferentes universidades norteamericanas y australianas, así como a los estándares nacionales publicados por el Consejo Asesor Externo del Centro Nacional de Evaluación para la Educación Superior (2000). Además, el diseño de SALIDEA considera el cumplimiento de los lineamientos del diseño universal de las evaluaciones que han sido propuestos por diferentes instancias (Center for Universal Design, 1997; Thompson y Thurlow, 2002; Thompson, Johnstone & Thurlow (2002), Johnstone, Altman y Thurlow, 2006), de manera que se atiendan necesidades de los diferentes usuarios de las pruebas, incluyendo a personas con discapacidades físicas o psicológicas.

Se considera que SALIDEA se administre desde una página web y que opere bajo la Licencia Creative Commons Atribución – No Comercial – Compartir Igual 4.0 Internacional, a la vez que agregue la funcionalidad del programa PARRERC©, el cual es una aplicación distribuida bajo la licencia “Reconocimiento – No Comercial – Compartir Igual 3.0 Unported”, según la clasificación de la organización Creative Commons. De igual modo, SALIDEA incorpora la funcionalidad de la aplicación PAGI© distribuida bajo la Licencia Creative Commons Reconocimiento – No Comercial – Compartir bajo la misma licencia 2.5 México License. Para ello, el diseño del sistema prevé que sea desarrollado en el lenguaje de programación JAVA.

Para incorporar los estándares que SALIDEA debe atender en su diseño, se realizó una amplia revisión de estándares propuestos por diversas instituciones. Por ejemplo el Código de prácticas justas de evaluación en educación (APA, AERA & NCME, 2004), que plantea que los desarrolladores de la prueba así como los administradores de la prueba, tienen derechos y responsabilidades. Por otra parte el Joint Committee on Testing Practices (2014) publicó recientemente la nueva versión de los estándares que han sido referentes importantes para las evaluaciones durante mucho tiempo. También se revisaron estándares nacionales desarrollados Consejo Asesor Externo del CENEVAL (2000) los cuales plantean algunos lineamientos para el buen funcionamiento de las pruebas a gran escala.

Además se consideró una necesidad el proteger la integridad de las pruebas, por lo que se revisaron las recomendaciones de la International Test Commission (2014) que

plantea la importancia de tener en cuenta los problemas vinculados a la seguridad y la integridad del test, así como la validez de las puntuaciones, cuando se aplica una evaluación en línea mediante métodos de evaluación convencionales. Por lo tanto, antes de tomar una decisión fue necesario revisar los estándares en cuestión de seguridad de las pruebas, propuestos por diversas organizaciones.

Para realizar el diseño de SALIDEA también se revisaron diversos tipos de software, para revisar la interfaz gráfica con la que estos contaban, es decir, los elementos visuales que utilizan como ventanas, iconos y menús para ejecutar comandos para abrir, borrar o mover archivos. Por otra parte se decidió incorporar lineamientos del diseño universal de las evaluaciones debido a que dicho movimiento tiene un enfoque para el diseño de las pruebas que en principio busca maximizar la accesibilidad para todos los examinados. El diseño universal, exige que los desarrolladores de la prueba sean claros en el constructo que se desea evaluar, hace posible participar en las evaluaciones a la más amplia gama posible de estudiantes e incluso reducir la necesidad de efectuar adaptaciones de los test o de evaluaciones alternativas, para ampliar el acceso a las pruebas.

Una parte medular del trabajo, es el análisis de requerimientos, el cual detalla los rasgos operacionales del software, es decir, permite dejar claro lo que el software debe hacer y bajo qué circunstancias debe hacerlo. Especificar los rasgos operacionales del software y los conecta con la interfaz gráfica del sistema. Para ello se especificó el requerimiento y se mostró el diseño del mismo.

Finalmente se elaboró una representación de la estructura del sistema en la que se muestran las funciones de SALIDEA y la incorporación de las aplicaciones externas PAGI y PARRERC y su funcionalidad dentro del sistema. Para ser más específico el diseño se optó que se realizara una retícula por cada usuario, la cual mostrará los elementos específicos y las opciones de navegación que tendrá cada usuario del sistema.

Palabras clave: Evaluación en línea, Estándares evaluativos, Diseño de software libre, Diseño universal, interfaz gráfica.

Capítulo I. Introducción.

1.1 Antecedentes

1.1.1 Conceptualizaciones de evaluación

Por evaluación entendemos cualquier método sistemático para obtener información, que sirve para derivar inferencias sobre características de las personas, objetos o programas; se trata de un proceso sistemático para medir o valorar las características o la ejecución de las personas, programas u otras entidades, para propósitos de derivar inferencias; en ocasiones se utiliza el término como sinónimo de prueba (AERA, APA, NCME, 2014).

Por su parte, la Organización para la Cooperación y el Desarrollo Económicos (OCDE), responsable de la prueba PISA, define a la evaluación como el proceso para determinar el valor de una actividad. Se trata de proporcionar información creíble y útil, que permita incorporar las enseñanzas aprendidas en el proceso de toma de decisiones, identificar el logro de los objetivos, así como la eficiencia, la eficacia, y el impacto del desarrollo (OCDE, 2013).

En México, la Secretaría de Educación Pública (SEP) define la evaluación de los aprendizajes como el proceso que permite obtener evidencias, elaborar juicios y brindar retroalimentación sobre los logros de aprendizaje de los alumnos a lo largo de su formación; por tanto, es parte constitutiva de la enseñanza y del aprendizaje (SEP, 2011).

Desde hace más de un siglo, en todo el mundo se empezaron a utilizar las pruebas estandarizadas como un medio para hacer más eficiente la evaluación y obtener resultados confiables, debido a que se evaluaba de la misma manera a todos los estudiantes (SEP, 2007).

Según Martínez (2009), Thorndike fue uno de los principales pioneros de las pruebas estandarizadas, él estaba convencido de que las escuelas tenían problemas de calidad, pues consideraba que las evaluaciones que desarrollaban los maestros tenían deficiencias graves. Por ello, Thorndike buscó elaborar instrumentos que permitieran establecer comparaciones entre los niveles de rendimiento de alumnos de diferentes escuelas.

Por su parte, como se citó en Martínez (2009), Shepard considera que el contenido de los textos sobre evaluación utilizados en las instituciones formadoras de maestros, prevalece la idea de las evaluaciones llevadas a cabo por los docentes deben aplicar en el aula tienen que reproducir las evaluaciones a gran escala. Por lo tanto, los maestros de las escuelas tendrían que aprender a elaborar preguntas estructuradas y a analizar estadísticamente los resultados de los instrumentos. Además, los docentes deberían de cuidar la validez y confiabilidad de tales instrumentos, en la misma forma en que debe hacerse a gran escala.

1.1.2 Instituciones de evaluación del aprendizaje a gran escala en México

En el siglo XX se desarrolló de manera global la psicometría, especialmente en Estados Unidos de América. Sin embargo en México permaneció ausente durante el siglo de oro de la psicometría, y no fue sino hasta mediados de los años noventa que en México surgió el interés por asuntos psicométricos que ayudaran a la evaluación rigurosa del aprendizaje a gran escala (Contreras, 2009).

Una de las instituciones pioneras que practica la evaluación estandarizada en México es el Centro Nacional de Evaluación de la Educación Superior (CENEVAL), que

fue creado en 1994 con el propósito elaborar exámenes para el ingreso a las instituciones de educación media superior y superior. CENEVAL busca mejorar la calidad de la educación mediante evaluaciones válidas, confiables y pertinentes de los aprendizajes, que contribuyan a la toma de decisiones fundamentadas (CENEVAL, 2013).

En 2002 fue creado el Instituto Nacional para la Evaluación de la Educación (INEE). La función principal de esta institución es la evaluación de la educación obligatoria: preescolar, primaria y secundaria, así como la educación media superior, tanto pública como privada. Además, contribuye a la evaluación de la formación inicial y continua de los docentes. Con el fin de obtener información para reflexionar acerca de la calidad de los aprendizajes y para la mejora educativa, a partir de 2005 se diseñaron los Exámenes de la Calidad y el Logro Educativos (Excale) (INEE, 2012). Cabe señalar que dicha institución fue elevada recientemente al rango de organismo público autónomo, con atribuciones para coordinar el Sistema Nacional de Evaluación Educativa, así como evaluar la calidad, el desempeño y los resultados del Sistema Educativo Nacional, en todas sus modalidades y servicios, así como generar los lineamientos de evaluación a que deberán ajustarse las autoridades educativas del país (INEE, 2013).

En el contexto de la Secretaría de Educación Pública (SEP), en 2005 la Dirección General de Evaluación cambió su designación a Dirección General de Evaluación de Políticas (DGEP) (SEP, 2012). A fines de ese mismo año, dicha dependencia dio inicio al programa evaluativo más ambicioso del país: el desarrollo de la Evaluación Nacional del Logro Académico en Centros Escolares (ENLACE), prueba

de aplicación censal en la educación básica diseñada para que los docentes, directivos, autoridades educativas, investigadores y escolares de todo el país, dispongan de una medida válida, objetiva y confiable, del estado actual del logro académico de los estudiantes de educación básica (SEP, 2013).

El desarrollo de dichas evaluaciones suscitó un enorme interés por las evaluaciones del aprendizaje en el país. Muestra de ello son la creación de institutos y unidades de evaluación del aprendizaje en varias entidades federativas, como son los casos del Estado de México, Sonora, Chiapas y Baja California, por mencionar algunos. De igual modo, dio un fortalecimiento de las dependencias de evaluación en los sistemas educativos estatales, no solo para apoyar la aplicación estandarizada de las evaluaciones nacionales como ENLACE y Excale o de las internaciones como PISA, sino para desarrollar y aplicar sus propios instrumentos.

Por ejemplo, el Instituto de Evaluación Educativa del Estado de México (IEEEM), tiene como objetivo evaluar al Sistema Educativo Estatal en todos los componentes que lo integran, con excepción de las instituciones de educación superior a las que la ley otorgue autonomía, generando información útil y relevante (IEEEM, 2013).

Otra organización que desarrolla evaluación a nivel estatal es el Instituto de Innovación y Evaluación Educativa del Estado de Sonora (IIEEES), el cual brinda a la comunidad los resultados obtenidos en las evaluaciones, con lo que se pretende mejorar la calidad de la educación que reciben los niños y jóvenes sonorenses (IIEEES, 2010).

Por su parte, el estado de Chiapas también estableció un organismo de evaluación, llamado Instituto de Evaluación e Innovación Educativa del estado de Chiapas (INEVAL), que tiene como función evaluar la calidad del Sistema Educativo Estatal, con respecto al desempeño de los actores, procesos y resultados (INEVAL, 2013).

En el caso de Baja California, se creó la Unidad de Evaluación Educativa (UEE) del Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California, la cual opera en conjunto con el Sistema Educativo Estatal de Baja California, y tiene como objetivo monitorear la calidad del aprendizaje que logran los estudiantes de preescolar, primaria y secundaria en Baja California, como parte de una estrategia evaluativa más general e integral, la cual permita contar con información válida y confiable para ayudar a la toma de decisiones del Sistema Educativo Estatal (UEE, 2010).

1.1.3 Evaluación del aprendizaje mediante el uso de la computadora

Tradicionalmente la evaluación del aprendizaje, tanto en el aula como a gran escala, ha sido mediante la aplicación de pruebas en la modalidad de lápiz y papel, en donde el examinado generalmente tiene que marcar o escribir las respuestas que a su juicio parecen pertinentes.

Con el tiempo, y a medida que avanzó el desarrollo tecnológico, se fueron creando diversos tipos de instrumentos de evaluación; entre ellos, los exámenes por computadora, que agregaron la funcionalidad de interacción con los estudiantes. La aplicación de dichos exámenes consistía inicialmente en insertar un CD-ROM en el

equipo de cómputo y al hacerlo se ejecutaba automáticamente un programa. Los alumnos solo debían seleccionar la respuesta correcta, con lo que se hacía más eficiente que el profesor calificará a cada estudiante, examen tras examen y que se redujera la subjetividad al hacerlo (Fuentes, 2007).

El desarrollo de la evaluación computarizada, que administra los exámenes convencionales formulados con preguntas de opción múltiple por computadora, presentó ventajas evidentes pues cualquier persona, independientemente de su experiencia en el uso de computadoras, podía contestar el examen y solamente era necesario tener la noción de un teclado normal de una máquina de escribir (Backhoff, 2002).

Posteriormente, el desarrollo de las computadoras y del Internet hicieron cada vez más viable la aplicación en línea de pruebas estandarizadas a gran escala; es decir, con programas que permiten administrar en tiempo real la aplicación de exámenes, analizar los resultados obtenidos y elaborar reportes de resultados a diferentes usuarios, de modo que se facilite el análisis y la realimentación oportunos de los procesos de enseñanza y aprendizaje de los estudiantes y profesores, mediante los resultados obtenidos.

Un ejemplo de ellos es el Examen de Habilidades y Conocimientos Básicos (EXHCOBA), que se ha diseñado desde 1992 en Baja California como proyecto de la Universidad Autónoma de Baja California (UABC), que evalúa las nociones y no las precisiones del conocimiento, así como la comprensión y la aplicación de los aprendizajes que se adquirieron en la educación básica y media superior. El examen inició con una versión en lápiz y papel y posteriormente se ha desarrollado en forma

computarizada. Así, el estudiante debe responder las preguntas que se le presentan en una pantalla de computadora, de la misma manera que lo haría en un examen de lápiz y papel (Backhoff, Rosas, Cuevas, Larrazolo y Velasco, 2002).

Una de las empresas más importantes, de origen estadounidense, que ha elaborado diversos tipos de software para desarrollar evaluaciones, es *Assessment Systems Corporation* (2012), misma que cuenta con un sistema avanzado para administrar la evaluación en línea. Se trata de una compañía de expertos en psicometría que desarrollado software para efectuar análisis de ítems y versiones de pruebas como el clásico ITEMAN, y que se dedican también a la elaboración de exámenes en diferentes ramas, como lo son la educativa y la empresarial. Los exámenes pueden entregarse impresos en papel, pero actualmente pone a disposición un software denominado *FastTEST Web* que permite la administración en línea de exámenes, el cual no se puede adquirir pues solo está disponible el servicio de evaluación.

1.1.4 Software para la evaluación del aprendizaje en línea

La evaluación del aprendizaje en línea es un área relativamente reciente que ha tenido un vertiginoso crecimiento y que involucra la aplicación de exámenes en pantalla, ya sea mediante el uso de un sistema local o un sistema basado en la Web (eAssessment Association, s/f). Usualmente implica transferir exámenes desarrollados en papel a la pantalla de la computadora, mediante alguna combinación de métodos

que crean situaciones virtuales en las cuales el examinado procesa información a fin de llegar a una respuesta.

De acuerdo con Rodríguez (2006) se han desarrollado tres tipos de software para apoyar la evaluación del aprendizaje en línea:

- *Entornos virtuales de formación*, los cuales ayudan a los docentes a operar o administrar un módulo o curso completo en línea, lo que incluye la distribución de contenidos, interacciones con los alumnos por correo electrónico, o chat y la aplicación de evaluaciones de los alumnos.
- *Herramientas de autor*, mismas que consisten en programas de cómputo destinados a la creación de ejercicios o tareas evaluativas.
- *Software específico más complejo*, que requiere el uso de un servidor para poner en la red a toda la institución, tanto para la creación y publicación de los exámenes, como para capturar y procesar los resultados de los estudiantes.

1. 2 Planteamiento del problema

Como se comentó en párrafos anteriores, nuestro país padeció durante mucho tiempo de una escasa cultura de la evaluación y no fue sino hasta en años recientes que surgió una preocupación al respecto en dependencias como la SEP y el INEE, y otras más en los estados, donde se han creado diversos institutos de evaluación educativa que aplican miles de exámenes cada año. No obstante, dichas aplicaciones se hacen con la modalidad de lápiz y papel, lo que las hace poco ecológicas y origina cuantiosos gastos que se deben pagar por la impresión de exámenes y otros materiales, la compra de hojas de respuesta ópticas, la capacitación al personal para la

aplicación estandarizada de las pruebas, así como cubrir los honorarios y otros gastos que ocasiona tanto su administración, como los eventos posteriores a la aplicación de los exámenes, como son la recolección y lectura de las hojas de respuestas, la estructuración de las bases de datos con las respuestas de los examinados, el análisis de la información y la elaboración de reportes de resultados para diferentes destinatarios, acciones que pueden llegar a ser también procesos complejos y tardados (Contreras, 2009; Castañeda, 2013).

Aunque los actuales sistema de cómputo para administrar pruebas en línea pueden ser una buena opción para resolver dichos problemas, la mayor parte de los programas que existen se encuentran en el idioma inglés, el servicio que brindan es costoso y en muchos casos el software no se vende, sino que se adquiere el servicio del desarrollo y administración de la prueba. Además, muchos de ellos están enfocados al sistema K-12 en Estados Unidos (equivalente a la educación básica en México) y no en las necesidades evaluativas de México.

Por otra parte, la mayoría de dichos sistemas no se encuentran apegados a estándares de calidad psicométrica que han sido propuestos por organizaciones internacionales (APA, AERA, NCME) y nacionales (CENEVAL) que norman esta clase de instrumentos; o bien no toman en cuenta algunos de los lineamientos del diseño universal de las evaluaciones, surgidos recientemente para orientar el desarrollo y administración de pruebas justas para los examinados, de manera que sean utilizados como herramientas de inclusión, al maximizar la accesibilidad y evitar realizar adaptaciones a los test.

Finalmente, cabe hacer notar la falta de tecnología en México para administrar pruebas en línea a gran escala, que esté disponible para varios actores del ámbito educativo nacional, como son las instituciones de evaluación, personal técnico, directivos o profesores en servicio, que deben elaborar sus propios exámenes y otros cuestionarios, con el fin de evaluar el progreso de sus estudiantes.

La problemática descrita en los párrafos anteriores, deja ver con claridad que existe un área de oportunidad para el diseño de un sistema de cómputo libre que permita la administración de pruebas en línea, que sea eficaz y eficiente y que tome en cuenta los estándares psicométricos y lineamientos del diseño universal de las evaluaciones.

1.3 Objetivos del estudio

Para superar las limitaciones y dificultades antes señaladas, en el presente estudio se propone lograr los siguientes objetivos:

1.3.1 Objetivo general

- Diseñar un sistema de cómputo de acceso libre, que permita administrar en línea y de manera estandarizada instrumentos para evaluar el aprendizaje a gran escala, el cual atienda los estándares internacionales y nacionales de calidad técnica y se apege a los lineamientos del diseño universal de las evaluaciones.

1.3.2 Objetivos específicos

- Efectuar un análisis de requerimientos para el diseño de un sistema de cómputo que permita administrar en línea pruebas de logro de referencia criterial y de respuesta seleccionada.
- Diseñar la interfaz gráfica del sistema de cómputo, de tal manera que cada usuario tenga acceso protegido y pueda transitar por los componentes y opciones que se diseñen de manera exclusiva para él.
- Diseñar los componentes necesarios para integrar al sistema las aplicaciones PAGI® (De Los Santos, 2010) y PARRERC® (Castañeda, 2012), las cuales permiten realizar análisis gráficos de ítems, así como análisis de los resultados generales y por área evaluada, y elaborar reportes de resultados acordes a las necesidades de información de diferentes usuarios.
- Prefigurar la funcionalidad del sistema con un ejemplo que represente los componentes de una prueba y las relaciones entre ellos, que se requieren para su administración mediante un sistema de cómputo.

1.4 Justificación del estudio

Uno de los problemas que con mayor frecuencia enfrentan las instituciones educativas para realizar un trabajo sistemático de evaluación del aprendizaje, consiste en una insuficiente cultura de la evaluación que prevalece entre los actores que participan en el proceso educativo, como son alumnos, profesores y directivos; lo que conduce a prácticas evaluativas que no son congruentes con los contenidos, enfoques y propósitos del modelo educativo, o bien son insuficientes o superficiales (SEP, 2011).

Al respecto, un sistema de cómputo de acceso generalizado y gratuito puede contribuir desarrollar una cultura evaluativa al hacer la evaluación más enfocada, significativa, útil y eficiente, puesto que los altos costos y los tiempos largos de espera entre la aplicación a gran escala de evaluaciones de lápiz y papel, y la obtención de los resultados y el análisis de los mismos, podrían superarse con el diseño de un software de acceso libre como el que se propone en este trabajo, el cual tras ser desarrollado permita administrar en línea, tanto escalas como pruebas de respuesta seleccionada, y al mismo tiempo haga posible analizar automáticamente los resultados obtenidos, así como contar de inmediato con reportes de la evaluación.

De manera relacionada, una ventaja del uso del sistema que se diseñó es la incorporación de dos programas de software con características similares, que ya están disponibles: el programa PARRERC (Castañeda, 2012), que proporciona análisis y reportes automáticos de los resultados de las pruebas y que con ello permite la retroalimentación inmediata a estudiantes, padres, maestros, y administradores escolares, lo que a su vez posibilita actuar de manera oportuna para la toma de decisiones para mejorar la calidad educativa, así como el programa PAGI (De los Santos, 2010) el cual elabora análisis gráficos de ítems, y con ello muestra de manera sencilla y significativa la calidad técnica de los ítems.

Otro rasgo que aporta el sistema propuesto, tiene que ver con una administración de exámenes en línea con mayor flexibilidad con respecto a la ubicación de los actores en el tiempo y en el espacio, debido a que eventualmente se pueden

aplicar y responder los test en cualquier momento y desde cualquier lugar donde se tenga una conexión a internet.

Una cualidad distintiva del sistema que aquí se propone es su claro compromiso con los estándares psicométricos nacionales (CENEVAL, 2000) e internacionales (AERA, APA, & NSME, 1999, 2014), particularmente los que tienen que ver con el desarrollo y administración de test, el análisis de resultados y la elaboración de reportes a los usuarios, así como el garantizar justicia a los examinados al evitar el sesgo y la ofensa en las pruebas y considerar los lineamientos del diseño universal de las evaluaciones (Thompson, Johnstone y Thurlow, 2002), de manera que propicien la participación efectiva de un rango más amplio de examinados

Finalmente, existen otras las razones por las cuales un sistema para la evaluación del aprendizaje en línea puede ser justificada. En su minuciosa exposición de argumentos al respecto, Ridgway, McCusker & Pead (2006) señalan entre otros los siguientes: puede contribuir a evitar el colapso de los actuales sistemas basados en lápiz y papel; permitir evaluar habilidades para la vida que resultan valiosas; puede ser mejor para los usuarios al proporcionar pruebas sobre demanda con realimentación y diagnósticos inmediatos y con resultados más exactos; puede mejorar la calidad técnica de los test al mejorar la confiabilidad de la calificación y propiciar una evaluación más auténtica, al hacer posible el uso de procesadores de palabras para producir ensayos, de hojas de cálculo, calculadoras y otros dispositivos que están disponibles; hace posible observar el cambio de variables en el tiempo; permite desarrollar interfaces para presentar datos complejos, probar las habilidades para la

solución de problemas y habilidades de proceso como la comprensión y representación de problemas, la generación y prueba de hipótesis y el descubrimiento de relaciones entre eventos, entre otros tipos de aprendizaje complejo que pueden evaluarse como las habilidades metacognitivas, la creatividad y el trabajo grupal; así como efectuar análisis automatizados de las ejecuciones de los estudiantes y mejorar la confiabilidad de las evaluaciones. Además de resultar ecológica y eficiente.

Capítulo II. Marco conceptual

En este capítulo se describen los inicios formales y desarrollos posteriores de la evaluación del aprendizaje a gran escala, particularmente la evaluación en línea. A continuación se exponen las principales críticas y ventajas que se han mencionado en la literatura especializada sobre la evaluación a gran escala. Por otra parte, se hace una revisión de los tipos de software para el desarrollo de pruebas en línea, así como de sistemas en línea para administrar pruebas estandarizadas. Además, se describen los estándares internacionales y nacionales relativos a la evaluación estandarizada de aprendizaje y en particular las características que deben poseer los desarrolladores y administradores de esta clase de pruebas y cuáles son los derechos y obligaciones que deben de tener los examinados, cuando se les administran línea pruebas de gran escala, así como los estándares de seguridad que se deben cumplir al administrar las pruebas. Posteriormente se señalan los lineamientos que han emergido recientemente para propiciar el diseño universal de las evaluaciones.

2.1 Evaluación del aprendizaje a gran escala

La evaluación a gran escala es una herramienta valiosa para apoyar los esfuerzos de mejora de la calidad educativa, de manera que pueda ofrecerse a cada estudiante la retroalimentación precisa sobre sus puntos fuertes y débiles, lo que es esencial para mejorar su aprendizaje. Sin embargo, se debe considerar el peligro de que las pruebas en gran escala produzcan consecuencias negativas para la calidad educativa, si se les comprende y utiliza mal (Martínez, 2009).

La evaluación del aprendizaje a gran escala se define como aquella que se administra a cientos o miles de examinados, a fin mostrar cómo se relaciona su ejecución con estándares o un dominio curricular definido; o bien para compararla con la de otros que también fueron evaluados. Por su escala, típicamente utiliza ítems de respuesta seleccionada y mide conocimiento en áreas de contenido amplias, más que habilidades específicas. En general, este tipo de evaluaciones tiene propósitos y audiencias fuera del aula, mismos que incluyen asumir responsabilidad, evaluación de programas, certificación de competencias y selección de estudiantes para programas especiales. Aunque la evaluación a gran escala puede informar y guiar las decisiones instruccionales en el aula, normalmente ese no es su propósito primario (Contreras, 2008).

A pesar de que la evaluación del aprendizaje a gran escala se practica desde tiempos remotos (considérense por ejemplo, los exámenes para el ingreso al servicio civil chino que se aplicaron desde el siglo VI [Berkshire Encyclopedia of China, 2010]), no es sino hasta mediados de 1980 que alcanzó un gran auge, principalmente en Estados Unidos de Norteamérica (De Pascale, 2003).

En la tabla 2.1 se documentan las principales críticas y limitaciones que diferentes autores han expresado sobre la evaluación del aprendizaje a gran escala.

Tabla 2.1
Críticas y limitaciones expresadas sobre la evaluación del aprendizaje a gran escala

Autor	Año	Crítica o limitación
Mehrens	1992	Mide sólo conocimiento declarativo y no procedimental
Mumford, Baughman, Supinski & Andersen	1998	Se centra en el resultado y no en el proceso
Mehrens, Wiggins	1992 1991	No cubre adecuadamente el dominio evaluado
Powell	1990	El formato de opción múltiple limita las evaluaciones y supone otras habilidades distintas a las evaluadas
Kane, Popham	2002 2001	Apoya decisiones políticas que no han sido adecuadamente examinadas y se basa en medidas con importantes limitaciones psicométricas
Canadian Teachers Federation	1999	Actúa como forma de control social y educativo que subvierte la autonomía profesional de los educadores
American Educational Research Association, Darling-Hammond	2000 2004	Responsabiliza a los profesores para obtener resultados con recursos no equitativos y contribuye a que se generen sentimientos de baja autoeficacia
McNeil, Valencia, Villarreal	2000 2003 2003	Hace caso omiso de factores claves que afectan el rendimiento de los estudiantes, tales como el nivel socio-económico, el estado y lengua de origen, así como la salud física y emocional.
Anderson y Postl, Glaser, Linn y Bohrnstedt, Hess y Brigham, Wideen, O'Shea, Pye y Ivany	2001 1997 2000 1997	Socava la calidad de la educación, al penalizar el pensamiento divergente, la creatividad y el trabajo intelectual en general.
Kohn, Smith y Fey, Volante,	2002 2000 2004	Estrecha y distorsiona el plan de estudios al fomentar "el enseñar para el examen".
Stiggins, Falk	2002 1998	Proporciona información que es de poca utilidad para mejorar la práctica en el aula.
Meaghan y Casas, Miller y Tovey	2001 1996	Daña el autoconcepto y el nivel de compromiso del estudiante.
Shepard	2003	Se utiliza para hacer comparaciones entre distintas escuelas de distintos distritos, discriminando a los que obtuvieron bajos puntajes.

Pellegrino y Glaser	2001	Evalúa solo ciertas habilidades, centrándose en lo que el alumno sabe y no en lo que el alumno no domina.
---------------------	------	---

Fuente: Elaboración propia, basada en Burger & Krueger (2003); y Volante (2006).

No obstante, diversos autores han destacado también ventajas importantes al practicar la evaluación a gran escala. La tabla 2.2 muestra las principales de ellas.

Tabla 2.2
Principales ventajas de la evaluación del aprendizaje a gran escala

Autor	Año	Ventaja
Anderson,	1990	Hace que los estudiantes trabajan más duro y los impulsa a desempeñarse mejor en las evaluaciones internacionales como TIMSS y PISA
Bishop	2001	
Cizek,	2001	Permite identificar las prácticas de enseñanza más exitosas de profesores competentes
Sanders y Horn	1998	
Goldberg y Roswell	2000	Mejora el pensamiento crítico y reflexivo de los docentes al planificar la instrucción
Earl y Torrance,	2000	Conduce a un incremento positivo del conocimiento de los educadores acerca de las pruebas y asuntos relacionados con la evaluación
Runte	1998	
Gambell y Hunter,	2004	Mejora las prácticas de instrucción y de evaluación de los docentes, particularmente cuando se involucran en la construcción de estas evaluaciones
Green	1998	
Wideman	2002	Estimula la investigación acción que sirve como base para el mejoramiento escolar
Anderson y Postl,	2001	Proporciona a las escuelas información valiosa sobre las consecuencias de sus prácticas pasadas y sobre la efectividad de su programa
Taylor y Tubianosa	2001	
Shepard	2003	Se utilizan los resultados de las evaluaciones para tomar decisiones importantes acerca del profesorado y los estudiantes.
		Se utiliza para hacer comparaciones entre distintas escuelas de distintos distritos.
		Diagnostica las fortalezas y debilidades de los planes

Autor	Año	Ventaja
		de estudios.
		Reconoce los objetivos del plan de estudio que se han alcanzado.

Fuente: Elaboración propia, basada en Volante (2006) y Shepard, (2003).

Respecto a la evaluación a gran escala en nuestro país, según Martínez (2009) el desarrollo de pruebas en gran escala para la educación básica inició en la década de 1970, y se desarrolló a partir de la de 1990 con las pruebas para evaluar el Factor Aprovechamiento Escolar del Programa de Carrera Magisterial. La tendencia resaltó una década después, con las pruebas Excale del Instituto Nacional para la Evaluación de la Educación (INEE) a partir de 2003, y en particular con las pruebas censales ENLACE que desarrolló la Dirección General de Evaluación de Políticas de la SEP y que se aplicaron desde 2006.

2.2 Evaluación del aprendizaje a gran escala en línea

2.2.1 Antecedentes de la evaluación a gran escala en línea

Los primeros antecedentes documentados de máquinas que se utilizaron para apoyar la enseñanza y la evaluación del aprendizaje datan del siglo XVI. En 1588, se dio a conocer un diseño para una "rueda de libros" (véase figura 2.1) que pretendía que el estudio fuera más rápido y conveniente, al simplificar el acceso a información relevante y con ello facilitar el aprendizaje (Szenaris, 2011).



Figura 2.1 Rueda de libros de Ramelli (1588)

Fuente: <http://www.futuropasado.com/>

Por su parte, Isaac Pitman fue pionero en la educación y la evaluación del aprendizaje a distancia. En 1840, en Gran Bretaña, Pitman comenzó a impartir cursos de taquigrafía en los cuales mandaba a sus estudiantes una lección mediante el correo y ellos la estudiaban, practicaban y respondían lo que les solicitaba; para enviar después de regreso la lección contestada por el mismo medio (Matthews, 1999).

Más adelante, Pressey desarrolló en Estados Unidos una máquina para facilitar la evaluación del aprendizaje mediante test con ítems de opción múltiple. La máquina (véase figura 2.2) mostraba la base de cada ítem y cinco posibles respuestas, y el estudiante debía presionar la tecla con el número que tenía la respuesta que juzgaba correcta. Al presionar la tecla con su respuesta, se perforaba una hoja de respuestas que contenía la clave de cada ítem y enseguida se mostraba el siguiente ítem. Para

conocer el resultado de la prueba solo debía deslizarse la hoja de respuestas que contenía las perforaciones que coincidían con la clave en la parte de atrás de la máquina y observar el puntaje en el contador (Benjamin, 1988).

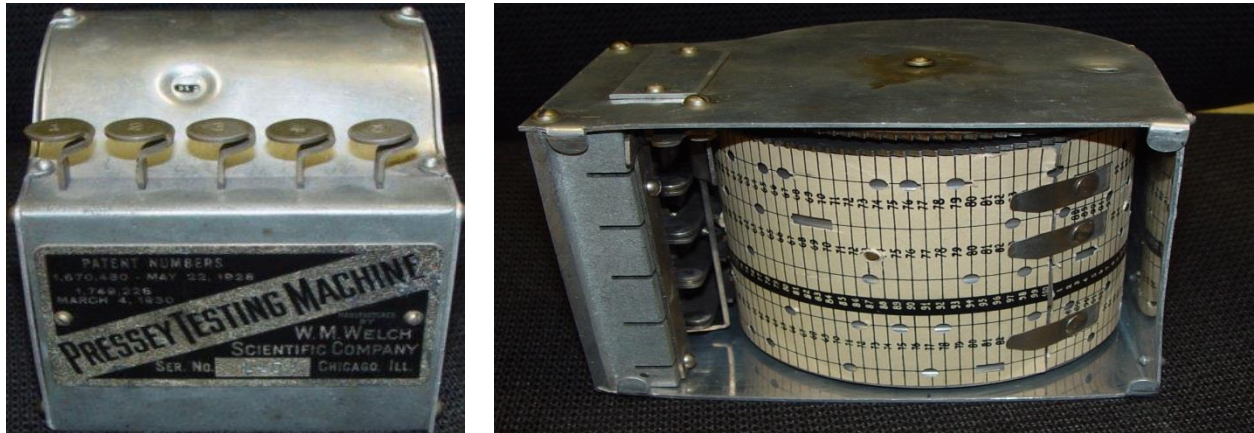


Figura 2.2 Máquina para aplicar pruebas de opción múltiple de Pressey (1928)

Fuente: http://en.wikipedia.org/wiki/File:Pressey_Testing_Machine_1.jpg

2.2.2 Desarrollo de la evaluación del aprendizaje a gran escala en línea.

Hamilton y Klein (1999) afirman que la evaluación del aprendizaje en línea es una alternativa prometedora debido a los avances tecnológicos, al incremento de computadoras en las escuelas y a que proporcionan oportunidades para explorar nuevas alternativas de evaluación.

Las computadoras y el Internet ofrecen nuevos enfoques para la aplicación de exámenes. En general, estas pruebas tienen tres características principales. En primer lugar, los ítems pueden ser administrados de manera adaptativa, lo que significa que el examen presenta al examinado ítems con mayor o menor nivel de dificultad dependiendo de si responde correctamente o no cada ítem. En segundo lugar, el sistema hace uso de diferentes tipos de ítems, incluyendo los de respuesta

seleccionada (por ejemplo, opción múltiple) y respuesta construida (por ejemplo, respuesta breve o ensayo). En tercer lugar, la evaluación se administra vía internet en vez de confiar solo en estaciones de trabajo independientes (Hamilton y Klein, 1999).

Bennette (1998) asegura que la evaluación a gran escala en línea ha pasado por distintas etapas; para ello describe tres generaciones en las que se distinguen el propósito de la prueba, su formato y contenido, el lugar de entrega de la prueba, y la medida en que las pruebas capitalizan las nuevas tecnologías.

Según dicho autor, en general la primera generación estableció la infraestructura básica para la prueba electrónica; en la segunda generación, las pruebas a gran escala se someten a cambio cualitativo, pero sus propósitos y los mecanismos de administración siguen siendo esencialmente los mismos; por su parte, la última generación brinda un replanteamiento de los objetivos y los mecanismos de la evaluación a gran escala.

2.2.3 Tipos de software que se emplean para desarrollar y administrar pruebas

Con el paso de los años, se han ido creando tipos de software para hacer la evaluación en línea de una manera más rápida y sencilla. La principal clasificación de estos instrumentos considera aquellos que están destinados a la evaluación en el aula y los que están destinados a la evaluación a gran escala, es decir a cientos o miles de examinados.

A continuación, en la tabla 2.3 se presenta una revisión de algunos de los principales programas de cómputo destinados a aplicar en línea exámenes para evaluar el aprendizaje.

Tabla 2.3
Principales programas de cómputo para administrar pruebas para evaluar el aprendizaje

Nombre	Empresa o autor	¿Qué hace?	Costo	Destinatario	¿Cómo evalúa?	URL
Achievement Series	Scantron	Permite a educadores desarrollar y administrar las pruebas en línea y en papel, capturar resultados de inmediato y producir informes basados en estándares.	-Costo según opciones. \$43 dólares el paquete mínimo	-Profesores del sistema k- 12 en Estados Unidos	- Se analizan las formas y se producen informes en línea inmediatamente después de una prueba en línea. - Informes de acceso basados en estándares para un estudiante, clase, escuela, distrito y más. - Analiza los datos por subgrupo y demografía.	http://www.scantron.com/achievementseries/
AnswerGarden	Creative Heroes	-Herramienta para retroalimentación minimalista. Puede utilizarse como una herramienta para intercambiar ideas en línea o incrustarlo en una web o blog como encuesta o libro de visitas.	-Gratuito	- Administradores de sitio, bloggers, profesores, equipos creativos y más	-Envía una notificación (correo) con los resultados de las personas que respondieron la prueba.	http://answergarden.ch/
Articulate Quizmaker '09	Articulate Global Inc.	-Es un software con funcionalidades para la creación de SCORM o concursos Flash, compatible con AICC mediante una interfaz de usuario.	-Muestra gratis por 30 días. Se puede rentar el servicio por 19.99 dólares al mes y comprar el programa por 1,398 dólares. La versión más completa cuesta 1,896 dólares	-Empresas -Maestros	-Las opciones incluyen ítems con asignación al azar, integración de contenidos audiovisuales, ítems ramificados basados en la respuesta anterior y la provisión de retroalimentación vía correo electrónico o LMS.	http://www.articulate.com/products/quizmaker.php
Assess	The Training Registry	-Es una herramienta basada totalmente en web que permite desarrollar y aplicar evaluaciones en línea, como encuestas y exámenes.	-Está disponible sobre una base anual con cuota de suscripción para uso ilimitado	-Empresas -Escuelas	-La herramienta de encuesta presenta resultados inmediatos, actualizados y con tanto o tan poco detalle como se requiera.	http://www.assess.biz/
Avancert	Avancert Limited	-Avancert puede utilizarse como herramienta de autor y entregar tareas de evaluación basadas en video y en formas convencionales. -Análisis de informes de rendimiento y aleatorización de ítems (no hay dos evaluaciones con el mismo contenido)	-Existe una versión gratuita la cual cuenta con pocas funciones, mientras que la versión de paga (2,499 dólares) cuenta con	-Organizaciones educativas que ofrecen capacitación -Profesionales de recursos humanos -Organismos que otorgan licencias de certificación	-Después de que se completan las evaluaciones, los resultados pueden enviarse automáticamente a cada estudiante y se mantiene automáticamente una transcripción actualizada de resultados.	http://www.avancert.com/

Tabla 2.3
Principales programas de cómputo para administrar pruebas para evaluar el aprendizaje

Nombre	Empresa o autor	¿Qué hace?	Costo	Destinatario	¿Cómo evalúa?	URL
			variedad de funciones.	-Profesionales de marketing -Centros de atención telefónica		
BrainCog	Blue Six	-El sistema de evaluación en línea de BrainCog apoya la construcción de pruebas de opción múltiple mediante una interfaz personalizable. Incluye opciones como asignación de ítems al azar y pruebas cronometradas.	-Prueba gratis (10 exámenes) y el paquete más costoso 399 dólares por 200 exámenes.	-Maestros	-Informes: Promedio de puntuaciones, número de aprobados y reprobados; o bien un informe detallado para cada estudiante.	http://brain-cog.com/intro.asp
Buildmytest	Edtech systems	-Una solución completa para bancos de ítems al nivel de distrito escolar, construcción y entrega de la prueba en línea, administración e informes de resultados de la prueba.	-Pide llamar para conocer los precios	-Maestros	-Los informes muestran solo información relevante sobre el test. Hay decenas de informes que se pueden utilizar para analizar la prueba del estudiante	http://www.edtechsystems.com/products/buildmytest.htm
ClassMarker	Classmarker PtyLtd	-Es un constructor de pruebas en línea. Es fácil de usar, personalizable para negocios, formación educativa, evaluación con pruebas y cuestionarios calificados al instante.	-Prueba gratis por 30 días; después el precio depende del paquete que se necesite	-Empresas y escuelas	-Se pueden compartir los exámenes entre varios administradores de pruebas.	http://www.classmarker.com/online-testing/how-to-create-online-quiz/
FasTest	Assessment Systems Corporation	-Es una empresa que se dedica a la construcción de pruebas en cualquier área, para distintos objetivos.	-Llamar para conocer precios	-Empresas y escuelas	-Cuentan con banco de ítems, con pruebas adaptativas. -Las pruebas pueden ser en línea o entregadas a domicilio.	http://www.fasttestweb.com/why-ftw/

Fuente: Elaboración propia basada en la revisión de los sistemas que se describen.

2.2.4 Sistemas para administrar evaluaciones en línea

Actualmente existen diversos sistemas para administrar evaluaciones del aprendizaje en línea; entre ellos SEA, ARCOTS, QUESTAR y OAS. En la tabla 2.4 se describen estos sistemas y se explican las características particulares que tienen.

Tabla 2.4
Sistemas para administrar en línea evaluaciones del aprendizaje.

Sistema	¿Qué es?	¿Cómo funciona?	Origen	URL
SAE	Es un sistema que utiliza evaluaciones con actividades diseñadas por docentes que, junto con técnicos en evaluación e inspectores escolares, proponen pruebas para obtener información sobre algunos aspectos del aprendizaje y para reflexionar sobre la enseñanza.	El docente habilita a sus alumnos para realizar las pruebas disponibles. Los alumnos ingresan al sitio, responden la prueba y luego ésta queda disponible para el maestro. Para la aplicación en el aula se diseñaron dos portales: uno para los maestros y otro para los alumnos.	Montevideo Uruguay	http://www.ane.edu.uy/sea/
ARCOTS	Permite al profesorado administrar pruebas y otros instrumentos de evaluación en línea. El sistema también tiene potencial para ofrecer en el futuro evaluaciones con tareas interactivas complejas, que darán a los maestros información más detallada sobre las habilidades requeridas y las estrategias a ser utilizadas por los estudiantes para resolver problemas de manera colaborativa.	ARCOTS solo está disponible para usuarios registrados en el proyecto de asociaciones de evaluación y aprendizaje y el proyecto de evaluación de la alfabetización. Las instituciones deben dar de alta sus escuelas y sus estudiantes. Así mismo, los maestros deben de incluir las pruebas que quieren que ARCOTS administre.	University of Melbourne	http://education.unimelb.edu.au/arc/arcots
OAS	OAS representa un recurso dedicado a las escuelas, los distritos, los maestros, estudiantes y padres de familia que permite la instrucción en el aula permanente y aprendizaje de los estudiantes.	Permite a los administradores crear pruebas. Mediante la asignación de evaluaciones a nivel el distrito; los maestros y administradores pueden compartir informes y diálogos comunes en las estrategias de enseñanza con los datos que se	Georgia Department of Education	https://www.georgiaoas.org/sevlet/a2l

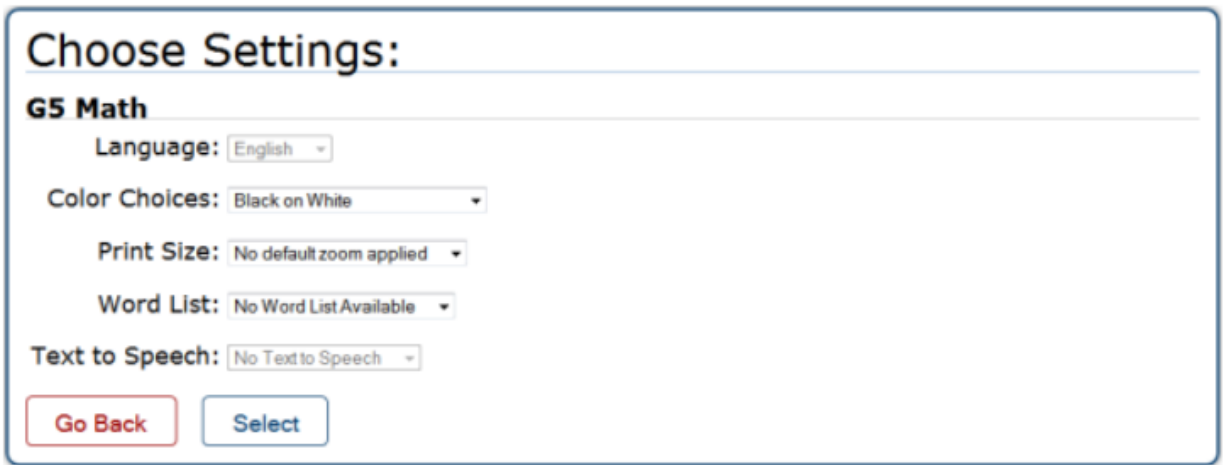
Sistema	¿Qué es?	¿Cómo funciona?	Origen	URL
		encuentren en todos los niveles de la administración.		
OAKS	Ayuda a las escuelas y los distritos escolares a desarrollar políticas y procedimientos que maximizan la utilidad de los datos de evaluación y minimizan el impacto en la enseñanza de la administración de exámenes.	Los administradores de la prueba crean sesiones de prueba y administran la prueba en línea al estudiante durante la sesión. Los estudiantes deben utilizar el explorador 2013- 2014, y deben de ser dados de alta por el administrador de la prueba.	Oregón, Estados Unidos	http://www.oaks.k12.or.us/portal/
SARAS	Es un sistema para la administración de pruebas estandarizadas, se puede utilizar en escuelas de nivel básico y superior	Se pueden crear grupos y roles dentro del mismo; el administrador puede crear 27 opciones de ítems diferentes; permite un reporte personalizado; provee de monitoreo durante el examen, y al terminar la prueba se realiza un análisis detallado del examen con base en la teoría de la respuesta al ítem.	Cambridge, Estados Unidos.	http://testing-assessments.excelinda.com/?gclid=C MvQ8L 3I6LUC Fag7M godpAo APA
HOSTED TEST	Es una aplicación de software en línea para la creación de pruebas y evaluaciones. Se pueden crear inmediatamente pruebas personalizadas sin la necesidad de instalación de software, descargas o programación. Todos los aspectos del desarrollo de la prueba, la administración y los informes se hacen en línea	Funciona en 4 pasos: 1. Crear la prueba introduciendo sus ítems. Puede tener cualquier longitud, número ilimitado de preguntas y respuestas. 2. Modificar y personalizar, añadiendo logo, imágenes, colores, fondos, botones de navegación personalizados y texto. 3. Poner a prueba. Los estudiantes pueden ver la prueba directamente en su sitio web o puede abrirlo	Mission Viejo, California, Estados Unidos	http://www.hostedtest.com/home.html

Sistema	¿Qué es?	¿Cómo funciona?	Origen	URL
		desde un hipervínculo. Lo verán exactamente como se diseñó. Los resultados se tabulan y se almacenan en una base de datos segura.		

Fuente: Elaboración propia basada en la revisión de los sistemas consultados.

Entre los sistemas para administrar pruebas que fueron revisados, el más desarrollado fue “Smarter Balanced” (página web disponible en: <http://www.smarterbalanced.org/smarter-balanced-assessments/>), el cual ofrece un entrenamiento virtual práctico y teórico a los administradores del sistema y a los estudiantes que serán evaluados. Este sistema está apegado a la mayoría de los estándares internacionales que se describen en la siguiente sección. Smarter Balanced contiene una guía en donde explica de manera detallada cada icono que se utiliza dentro del sistema.

Además, permite configurar el sistema según las preferencias de los usuarios. Para ello, muestra las opciones que están disponibles al momento de acceder al sistema, como son el idioma, el color del texto y del fondo de la pantalla, el tamaño de la letra, y un glosario en diversos idiomas. Cuando se seleccionan las opciones que corresponden a las necesidades del usuario, ya sea como administrador o como estudiante, aparecerá una pantalla que resume las opciones que fueron seleccionadas, en caso que se desee cambiar alguna.



Choose Settings:

G5 Math

Language: English ▾

Color Choices: Black on White ▾

Print Size: No default zoom applied ▾

Word List: No Word List Available ▾

Text to Speech: No Text to Speech ▾

Go Back **Select**

Figura 2.3 Pantalla para elegir opciones
Fuente: Smarter Balance

En la siguiente pantalla de Smarter Balanced se muestra un ítem de una prueba, y se identifican las opciones que están disponibles para los examinados al momento de responder, como son una calculadora, ampliar o disminuir el tamaño de la imagen, guardar el examen, ver un tutorial acerca del ítem, ir al siguiente ítem o regresar y marcar el ítem para su revisión.

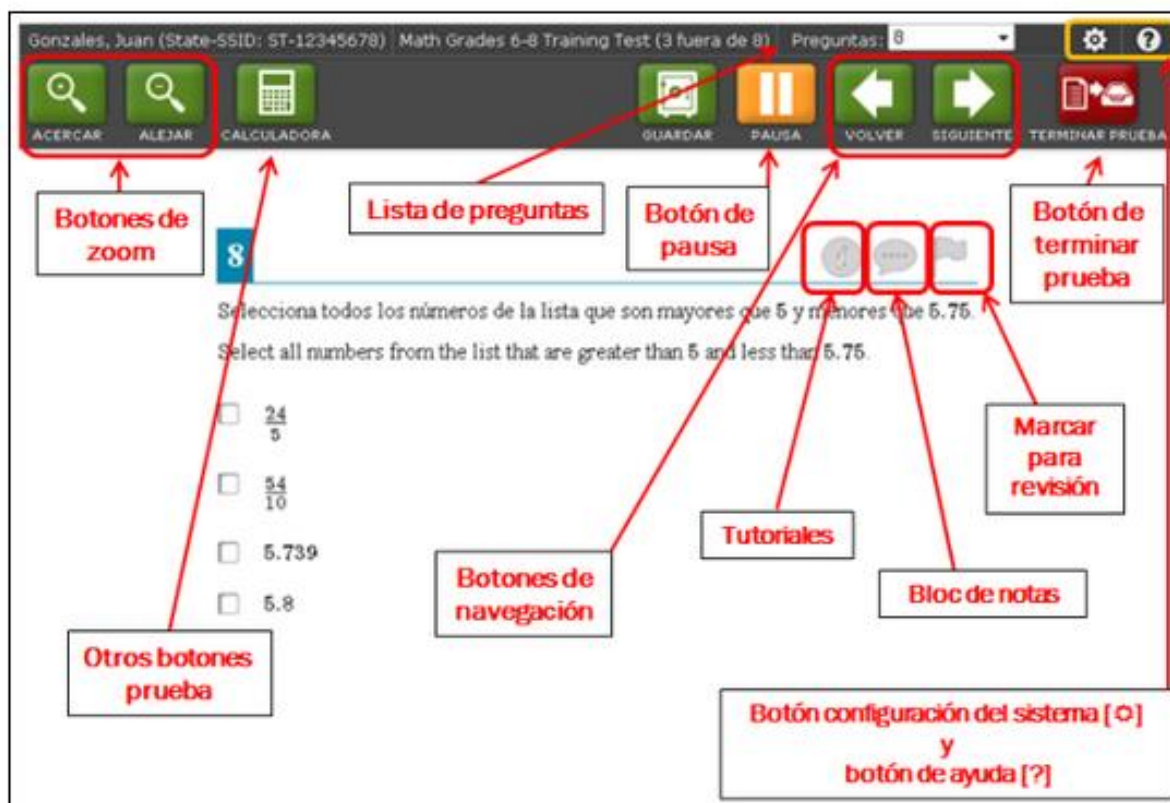


Figura 2.4 Ejemplo de interfaz gráfica de la prueba
Fuente: Smarter Balance

2.3 Estándares para la administración en línea de pruebas de gran escala en línea.

Dado el poderoso impacto que tienen los resultados de las evaluaciones a gran escala sobre las vidas de quienes responden los test, sus padres, profesores y autoridades educativas, ha sido necesario normar su diseño, elaboración, administración, calificación y el análisis y reporte de sus resultados. Al respecto, la comunidad internacional de evaluadores se ha orientado por los estándares desarrollados por el *Joint Committee on Testing Practices* (1999; 2004; 2014), integrado por la Asociación Norteamericana de Psicología (APA), la Asociación Norteamericana de Investigación Educativa (AERA) y el Consejo Nacional para la

Evaluación en Educación (NCME). Para el caso de México, el Consejo Asesor Externo del CENEVAL adaptó los estándares del mencionado comité conjunto de 1999, para su uso en nuestro país (Consejo Asesor Externo del CENEVAL, 2000).

Para que las puntuaciones en las pruebas puedan ser útiles e interpretables, se requiere que la prueba sea administrada y calificada de acuerdo con las instrucciones del desarrollador del instrumento. Cuando las instrucciones sobre las condiciones de administración de la prueba y su calificación siguen los mismos procedimientos detallados para todos los examinados, se dice que la prueba está estandarizada. La estandarización permite asegurar que todos los examinados tienen la misma oportunidad para demostrar sus competencias; por otro lado, mantener condiciones de seguridad del test también permite asegurar que nadie tenga una ventaja injusta. La adherencia a los procedimientos de administración estandarizada resulta más importante a medida que se incrementa el impacto de los resultados de la prueba (APA, AERA & NCME, 2014).

Asimismo, la preocupación por la equidad en el tratamiento puede requerir, para algunas pruebas, que todos los examinados tengan administradores de la prueba calificados con los que puedan comunicarse y sentirse cómodos en la medida de lo posible. Cuando interviene la tecnología, es importante que los examinados tengan una experiencia a priori del equipo que se le proporcionará, y cuenten con equipo con velocidades de procesamiento similares y que se les proporcionen imágenes con claridad y tamaño similares, entre otros aspectos de los medios de comunicación. Los procesos para la estandarización de la administración de una prueba deben ser

cuidadosamente documentados por el desarrollador de la prueba y seguidos escrupulosamente por el administrador de la prueba. Aunque la estandarización ha sido un principio fundamental para asegurar que todos los examinados tengan la misma oportunidad de demostrar su nivel de logro con respecto al constructo que se pretende medir en una prueba, a veces se necesita flexibilidad para ofrecer a algunos examinados oportunidades equivalentes al responder la prueba. En estos casos, los procesos regulares que conllevan las pruebas estandarizadas, que no suponen ningún desafío particular para la mayoría de los examinados, pueden impedir que grupos o individuos específicos demuestren con precisión su nivel con respecto al constructo de interés (APA, AERA & NCME, 2014).

Disposiciones como las mencionadas, que están contenidas en los documentos normativos, contienen dos tipos de referencias importantes para la administración de la evaluación a gran escala que se aplica en línea: las que definen los rasgos que deben poseer quienes administran en línea las pruebas, así como las que consideran los derechos y responsabilidades tanto de quienes administran las pruebas, como quienes las responden. Enseguida se describen brevemente ambos tipos de disposiciones.

2.3.1 Características de los administradores de exámenes

De acuerdo con el Joint Committee on Testing Practices (2004), los administradores de las evaluaciones deben contar con ciertas características, mismas

que se requieren para la buena administración de las pruebas en línea; las principales son:

- Proporcionar evidencia de lo que mide el examen.
- Describir cómo fueron seleccionados los contenidos del examen y cómo fue desarrollado.
- Realizar las modificaciones o adaptaciones a la administración de la prueba que resulten apropiadas para hacer justicia a todos los evaluados.
- Desarrollar e implementar procedimientos para asegurar la confidencialidad de los resultados.
- Asesorar a los usuarios de las pruebas sobre los beneficios y limitaciones de los resultados de la prueba y su interpretación apropiada.
- Proporcionar resultados de las pruebas de manera oportuna y de forma que se entiendan por parte del examinado.
- Proteger los resultados de las pruebas de accesos y publicación no autorizados.

2.3.2 Derechos y responsabilidades de quienes administran y responden las pruebas estandarizadas

Según con el Código de prácticas justas de evaluación en educación (APA, AERA & NCME, 2004), los desarrolladores de la prueba así como los administradores de la prueba, tienen derechos y responsabilidades. A continuación, se mencionan los

principales derechos y responsabilidades que corresponden a cada uno de ellos, al momento de desarrollar y seleccionar las pruebas adecuadas (Ver tabla 2.5); al momento de administrar la prueba y elaborar el reporte de resultados (Ver tabla 2.6); y al momento de reportar e interpretar los resultados del examen (Ver tabla 2.7).

Tabla 2.5
Derechos y responsabilidades de los desarrolladores y administradores de pruebas al momento de desarrollar y seleccionar las pruebas adecuadas

Desarrollador de la prueba	Administrador de la prueba
Proporciona evidencia de lo que mide el test, los usos recomendados, los examinados que se pretende evaluar, las fortalezas y limitaciones de la prueba, incluido el nivel de precisión de las puntuaciones del examen.	Define el propósito de la prueba, el contenido y habilidades que serán evaluadas y los examinados que se pretende evaluar. Selecciona y usa el test más apropiado con base en una revisión exhaustiva de la información disponible.
Describe cómo se seleccionaron los contenidos y habilidades que serán evaluados y cómo se desarrolló el test.	Revisa y selecciona pruebas con base en la relevancia del contenido del test, las habilidades evaluadas, y la cobertura del contenido para el propósito previsto.
Comunica la información sobre características de la prueba a un nivel de detalle apropiado para los usuarios previstos del examen.	Revisa los materiales proporcionados por los desarrolladores de la prueba y selecciona pruebas para las cuales se proporciona información clara, precisa y completa.
Proporciona orientación acerca de los niveles de habilidades, conocimientos y formación que son necesarios para la revisión, selección, y administración apropiadas de las pruebas.	Selecciona las pruebas a través de un proceso que incluye a personas con conocimientos, habilidades y formación adecuados.
Proporciona evidencia de que la calidad técnica del test, incluyendo la confiabilidad y la validez, satisface los propósitos pretendidos.	Evalúa la evidencia de la calidad técnica de la prueba proporcionada por el desarrollador del test y en su caso de revisores independientes.
Proporciona especialistas muestras representativas de ítems o pruebas de práctica, instrucciones, hojas de respuesta, manuales, y reportes de resultados.	Evalúa muestras representativas de ítems o pruebas de la práctica, instrucciones, hojas de respuesta, manuales, y reportes de resultados, antes de seleccionar una prueba.
Evita el contenido y el lenguaje potencialmente ofensivos al desarrollar los ítems y materiales de la prueba.	Evalúa los procedimientos y materiales utilizados por los desarrolladores de prueba, así como la prueba resultante, para asegurar que se evita lenguaje y contenido potencialmente ofensivos.
Hace versiones de las pruebas adaptadas apropiadamente, así como procedimientos para su administración a examinandos con discapacidad que necesitan arreglos	Selecciona pruebas con versiones modificadas apropiadamente o con procedimientos de administración para examinandos con discapacidad que necesitan

Desarrollador de la prueba	Administrador de la prueba
especiales.	arreglos especiales.
Obtiene y aporta evidencia sobre el desempeño de examinados de diversos subgrupos, haciendo esfuerzos para obtener tamaños de muestra que sean adecuados para el análisis de subgrupos. Evalúa la evidencia para asegurar que las diferencias en el rendimiento están relacionadas con las habilidades que están siendo evaluadas.	Evalúa la evidencia disponible sobre el rendimiento de los examinados de diversos subgrupos. Determina, en la medida de lo posible, qué diferencias en el rendimiento pueden haber sido ocasionadas por factores ajenos a las habilidades que están siendo evaluadas.

Fuente: Joint Committee on Testing Practices (2004)

Tabla 2.6

Derechos y responsabilidades de los desarrolladores y administradores de pruebas al momento de administrar la prueba y el reporte de resultados

Desarrollador de la prueba	Administrador de la prueba
Proporciona descripciones claras de procedimientos detallados para administrar pruebas de manera estandarizada.	Sigue los procedimientos establecidos para la administración de pruebas de manera estandarizada.
Proporciona pautas sobre procedimientos razonables para la evaluación de las personas con discapacidades que necesitan adaptaciones especiales o los que tienen diversos orígenes lingüísticos.	Provee y documenta procedimientos adecuados para examinados con discapacidad que necesitan adaptaciones especiales o para quienes tienen diversos orígenes lingüísticos. Algunas adecuaciones pueden ser requeridas por ley o reglamento.
Proporciona información a los examinados o a otros usuarios de la prueba sobre los formatos y procedimientos para responder a los ítems, incluida la información sobre el uso de cualquier material y el equipo necesario.	Proporciona a los examinados la oportunidad de familiarizarse con formatos de preguntas de prueba y materiales o equipos que se pueden utilizar durante la prueba.
Establece e implementa procedimientos para garantizar la seguridad de los materiales de prueba durante todas las fases de su desarrollo, administración, calificación, y elaboración de informes.	Protege la seguridad de los materiales de prueba, incluido el respeto de los derechos de autor y elimina posibles oportunidades para los examinados de obtener puntajes por medios fraudulentos.
Proporciona procedimientos, materiales y directrices para calificar la prueba, y para monitorear la exactitud del proceso de puntuación. Si la calificación de la prueba es responsabilidad del desarrollador, se debe proporcionar un entrenamiento adecuada a los calificadores.	En caso de que la calificación del examen sea responsabilidad del desarrollador, se debe proporcionar una capacitación adecuada a los calificadores y garantizar y controlar la exactitud del proceso de calificación.
Corrige los errores que afectan la interpretación de las puntuaciones y comunica los resultados corregidos con prontitud.	Corrige los errores que afectan la interpretación de las puntuaciones y comunica los resultados corregidos con prontitud.

Desarrollador de la prueba	Administrador de la prueba
Desarrolla e implementa procedimientos para garantizar la confidencialidad de las calificaciones.	Desarrolla e implementa procedimientos para garantizar la confidencialidad de las calificaciones.

Fuente: Joint Committee on Testing Practices (2004)

Respecto a la administración de la prueba, se sabe que incluso los examinados que están familiarizados con las computadoras pueden necesitar alguna breve explicación o práctica para manejar los detalles específicos, tales como la interfaz de la prueba. Ciertas prácticas especiales para administrar el ambiente de la prueba pueden reducir la varianza al constructo, tales como evitar reflejos de luz en la pantalla del ordenador que interfieren con la legibilidad, o el mantenimiento de un entorno tranquilo cuando los examinados empiezan o terminan la prueba en momentos diferentes. Quienes administran pruebas informatizadas deberían recibir formación para que puedan hacer frente a los problemas de hardware, software, o la administración del examen. Las pruebas administradas por computadora en aplicaciones basadas en Web pueden requerir también otros apoyos para mantener entornos normalizados (APA, AERA & NCME, 2014).

Tabla 2.7

Derechos y responsabilidades de los desarrolladores y administradores de pruebas al momento de reportar e interpretar los resultados

Desarrollador de la prueba	Administrador de la prueba
Proporciona información para apoyar interpretaciones pretendidas de los resultados, referidas al contenido, grupos de comparación, y otras evidencias técnicas. Asesora a usuarios de la prueba de los beneficios y limitaciones de sus resultados e interpretación. Advierte contra asignar una mayor precisión que la que se justifica.	Interpreta el significado de los resultados de la prueba, teniendo en cuenta la naturaleza de los contenidos, normas o grupos de comparación, y otras evidencias técnicas y de los beneficios y limitaciones de los resultados del test.
Proporciona orientación para interpretar los resultados de pruebas administradas con modificaciones. Informa a los usuarios de	Interpreta resultados de la prueba y los procedimientos de administración modificados dado el impacto que pudieron haber tenido

Desarrollador de la prueba	Administrador de la prueba
problemas al interpretar los resultados, cuando se ha modificado el test o los procedimientos para su administración.	esas modificaciones en los resultados del test.
Especifica los usos apropiados de los resultados de las pruebas y advierte a los usuarios sobre malos usos potenciales.	Evita utilizar pruebas para fines distintos de los recomendados por el desarrollador del test a menos que exista evidencia para apoyar el uso o interpretación pretendida.
Cuando los desarrolladores de prueba establecen estándares proporcionan el fundamento, procedimientos y evidencias para establecer estándares de desempeño o calificaciones aprobatorias. Evitan el uso de etiquetas estigmatizantes.	Revisa los procedimientos para el establecimiento de estándares de desempeño o calificaciones aprobatorias. Evita el uso de etiquetas estigmatizantes.
Anima a los usuarios de prueba para tomar decisiones acerca de los examinados con base en múltiples fuentes de información adecuada, no en un solo examen.	Evita usar al examen como único determinante para decidir sobre los examinados. Interpreta los resultados de las pruebas en conjunto con otra información sobre individuos.
Proporciona información que permite a los usuarios de prueba interpretar y reportar con exactitud los resultados obtenidos por grupos de examinados, incluida la información sobre quiénes fueron o no incluidos en los diferentes grupos que se comparan, e información sobre los factores que pudieron influir en la interpretación de los resultados.	Establece la interpretación y el uso previsto de los resultados de las pruebas para grupos de examinados. Evita agrupar resultados de pruebas para fines no recomendados por el desarrollador del test, a menos que exista evidencia para apoyar un uso pretendido. Reporta el procedimiento para determinar quién fue o no incluido en los grupos que se comparan y describe factores que pudieron influir en la interpretación de los resultados.
Proporciona con oportunidad los resultados de la prueba y de manera que sean comprendidos por el examinado.	Comunica con oportunidad los resultados de la prueba y de manera que sean comprendidos por el examinado.
Proporciona orientación a los usuarios sobre cómo monitorear el grado en que la prueba está cumpliendo con los fines previstos.	Desarrolla e implementa procedimientos para monitorear el uso de la prueba, incluyendo la consistencia con los fines previstos del test.

Fuente: Joint Committee on Testing Practices (2004)

El Joint Committee on Testing Practices (2014) publicó recientemente la nueva versión de los estándares que se revisaron anteriormente en este apartado. A continuación, en la tabla 2.8 se muestran los estándares que incluye dicha versión relacionados con la administración, calificación, reporte e interpretación de los resultados de las pruebas.

Tabla 2.8 Nuevos estándares relacionados con la administración, calificación, reporte e interpretación de los resultados de las pruebas (Joint Committee on Testing Practices, 2014)

Administración de la prueba	Calificación de la prueba	Reporte e interpretación de resultados de la prueba
6.1 Los administradores de la prueba deben seguir cuidadosamente los procedimientos estandarizados para la administración y calificación especificados por el desarrollador de prueba, así como las instrucciones de los usuarios del test.	6.8 Los responsables de calificar la prueba deben establecer protocolos de calificación. La calificación de test que implique juicios debe incluir rúbricas, procedimientos y criterios para calificar. Al calificar respuestas complejas por computadora, debe documentarse la exactitud de los algoritmos y procesos.	6.10 Cuando se publique información sobre los resultados de la prueba, los responsables del programa de pruebas deben proporcionar interpretaciones adecuadas a la audiencia. Las interpretaciones deben describir en un lenguaje sencillo, el dominio que cubre el test, lo que representan las puntuaciones, la precisión/confiabilidad de las puntuaciones y cómo se pretende que sean utilizadas las calificaciones.
6.2 Cuando se han establecido procedimientos formales para solicitar y recibir acomodaciones, los examinados deben ser informados de estos procedimientos antes de la prueba.	6.9 Los responsables de los puntajes del test deben establecer y documentar el control y calidad de los procesos y criterios. Debe proporcionarse entrenamiento adecuado. La calidad de la calificación debe monitorearse y documentarse. Cualquier fuente de error sistemático de puntuación debe ser documentada y corregida.	6.11 Cuando se reportan interpretaciones generadas automáticamente de protocolos de respuesta o de la ejecución en el test, deben estar disponibles las fuentes, el fundamento y base empírica de tales interpretaciones y deben ser descritas sus limitaciones.
Administración de la prueba	Calificación de la prueba	Reporte e interpretación de resultados de la prueba

6.3 Los cambios o interrupciones en procedimientos estandarizados para administrar o calificar los tests deben ser documentados y reportados al usuario de prueba.	6.12 Cuando se obtenga información a nivel de grupo, al agregar resultados parciales de test aplicados a individuos, debe reportarse evidencia de validez y confiabilidad/precisión para el nivel de agregación al cual se reportan los resultados. Las calificaciones no deben reportarse para individuos sin evidencia apropiada que soporte las interpretaciones para los usos pretendidos.
6.4 El ambiente de prueba debe proporcionar comodidad razonable con mínimas distracciones para evitar varianza irrelevante al constructo	6.13 Cuando se encuentra un error en las calificaciones o en otra información importante emitida por la organización responsable del test o por otra institución, tal información o la puntuación corregida deberá ser distribuida tan pronto como sea posible a todos los destinatarios conocidos, quienes de lo contrario podrían utilizar los resultados erróneos como base para tomar decisiones. El informe corregido debe etiquetarse como tal y documentar lo que se hizo para corregir el error. También debe quedar aclararse a los destinatarios del informe porqué se corrigió un puntaje,
6.5 Debe proporcionarse a los examinandos instrucciones y prácticas , así como otros apoyos necesarios para reducir la varianza irrelevante al constructo	6.14 Las organizaciones que informan sobre las puntuaciones de la prueba, identificable individualmente, deben tener directrices de política claras sobre la disponibilidad y uso temporal de los datos de un registro de un individuo para efectos de investigación u otros fines. La política debe documentarse y estar disponible para el examinado. Los usuarios del test deben mantener una seguridad de los datos que sea apropiada.
6.6 Debe asegurarse la integridad de las puntuaciones, eliminando las oportunidades para que los examinados logren resultados por medios fraudulentos o engañosos.	6.15 Cuando se conserven datos individuales sobre una prueba, tanto el protocolo de la prueba como cualquier informe escrito deben también conservarse en alguna forma.
6.7 Los usuarios de la prueba tienen la responsabilidad, en todo momento, de proteger la seguridad de los materiales de prueba.	6.16 La transmisión de resultados en las pruebas identificados individualmente, a personas o a instituciones autorizadas, debe hacerse de una manera que proteja la naturaleza confidencial de los resultados y de la información adicional pertinente.

2.3.3 Derechos de los examinados

El Código de prácticas justas de evaluación en educación (APA, AERA & NCME, 2004) menciona que los examinados tiene derechos al momento de responder la prueba. A continuación se mencionan algunos de ellos:

- Informar a los examinados antes de la administración de la prueba sobre la cobertura de la prueba y los tipos de formatos de pregunta. Se debe poner dicha información a disposición de los examinandos.
- Cuando una prueba es opcional, se proporciona a los examinados, padres o tutores información para ayudarles a decidir si se debe responder una prueba — incluidas las indicaciones de las consecuencias que puedan derivarse de no tomar el examen (por ejemplo, no ser elegible para competir por una beca especial) — y si existe una alternativa disponible para la prueba.
- Proporcionar a los examinandos, o a sus padres o tutores, información sobre los derechos que pueden tener los examinandos para obtener copias de las pruebas y hojas de respuestas completas, para volver a tomar las pruebas, para que las pruebas se vuelvan a calificar, o para que las calificaciones se puedan declarar nulas.
- Proporcionar a los examinandos, o a sus padres o tutores, información sobre las responsabilidades que tienen los examinandos, como ser conscientes de la finalidad y usos de la prueba, seguir instrucciones, no revelar los ítems de la prueba o no interferir con otros examinados que también responden la prueba.

-
- Informar a los evaluados, o a sus padres o tutores, cuánto tiempo se mantendrán en archivo las calificaciones e indicar a quién, en qué circunstancias y en qué forma serán o no serán liberados los puntajes de las pruebas y la información relacionada. Proteger los resultados de las pruebas de accesos no autorizados.
 - Describir los procedimientos para la investigación y resolución de circunstancias que puedan dar lugar a la cancelación o la retención de calificaciones, tales como el incumplimiento de los procedimientos de prueba especificados.
 - Describir los procedimientos que los examinandos, o sus padres o tutores y otras partes interesadas pueden utilizar para obtener más información acerca de la prueba, o para registrar quejas, o para resolver problemas.

2.4 Estándares de seguridad para las evaluaciones en línea

La necesidad de asegurar los test, exámenes, y otras formas de evaluación ha aumentado en importancia, a la par del crecimiento del uso de las pruebas, así como con el creciente rol de la tecnología en la implementación, administración, y calificación de las pruebas, en particular a través de internet (International Test Commission, 2014).

En efecto, un problema relevante de los sistemas que administran pruebas de gran escala en línea, es la seguridad. Diferentes organizaciones, entre ellas la Universidad de Australia (2013) y el sistema educativo del estado de Georgia en Estados Unidos (2013), así como la International Test Commission (2014), plantean que es de suma importancia tener en cuenta los problemas vinculados a la seguridad y

la integridad del test, así como la validez de las puntuaciones, cuando se aplica una evaluación en línea mediante métodos de evaluación convencionales. Por lo tanto, antes de tomar una decisión para llevar a cabo una evaluación en línea, siempre es necesario evaluar los riesgos involucrados y considerar si se desea o no utilizar otras opciones. Si se toma la decisión de proceder con la evaluación en línea, dichas organizaciones plantean la necesidad de cumplir con los siguientes estándares:

- Incorporar la introducción de contraseña para ingresar al examen, proporcionar la contraseña a los estudiantes solo durante el tiempo del examen, y asegurarse que la contraseña ya no está activa cuando los estudiantes dejan el examen.
- Asignar límites de tiempo y de acceso para el período de evaluación, con supervisores de la aplicación, e incluir medidas alternativas para los estudiantes a distancia (por ejemplo, aplicar la prueba en la biblioteca).
- Considerar el uso de un banco de ítems con su asignación al azar en la prueba.
- Los examinados deben entender las reglas de seguridad y las consecuencias de su violación, antes de registrarse y programarse para responder la prueba.
- Tratar de usar un formato de examen en el que los alumnos respondan una pregunta a la vez y donde se respondan los ítems según su tipo de respuestas dentro de un marco de tiempo limitado.
- Después de 20 minutos de no contestar el examen, la sesión debe de cerrarse automáticamente.

-
- La página debe de estar diseñada de modo que no se puedan abrir diferentes navegadores durante el tiempo en que el examinado este respondiendo la prueba.
 - El administrador de la prueba debe de aprobar a los examinados y corroborar que son los estudiantes que deben contestar esa prueba.
 - Cuando se recogen los resultados de prueba digital desde los servidores remotos, la transferencia de datos debe ocurrir inmediatamente después de la terminación de una prueba o después de la terminación de cada sección en línea. Los datos deben estar protegidos por procedimientos fuertes de acceso mientras residen en un servidor remoto y por encriptación fuerte durante la transmisión.

Por otra parte, las prácticas de copiar en las pruebas, robar la prueba y otras infracciones similares pueden ocurrir incluso al programa más seguro. Sin embargo, un programa de gestión de seguridad activa ayudará a asegurar que las infracciones sean menos numerosas, y que los daños sean limitados (International Test Commission, 2014).

2.5 Principios relativos al diseño universal de las evaluaciones

El diseño universal de las evaluaciones es un enfoque para el diseño de las pruebas que en principio busca maximizar la accesibilidad para todos los examinados. El diseño universal, exige que los desarrolladores de la prueba sean claros en el

constructo (s) que se desea evaluar. (APA, AERA & NCME, 2014). El diseño universal, como se aplica a las evaluaciones, es un concepto que permite a la más amplia gama posible de estudiantes a participar en las evaluaciones y puede incluso reducir la necesidad de efectuar adaptaciones de los test y las evaluaciones alternativas, para ampliar el acceso a las pruebas al expandir el acceso a ellas.

En Estados Unidos, el acta No Child Left Behind de 2001 enfatizó la idea de rendir cuentas en el ámbito de la evaluación a gran escala. Por ello, dado el alto impacto que pueden tener las evaluaciones a gran escala sobre los estudiantes, sus padres, docentes y otros agentes educativos, se requiere diseñarlas y administrarlas de maneras que permitan la participación del más amplio rango de examinados, incluso aquellos que tradicionalmente han sido excluidos de la evaluación por su condición de fragilidad física o psicológica (Thompson, Johnstone y Thurlow, 2002)

Por ejemplo, pueden surgir problemas debido a la discapacidad de un examinado, sus antecedentes culturales, su contexto lingüístico, la raza, el origen étnico, el nivel socioeconómico, o bien las limitaciones que puedan surgir con el envejecimiento, o alguna combinación de estos u otros factores. En algunos casos, puede ser alcanzada una mayor comparabilidad de las calificaciones, si los procedimientos estandarizados sufren cambios para hacer frente a las necesidades de grupos específicos o individuos, pero sin que tengan un efecto adverso sobre la validez o la confiabilidad de los resultados obtenidos. Por ejemplo, pueden proporcionarse a personas con algunas discapacidades visuales una versión de la prueba en braille, una

hoja de respuesta con letras y burbujas de gran tamaño, o un lector de pantalla, a fin de obtener un acceso más equitativo a contenido de la prueba (APA, AERA & NCME, 2014).

Para asegurar que las pruebas se diseñen de modo que sean accesibles para todos los examinados y que atiendan las necesidades especiales de los más vulnerables, diversos autores y organizaciones (Center for Universal Design, 1997; Thompson y Thurlow, 2002; Thompson, Johnstone y Thurlow, 2002; Thompson, Thurlow, Quenemoen y Lehr, 2002; Axelson, 2005; Johnstone, Altman y Thurlow, 2006) han propuesto los siguientes principios y acciones para desarrollar evaluaciones incluyentes basadas en el diseño universal:

- Proveer evaluaciones que incorporen a todo tipo de examinados
- Medir lo que se pretende medir
- Reducir el sesgo al mínimo
- Establecer instrucciones y procedimientos claros y comprensibles
- Asegurar acomodaciones o adaptaciones flexibles
- Emplear lenguaje comprensible
- Ser legibles

La meta de las evaluaciones desarrolladas bajo los principios del diseño universal es proporcionar evaluaciones lo más válidas posible, para el mayor número de examinados, incluyendo aquellos con discapacidad. En consecuencia, dichas evaluaciones están destinadas a incrementar el acceso, no a reducir los estándares de

ejecución de los examinados. En términos llanos, diseño universal significa diseñar para todos y estos principios son aplicables por igual a evaluaciones de lápiz y papel, por computadora y en línea (Johnstone y colaboradores, 2006).

Lo anterior incluye también cosas como la reducción de la carga de lenguaje en áreas de contenido que no sean de lectura, aumentar el tamaño de la fuente, incluir menos ítems por página y enmarcar los ítems para apoyar el enfoque visual. Los exámenes de lectura deben tener un vocabulario al nivel del grado que corresponda. De hecho, en todas las demás pruebas, el nivel de vocabulario estará por debajo del grado están evaluando, excepto por las palabras que se refieren a los contenidos cuyo dominio se evalúa (Oklahoma State Department of education, 2013).

En síntesis, la idea fundamental que surge del diseño universal de las evaluaciones es que los conocimientos, habilidades y actitudes que se midan se definan con claridad, de modo que se reduzcan las barreras cognitivas, sensoriales, emotivas o físicas que son irrelevantes a los constructos que se miden.

El diseño universal no elimina la necesidad de efectuar adaptaciones al administrar las pruebas, pero puede reducirla significativamente. Como se verá más adelante en el capítulo III, el diseño de SALIDEA no atiende todos los aspectos considerados por el movimiento de evaluación universal. En la tabla 2.9 se presentan los más relevantes en virtud de que constituyen un referente importante en la evaluación del aprendizaje a gran escala en línea.

Tabla 2.9
Elementos de la evaluación basada en el diseño universal

Elemento de la evaluación	Ejemplo
Instrucciones y procedimientos simples, claros e intuitivos	Todas las instrucciones y procedimientos son presentados en lenguaje comprensible, claro y sencillo.
Máximo de comprensibilidad	Se sigue una variedad de directrices sobre la legibilidad y la lengua llana (por ejemplo, la longitud de las oraciones y el número de palabras difíciles se mantienen al mínimo) para producir texto legible y comprensible.
Máximo de legibilidad	Se aplican al texto, tablas, figuras e ilustraciones con características que aseguran una fácil interpretación.
Uso sin sesgo	El diseño es útil y preferido por personas con habilidades diferentes.
Flexibilidad en el uso	El diseño adapta para una amplia gama de habilidades y preferencias individuales.
Uso simple e intuitivo	El uso del diseño es fácil de entender, independientemente de la experiencia, conocimientos, habilidades lingüísticas o nivel de concentración actual de los usuarios.
Información que puede percibirse	El diseño comunica eficazmente al usuario la información necesaria, independientemente de las condiciones ambientales o sus capacidades sensoriales.
Tolerancia al error	El diseño minimiza los riesgos y las consecuencias adversas de acciones accidentales o imprevistas.
Esfuerzo físico reducido	El diseño se puede utilizarse de manera eficiente y cómoda con un mínimo de fatiga.
Tamaño y espacio para facilitar la aproximación y el uso	Se proporcionan espacio y tamaño adecuados para aproximarse, alcanzar, manipular y usar sin importar el tamaño del cuerpo, postura o movilidad del usuario.

Fuente: basada en Thompson, Johnstone & Thurlow (2002).

De particular relevancia para la evaluación en línea, son los principios del diseño universal que norman la edición de los párrafos y que se presentan en la tabla 2.10.

Tabla 2.10
Dimensiones y características de máxima legibilidad en el párrafo

Dimensión	Características de legibilidad
Contraste	La letra de color negro en fondo de color pastel o blanco favorece la legibilidad y disminuye la fatiga visual.
Tamaño de la letra	Los tamaños de letra grandes son más eficaces para los estudiantes pequeños que están aprendiendo a leer, para alumnos con dificultades visuales y personas con problemas de fatiga ocular. El tamaño legal para el texto en letra es de 14 puntos.
Espaciado	La cantidad de espacio entre cada caracter puede afectar a la legibilidad. El espaciado debe ser amplio entre letras y palabras. Las fuentes con espacio fijo parecen ser más legibles para algunos lectores que las fuentes de espacio proporcional.
Interlineado	La cantidad de espacio vertical entre los renglones debe ser suficiente para evitar que la letra se vea borrosa. La cantidad necesaria varía con el tamaño de la letra; por ejemplo, la letra con tamaño de 14 puntos necesita de 3 a 6 puntos de interlineado.
Tipo de letra	El tipo de letra estándar, con mayúsculas y minúsculas, es más fácil de leer que la letra cursiva, inclinada, con minúsculas, o todo en mayúsculas.
Justificado	El texto sin justificar, con margen derecho, es más fácil de ver y analizar que el texto justificado; en particular para lectores con dificultades para leer.
Longitud del renglón	La longitud óptima es de alrededor de 4 pulgadas o de 8 a 10 palabras por línea. Esta longitud evita la fatiga de lector y la dificultad para localizar el inicio del siguiente renglón, lo que hace que el lector se desubique.
Espacio en blanco	Una regla general es que el texto ocupe sólo la mitad de una página. El espacio en blanco ancla al texto en el papel y aumenta la legibilidad.
Gráficas y tablas	Los símbolos utilizados en los gráficos tienen que ser altamente discriminables. Las etiquetas deben colocarse directamente junto a las barras o líneas de una gráfica, para que la información se pueda encontrar de forma rápida y no se requiera el uso de la memoria a corto plazo.
Ilustraciones	Cuando se usa una ilustración, debe estar directamente al lado de la pregunta que la requiera. Debido a que las ilustraciones crean numerosos problemas visuales y de distracción, y pueden interferir con el uso de algunas acomodaciones como lupas, deben usarse sólo cuando contienen información cuyo dominio se evalúa.
Formatos de respuesta	Las opciones de respuesta deben incluir círculos más grandes (para las pruebas que se responden en una burbuja).

Fuente: basada en Thompson, Johnstone & Thurlow (2002).

2.6 Diseño de la interfaz gráfica

Una interfaz gráfica del usuario (GUI por su sigla en inglés) utiliza elementos visuales que presentan la información almacenada en una computadora de una manera fácil de entender. Estos elementos hacen que sea fácil para las personas utilizar programas de cómputo. Una interfaz gráfica del usuario utiliza ventanas, iconos y menús para ejecutar comandos, como abrir archivos, borrar archivos y mover archivos. Los elementos gráficos más comunes están representados por: ventana, icono o control, menú, y el cursor (Zandbergen, 2014).

El diseño de una buena interfaz de usuario es esencial para el éxito de un sistema. Una buena interfaz de usuario anima a una interacción fácil y natural entre un usuario y el sistema. Lo ideal sería que un usuario olvide que está utilizando una computadora y siga adelante con lo que quiere hacer en ella (Zandbergen, 2014).

De hecho, las principales preocupaciones al diseñar una interfaz gráfica son la usabilidad y la accesibilidad, las cuales están íntimamente relacionadas. La usabilidad busca asegurar que el diseño resulta amigable, intuitivo y fácil de comprender. La accesibilidad busca asegurar que la interfaz no crea barreras para quienes tienen limitaciones o necesitan formas alternativas para acceder a los recursos del sistema (Jisc Digital Media, 2014).

De acuerdo con Barnum (2002), al diseñar la usabilidad de un sistema debe asegurarse:

-
- Facilidad de aprendizaje: el sistema deberá ser fácil de aprender para que los usuarios puedan utilizarlo de inmediato.
 - Eficiencia: el sistema deberá ser eficiente, de manera que una vez que se aprende, el usuario pueda alcanzar un alto nivel de productividad.
 - Facilidad para memorizarse: el sistema deberá ser fácil de recordar, de manera que el usuario pueda utilizar el sistema una y otra vez sin tener que volver a aprender a utilizarlo.
 - Tolerancia al error: el sistema deberá de tener un bajo rango de error, de esta manera los usuarios pueden cometer algunos errores y recuperarse fácilmente.
 - Satisfacción: el sistema deberá ser amigable, para que los usuarios estén satisfechos cuando lo utilicen y les guste hacerlo.

Además, según Martin (s/f), una interfaz gráfica debe de estar bien organizada, con una estructura clara y consistente; tener un funcionamiento económico, al permitir un máximo de funcionalidad con el menor número de iconos; y tener capacidad comunicativa, al mantener un equilibrio entre la legibilidad, la tipografía, el simbolismo y el color. Para ilustrar dichos conceptos, dicha autora muestra ejemplos de interfaces graficas que están mal o bien organizadas, con relaciones entre elementos que son inapropiadas o apropiadas, o con íconos ambiguos o que resultan claros. En la figura 2.5 se ilustran a la izquierda las condiciones inapropiadas y a la derecha las que están bien diseñadas.

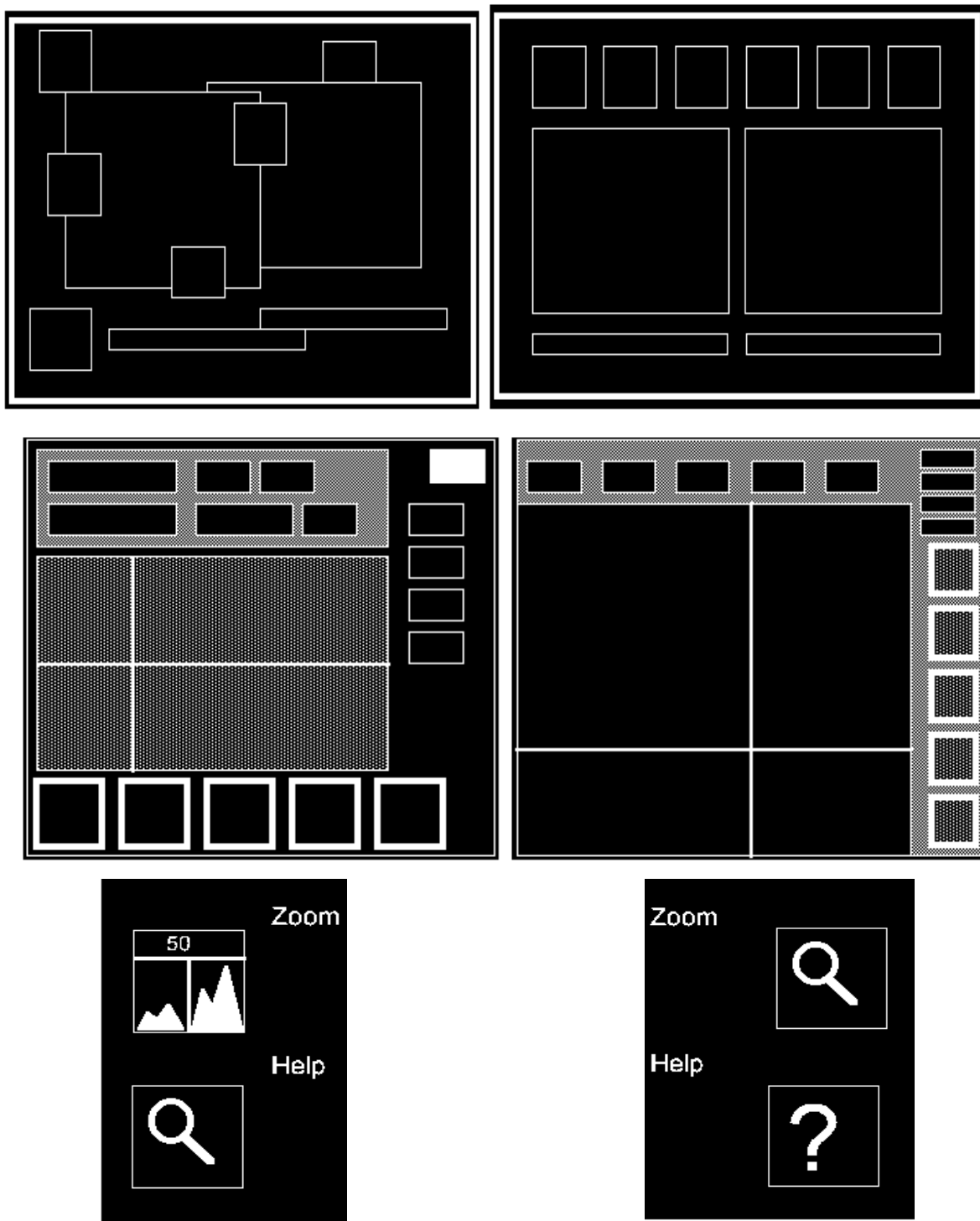


Figura 2.5 Ejemplos de interfaces graficas con organización, relaciones entre elementos e íconos, inapropiados (izquierda) y apropiados (derecha).
Fuente: http://web.cs.wpi.edu/~matt/courses/cs563/talks/smartin/int_design.html

Con base en los aspectos conceptuales descritos en las secciones anteriores, se realizó el diseño del sistema para administrar en línea instrumentos de evaluación del aprendizaje, que fue el objeto del presente trabajo. Para ello se buscó cumplir, siempre que fue posible, con los estándares de calidad técnica para la elaboración, administración, calificación, análisis y reporte de resultados de las pruebas, así como con los principios y lineamientos del diseño universal de las evaluaciones. En el siguiente capítulo se describen los procedimientos que fueron utilizados para lograrlo.

Capítulo III. Método.

En el presente capítulo se describen los procedimientos que se siguieron al diseñar SALIDEA, a fin atender los estándares para la administración de pruebas a gran escala y los correspondientes para la aplicación de pruebas en línea, establecidos tanto por el Joint Committee on Testing Practices (APA, AERA y NCME, 1999; 2004, 2014), como por el Centro Nacional de Evaluación de la Educación Superior (CENEVAL, 2000), los cuales fueron considerados junto con los principios y recomendaciones relativos al diseño universal de las evaluaciones, para el diseño de una aplicación a la cual se denomina ***Sistema para administrar en línea instrumentos de evaluación del aprendizaje (SALIDEA)***, lo cual fue el objeto del presente trabajo. Además, se describen los procedimientos que se siguieron inicialmente para analizar los requerimientos de dicho sistema de evaluación y prefigurar su funcionalidad, en particular para diseñar la interfaz de cada uno de los usuarios de SALIDEA.

En la figura 3.1 se muestra un diagrama con las etapas generales que se siguieron para el diseño de SALIDEA. Como puede observarse en el diagrama, lo que primero se hizo fue considerar las necesidades de los usuarios del sistema, en este caso representadas por los estándares psicométricos y principios del diseño universal de las evaluaciones. Enseguida se procedió a especificar los requerimientos necesarios para identificar las funciones del sistema. A partir de los requerimientos funcionales del sistema identificados se procedió a elaborar un diseño preliminar del sistema de evaluación que produjo las ventanas de la interfaz, con sus controles y funciones de

navegación. Finalmente se hizo un diseño detallado de estos componentes para cada uno de los usuarios de SALIDEA, mismo que fue representado gráficamente mediante una retícula que muestra los componentes y relaciones entre ellos, que fueron considerados para cada usuario.

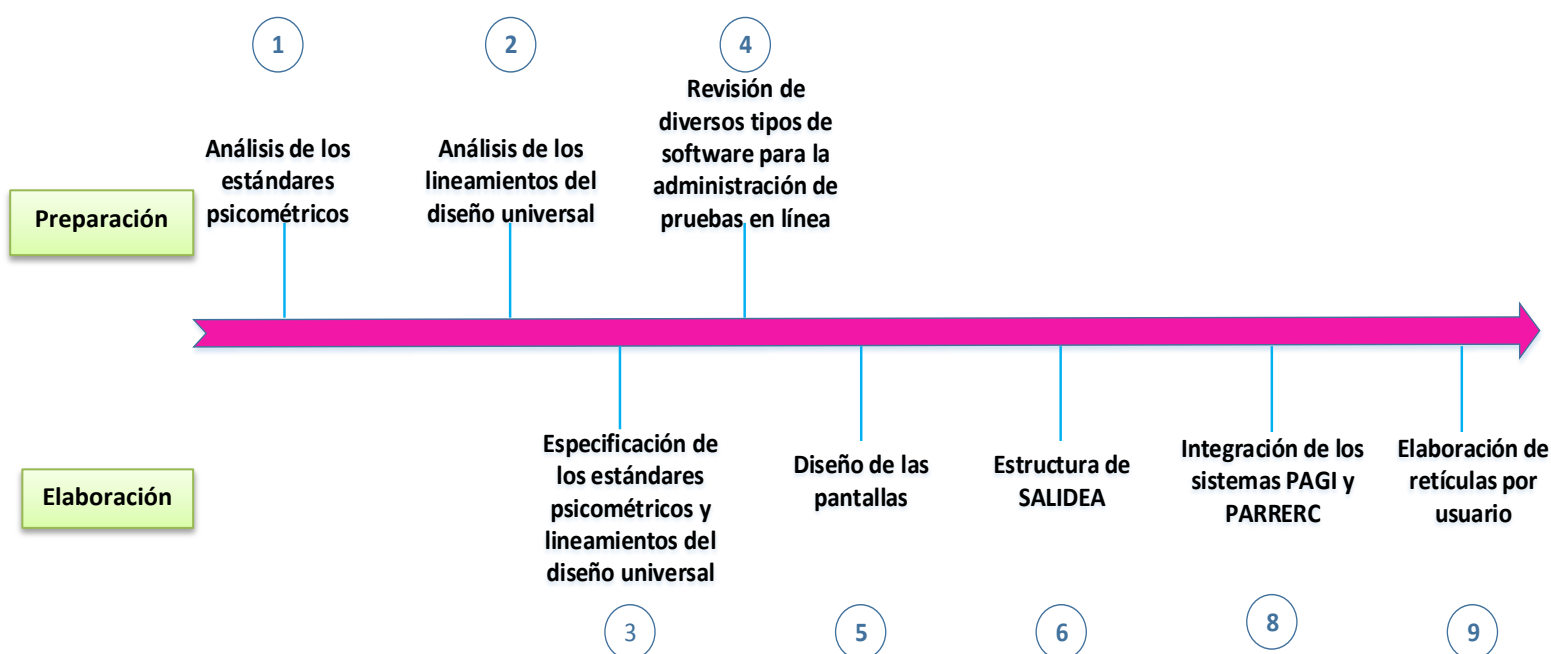


Figura 3.1 Etapas que se siguieron para diseñar SALIDEA

3.1 Apego a la normatividad para la administración estandarizada en línea de exámenes aplicados a gran escala.

Como ya se comentó en el capítulo anterior, los exámenes que se aplican a gran escala tienen un impacto importante sobre los estudiantes que los responden y sobre sus familias, profesores y escuelas. Ello ha ocasionado que en los contextos

internacional y nacional se normen tanto su desarrollo como su administración. Tal es el caso de los estándares publicados en Estados Unidos por el Joint Committee on Testing Practices (APA, AERA, NCME, 1999; 2004, 2014), en México por el CENEVAL (2000) y por otros autores y organizaciones como el Center for Universal Design (1997), Thompson y Thurlow (2002), y Thompson, Johnstone y Thurlow (2002).

3.1.1 Incorporación de los estándares psicométricos de APA-AERA-NCME y CENEVAL.

En las tablas que se presentan a continuación se muestran los estándares para el desarrollo de las pruebas (tabla 3.1), para la administración de pruebas (Tabla 3.2), y para la calificación, interpretación y elaboración de informes de resultados de las pruebas (Tabla 3.3) propuestos por ambas organizaciones, mismos que fueron tomados en cuenta para diseñar SALIDEA. Cabe señalar que en las tablas solo aparecen los estándares que resultan significativos para la aplicación de las pruebas en línea, por lo que no se consideran otros que solo hacen sentido para la aplicación de pruebas de lápiz y papel.

Tabla 3.1
Estándares psicométricos relativos al desarrollo de la prueba, que fueron atendidos al diseñar SALIDEA.

Organismo	Año	Estándar	SALIDEA	Componente
Joint Committee on Testing Practices	1999	Proporciona a todos los examinados oportunidad comparable para demostrar su status respecto a los constructos que el test pretende medir. Este tratamiento incluye proporcionarles condiciones apropiadas para responder, igual oportunidad de familiarizarse con el test, material de práctica, etc.	Presenta componentes diseñados para eliminar o reducir barreras cognitivas, sensoriales, emotivas o físicas que son irrelevantes a los constructos que se miden en la prueba	Conjunto de componentes de SALIDEA
	1999; 2004	Proporciona evidencia de lo que mide el examen	Presenta sección con los propósitos de la prueba, descripción e importancia de lo que pretende medir. Indica cómo se seleccionaron los contenidos, cómo se evaluará y ejemplos. Además tiene una sección donde muestra las propiedades psicométricas de los ítems y versiones de la prueba	Sección de inicio de la prueba Manual del instrumento
	1999; 2004	Describe cómo fueron seleccionados los contenidos del examen y cómo fue desarrollado el instrumento.	Una sección del manual de la prueba muestra la estructura, especificaciones de ítems, y el proceso para su elaboración y validación.	Manual del instrumento
Joint Committee on Testing Practices	1999; 2004	Evita el contenido o lenguaje potencialmente ofensivos cuando se desarrollan los ítems de la prueba y los materiales relacionados.	Los ítems son revisados por el desarrollador del examen y el administrador de SALIDEA para evitar penalizar u ofender a determinados estudiantes (mujeres, indígenas, etc.)	- Desarrollador del examen - Administrador de SALIDEA
CENEVAL	2000	Presenta el manual de la prueba	Presenta la documentación del examen y ejemplos de los ítems que forman la prueba (opción múltiple, respuesta breve, etc.)	Manual del instrumento
		Anticipa y hace explícito el desarrollo de la aplicación	Presenta sección de instrucciones sobre cómo se llevará a cabo la aplicación del examen (tiempo, forma, condiciones, etc.)	Instructivo para el examinado

Fuente: Elaboración propia basada en Joint Committee on Testing Practices (1999; 2004) Ceneval (2000)

Cabe señalar que, aunque el diseño de SALIDEA prevé que sea un sistema para la administración de pruebas y que por ello no incluye un componente para el desarrollo de exámenes, se considera que debe atender también esta clase de estándares que especifican que la administración debe incluir información a los examinados sobre tales cuestiones. Lo anterior significa que, cuando eventualmente fuera utilizado SALIDEA por alguna persona o institución para dar de alta y administrar un examen, deberá incorporar la información correspondiente a lo que se especifica en la columna de la derecha de la tabla anterior, en particular el manual del examen.

Tabla 3.2
Estándares psicométricos relativos a la administración de las pruebas, que fueron atendidos al diseñar SALIDEA.

Organismo	Año	Estándar	SALIDEA
Joint Committee on Testing Practices	1999; 2004	Proporciona pautas sobre procedimientos razonables para la evaluación de personas con discapacidades que necesitan adaptaciones especiales.	Atiende las necesidades de personas con dificultades visuales mediante el uso de fuente especial APFont. Además se cuenta con botones para aumentar el tamaño de las figuras y el de la letra en los textos, mediante la tecla Ctrl y la rueda del mouse.
Joint Committee on Testing Practices	1999; 2004	Proporciona información sobre los formatos y procedimientos para responder las preguntas de la prueba, información sobre el uso de material y equipo necesario.	Presenta una sección donde aparece la prueba y el instructivo para responderla.
Joint Committee on Testing Practices	1999; 2004	Establece e implementa procedimientos que garantizan la seguridad de los materiales durante el desarrollo, administración, puntuación e informes de la prueba.	Cuenta con candados de seguridad para darse de alta y para ingresar al sistema; evita que personas ajenas puedan ver la prueba o los resultados. Sección donde los diferentes destinatarios pueden consultar, previo ingreso protegido, el reporte de resultados del examen que le corresponde a cada uno de ellos.

Fuente: Elaboración propia basada en Joint Committee on Testing Practices (1999; 2004) Ceneval (2000)

El diseño de SALIDEA contempla que sea un sistema para la administración en línea de pruebas, que da especial importancia al cumplimiento de los estándares que aparecen en la tabla anterior. En particular, el diseño incluye como componentes los instructivos para el desarrollador del examen (Véase apéndice 1), para el administrador de la prueba (Véase apéndice 2) y para el examinado (Véase apéndice 3).

En el primero de ellos se describen las acciones y elementos asociados a los componentes que debe aportar el desarrollador, a fin de que SALIDEA pueda administrar la prueba. Estos incluyen las versiones del examen que se aplicará, el manual técnico de la prueba, la revisión del sesgo y la ofensa en los ítems y versiones, entre otros.

En el segundo instructivo se describen las funciones y responsabilidades del administrador, tanto en lo relacionado con la operación del sistema como con la administración del examen, así como los procedimientos y acciones específicos que debe llevar a cabo para preservar la validez e integridad del examen.

Por su parte, el instructivo para el examinado proporciona información relacionada con los derechos y obligaciones de los examinados, las formas de responder la prueba, y aporta ejemplos de ítems del examen.

Tabla 3.3

Estándares psicométricos relativos a la calificación de la prueba, la interpretación de sus resultados y elaboración de informes de resultados de la prueba, que fueron atendidos al diseñar SALIDEA.

Organismo	Año	Estándar	SALIDEA
Joint Committee on Testing Practices	1999; 2004	Proporciona resultados de las pruebas de manera oportuna y de forma que se entiendan por parte del examinado	Los reportes de SALIDEA presentan la información apropiada para cada usuario, de manera inmediata, simple y significativa; tanto los resultados generales en la prueba, como por área evaluada y su relación con variables de contexto exploradas, así como en general los de todos los componentes que incluye el sistema.
		Protege los resultados de las pruebas de accesos y publicación no autorizados.	SALIDEA tiene un acceso protegido mediante un usuario y una clave de acceso únicos, para cada usuario especificado y proporciona a cada uno de ellos el informe particular de resultados del examen que le corresponde.
Joint Committee on Testing Practices	1999; 2004	Desarrolla e implementa procedimientos para asegurar la confidencialidad de los resultados.	- Ofrece aviso de privacidad - Ofrece un usuario para registro al sistema - Ofrece una clave de acceso única.
Joint Committee on Testing Practices	1999; 2004	Asesora a los usuarios de pruebas sobre los beneficios y limitaciones de los resultados de la prueba.	Presenta en el manual del instrumento sugerencias sobre los usos y consecuencias válidos de los resultados del examen.

Fuente: Elaboración propia basada en Joint Committee on Testing Practices (1999; 2004) Ceneval (2000)

De este modo, tras responder los examinados la prueba, se prevé que SALIDEA analice su ejecución y gestione reportes de resultados especializados, mismos que pueden ser tan detallados como lo requieran las necesidades informativas de los diferentes usuarios. Además, los informes proporcionados deben ofrecer seguridad, por lo que se suministra a cada usuario una clave de acceso única al sistema. Asimismo, informa a los usuarios sobre los principales usos y consecuencias válidos de los

resultados de la prueba, de modo que cuenten con elementos de información apropiados para la toma de decisiones que a cada uno de ellos le corresponde.

3.1.2 Incorporación de principios del diseño universal de las evaluaciones

De acuerdo con Burgstahler (2007) el diseño de cualquier producto o entorno implica considerar factores como la estética, opciones de ingeniería, problemas de seguridad y cumplimiento de estándares. Por ello, al desarrollar un producto deben atenderse no solo las necesidades de usuarios regulares, sino también las necesidades de las personas con limitaciones o discapacidades tanto físicas como psicológicas.

Con el propósito de minimizar adecuaciones o adaptaciones a los test debidas a quejas o sugerencias de los usuarios, desde un inicio se consideró que SALIDEA debe estar basado en los principios y lineamientos del diseño universal de las evaluaciones.

Además de tomar como guía los lineamientos comentados en la sección 2.4 del capítulo anterior, para lograr un sistema congruente con el diseño universal de las evaluaciones, se siguieron los tres principios generales que propone Burgstahler (2007): a) Propiciar un uso equitativo para todo tipo de personas, incluyendo aquella con limitaciones; b) Proporcionar tamaño y espacio adecuados para un mejor enfoque, alcance, manipulación y uso del sistema de evaluación; c) Establecer expectativas claras al examinado. Asimismo, en el diseño se incorporaron varios elementos de la evaluación basada en el diseño universal propuestos Thompson y Thurlow (2002).

Así, en la tabla 3.4 se muestran las fases del procedimiento general que se siguió para incorporar en el diseño de SALIDEA los principios del diseño universal.

Tabla 3.4
Etapas del procedimiento para incorporar el diseño universal de las evaluaciones al diseñar SALIDEA

Etapas	SALIDEA
Se seleccionó el contenido a incluir en el sistema y después se realizó el diseño.	SALIDEA cuenta con una fase de selección de contenidos para la interfaz de la página web, la cual considera el resto de las etapas y después se inicia el diseño de la misma. (véase la estructura resultante en la figuras que se presentan en el capítulo 4)
Se buscó establecer una navegación sencilla y consistente	Los menús se encuentran organizados de forma lógica, clara y concisa; en los enlaces se utiliza texto breve y significativo.
Se ponderaron con cuidado las herramientas que estarían disponibles a fin de propiciar un uso sin sesgo	Se definió que SALIDEA fuera flexible para personas con habilidades diferentes. Para ello se consideró dar tiempo extendido para contestar la prueba y se decidió proveer de herramientas para apoyar la movilidad reducida, la visión limitada y las dificultades de aprendizaje.
Se seleccionó cuidadosamente el color del fondo de la interfaz y los menús.	Se resolvió maximizar el contraste: utilizar un fondo color blanco y en los menús un gris tenue; utilizar letras de color negro, para que se puedan leer y ubicar las opciones.
Se buscó una fuente de letra que evitara la distorsión de la imagen al aumentar el zoom.	Se estableció que SALIDEA debe utilizar la fuente APHont , recomendada por la Imprenta Americana para los Ciegos, con la cual no se distorsiona el texto al momento de ampliarlo.
Se examinó con cuidado los tipos de ítems que se facilitarían para estudiantes con distintos tipos de aprendizaje.	Se decidió propiciar el uso de ítems de respuesta seleccionada y respuesta breve que, sin perder validez, contengan imágenes simples y textos breves para apoyar a estudiantes con escasa movilidad, reducida visión y dificultad de aprendizaje.
Se buscó que la interfaz tuviera un uso simple e intuitivo, con instrucciones y procedimientos simples, claros e intuitivos	Se determinó que el uso del sistema fuera fácil de entender por los usuarios, independientemente de su experiencia, conocimientos, habilidades lingüísticas o nivel de concentración; con instrucciones y procedimientos en lenguaje comprensible, claro y sencillo.
Se buscó maximizar la comprensibilidad y legibilidad textuales	Se decidió utilizar lengua llana y mantener al mínimo la longitud de las oraciones y el número de palabras difíciles, para producir texto legible y comprensible; y se aplican al texto, tablas, figuras e ilustraciones con características que aseguran una fácil interpretación.
Se exploraron formas de admitir tolerancia al error	Se incorporaron elementos para reducir errores por parte de examinados, tales como la opción para hacer zoom a la pantalla en caso de requerirlo, burbujas más grandes en las opciones de respuesta, para ayudar a personas con limitaciones en la movilidad o la visión.
Se estudiaron el tamaño y espacio para facilitar la aproximación y el uso	Se optó por utilizar botones grandes en los menús del sistema.

Fuente: Elaboración propia basada en Thompson, Johnstone & Thurlow (2002).

Como puede observarse en la tabla anterior, las áreas del sistema para las cuales se desarrollaron procedimientos y acciones específicos fueron las de presentación, navegación, respuesta, escenario, programación, duración y uso del equipo y la interfaz. En particular, se enfatizó que el texto y el formato de los menús y los ítems fueran claros, concisos y legibles; por ejemplo el tipo y tamaño de fuente, el alto contraste entre el texto y el fondo, el uso preferente de ítems de selección y el tamaño apropiado de las burbujas para responder, la eliminación de una hoja de respuestas separada del examen, y la duración de la prueba mayor a la estimada (Véase el apéndice 4 donde se ilustran algunos de estos aspectos).

Cabe señalar que el diseño universal no elimina la necesidad de efectuar adaptaciones al administrar las pruebas. Por ejemplo, si un examinado es ciego y debe responder la prueba tendrá que diseñarse una versión en braille o emplear una modalidad que presente los ítems de manera oral. Sin embargo, la idea es que el diseño universal pueda reducir de manera significativa tales acomodaciones o adaptaciones.

En la tabla 3.5 se presentan las dimensiones y características para propiciar la máxima legibilidad del texto en SALIDEA, que fueron retomadas de la propuesta de Thompson y Thurlow (2002) que se presentó en la tabla 2.7.

Tabla 3.5
Características de máxima legibilidad textual consideradas al diseñar SALIDEA

Dimensión	Características de legibilidad
Contraste	Letra de color negro en fondo de color pastel o blanco para favorecer la legibilidad y disminuir la fatiga visual.
Tamaño de la letra	Tamaño grande de letra para apoyar a los niños que están aprendiendo a leer, y a los alumnos con dificultades visuales y personas con problemas de fatiga ocular. En SALIDEA el tamaño legal previsto para el texto es de 14 puntos.
Tipo de letra	Tipo de letra estándar, con mayúsculas y minúsculas, para facilitar la lectura.
Justificación	Texto alineado a la izquierda, para apoyar a lectores con dificultad para leer.
Longitud del renglón	Longitud del renglón de 4 pulgadas (8 a 10 palabras por línea) para evitar la fatiga de lector y la dificultad para localizar el inicio del siguiente renglón.
Gráficas y tablas	Símbolos altamente discriminables en los gráficos, etiquetas colocadas junto a las barras o líneas de las gráficas, para su fácil localización.
Ilustraciones	Ilustraciones ubicadas al lado de la pregunta que las requieran y siempre que incluyan información cuyo dominio se evalúa.
Formatos de respuesta	Opciones de respuesta con burbujas más grandes en los ítems de respuesta seleccionada.

Fuente: Elaboración propia basada en Thompson, Johnstone & Thurlow (2002).

Para ilustrar la aplicación de estas características al texto, en el apéndice 4 se muestra el aspecto que tendrían los textos en un ítem administrado por SALIDEA.

3.2. Análisis de requerimientos

Para establecer cómo cubriría SALIDEA el espacio entre la definición del software a nivel sistema y el propio diseño del programa, se llevó a cabo un análisis de requerimientos. Esta técnica analítica permite especificar los rasgos operacionales del software (función, datos y rendimientos), conectar la interfaz con otros elementos del sistema y establecer las restricciones que debe cumplir el software (Gómez, 2011). En esencia, es un proceso de descubrir, analizar, documentar, especificar y verificar los requisitos del software. En esta etapa del diseño se busca aclarar lo que desea el usuario y la forma en que se va a presentar la solución buscada. Es decir, permite dejar claro lo que el software debe hacer y bajo qué circunstancias debe hacerlo.

Tabla 3.6
Análisis de requerimientos para el diseño de SALIDEA.

Requisito	Descripción	Elementos, datos o secciones
R1. Registrar administrador	La cuenta de administrador debe ser creada por <i>default</i> . Es decir, el sistema debe de tener un primer usuario predeterminado al momento en que alguien desee administrar la prueba.	Usuario: admin Contraseña: 123456
R2. Actualizar datos del administrador	Una vez registrado el administrador, aparecerán unos cuadros de texto, los cuales deben de ser llenados para actualizar la información de dicho usuario.	Nombre completo Fecha de nacimiento Correo electrónico Correo electrónico alternativo
R3. Registrar usuarios (examinados, profesores, padres de familia, directivo y autoridad educativa)	Para darse de alta en el sistema es necesario hacer clic en el botón “crear cuenta”; aparecerá un listado de roles y se podrá seleccionar la opción que corresponde a cada usuario. Enseguida aparecerán cuadros de texto que deben de ser llenados por el usuario correspondiente. Al momento de registrarse un usuario, el administrador de SALIDEA recibirá una notificación que dicho usuario desea darse de alta, por lo que el administrador cotejará con su base de datos si es factible que este usuario pueda darse de alta en el sistema. En caso de que se acepte la solicitud, el administrador dará clic en el botón aceptar solicitud. Al hacer esto, se le enviará un correo al usuario indicándole que su solicitud ha sido aceptada.	Nombre completo Fecha de nacimiento Correo electrónico Correo electrónico alternativo
R4. Iniciar sesión	Se deberá dar clic en el botón “iniciar sesión” que se encuentra en el menú principal de SALIDEA. Se llenarán los cuadros de texto e inmediatamente se dará clic en la opción “entrar”.	Usuario Contraseña
R5. Recordar contraseña	En caso de que un usuario olvide su contraseña, se dará clic en el hipervínculo con el nombre de “olvidó contraseña” y se dará clic en el botón recuperar contraseña. Aparecerá una ventana donde el usuario deberá escribir su correo electrónico y luego debe dar clic en “recordar contraseña”. Inmediatamente se enviará un correo a la dirección de correo registrada, con una contraseña nueva.	Correo electrónico con una contraseña nueva.
R6. Pantalla principal	Al iniciar sesión, el usuario podrá ver el menú principal: El menú tendrá una configuración particular dependiendo el usuario.	Menú de inicio
R7. Pantalla administrador	Para el administrador aparecerá la pantalla con un menú en el cual se encuentran botones para modificar la administración de la prueba.	Página de inicio Agregar prueba

Requisito	Descripción	Elementos, datos o secciones
		Agregar manual técnico
R8. Pantalla examinado	Para el examinado aparecerá la pantalla con el menú donde se encuentran los manuales y la opción iniciar prueba.	Manual de la prueba Iniciar prueba
R9. Pantalla padre de familia	Para el padre de familia aparecerá el menú con las opciones de ver el manual técnico de la prueba y ver los resultados de la prueba de su hijo.	Manual de la prueba Resultados
R.10 Pantalla directivos	Aparecerá en la pantalla el menú con las opciones de ver el manual técnico de la prueba y ver los resultados de los estudiantes por escuela y zona.	Manual de la prueba Resultados
	Antes de poder administrar una prueba en SALIDEA es necesario que se registren los datos generales del instrumento. Para ello, el administrador dará clic en la opción “Agregar prueba” que se encuentra en el menú principal del administrador y enseguida aparecerán unos cuadros de texto que deben de ser llenados por el administrador.	Nombre de la prueba Áreas del conocimiento evaluar Escuelas a las que se administrará la prueba Grado escolar de los examinados
R11.Datos generales de la prueba	Al dar clic en guardar, aparecerá una nueva pantalla de captura de información relacionada con el contenido a evaluar con la prueba. En la parte inferior de la pantalla aparecerán dos botones: “aceptar” y “vista previa”. El primero guarda la información, mientras que el segundo permite visualizar la información cargada para poder detectar algún error en la captura.	Propósitos de la prueba ¿Qué pretende medir? ¿Cómo se seleccionaron los contenidos? Usos válidos de la prueba. Aviso de privacidad. Importancia de la prueba. Instrucciones de la prueba. Ejemplo de la prueba (ítem muestra).
R12. Documentación de la prueba	Para contar con evidencias de validez de la prueba, el administrador deberá agregar los documentos del desarrollo y validación del instrumento preparados por los autores de la prueba, mismos que estarán disponibles para todos los usuarios. Para cargar los documentos, dar clic en el botón “documentos de validación de la prueba”. Aparecerán recuadros para subir dichos documentos. En la parte inferior de cada recuadro aparecerán los botones “buscar” y “subir” que permitirán localizar e incorporar a la base de datos dicha documentación. Al dar clic, se dará aviso al administrador, en caso de éxito	Materiales para capacitar el comité desarrollador del examen. Documento de justificación del contenido importante a evaluar. Retícula o tabla que articula el dominio del contenido a evaluar. Documento de justificación del contenido importante a evaluar Tabla de especificaciones del

Requisito	Descripción	Elementos, datos o secciones
	☒ o fracaso ☐ . Debajo de este aviso habrá un botón para aceptar, mismo que llevará al usuario a la pantalla principal de la prueba.	examen. Documento con las especificaciones de ítems. Ítems y versiones calibrados tras prueba empírica.
R13.Dar formato y subir prueba	Fase 1. Para dar formato a la prueba, de manera que se incorpore al sistema, es necesario descargar y cumplimentar el formato Excel llamado “Formato prueba” que proporciona SALIDEA, el cual se encuentra en el menú principal de la prueba. - Los ítems no tendrán un máximo o mínimo de palabras, debido a que estos ya están previamente probados y calibrados. Fase 2. Al concluir la captura de la prueba, se debe guardar el archivo y subirlo a SALIDEA. Esto se logra dando clic en el botón “subir archivo de examen”. Fase 3. SALIDEA detectará aquellos archivos cuyo nombre incluye una extensión que identifica gráficos, como .jpeg, .img u otros. Aparecerá un directorio específico en el que se subirán las imágenes correspondientes a los ítems.	Formato de la prueba en el que se consignan para cada ítem: las instrucciones específicas para responderlo, la información contextual para responderlo, la base del ítem, las opciones de respuesta, el texto o gráfica que en su caso acompaña a cada ítem, así como la clave de respuesta.
R14.Configuración de la prueba	Una vez que el examen se encuentra en SALIDEA, dar clic en configurar la administración de la prueba. Aparecerá un menú con botones que representan opciones que el administrador puede activar o desactivar según determinados criterios.	Opción para aleatorizar ítems. Opción para aleatorizar clave y distractores. No omitir ítems. Marcar para revisión cuando los ítems que presentan alta dificultad. Revisar los ítems anteriores y cambiar respuesta.
R15. Reporte de resultados	En esta sección es donde se integran las aplicaciones PARRERC y PAGI. Dichos programas de software fueron elaborados para analizar los resultados de la prueba y generar reportes dirigidos a diversos tipos de usuarios. Además, con ellos se puede obtener evidencia sobre la calidad técnica de la prueba mediante el cálculo de la dificultad, discriminación y funcionamiento de distractores de los ítems, así como la confiabilidad y correlación de los ítems con el total de la prueba. Dependiendo del tipo de	Reportes de resultados, según usuario

Requisito	Descripción	Elementos, datos o secciones
	usuario con el que se haya iniciado la sesión, será el formato y contenido del reporte.	
R16. Reporte de resultados: examinados	El reporte debe contener el total y el porcentaje de aciertos que obtuvo en general y en cada unidad de aprendizaje, así como el porcentaje promedio de aciertos del grupo al que pertenece, para que ubique su resultado entre el obtenido en promedio por el resto de compañeros.	Reporte de resultados para los examinados
R17. Reporte de resultados: docentes	Informe sobre el desempeño global y por unidad de aprendizaje del grupo, así como el desempeño individual de cada estudiante, en general y por unidad de aprendizaje. Asimismo, debe contar con datos que permitan ubicar el desempeño general por asignatura y por unidad de aprendizaje del grupo, en relación con el desempeño de otros grupos del mismo grado; así como el promedio general y por unidad de aprendizaje por grupo.	Reporte de resultados a docentes
R18. Reporte de resultados: padres de familia	La información que se les brinda en el reporte de resultados a los padres de familia es igual a la de los alumnos. Incluye el total y el porcentaje de aciertos general y por unidad de aprendizaje del estudiante y el porcentaje promedio de aciertos del grupo, para ubicar el porcentaje promedio de su hijo en relación con el desempeño promedio del resto de sus compañeros.	Reporte de resultados a padres de familia
R.19 Reporte de resultados: directivos	En este caso el reporte es diferente. Lo que debe contener el informe al director es el resultado general y por unidad de aprendizaje por materia y grupo evaluados. De igual manera, se entrega el promedio y el porcentaje promedio de aciertos en la prueba y por unidad de aprendizaje de toda la escuela. Adicionalmente, los inspectores y asesores técnico pedagógicos reciben información con la que puedan comparar el desempeño en las escuelas a las que prestan servicio, así como con el de las escuelas del municipio o municipios con circunstancias similares.	Reporte de resultados a directivos
R20. Responder prueba	Para que los examinados respondan la prueba, es necesario que inicien sesión y vayan al menú principal y dar clic en el botón “Prueba” aparecerá una pantalla nueva en donde se podrán ver los datos generales de la prueba; además, se verá el botón de “iniciar prueba” y al dar clic en él aparecerán las instrucciones generales y el primer ítem. Una vez que el	Zoom Marcar para revisión Datos de los ítems: Ítems terminados Ítems faltantes

Requisito	Descripción	Elementos, datos o secciones
	examinado responda el primer ítem aparecerá un icono en forma de flecha en la esquina inferior derecha, al dar clic se pasará al siguiente ítem; asimismo aparecerá una flecha en la esquina inferior izquierda, al dar clic se regresará al ítem anterior. La prueba tendrá elementos informativos. Nota: Si llegará ocurrir algún problema con la conexión a internet o el alumno decidiera cerrar sesión sin haber terminado el examen, no se podrán guardar las respuestas que se hayan elegido.	Ítems marcados para revisión Tiempo restante de la prueba
R21. Finalizar prueba	Al llegar al último ítem aparecerá la opción de “terminar prueba” al dar clic aparecerá un texto en donde se pregunta ¿está seguro que quiere terminar la prueba? y las opciones “Sí” y “No”. Si se da clic en “Sí” la prueba finalizará. Si se selecciona “No” se regresará al examen.	Terminar prueba
R22. Cerrar sesión	Desde el momento en el que se inicia sesión, aparecerá un texto en la parte superior izquierda con el nombre del usuario y a un lado la opción “cerrar sesión”. Al dar clic aparecerá el texto “Seguro que desea cerrar sesión en SALIDEA” y dos botones “Aceptar” y “Cancelar”.	Cerrar sesión
R23. Seguridad	El administrador de SALIDEA evita que alguna persona ajena pueda ver la prueba o los resultados, debido a que no cualquier persona puede darse de alta en SALIDEA sin previa aprobación del administrador.	Acceso protegido

Fuente: Elaboración propia.

Capítulo IV. Resultados.

En este apartado se muestran los principales resultados del diseño de SALIDEA, en particular del análisis de requerimientos que fue necesario hacer para su diseño. También, se describen las funciones que SALIDEA permitirá realizar a cada usuario, todo esto apegado a los estándares de calidad técnica de las evaluaciones que fueron definidos en los capítulos 2 y 3. Asimismo, se presenta la estructura general del sistema en un esquema en el que se explica cómo se integraron a SALIDEA las aplicaciones PAGI y PARRERC. Además, se prefigura la funcionalidad del sistema con un ejemplo que presenta los componentes de una prueba y las relaciones entre ellos, que se requieren para su administración estandarizada en línea. El ejemplo muestra las interfaces de usuario desarrolladas para el administrador, el examinado y los demás usuarios de SALIDEA. A continuación se muestran y describen dichos resultados.

4.1 Resultados del análisis de requerimientos

4.1.1 Referencias

Se especificó que SALIDEA opere bajo la Licencia Creative Commons Atribución–No Comercial–Compartir Igual 4.0 Internacional, a la vez agregue la funcionalidad del programa PARRERC, el cual es una aplicación distribuida bajo la licencia “Reconocimiento – No Comercial – Compartir Igual 3.0 Unported”, según la clasificación de la organización Creative Commons y desarrollada en el lenguaje de

programación JAVA. PARRERC© fue diseñado por Cristian Castañeda para obtener su grado de maestría con la tesis denominada “Desarrollo de un software para el análisis de datos y la emisión de informes de resultados de exámenes de referencia criterial alineados con el currículum” (2013). Además se especificó que integre el software libre PAGI©, creado por María Guadalupe De Los Santos, para obtener su grado de maestría con la tesis “Desarrollo, operación y evaluación de un módulo para capacitar docentes en servicio para que mejoren sus exámenes de opción múltiple mediante el análisis grafico de ítems” (2010)

Como ya se comentó se consideró también que SALIDEA esté apegado a estándares de calidad para instrumentos de evaluación propuestos el Joint Committee on Testing Practices en el documento “Code of Fair Testing Practices in Education” y el Consejo Asesor Externo del Centro Nacional de Evaluación para la Educación Superior de México del CENEVAL “Estándares de Calidad para Instrumentos de Evaluación Educativa”. Además, se diseñó para atender los estándares del diseño Universal de las evaluaciones propuestos por la Universidad de Minnesota en el documento “A state guide to the development of universally designed assessments” y “Universally designed assessments: Better tests for everyone”

4.1.2 Interfaces con el Software

El usuario no requiere un sistema operativo en particular para el correcto funcionamiento de la aplicación Web. Asimismo, el sistema tendrá la posibilidad de ser ejecutado en cualquier tipo de navegador. Se integrará el programa PARRERC© para

realizar el análisis y reporte de resultados y el programa PAGI© para efectuar análisis gráfico de ítems.

4.1.3 Proveedor

Se tiene considerado que el encargado de elaborar dicho sistema sea el programador de la Unidad de Evaluación Educativa de la UABC.

4.1.4 Funciones del sistema

Se prevé que el software permita la administración de evaluaciones en línea, por lo que lo primero que debe de hacer es dar de alta a los diferentes tipos de usuarios.

4.1.5 Requisitos de funcionalidad

R1. Registrar al administrador


SALIDEA
Sistema para administrar en línea instrumentos de evaluación del aprendizaje

Crear cuenta

Inicar sesión

Servicios

¿Qué es?

Nombre usuario

Admin

Contraseña

123456

R2. Actualizar datos del administrador



Nombre	
Fecha de nacimiento	
Correo electronico	
Nombre usuario	
Contraseña	
Entrar	



R3. Registrar usuarios (examinados, profesores, padres de familia, equipos directivos)



Usuario	
Nombre	
Fecha de nacimiento	
Correo electronico	
Nombre usuario	
Contraseña	
Entrar	

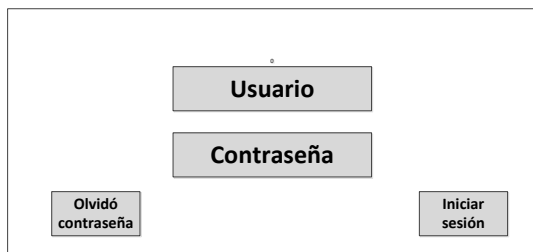
Examinados
Profesores
Padres de familia
Equipos directivos



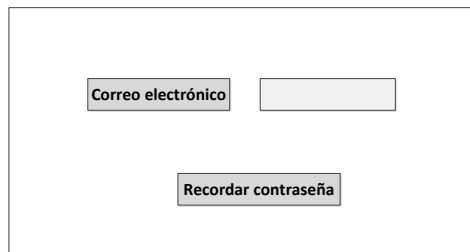
R.4 Iniciar sesión



R5. Recordar contraseña



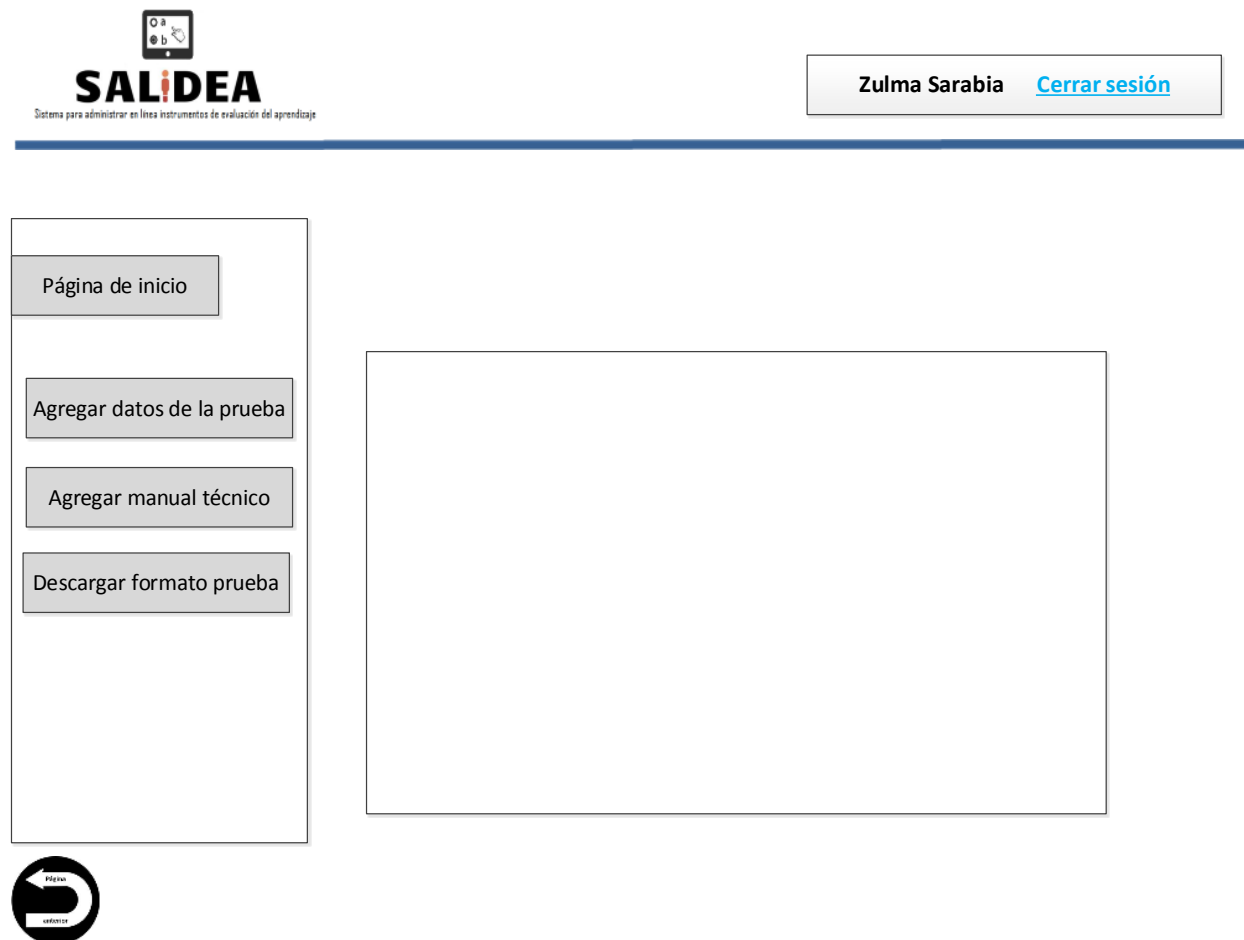
A form for password recovery with four input fields. The first field is labeled "Usuario" and the second is labeled "Contraseña". Below the "Contraseña" field is a button labeled "Iniciar sesión". To the left of the "Contraseña" field is a button labeled "Olvidó contraseña".



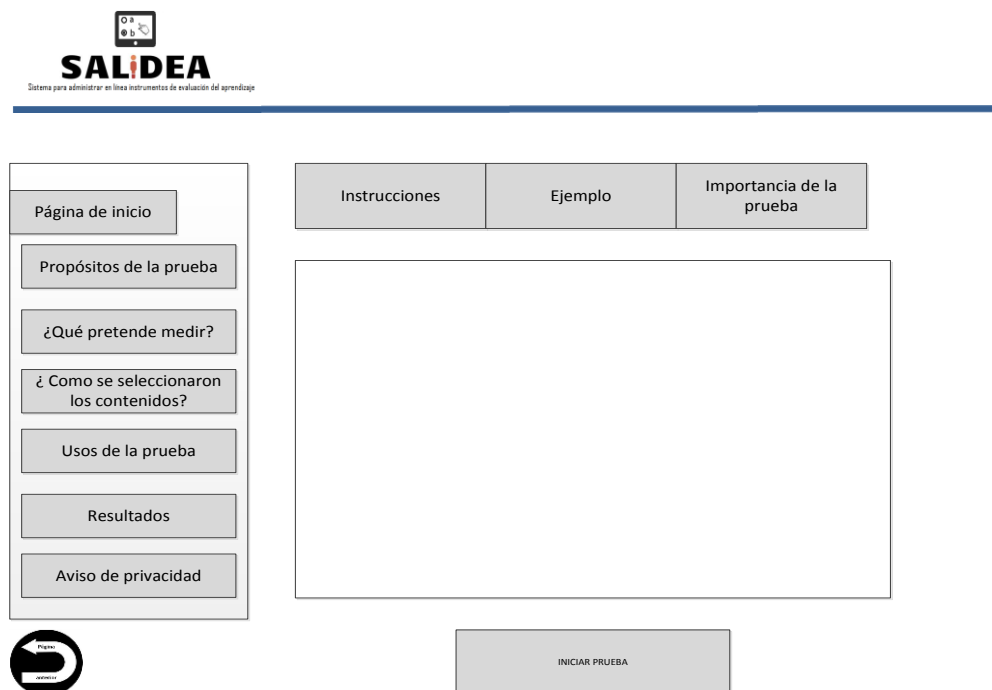
A form for password recovery with two input fields. The first field is labeled "Correo electrónico" and the second is an empty field. Below the second field is a button labeled "Recordar contraseña".



R6. Pantalla principal administrador



R7. Pantalla examinados



R8 / R9 / R10. Pantalla padres de familia, docentes y directivos



Página de inicio
Propósitos de la prueba
¿Qué pretende medir?
¿ Como se seleccionaron los contenidos?
Usos de la prueba
Reporte de resultados



Instrucciones	Ejemplo	Importancia de la prueba
---------------	---------	--------------------------

R11. Datos generales de la prueba

Página de inicio Propósitos de la prueba ¿Qué pretende medir? ¿ Como se seleccionaron los contenidos? Usos de la prueba Resultados Aviso de privacidad	Instrucciones Ejemplo Importancia de la prueba
---	---

Propósitos de la prueba

Agregar texto



R12. Documentos con evidencias de validez de la prueba

Página de inicio Editar examen Manual técnico	Materiales para capacitar el comité desarrollador del examen Buscar Subir ☒ ☐ Documento de justificación del contenido importante a evaluar Buscar Subir ☒ ☐ Retícula Buscar Subir ☒ ☐ Documento de justificación del contenido importante a evaluar Buscar Subir ☒ ☐ Tabla de especificaciones del examen Buscar Subir ☒ ☐ Documento con especificaciones de ítems Buscar Subir ☒ ☐ Ítems y modelos calibrados tras prueba empírica Buscar Subir ☒ ☐
--	--



R13. Elaborar y subir prueba

**SALIDEA**
Sistema para administrar en línea instrumentos de evaluación del aprendizaje

Zulma Sarabia [Cerrar sesión](#)

Descargar archivo ↓

Subir archivo ↑



Nota: Es necesario que el usuario cuente con Microsoft Access



Configurar prueba

R14. Configurar prueba

**SALIDEA**
Sistema para administrar en línea instrumentos de evaluación del aprendizaje

Zulma Sarabia [Cerrar sesión](#)

Configurar prueba

☐ Aleatorizar ítems

☐ Aleatorizar clave y distractores

☐ No omitir ítems

☐ Responder todos los ítems de la página antes de ir a la página siguiente.

☐ Los estudiantes pueden seleccionar la opción "marcar para revisión" para aquellos ítems que presenten dificultad alta.

☐ Revisar los ítems anteriores y cambiar respuesta



R15 / R16 Reporte de resultados alumnos

REPORTE DE RESULTADOS

El examen de Español aplicado contenía 44 ítems. El porcentaje obtenido por el alumno fue de 90%. A continuación se muestra el informe detallado:

Porcentaje de aciertos	Unidades de aprendizaje		
	Unidad 1	Unidad 2	Unidad 3
90%	30%	30%	30%
Porcentaje grupal			
60%	20%	20%	20%

No. ítems	Unidades de aprendizaje		
	Unidad 1	Unidad 2	Unidad 3
33	11	11	11



R17. Reporte de resultados docentes

REPORTE DE RESULTADOS GRUPO

Los estudiantes del grupo B obtuvieron un puntaje promedio de **27.818** el cual es superior al puntaje promedio de todos los estudiantes examinados en esta escuela, el cual se muestra en la tabla 2. El puntaje más alto de este grupo fue de 35 y el más bajo fue de 20.

Tabla 1. Resultados generales del grupo B

Grupo	No. de examen	Puntaje	Porcentaje
B	130	35	79.545%
	124	33	75.000%
	133	33	75.000%
	127	31	70.455%
	129	27	61.364%
	132	27	61.364%
	128	26	59.091%
	134	26	59.091%
	126	25	56.818%
	131	23	52.273%
	125	20	45.455%
Total grupo:		27.818	63.223%

Tabla 2. Resultados generales de la escuela

Turno	Grupo	Promedio	Porcentaje
1	B	27.818	63.223%
	C	24.474	55.622%
	A	28.333	64.394%
Total escuela:		26.139	59.407%



R18. Reporte de resultados padres de familia



Padre de familia [Cerrar sesión](#)

REPORTE DE RESULTADOS DEL ALUMNO X

El examen de Español aplicado contenía 44 ítems. El porcentaje obtenido por el alumno fue de 90%. A continuación se muestra el informe detallado:

Porcentaje de aciertos	Unidades de aprendizaje		
	Unidad 1	Unidad 2	Unidad 3
90%	30%	30%	30%
Porcentaje grupal			
60%	20%	20%	20%

No. ítems	Unidades de aprendizaje		
	Unidad 1	Unidad 2	Unidad 3
33	11	11	11



R19. Reporte de resultados directivos



Director [Cerrar sesión](#)

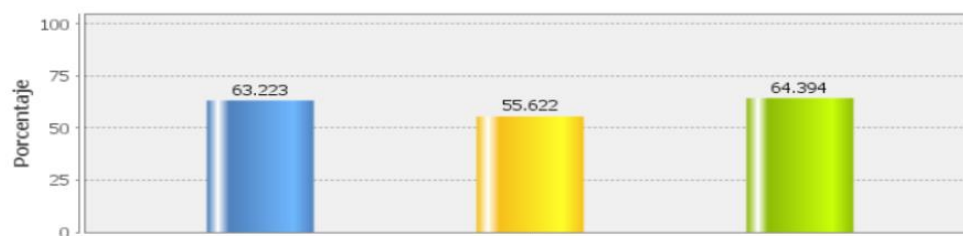
REPORTE DE RESULTADOS DE LA ESCUELA

La Escuela tiene una media de aciertos de 26.139, la cual es superior a la media de su zona escolar.

Tabla 1. Resultados generales de la escuela

Turno	Grupo	Promedio	Porcentaje
1	B	27.818	63.223%
	C	24.474	55.622%
	A	28.333	64.394%
	Total turno	26.139	59.407%
Total escuela:		26.139	59.407%

Escuela



R20. Responder prueba



Items
terminados
1

Items
faltantes
30

Items
marcados
para revisión
0

Tiempo
restante
33 min

Instrucciones. Enseguida se presenta la parte inicial de varios textos que corresponden a ciertos géneros literarios populares. En cada caso, selecciona la opción que indica de qué género se trata:

2. "Dicen que existió un pueblo nómada. . ." Se trata de:

- Una leyenda
- Una adivinanza
- Una copla
- Un dicho



R21. Finalizar prueba



ALUMNO X

[Cerrar sesión](#)

¿ Seguro qué desea terminar la prueba?

SI

NO



R22. Cerrar sesión

**SALIDEA**
Sistema para administrar en línea los instrumentos de evaluación del aprendizaje

Zulma Sarabia [Cerrar sesión](#)

Items terminados
1

Items faltantes
30

Items marcados para revisión
0

Instrucciones. Enseguida se presenta la parte inicial de varios textos que corresponden a ciertos géneros literarios populares. En cada caso, selecciona la opción que indica de qué género se trata:

2. "Dicen que existió un pueblo nómada. . ." Se trata de:

- Una leyenda
- Una adivinanza
- Una copla
- Un dicho





¿ Seguro que desea cerrar sesión en SALIDEA?

aceptar

cancelar

4.2 Resultados del diseño de SALIDEA

4.2.1 Estructura general del diseño de SALIDEA

Uno de los resultados más importantes del presente trabajo se presenta en la figura 4.1, en la cual se muestran los componentes y elementos que se diseñaron para SALIDEA, así como las relaciones entre ellos que muestran las funcionalidades del sistema que se derivaron del análisis de requerimientos cuyos resultados se mostraron en el punto anterior. No obstante, también integra otros componentes y elementos mencionados a lo largo del trabajo, como la base de datos relacional que acopia la información y alimenta con ella al sistema; o bien la incorporación de las aplicaciones PAGI y PARRERC.

En la figura 4.1 se ilustran los requerimientos del sistema que se describieron en el capítulo III, desde registrar los usuarios hasta entregarles el correspondiente reporte de resultados. También se muestra la forma en que se integran al sistema los programas analíticos PAGI y PARRERC; es decir, se observa cómo al momento en que un estudiante termina la prueba, se creará una base de datos que se dividirá en: datos generales sobre los estudiantes que respondieron la prueba y datos sobre su ejecución en ella (las respuestas de los examinados), los cuales a su vez se dividirán en Datos para análisis psicométricos, Datos para analizar la ejecución por dominio y subdominio evaluados y finalmente en Datos para analizar variables de contexto y su relación con la ejecución, los cuales serán tomados por las aplicaciones PAGI y PARRERC, para analizar los resultados y elaborar los reportes de resultados.

En ausencia de una aplicación informática, con las características y condiciones para la administración de pruebas a gran escala, que se comentaron al inicio del presente documento, inicialmente fue necesario crear una simulación de la página web con sus pantallas, controles e hipervínculos, lo cual se hizo con el apoyo del programa Microsoft Visio 2010. Sin embargo, cuando se desarrolló posteriormente el análisis de requerimientos se notó que existían varios usuarios, por lo que crear los hipervínculos correspondientes complicaría el diseño y no se tendría una visión clara y simple del sistema. Ello ocasionaría también que se perdiera una visión estructurada del mismo.

En consecuencia, se decidió crear retículas en Microsoft Visio 2010 con dichos elementos para los diferentes usuarios del sistema, en donde se apreciara la estructura particular formada por las pantallas y controles, así como las correspondientes

relaciones entre ellos representadas mediante flechas que los unen, mismas que también pueden representar la manera en que se atienden los estándares de calidad técnica o la funcionalidad de los controles de SALIDEA.

Las retículas de los diferentes usuarios que forman parte del diseño de SALIDEA se presentan en el apéndice 5 (contenido en el CD adjunto a este documento) y se describen a continuación de manera general.

4.2.2 Interfaz con el administrador (Ver apéndice 5)

Al administrador es el usuario principal de SALIDEA. Lo primero que debe hacer es darse alta como administrador. SALIDEA le dará un usuario predeterminado, inmediatamente tendrá que actualizar los datos. Al hacer esto aparecerá un menú con las opciones “Agregar los datos de la prueba” “Agregar manual técnico de la prueba” y “Descargar el formato de la prueba”. En la opción agregar datos de la prueba, el administrador debe proporcionar los datos generales de la prueba, tales como:

- Propósitos de la prueba
- ¿Qué pretende medir?
- ¿Cómo se seleccionaron los contenidos?
- Usos de la prueba
- Aviso de privacidad
- Importancia de la prueba
- Instrucciones de la prueba
- Ejemplo de la prueba (ítem muestra)

En la opción de Agregar manual técnico de la prueba, el administrador deberá insertar las evidencias y documentos probatorios de la validez de la prueba que aportó el desarrollador del instrumento:

- Materiales para capacitar el comité desarrollador del examen
- Documento de justificación del contenido importante a evaluar
- Retícula o tabla que presenta la estructura del dominio a evaluar
- Documento de justificación del contenido importante a evaluar
- Tabla de especificaciones del examen
- Documento con especificaciones de ítems
- Ítems y versiones calibrados tras prueba empírica

Una vez agregados estos datos, el administrador deberá descargar el formato de la prueba. Ahí es donde debe insertar los ítems del examen y la clave. Tras llenar el formato deberá subirlo a SALIDEA. Aparecerá la opción de configurar prueba, en donde se podrán activar o desactivar las siguientes opciones:

- Configurar prueba
- Aleatorizar ítems
- Aleatorizar clave y distractores
- No omitir ítems
- Responder todos los ítems de la página antes de ir a la página siguiente.
- Los estudiantes pueden seleccionar la opción “marcar para revisión” para aquellos ítems que presenten dificultad alta.
- Revisar los ítems anteriores y cambiar respuesta.

Cuando el total de examinados termina la prueba, el administrador podrá obtener información sobre su calidad técnica mediante la obtención de la dificultad, discriminación y funcionamiento de distractores de los ítems, así como la

confiabilidad y correlación ítem – total en la prueba. También puede obtener el análisis de varianza de un factor (ANOVA) para conocer si existen diferencias significativas en el logro según los resultados de los grupos conformados y otras variables (Castañeda, 2012).

4.2.3 Interfaz con el examinado (Ver apéndice 5)

Lo primero que deben de hacer los examinados es crear una cuenta en SALIDEA. Dicho usuario debe esperar por la confirmación del administrador de la prueba para poder entrar al sistema. Se le mandará un correo electrónico indicándole la confirmación de su cuenta. Una vez iniciada la sesión aparecerá un menú principal, en donde se encuentran las opciones “Examen”, “Validez de la prueba” y “Reporte de resultados”. Al dar clic en Examen se mostrará un menú diferente, en donde el examinado podrá conocer los datos generales de la prueba, como lo son:

- Propósitos de la prueba
- ¿Qué pretende medir?
- ¿Cómo se seleccionaron los contenidos?
- Usos de la prueba
- Aviso de privacidad
- Importancia de la prueba
- Instrucciones de la prueba
- Ejemplo de la prueba (ítem muestra)

Además aparecerá el botón “Iniciar prueba”; al dar clic se iniciará el examen. Al lado derecho de la pantalla habrá un recuadro en el que se indica: los ítems terminados, ítems faltantes y los ítems que fueron marcados para revisión. Abajo de la pantalla

aparecerán cuatros íconos; uno es para hacer la pantalla más grande y el otro para reducir el tamaño de la pantalla. Los otros dos son para regresar un ítem atrás y el otro para avanzar al siguiente ítem. Al llegar al último ítem aparecerá la opción de Finalizar la prueba. Al dar clic en Finalizar prueba se mostrarán otros dos botones; el primero de ellos para Cerrar sesión, mientras que el segundo es para Mostrar los resultados de la prueba, el cual mostrará el puntaje total y el porcentaje de aciertos que obtuvo, en general y en cada unidad de aprendizaje.

4.2.4 Interfaz con los padres de familia (Ver apéndice 5)

Al igual que en el caso de los demás usuarios, lo primero que deben hacer los padres de familia es crear una cuenta en SALIDEA. Dicho usuario debe esperar por la confirmación del administrador de la prueba para poder entrar al sistema. Para ello, se le enviará un correo electrónico indicándole la confirmación de su cuenta. Una vez iniciada la sesión aparecerá un menú principal, en donde se encuentran las opciones: “Examen”, “Validez de la prueba” y “Reporte de resultados”. Al dar clic en Examen se mostrará un menú, en donde el padre de familia podrá conocer los datos generales de la prueba, como lo son:

- Propósitos de la prueba
- ¿Qué pretende medir?
- ¿Cómo se seleccionaron los contenidos?
- Usos de la prueba
- Aviso de privacidad
- Importancia de la prueba
- Instrucciones de la prueba

-
- Ejemplo de la prueba (ítem muestra)

Además los padres de familia podrán ver el puntaje total y por área evaluada que su hijo obtuvo tras responder la prueba. Para ello, hará clic en la opción Ver resultados de la prueba que aparece en el menú principal.

4.2.5 Interfaz con los profesores (Ver apéndice 5)

Lo primero que deben de hacer los docentes es también crear una cuenta en SALIDEA. Dicho usuario debe esperar por la confirmación del administrador quien le mandará un correo electrónico indicándole que se ha creado su cuenta. Una vez iniciada la sesión aparecerá un menú principal, en donde se encuentran las opciones “Examen”, “Validez de la prueba” y “Reporte de resultados”. Al dar clic en Examen se mostrará un menú, en donde el profesor podrá conocer los datos generales de la prueba. Además, los maestros podrán ver los resultados de sus estudiantes al dar clic en la opción “Reporte de resultados”. Cabe señalar que los docentes solo podrán ver las puntuaciones que obtuvieron sus estudiantes, pero no podrán identificar a cada uno de ellos, debido a que el sistema fue diseñado para preservar el anonimato de los estudiantes.

4.2.6 Interfaz con los directores (Ver apéndice 5)

Lo primero que deben de hacer los directores es crear una cuenta en SALIDEA. Este usuario también debe esperar por la confirmación del administrador de la prueba para poder entrar al sistema. Para ello, se le mandará un correo electrónico indicándole la confirmación de su cuenta. Una vez iniciada la sesión aparecerá un menú principal, en donde se encuentran las opciones “Examen”, “Validez de la prueba” y “Reporte de

resultados”. Al dar clic en Examen se mostrará un menú, en donde el director podrá conocer los datos generales de la prueba obtenidos por los estudiantes del plantel.

El director podrá ver los resultados de los estudiantes por grupo, al dar clic en la opción “Reporte de resultados”; al igual que los docentes, los directores solo podrán ver los puntajes de los estudiantes sin identificar el nombre de cada uno de ellos.

4.2.7 Interfaz con autoridades educativas (Ver apéndice 5)

Al igual que todo usuario del sistema, los funcionarios que son autoridades educativas deben primero crear una cuenta en SALIDEA. Harán la solicitud y esperarán la confirmación del administrador quien confirmará la creación de su cuenta vía correo electrónico. Una vez iniciada la sesión también aparecerá el menú principal, en donde se encuentran las opciones “Examen”, “Validez de la prueba” y “Reporte de resultados”. Al dar clic en Examen se mostrará un menú, en donde el funcionario podrá conocer los resultados generales de la prueba.

Las autoridades educativas solo podrán ver los resultados de los estudiantes por grupo, escuela, zona escolar y municipio, sin poder identificar los nombres de cada uno de los estudiantes.

Capítulo V. Discusión

Como se estableció en el Capítulo I, para darle sentido y orientación a este trabajo se propuso un objetivo general y cuatro objetivos específicos. En esta última sección del documento se analizarán y discutirán los resultados obtenidos que se presentaron en el capítulo anterior, para observar en qué medida con ellos se lograron los objetivos propuestos. Así, se discutirá la manera en que se analizaron los requisitos de funcionalidad del sistema, es decir, la manera en la que operará SALIDEA y de qué manera tomó en cuenta los lineamientos del diseño universal y los estándares psicométricos propuestos por diversos organismos (primer objetivo específico). Se examinará la interfaz gráfica que fue utilizada en SALIDEA, formas, colores, menús, tipo de letra, fondos (segundo objetivo específico). Por otra parte, se explicará cómo se prevé que sean incorporadas al sistema las aplicaciones PAGI y PARRERC, que permiten elaborar análisis y reportes de resultados de las pruebas (tercer objetivo específico). Se analizará en qué medida el diseño de SALIDEA pudo anticipar la funcionalidad del sistema mediante ejemplos que ilustraran la representación de los componentes de una prueba y las relaciones entre ellos, que requiere un sistema de cómputo para la administración de una prueba(cuarto objetivo específico). Cabe señalar que, en cada caso se comentan también los aspectos y condiciones que limitaron el logro de cada objetivo y se sugieren líneas de acción, tanto para superar esos retos como para mejorar los componentes del sistema o elaborar otros que lo hagan más completo y funcional.

- **Análisis de requerimientos para el diseño de SALIDEA**

El primer objetivo específico que se propuso al inicio del trabajo, fue que el análisis de requerimientos atendiera los principales estándares psicométricos que orientan el desarrollo y aplicación de las pruebas; particularmente los correspondientes a la administración en línea de pruebas al evaluar el aprendizaje a gran escala. La idea fue incorporar al diseño del sistema tales recomendaciones, a fin de incrementar la validez de la prueba y reducir posibles sesgos cuando los estudiantes la respondan. En cuanto a la manera específica en que se atendieron las recomendaciones de los organismos que norman las evaluaciones, tenemos los siguientes componentes considerados en el diseño de SALIDEA:

- Respecto a los estándares psicométricos relativos al desarrollo de la prueba, cabe recordar que el foco de SALIDEA es la administración, no el desarrollo de pruebas. No obstante, en su diseño se consideró necesario que el sistema aportara información y evidencias de validez relacionadas con el contenido de la prueba. En consecuencia, el diseño de SALIDEA incluyó un componente para que el usuario **Administrador** incorpore las versiones del examen que se aplicarán, las secciones del manual técnico de la prueba, la documentación sobre la revisión del sesgo y la ofensa, así como el análisis psicométrico de los ítems y versiones, entre otros documentos que debe aportar la instancia que desarrolló la prueba (véase la retícula del Administrador en el apéndice 5). En consecuencia, el diseño de SALIDEA incorpora una sección al inicio de la prueba que proporciona al examinado información sobre los propósitos de la prueba, una descripción sobre la importancia de lo que se pretende medir, sobre cómo se seleccionaron los contenidos; o bien proporciona instrucciones para responder o ejemplos de ítems que aparecen en la

prueba, entre otros elementos. Con ello, una vez que el sistema se desarrolle estará en condiciones de ofrecer al examinado, no solo una prueba a responder, sino también información relacionada con ella que le resulte significativa y propicie su interés para responder el examen con seriedad (véanse por ejemplo el instructivo para el examinado en el apéndice 3 y la retícula del examinado en el apéndice 5). Lo mismo sucede con otros usuarios del sistema, como padres, profesores y directivos escolares, a quienes se ofrece información general y técnica sobre el instrumento (véanse las retículas correspondientes a esos usuarios en el apéndice 5).

- En cuanto a la atención de los estándares psicométricos relativos a la administración de las pruebas, que es el foco principal de este trabajo, el diseño de SALIDEA consideró desde un inicio que el Administrador es el usuario central del sistema, y que el cumplimiento de este tipo de estándares resulta indispensable. Para ello, el diseño incluyó como componentes principales: a) el instructivo para los desarrolladores del examen que aparece en el apéndice 1, donde se les especifica la documentación que deben proporcionar a fin de que el Administrador de SALIDEA pueda incorporarla al sistema y proceder a su administración; b) el propio instructivo para el Administrador (ver apéndice 2), que especifica sus funciones, responsabilidades y acciones específicas que debe realizar para operar el sistema, administrar el examen y preservar su integridad, así como en general cuidar la validez del instrumento; c) el instructivo para el examinado, que por su naturaleza específica también deben elaborar los desarrolladores del instrumento, y después

incorporar al sistema el Administrador, donde se le proporcionan las instrucciones específicas para responder la prueba y ejemplos de reactivos empleados en ella.

- Respecto a la atención de los estándares psicométricos relativos a la calificación, análisis, interpretación y comunicación de resultados de la prueba, en principio dicha normatividad establece que el desarrollador de la prueba debe proporcionar la información necesaria para orientar tanto los análisis de los datos derivados de la prueba, como la elaboración de los reportes de resultados obtenidos por grupos de examinados, así como la interpretación de los mismos; incluida la información sobre quiénes fueron o no incluidos en los diferentes grupos que se comparan, y la información sobre las variables de contexto exploradas que pudieron influir en los resultados y en su interpretación. Además, al desarrollador del examen le corresponde especificar los usos apropiados de los resultados de la prueba y advertir a los usuarios sobre malos usos potenciales. Esta información debe ser oportuna, por lo que debe ser proporcionada antes de que se publiquen los resultados de la prueba, y estar escrita de manera simple y clara, para que sea comprendida por el examinado y los demás usuarios de la información. Además, el desarrollador proporciona orientación a los usuarios sobre cómo monitorear el grado en que la prueba está cumpliendo con los fines previstos. En consecuencia, toda esta información deberá ser aportada junto con la demás documentación que ya hemos mencionado (examen, manual técnico, etc.), a fin de pueda ser incorporada al sistema.

Para ello, en el diseño de SALIDEA se especificó que el Administrador incorpore la información que corresponda a esos aspectos en tres secciones que fueron

diseñadas: Documentos de validación de la prueba, Resultados y Aviso de privacidad, de modo que todos los usuarios puedan consultarlos en las secciones donde son publicados los resultados.

Respecto a los tipos de análisis y reportes de resultados que se consideraron en el diseño de SALIDEA, en la tabla 3.6 se especificaron varios requerimientos. Por ejemplo, en el requerimiento R15. Análisis y reporte de resultados, se establece que en esta sección es donde se integran las aplicaciones PAGI y PARRERC que llevan a cabo tres procesos: a) califican el examen; b) calculan los parámetros tanto de los ítems (dificultad, discriminación, correlación punto biserial, funcionamiento de distractores, así como su representación gráfica), como de las versiones de la prueba (confiabilidad, distribución de la dificultad y la discriminación, nivel de ajuste a un modelo TRI, etc.); y c) realizan los análisis de la ejecución general en la prueba y en cada una de las áreas de contenido que fueron evaluadas por ella, y además los que relacionan dichas ejecuciones con las variables de contexto que también fueron exploradas.

Por su parte, el diseño de los requisitos R16, R17, R18, y R19, que aparecen en la misma tabla, detalla las clases de datos particulares que deben ser incluidos en cada uno de los reportes de resultados que SALIDEA proporciona de manera diferencial a cada uno de los usuarios del sistema. En la sección 4.1.4 del capítulo de resultados, se ilustran los componentes de la interfaz gráfica que fueron diseñados para representar las clases de resultados que el sistema debe presentar a los alumnos, docentes, padres de familia, directivos escolares y autoridades

educativas. Además, en el apéndice 5 se muestran las retículas correspondientes a las interfaces gráficas que fueron diseñadas para dichos usuarios, en las cuales aparecen también las clases de datos que les corresponden.

En consecuencia, podemos decir que el diseño de SALDEA contempla los aspectos básicos considerados en los estándares que norman la calificación, análisis, interpretación y comunicación de resultados de la prueba, restando solo que el sistema esté funcional y que con ello las aplicaciones PAGI y PARRERC puedan analizar y reportar los resultados de los ítems y de la prueba.

- En cuanto a la atención de los principios y lineamientos del diseño universal de las evaluaciones, algunas de las normativas que se han comentado a lo largo del presente documento debieron ser consideradas previamente por el desarrollador de la prueba. En apéndice 1 se enuncian varias de ellas. Por ejemplo que el lenguaje que se utilizó en los ítems esté acorde con la edad y grado escolar de los examinados, que los textos en los ítems sean breves, claros y significativos, o que se aplique al texto, tablas, figuras e ilustraciones características que aseguran una fácil interpretación de los ítems, así como que se mantengan al mínimo la longitud de las oraciones y el número de palabras difíciles, para producir texto legible y comprensible.

En cuanto a los aspectos que fueron considerados en el diseño de SALIDEA, podemos mencionar que se buscó diseñar una navegación simple y consistente, en donde los menús se encuentran organizados de forma lógica, clara y concisa; con enlaces que utilizan texto breve y significativo y a la vez son flexibles y funcionales para personas con habilidades diferentes.

Además, las herramientas y componentes de la interfaz que están disponibles en SALIDEA se diseñaron cuidadosamente. Por ejemplo se maximizó el contraste: utilizando un fondo color blanco y en los menús un gris tenue; y letras de color negro, para que se puedan leer y ubicar las opciones con facilidad. Se utilizó la fuente APHont, de 14 puntos, recomendada por la Imprenta Americana para Ciegos.

Estos y otros aspectos del diseño del sistema que aparecen en la tabla 3.4 y que se ilustran en el apéndice 4 y en las retículas que aparecen en el apéndice 5, muestran que se atendieron adecuadamente los principales lineamientos del diseño universal de las evaluaciones. Al respecto, cabe señalar que esta es un área emergente y prometedora en el campo del desarrollo de pruebas, la cual ha recibido atención y respaldos muy importantes por parte de los psicómetras, como lo es el hecho de haber sido incorporados sus planteamientos en los recientes estándares emitidos por el comité conjunto que norma las prácticas evaluativas, integrado por la AERA, la APA y el NCME (2014).

De esta manera, a partir de lo que se comentó en los incisos anteriores, puede concluirse respecto al logro del primer objetivo propuesto en el presente estudio, que el diseño de SALIDEA logró atender e incorporar varios de los estándares que se proponen en la literatura especializada para la administración de las pruebas, y que en buena medida se atendieron las recomendaciones que se formulan en el contexto del diseño universal de las evaluaciones. Por lo menos se consideraron las más básicas y las que fue relativamente fácil plasmar en el diseño, como por ejemplo las características de máxima legibilidad de los textos que aparecen en el apéndice 4 o los

controles para ampliar las imágenes y textos que se muestran en ciertas secciones de la interfaz gráfica.

Por otra parte, resultó más difícil incorporar en el diseño las recomendaciones derivadas de los estándares psicométricos para el desarrollo de las pruebas, y los correspondientes a la calificación, análisis, interpretación y reporte de resultados del examen. Sobre tales aspectos, el diseño de SALIDEA solo considera los componentes para integrar los documentos que los desarrolladores deben aportar al usuario Administrador, la sección del sistema donde éste debe incorporarlos y las secciones de la interfaz gráfica donde finalmente pueden consultarlos los demás usuarios. Lo mismo sucede con los componentes para efectuar los análisis y reportes de resultados de la prueba, sobre los cuales únicamente se detallan los tipos de análisis básicos que deben realizarse, tanto los psicométricos como los que muestran el logro de los examinados en la prueba, referido a los dominios y subdominios que son evaluados. Sin embargo, nada se dice sobre la estructura de la base de datos relacional que se requiere, ni sobre la interfaz gráfica donde se especifiquen los análisis a realizar o los tipos de datos que integrarán los reportes de resultados que se entregarán a los diferentes usuarios. Estos y otros aspectos que no fueron considerados deberán ser especificados en un diseño más avanzado que ya esté cercano a la versión operativa de SALIDEA.

- **Interfaz gráfica del sistema de cómputo, de tal manera que cada usuario tenga acceso protegido y pueda transitar por los componentes y opciones que se diseñen de manera exclusiva para él**

Para cumplir con la interfaz gráfica se realizó un diseño a partir del análisis de requisitos. Para representar visualmente estos elementos que fueron diseñados, y dada la ausencia del software operativo que los hiciera funcionales, se elaboró una aproximación al diseño del sistema utilizando el programa de diseño gráfico Microsoft Visio 2010, mediante el cual se elaboraron retículas que muestran las formas y controles que se diseñaron para SALIDEA y que mediante flechas simulan los hipervínculos entre ellos y en general los canales de navegación del sistema. Por otra parte, puesto que SALIDEA cuenta con varios usuarios, se complicó diseñar una interface gráfica general que los incluyera a todos ellos, dado que cada usuario tiene componentes y opciones de navegación diferentes. Por ello, se optó por diseñar una retícula diferente para cada usuario (ver apéndice 5), en la que se mostrarán las forma, controles y elementos de navegación que le son propios. Desde luego, la manera apropiada de superar tales limitaciones e integrar los elementos diseñados será desarrollar una versión operativa de SALIDEA.

En el análisis de requisitos se proporcionaron los pasos que cualquier usuario debe de seguir al momento de darse de alta en el sistema y actualizar sus datos (R1,R2), se especificaron los componentes necesarios para garantizar la integridad del sistema, del instrumento y de la información que aporta el examinado, los cuales incluyeron los procedimientos para el registro de los usuarios al sistema, claves para el ingreso protegido, la previsión de informes de resultados apropiados para cada usuario.

Por otra parte las retículas que se crearon, muestra la navegación de cada usuario por el sistema, por ejemplo, muestra los pasos que el administrador debe de seguir al momento de utilizar SALIDEA, lo primero que debe de hacer es darse de alta en el

sistema, el cual, le dará un usuario y contraseña por default, después tendrá que entrar a la opción de actualizar datos, enseguida el administrador podrá agregar una prueba, y los documentos de validez de la prueba, además de poder configurar la prueba, dando de alta opciones de examen que se consideren pertinentes.

- **Integración de los programas PAGI y PARRERC al SALIDEA**

El diseño de SALIDEA prevé la incorporación de dos software desarrollados en la Unidad de Evaluación Educativa (UEE). El primero de ellos, PAGI, realiza un análisis gráfico de ítems que permite analizar de manera simple, directa y visual la calidad técnica de los ítems, al mostrar cómo responden los examinados los ítems de respuesta seleccionada. En consecuencia los usuarios de SALIDEA podrán conocer las variaciones en las tendencias de la respuesta correcta y los distractores en cada ítem de respuesta seleccionada una vez que se califique la prueba. Con esto se pretende que los usuarios puedan observar la calidad de ítems de respuesta seleccionada que integraron la prueba aunque estos no tengan conocimientos psicométricos.

Por otra parte, en el diseño se consideró integrar también el software PARRERC, con el cual se pueden obtener, a partir de las respuestas de los examinados, estadísticos descriptivos del total de aciertos de los alumnos que respondieron la prueba y el total que obtuvieron en cada una de las partes cuyo dominio se evalúa en el instrumento; en ambos casos desagregados tales resultados por alumno, grupo, turno, escuela y municipio, así como relacionar la ejecución con las variables de los contextos, personal, familiar y escolar de los estudiantes (véase en el menú principal la opción “reporte de resultados”). También incluye el análisis de la varianza de un factor

(ANOVA), para conocer si existen diferencias significativas en el logro según los resultados de los grupos conformados y otras variables.

Los componentes específicos que se diseñaron para incorporar a SALIDEA estas dos funcionalidades ya se describieron en el tercer inciso del punto anterior, donde se comentaron los requerimientos R15 al R19 relativos al análisis y reporte de la información derivada del examen y donde se refirieron los componentes de la interfaz gráfica que se diseñaron para representar los tipos de resultados que SALIDEA debe presentar a los usuarios de la información.

En este punto, solo se enfatiza que en el diseño del sistema se desarrollaron los componentes básicos de la interfaz que son necesarios para integrar ambos programas de software a la versión operativa de SALIDEA. Sin embargo, aún son necesarios otros componentes; en particular los que permiten conectar la sección de la interfaz donde los examinados responden la prueba, con la base de datos relacional donde se registran sus respuestas al examen; y entre esta base de datos y los elementos que requieren las aplicaciones PAGI y PARRERC para ponerse en operación para realizar los análisis y emitir los reportes de información. Dichos componentes requieren el diseño de la base de datos relacional, la creación de archivos especiales que presenten los datos de la prueba con los formatos particulares que requieren ambos programas para funcionar, entre otros más que le corresponde desarrollar al proveedor de los servicios de cómputo que desarrolle programáticamente la versión operativa de SALIDEA.

En el esquema 4.2.1 se muestra la estructura del sistema y en ella la integración de PAGI y PARRERC. Se muestra cómo al momento de terminar la prueba los

examinados, sus respuestas se envían a una base de datos que incorpora los datos generales sobre los examinados que respondieron la prueba y datos sobre sus respuestas a los ítems de logro, los cuales a su vez se dividen en datos para análisis psicométricos, datos para analizar la ejecución por dominio y subdominio evaluados y en datos para analizar variables de contexto, los cuales son enviados a PAGI y PARRERC y de esta manera obtener los reportes de resultados para los diferentes usuarios.

- **Prefigurar la funcionalidad del sistema con un ejemplo que represente los componentes de una prueba y las relaciones entre ellos, que se requieren para su administración mediante un sistema de cómputo.**

Por último se necesitó un diagrama en el que se plasmara la integración de los estándares psicométricos y los lineamientos del diseño universal, además de las funcionalidades del sistema y la integración de los sistemas PAGI y PARRERC (ver página 81).

Se hizo un diagrama, el cual mostraba el interior del software y no solamente el diseño. Por ejemplo, cuando algún examinado trata darse de alta, le envía una solicitud al administrador, en este caso, el diagrama, indica a quién se le pidió la información y una vez que le llega dicha información, el administrador tomará la decisión de confirmar la solicitud o rechazarla.

Sin embargo, la parte más importante del esquema, es la manera en la que opera la integración de los estudiantes, por lo que, para empezar, debe de existir una prueba

que el administrador debe dar de alta y configurarla, inmediatamente los examinados contestarán la prueba y sus resultados se almacenarán en una base de datos, la cual se separará en Datos para análisis psicométricos, Datos para analizar la ejecución por dominio y subdominio evaluados y Datos para analizar variables de contexto (VC), las cuales se mandarán a las aplicaciones PAGI y PARRERC respectivamente, y como resultado se obtendrán, análisis gráficos de ítems y reportes de resultados (dependiendo el usuario), por lo que el diagrama servirá como una guía para el programador del sistema.

5.1 Limitaciones y comentarios finales

Sin duda, la falta de infraestructura computacional en las instituciones educativas del país constituye una limitante general para que pueda utilizarse un sistema como SALIDEA para administrar en línea pruebas para evaluar el logro educativo. Por otra parte, aunque una escuela o grupo de escuelas lograra instalar los componentes necesarios para operar el sistema de evaluación, aun así se tendrían dificultades relacionadas con la falta de una cultura evaluativa o computacional por parte de los examinados y sus escuelas, lo que conlleva riesgos para la integridad y validez del instrumento, como serían la falta de familiaridad con equipos de cómputo o la existencia de muchas prácticas inconvenientes como las que reveló la experiencia de la aplicación censal de la prueba ENLACE en las escuelas.

Aunque SALIDEA puede aportar muy poco para resolver ese tipo de dificultades, se considera que su diseño constituye primera aproximación relevante que puede

contribuir al desarrollo de sistemas de cómputo que hagan posible administrar en línea pruebas de gran escala. En particular, hay varias características de su diseño que hacen que el sistema sea eficiente, ecológico e incluyente:

- La administración de pruebas en línea mediante un sistema como SALIDEA hace posible ahorrar costos diversos como los de impresión de cuadernillos y compra de hojas de respuestas, servicio de lector óptico de hojas de respuestas, personal para la aplicación, captura, calificación, análisis y reporte de resultados, entre otros más, lo que a su vez implica reducir errores y agilizar procesos como la realimentación inmediata a estudiantes, padres, maestros, y administradores escolares, a fin de que ellos puedan actuar de manera oportuna para la toma de decisiones. Además, como se mencionó en el capítulo II, aunque existen diversos tipos software que brindan el servicio de administración de pruebas en línea, tienen un costo elevado y muchos de ellos están en otros idiomas. En contraste el diseño de SALIDEA considera que sea un sistema gratuito y de código abierto para que pueda mejorarse.
- Los exámenes que se administran en línea requieren componentes y controles que propicien la seguridad para el propio sistema de evaluación, para el examen que administra, y para los usuarios. Al respecto, el diseño de SALIDEA consideró componentes para proteger la integridad de los materiales de la prueba, aleatorizar la presentación de los ítems de la prueba y aleatorizar la ubicación de respuesta correcta en cada ítem, así como la adopción del formato de presentación de un ítem a la vez, de manera que resulte muy difícil que dos

alumnos que estén juntos tengan en pantalla la misma prueba. Además, se diseñaron componentes para darse de alta en el sistema de manera válida; para el ingreso protegido al sistema mediante una clave de acceso única; para preservar la confidencialidad de los examinados y la validez de sus resultados en la prueba, y para controlar el tiempo efectivo de examen desde que el estudiante hace clic a su ingreso; así como en general para propiciar que todos los componentes del sistema proporcionen a todos los examinados, las mismas oportunidades de demostrar su nivel real con respecto a las competencias cuyo dominio se evalúa en la prueba.

Para una segunda versión del diseño del sistema, se considera conveniente agregar los siguientes elementos: a) nuevos usuarios, como los propios desarrolladores de la prueba que pueden incorporar directamente la prueba y la documentación técnica que la acompaña, recibir realimentación psicométrica y de otros tipos que sea de utilidad para el rediseño del instrumento. b) También puede facilitarse los trabajos de jueceo de comités de examen como los que convalidan en línea las especificaciones de los ítems o quienes califican en línea las respuestas a ítems de ensayo; c) Incorporar un componente que permita reducir la presencia de personas que monitorean la aplicación del examen en localidades remotas, pero preservando la seguridad mediante un monitoreo con cámaras web y que hagan posible que el administrador pueda resolver las dudas inmediatamente. d) Se considera deseable atender más lineamientos del diseño universal, como incorporar audio a los ítems a fin de que estudiantes con limitaciones puedan leer y escuchar el ítem las veces que sea necesario; e) Cuando el sistema diseñado esté funcional, resulta necesario efectuar un

pilotaje previo a su operación, el cual puede incluir una encuesta de salida a los examinados y los demás usuarios, para asegurar que los elementos diseñados con base en la normatividad mencionada atienden sus necesidades de información y resultan funcionales y efectivos; f) con el desarrollo y la inversión necesarias se podría considerar un proyecto mucho más ambicioso, el cual pueda proveer de equipo de cómputo móvil dedicado y dispositivos como clikers, para administrar la prueba en las localidades donde se tenga instalado SALIDEA.

Referencias bibliográficas.

- Assessment Systems Corporation (2012). FastTest. Disponible en <http://www.fasttestweb.com/>
- Axelson, M. (2005). *Maximizing the Effectiveness of Online Accountability Assessments for Students with Disabilities*. Appalachia Educational Laboratory at Edvantia. Policy Briefs. Washington: Institute of Education Sciences.
- 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). Consultado el 10 de febrero de 2013 en <http://redie.uabc.mx/vol2no1/contenido-backhoff.html>
- Backhoff E., Rosas M., Cuevas R., Velasco A. (2002) Evaluación del aprendizaje por computadora: una década de innovación educativa en la UABC. 3er Congreso Nacional y 2º Internacional: Retos y Expectativas de la Universidad. Consultado el 02 de abril de 2013 en http://www.exhcoba.mx/pdf/2002_Evaluacion_de_una_decada_de_aprendizaje_por_computadora.pdf
- Barnum, C. (2002). Usability testing and research. What Is usability and What is Usability Testing.
- Benjamin, L. (1988). A History of Teaching Machines. *American Psychologist*(*American Psychological Association*).Vol. 43, No. 9, 703-712.

-
- Berkshire Publishing. (2010). Berkshire Encyclopedia of China. Modern and Historic Views of the World's Newest and Oldest Global Power. Great Barrington, MA: Berkshire, pp. 405–410.
- Burger, J. M., & Krueger, M. (2003). A balanced approach to high-stakes achievement testing: An analysis of the literature with policy implications. *International Electronic Journal for Leadership in Learning*, 7(4). Available [online] <http://www.ucalgary.ca/~iejll>.
- Burgstahler, S. (2007) Universal Design of Instruction (UDI): Definition, Principles, Guidelines, and Examples. University of Washintong
- Castañeda, C. E., (2013). *Desarrollo de un software para el análisis de datos y la emisión de informes de resultados de exámenes de referencia criterial alineados con el currículum*. Tesis de Maestría inédita. Universidad Autónoma de Baja California. Ensenada. Disponible en <http://iide.ens.uabc.mx/images/pdf/tesis/MCE/Tesis%20MCE%20Cristian%20Ernesto%20Casta%C3%B1eda%20S%C3%A1nchez.pdf>
- Center for Universal Design. (1997). What is universal design? North Carolina State University. Recuperado de: <http://www.design.ncsu.edu>.
- Chiarani M, Ponce V, Zangla M & Allendes P., (2008). *Módulos de evaluación en las plataformas de código abierto. Calidad y Materiales educativos y Herramientas Tecnológicas en Educación a Distancia*. San Luis, Argentina. Universidad Nacional de San Luis.

Consejo Asesor Externo del Centro Nacional de Evaluación para la Educación Superior (2000). *Estándares de Calidad para Instrumentos de Evaluación Educativa*. pp. 1-51. México: CENEVAL.

Contreras, L. A. (2008). Glosario de evaluación del aprendizaje. Documento inédito.

Contreras, L. A. (2009). *Desarrollo y validación de un modelo para elaborar exámenes criterios de gran escala alineados con el currículum*. (Tesis inédita de doctorado). Universidad Autónoma del Estado de Morelos. Cuernavaca.

Contreras, L. A. (2000). Desarrollo y Pilotaje de un Examen de Español para la Educación Primaria en Baja California. Tesis para optar por el grado de Maestro en Ciencias Educativas. México. Baja California. UABC. En <http://iide.ens.uabc.mx/images/pdf/tesis/MCE/Tesis%20MCE%20Luis-Angel-Contreras-Nino.pdf>

De Los Santos, G. (2010). Desarrollo, operación y evaluación de un módulo para capacitar a docentes en servicio para que mejoren sus exámenes de opción múltiple mediante el análisis gráfico de ítems. Tesis para optar por el grado de Maestro en Ciencias Educativas. México. Baja California. UABC. En <http://iide.ens.uabc.mx/images/pdf/tesis/MCE/Tesis%20MCE%20Guadalupe%20de%20los%20Santos%20Lazaro.pdf>

De Pascale, C. (2003). The Ideal Role of Large-Scale Testing in a Comprehensive Assessment System. Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, Illinois.

eAssessment Association (s/f). *e-Testing*. Disponible en

<http://www.e-assessment.com/resources/e-testing/>

Fuentes, M. A. (2007) *Sistema Autoejecutable de Exámenes Interactivos por Computadora*. Universidad Autónoma del Estado de Hidalgo.

International test commission (2014). The Security of Tests, Examinations, and Other Assessments. Document reference: ITC-G-TS-20140706

Instituto de Evaluación Educativa del Estado de México [IEEEM]. Consultado el 10 de marzo de 2013 en:

<http://www.edomexico.gob.mx/evaluacioneducativa/preguntas.html>

Instituto de Innovación y Evaluación Educativa del Estado de Sonora [IIEEES].

Consultado el 23 de marzo de 2013 en: <http://www.ieees.gob.mx/>

Instituto Estatal de Evaluación e Innovación Educativa [INEVAL] Disponible en:

<http://www.ineval.chiapas.gob.mx/>

Instituto Nacional para la Evaluación de la Educación [INEE]. Explorador Excale.

<<http://www.inee.edu.mx/explorador>>, [Consultado el 10 de marzo de 2013].

Instituto Nacional para la Evaluación de la Educación [INEE]. La construcción del INEE

autónomo. En: <<http://www.inee.edu.mx/index.php/el-inee-en-la-reforma-educativa/520-construccion-del-inee-autonomo/1613-construccion-del-inee-autonomo>>, [Consultado el 16 de agosto de 2013]

-
- Jisc Digital Media. (2014). Graphical User Interface Design: Developing Usable and Accessible Collections. Autor. Disponible en <http://www.jiscdigitalmedia.ac.uk/guide/graphical-user-interface-design-developing-usable-and-accessible-collection>
- Johnstone, C. J., Altman, J., & Thurlow, M. (2006). *A state guide to the development of universally designed assessments*. National Center on Educational Outcomes. Minneapolis: University of Minnesota.
- Johnstone, C., Altman, J., Thurlow, M., & Moore, M. (2006). Universal design online manual. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Joint Committee on Testing Practices (1999). American Educational Research Association (AERA), American Psychological Association (APA) y National Council on Measurement in Education (NCME). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- Joint Committee on Testing Practices (2014). American Educational Research Association (AERA), American Psychological Association (APA) y National Council on Measurement in Education (NCME). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, D.C.: American Psychological Association.

Hamilton, L. & Klein, S. (1999). *Large-Scale Testing. Current practices and new directions*. Santa Monica, Ca.: Rand Education.

Martin, S. (s/f). Effective Visual Communication for Graphical User Interfaces. Disponible en http://web.cs.wpi.edu/~matt/courses/cs563/talks/smartin/int_design.html

Martínez, F. (2009). Evaluación formativa en aula y evaluación a gran escala: hacia un sistema más equilibrado. *Revista Electrónica de Investigación Educativa*, 11 (2). Consultado el día 01 de Agosto de 2012 en: <http://redie.uabc.mx/>

Matthews, D. (1999). The Origins of Distance Education. *T.H.E. Journal*. Vol. 27, No. 2. Sept. 99. 56-66.

Oklahoma State Department of education (2013). Test and item specification. Oklahoma school testing program. Recuperado 02 de septiembre de 2014. http://ok.gov/sde/sites/ok.gov.sde/files/documents/files/OCCT_G3_Rdg_ItemSpecs_2013-14_FINAL_0.pdf

Ridgway, J., McCusker, S. & Pead, D. (2006). Literature Review of E-assessment. <http://hal.archives-ouvertes.fr/docs/00/19/04/40/PDF/ridgway-j-2004-r10.pdf>

Rodríguez, M. (2006). Aplicación de las TICs en evaluación de alumnos universitarios. *Revista de la Universidad de Salamanca*. En http://campus.usal.es/~teoriaeducacion/rev_numero_06_2/n6_02_art_rodriguez_conde.htm

Secretaría de Educación Pública [SEP]. (2011). *Acuerdo número 592 por el que se establece la articulación de la Educación Básica*. Recuperado el 22 de Enero de 2013. En

http://basica.sep.gob.mx/reformasecundaria/doc/sustento/Acuerdo_592_completo.pdf

Secretaría de Educación Pública [SEP]. (2012). *Manual de Organización de la Dirección General de Evaluación de Políticas*. Recuperado el 22 de Enero de 2013 en:

http://normatecainterna.sep.gob.mx/work/models/normateca/Resource/272/2/images/mo_dgep.pdf

Secretaría de Educación Pública [SEP]. (2013). *Evaluación Nacional del Logro Académico en Centros Escolares. Educación Básica. Manual Técnico ENLACE 2013*.

Shepard, L.A. (2003). *Reconsidering Large-Scale Assessment to Heighten Everyday Assessment in the Science Classroom*. Arlington, VA: National Science Teachers Association.

Szenaris GmbH (2011). *Letter I/2011*. Autor. Disponible en: <http://szenaris.com/271-1-szenaris-Letter-I2011.html#top4>

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments (Synthesis Report 44)*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Recuperado el 13 diciembre de 2013, de: <http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html>

Thompson, S. J., Thurlow, M. L., Quenemoen, R. F., & Lehr, C. A. (2002). Access to computer-based testing for students with disabilities (Synthesis Report 45). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Recuperado el 12 diciembre de 2013, de <http://education.umn.edu/NCEO/OnlinePubs/Synthesis45.html>

Thompson, S., & Thurlow, M. (2002). Universally designed assessments: Better tests for everyone! (Policy Directions No. 14). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Recuperado 21 de diciembre de 2013 de <http://education.umn.edu/NCEO/OnlinePubs/Policy14.htm>

Unidad de Evaluación Educativa (2010). Página Web de la UEE. En <http://uee.uabc.mx/uee/>

Volante, L. (2006). An Alternative Vision for Large-scale Assessment in Canada. Journal of Teaching and Learning, 6, Vol. 4, No. 1.

Zandbergen, P. (2014) Education Portal.

<http://education-portal.com/academy/lesson/developing-a-graphical-user-interface-gui.html#lesson>

Apéndices

Apéndice 1



SALIDEA

Sistema para administrar en línea instrumentos de evaluación del aprendizaje

Lineamientos para el desarrollador de la prueba

SALIDEA es un sistema que permite la administración en línea de exámenes de referencia criterial. Sin embargo, para poder cumplir con su función se requiere contar con la correspondiente documentación que avala la calidad técnica del instrumento, de manera que sea posible llevar a cabo su aplicación de manera estandarizada y, posteriormente al concluir la evaluación, analizar y reportar los resultados de la evaluación de manera inmediata, simple y significativa, tanto los resultados generales y como específicos, para los diferentes usuarios de la información. Para realizar lo anterior, SALIDEA fue diseñado de manera que la administración de una prueba se ajuste a estándares psicométricos internacionales propuestos por el *Joint Committee on Testing Practices* (AERA, APA y NCM, 1999, 2014), por diferentes universidades norteamericanas y australianas, así como por los estándares nacionales publicados por el Consejo Asesor Externo del Centro Nacional de Evaluación para la Educación Superior (2000). Además, el diseño de SALIDEA considera el cumplimiento de los lineamientos del diseño universal de las evaluaciones que han sido propuestos por diferentes instancias (Center for Universal Design, 1997; Thompson y Thurlow, 2002;

Thompson, Johnstone & Thurlow (2002), Johnstone, Altman y Thurlow, 2006), de manera que se atiendan necesidades de los diferentes usuarios de las pruebas, incluyendo a personas con discapacidades físicas o psicológicas.

Por lo anterior, para que SALIDEA esté en condiciones de administrar una prueba es necesario que el desarrollador de la prueba tenga en cuenta los siguientes aspectos de la documentación que debe ser entregada y que son críticos para que pueda ser integrada el sistema.

- Proporcionar evidencia de lo que mide el test; los usos recomendados de sus resultados; los examinados a quienes se pretende evaluar; las fortalezas y limitaciones de la prueba, incluido el nivel de precisión de las puntuaciones del examen.
- Describir cómo se seleccionaron los contenidos y las habilidades que serán evaluados y cómo se desarrolló el test.
- Comunicar la información sobre características de la prueba con un nivel de detalle apropiado y con claridad para los usuarios previstos del examen.
- Proporcionar evidencia sobre la calidad técnica del test y sus ítems, incluyendo la confiabilidad y la validez, y si satisface los propósitos pretendidos.
- Evidencia de que se cuidó que el contenido y el lenguaje empleados en los materiales de la prueba no fueran potencialmente ofensivos para los examinados.

El desarrollador de la prueba debe de tomar en cuenta también los siguientes aspectos relacionado con la administración de la prueba y con el análisis y reporte de resultados:

- Proporcionar descripciones claras de procedimientos detallados para administrar pruebas de manera estandarizada.

-
- Proporcionar pautas sobre procedimientos razonables para la evaluación de las personas con discapacidades que necesitan adaptaciones especiales o los que tienen diversos orígenes lingüísticos.
 - Proporcionar información a los examinados o a otros usuarios de la prueba sobre los formatos y procedimientos para responder a los ítems, incluida la información sobre el uso de cualquier material y el equipo necesario.
 - Establecer e implementar procedimientos para garantizar la seguridad de los materiales de prueba durante todas las fases de su desarrollo, administración, calificación, y elaboración de informes.
 - Proporcionar procedimientos, materiales y directrices para calificar la prueba, y para monitorear la exactitud del proceso de puntuación. Si la calificación de la prueba es responsabilidad del desarrollador, como es el caso de emplear rúbricas, se debe considerar un entrenamiento adecuada a los calificadores.
 - Corregir los errores que afectan la interpretación de las puntuaciones y comunicar los resultados corregidos con prontitud.
 - Desarrollar e implementar procedimientos para garantizar la confidencialidad de las calificaciones.

A continuación se mencionan los derechos y responsabilidades que establecen los estándares mencionados para los desarrolladores de pruebas cuando se trata de las funciones de reportar e interpretar los resultados.

- Proporcionar información para apoyar interpretaciones pretendidas de los resultados, referidas al contenido, grupos de comparación, y otras evidencias técnicas. Advertir a los usuarios de la prueba de los beneficios y limitaciones de sus resultados e interpretación.
- Proporcionar orientación para interpretar los resultados de pruebas administradas con modificaciones. Informar a los usuarios de problemas al interpretar los resultados, cuando se ha modificado el test o los procedimientos para su administración.

-
- Especificar los usos apropiados de los resultados de las pruebas y advertir a los usuarios sobre malos usos potenciales.
 - Proporcionar oportunamente los resultados de la prueba y de manera que sean comprendidos por el examinado.

Apéndice 2



SALIDEA

Sistema para administrar en línea instrumentos de evaluación del aprendizaje

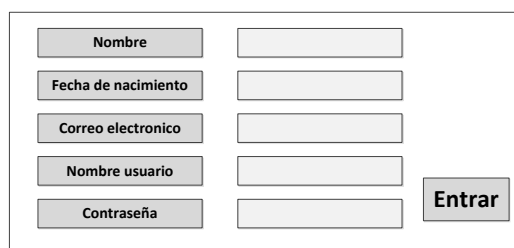
Manual para el Administrador de SALIDEA

SALIDEA es un sistema que permite la administración de exámenes de referencia criterial y la correspondiente documentación que avala su calidad técnica, de manera que sea posible llevar a cabo su aplicación de manera estandarizada y, posteriormente al concluir la evaluación, analizar y reportar de manera inmediata, simple y significativa los resultados generales y específicos, para los diferentes usuarios de la información. Todo lo anterior apegado a estándares psicométricos internacionales propuestos por el *Joint Committee on Testing Practices* (AERA, APA y NCM, 1999, 2014, así como por los estándares nacionales publicados por el Consejo Asesor Externo del Centro Nacional de Evaluación para la Educación Superior (2000). Además, el diseño de SALIDEA considera los lineamientos del diseño universal de las evaluaciones que han sido propuestos por diferentes instancias (Center for Universal Design, 1997; Thompson y Thurlow, 2002; Thompson, Johnstone & Thurlow (2002), Johnstone, Altman y Thurlow, 2006), de manera que se atiendan necesidades de los diferentes usuarios de las pruebas, incluyendo a personas con discapacidades físicas o psicológicas.

El administrador debe darse de alta en SALIDEA, tendrá un usuario y contraseña predeterminados, como se muestra en la siguiente imagen.



Una vez que iniciaron sesión es necesario que se actualicen los datos del administrador.



Al entrar al sistema aparecerá la pantalla principal del administrador, la cual incluye las opciones, “Agregar datos de la prueba prueba”, “Agregar manual técnico” y “Descargar formato de la prueba”

Página de inicio

Agregar datos de la prueba

Agregar manual técnico

Descargar formato prueba



En la primera opción, “Agregar datos de la prueba” el administrador incluye lo siguiente:

- Una sección con los propósitos de la prueba, una descripción de lo que pretende medir y la importancia para el aprendizaje del estudiante, e indica cómo se seleccionaron los contenidos, cómo se va evaluar y ejemplos de reactivos.
- Sección donde se muestran las propiedades psicométricas de los ítems y versiones de la prueba
- Sección del manual de la prueba que muestra su estructura, las especificaciones de ítems, así como el proceso para su elaboración y validación.
- Proporciona información sobre los formatos y procedimientos de para responder a preguntas de la prueba, incluida la información sobre el uso de cualquier material y el equipo necesario.

SALiDEA
Sistema para administrar en línea instrumentos de evaluación del aprendizaje

Página de inicio

Propósitos de la prueba

¿Qué pretende medir?

¿ Como se seleccionaron los contenidos?

Usos de la prueba

Resultados

Aviso de privacidad

Instrucciones

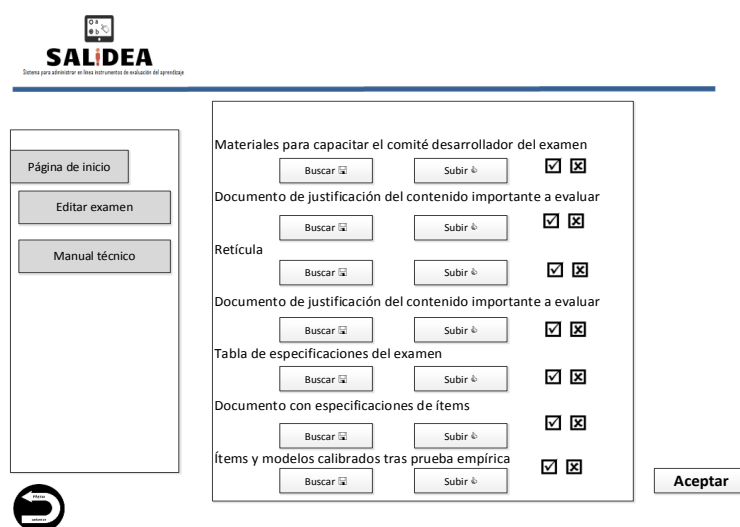
Ejemplo

Importancia de la prueba



INICIAR PRUEBA

En la segunda opción, “Agregar manual técnico”, el administrador debe de contener las evidencias de validez de la prueba que se muestran a continuación.



SALIDEA
Sistema para administrar la prueba y el contenido de la evaluación del aprendizaje

Página de inicio
Editar examen
Manual técnico

Materiales para capacitar el comité desarrollador del examen
Buscar Subir ☒ ☒

Documento de justificación del contenido importante a evaluar
Buscar Subir ☒ ☒

Reticula
Buscar Subir ☒ ☒

Documento de justificación del contenido importante a evaluar
Buscar Subir ☒ ☒

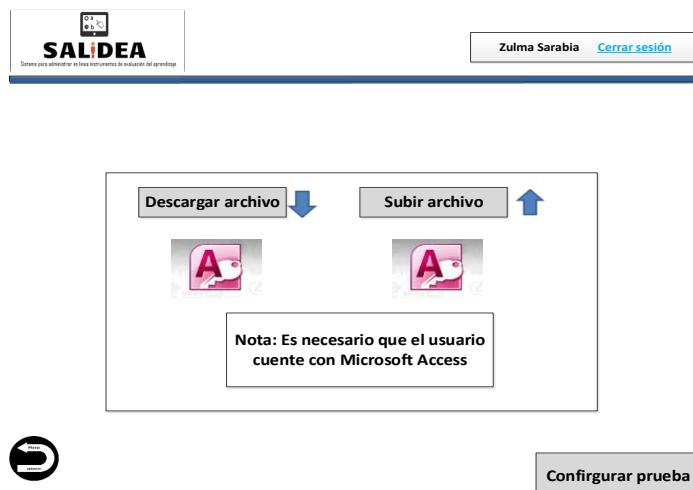
Tabla de especificaciones del examen
Buscar Subir ☒ ☒

Documento con especificaciones de ítems
Buscar Subir ☒ ☒

Ítems y modelos calibrados tras prueba empírica
Buscar Subir ☒ ☒

Aceptar

La tercera opción “Elaborar y subir prueba”, el administrador debe descargar el formato en Microsoft Access, y llenarlo con la prueba que se desea administrar, una vez que se tenga el formato listo, se debe subir a SALIDEA.



SALIDEA
Sistema para administrar la prueba y el contenido de la evaluación del aprendizaje

Zulma Sarabia Cerrar sesión

Descargar archivo Subir archivo

Nota: Es necesario que el usuario cuente con Microsoft Access

Configurar prueba

Una vez que se subió la prueba al Sistema, se podrá realizar la configuración de la prueba.

**SALIDEA**
Sistema para la Administración de la Evaluación de la Educación

Zulma Sarabia [Cerrar sesión](#)

Configurar prueba

☐ Aleatorizar ítems

☐ Aleatorizar clave y distractores

☐ No omitir ítems

☐ Responder todos los ítems de la página antes de ir a la página siguiente.

☐ Los estudiantes pueden seleccionar la opción "marcar para revisión" para aquellos ítems que presenten dificultad alta.

☐ Revisar los ítems anteriores y cambiar respuesta



En caso de que algún usuario olvide la contraseña, el administrador deberá mandar un correo con la nueva contraseña. Es necesario que el Administrador dedique un tiempo considerable a la administración de la prueba en el periodo de aplicación y entrega de resultados.

Apéndice 3

Instructivo para el alumno

A continuación se presenta el examen de español, que tiene 44 preguntas de opción múltiple y una pregunta de ensayo; dicho examen será aplicado a los niños que se encuentran en el 6to grado de primaria en el Estado de Baja California.

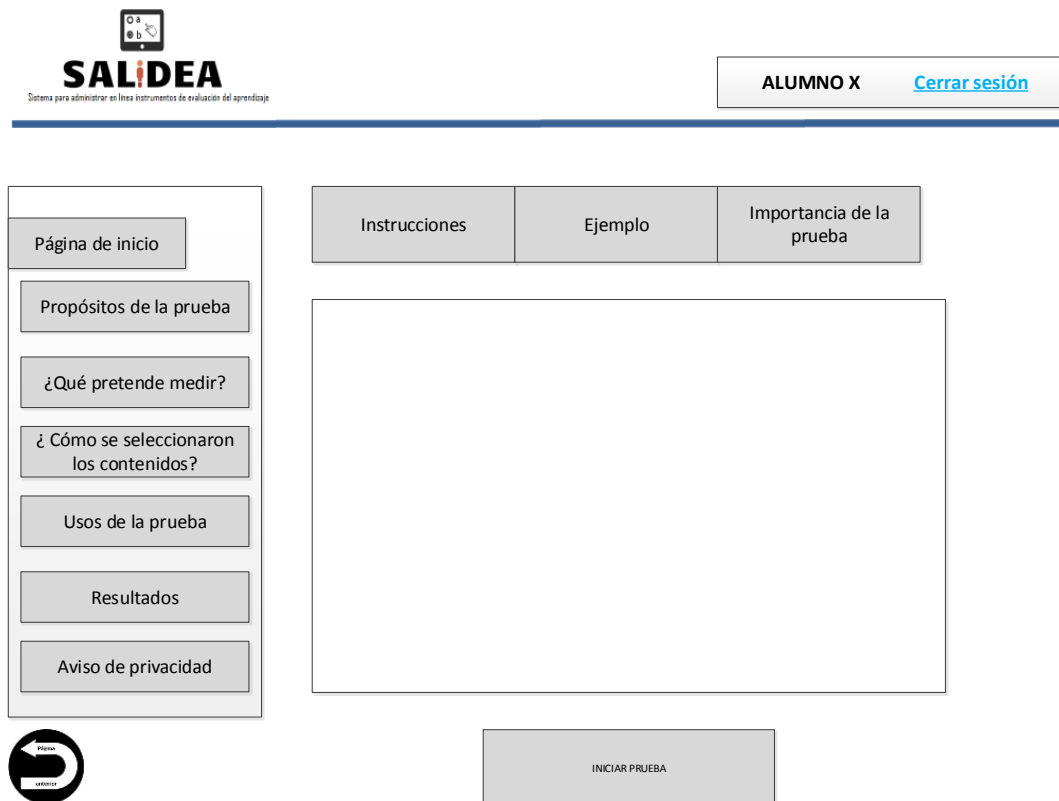
El examen evalúa 4 áreas de la materia de español:

- Lengua hablada
- Lengua escrita
- Recreación literaria
- Reflexión sobre la lengua

Por favor, no contestes lo primero que te venga a la cabeza. Reflexiona con cuidado las respuestas que elegirás.

Ten en mente que tus respuestas, junto con las de los demás compañeros que también responderán la prueba, serán utilizadas por las autoridades educativas de Baja California puedan hacer mejoras a los contenidos del curso de español, para que en el futuro niños como tu puedan aprender más.

Una vez que hayas entrado al sistema, da clic en la opción de Examen español y te aparecerá la siguiente pantalla.



Verás que en esta pantalla aparecen varias opciones para conocer detalles sobre la prueba que vas a responder y que harán que tu participación en ella resulte más significativa y te motives a realizar su mejor esfuerzo, como los propósitos que tiene la prueba, qué es lo que se pretende medir con ella, cómo fueron seleccionados los contenidos que se evalúan, el uso que se dará a los resultados de la prueba, la importancia que tienen tus respuestas en la prueba y un ejemplo del tipo de preguntas que aparecerán en tu examen. Así, es importante que revises cada una de las opciones, para que conozcas más acerca del examen que a continuación contestarás. Una vez que hayas leído, da clic en el botón iniciar prueba. Enseguida de aparecerá la siguiente pantalla.

Items
terminados
1

Items
faltantes
43

Items
marcados
para revisión
0

Instrucciones. Enseguida se presenta la parte inicial de varios textos que corresponden a ciertos géneros literarios populares. En cada caso, selecciona la opción que indica de qué género se trata:

2. "Dicen que existió un pueblo nómada. . ." Se trata de:

- Una leyenda
- Una adivinanza
- Una copla
- Un dicho



Al iniciar el examen te aparecerá la primera pregunta del examen y las instrucciones para responder la prueba. Al lado izquierdo se muestra una barra la cual te marcará los ítems o preguntas que has terminado, los que te falta responder y el tiempo restante que tienes para completar la prueba. También aparecerán los ítems que dejaste marcado para su revisión posterior; al dar clic en esa opción automáticamente aparecerá en la barra de la izquierda el número de ítems dejados para revisión. En la parte inferior aparecen los botones de aumentar o disminuir el tamaño de la pantalla en caso de que lo requieras. Si deseas regresar o pasar a otra

pregunta da clic en el botón Regresar o Siguiente que están en los extremos abajo de la pantalla.

Cuando llegues al último ítem aparecerá la opción de “Terminar prueba” y al darle clic en ella aparecerá un recuadro en el cual deberás confirmar si quieres terminar la prueba o regresar a ella.

Items
terminados
44

Items
faltantes
0

Items
marcados
para revisión
0

Instrucciones. Lee con atención el cuento que aparece a continuación.

El Hada Madrina

Esta era una vez una hada madrina que caminaba por el bosque. Cuando iba caminando, a una niña se le cayó un diente y el hada madrina fue a la casa de la niña y con su varita mágica apareció dinero y se lo dejó a la niña en su almohada (cuento de Angélica de Luna Torres, de 8 años. Jesús María, Aguascalientes).

Ahora, selecciona la opción que presenta la mejor secuencia de ilustraciones para crear una historieta a partir del cuento.

- 1, 3, 4, 2.
- 4, 3, 1, 2.
- 4, 1, 3, 2.
- 1, 4, 3, 2.



Terminar prueba

¿ Seguro qué desea terminar la prueba?

SI

NO



Una vez terminada la prueba, tendrás la opción de ver el reporte de resultados de la prueba que contestaste o salir del sistema.

Ver resultados

Cerrar sesión



Con eso se finaliza la prueba. ¡Muchas gracias por tu colaboración!

Apéndice 4

Dimensiones y características de máxima legibilidad en el párrafo

A continuación se comenta sobre el aspecto que deben tener los textos y los ítems que sean administrados por SALIDEA, cuando se atienden los principios y lineamientos del diseño universal de las evaluaciones. Así, al seguir las recomendaciones propuestas por Thompson & Thurlow (2002) sobre las dimensiones y características de máxima legibilidad en el párrafo, el diseño de SALIDEA consideró las que se ilustran en seguida:

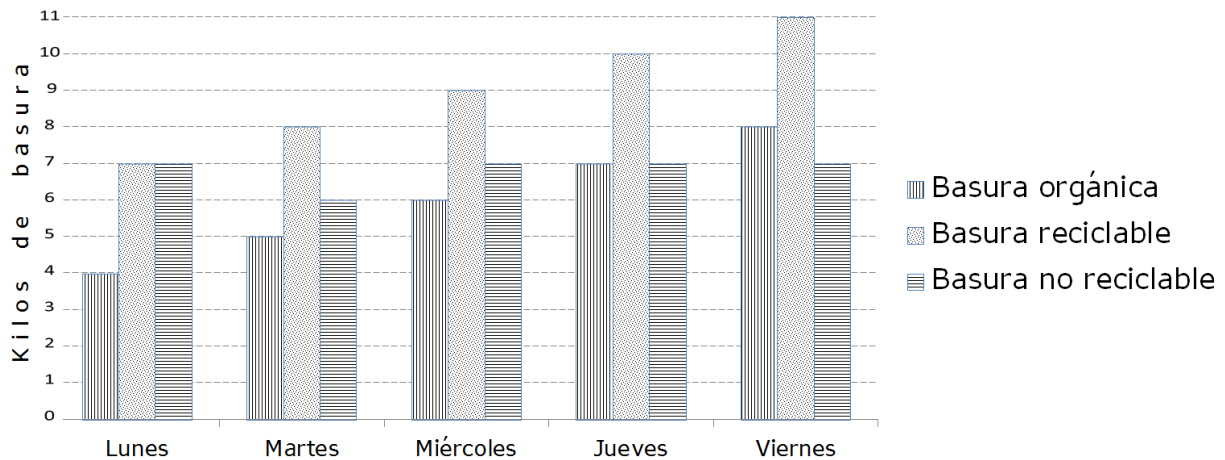
El tipo de letra estándar, con mayúsculas y minúsculas, es más fácil de leer que la letra cursiva, inclinada, con minúsculas, o todo en mayúsculas.

El texto sin justificar (con margen derecho) es más fácil de ver y analizar que el texto justificado, especialmente para lectores con dificultades para leer.

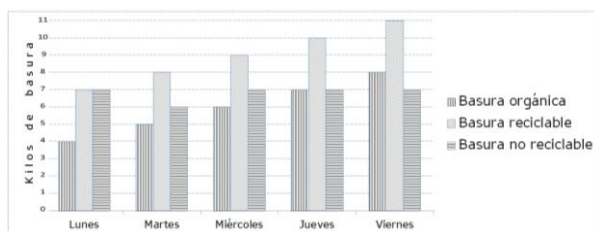
La longitud óptima es de alrededor de 4 pulgadas o de 8 a 10 palabras por línea. Esta longitud evita la fatiga de lector y la dificultad para localizar el inicio del siguiente renglón, lo que hace que el lector se desubique.

Los símbolos utilizados en los gráficos tienen que ser altamente discriminables. Las

etiquetas deben colocarse directamente junto a las barras o líneas de una gráfica, para que la información se pueda encontrar de forma rápida y no se requiera el uso de la memoria a corto plazo.



Cuando se usa una ilustración, debe estar directamente al lado de la pregunta que la requiera. Debido a que las ilustraciones crean numerosos problemas visuales y de distracción, y pueden interferir con el uso de algunas acomodaciones como lupas, deben usarse sólo cuando contienen información cuyo dominio se evalúa.



27. Si comparamos la cantidad de basura reciclable, con la cantidad de basura no reciclable que recolectaron el lunes Ana y Luis, ¿Cuál es el resultado?

- ☐ La misma cantidad de basura.
- ☐ Menos basura no reciclable.
- ☐ Más basura reciclable.

Las opciones de respuesta deben incluir círculos más grandes (para las pruebas que se responden marcando una burbuja).

Apéndice 5

Retículas de los usuarios de SALIDEA, incluidas en el CD adjunto al presente documento.