



**Universidad Autónoma de Querétaro**

**Facultad de Informática**

Evaluación estandarizada de los aprendizajes en la UABC: innovación  
desde el análisis psicométrico

**Tesis**

Que como parte de los requisitos  
para obtener el Grado de

**Doctor en Innovación en Tecnología Educativa**

Presenta

**Jorge Gustavo Gutiérrez Benítez**

Dirigido por:

Dr. Luis Alan Acuña Gamba

Querétaro, Qro. agosto de 2021



**Universidad Autónoma de Querétaro**  
**Facultad de Informática**  
**Doctorado en Innovación en Tecnología Educativa**

Evaluación estandarizada de los aprendizajes en la UABC: innovación desde el  
análisis psicométrico

**Tesis**

Que como parte de los requisitos para obtener el Grado  
Doctor en Innovación en Tecnología Educativa

Presenta

Jorge Gustavo Gutiérrez Benítez

Dirigido por:

Dr. Luis Alan Acuña Gamboa

Dr. Luis Alan Acuña Gamboa  
Presidente

Dra. Teresa Guzmán Flores  
Secretario

Dra. Rocío Adela Andrade Cázares  
Vocal

Dra. Rocío Edith López Martínez  
Suplente

Dr. Hugo Moreno Reyes  
Suplente

Centro Universitario, Querétaro, Qro.  
agosto, 2021  
México

## RESUMEN

Para cualquier institución educativa es importante obtener información del rendimiento de sus estudiantes, estimar los aprendizajes comparándolas con las metas establecidas. La evaluación ha tenido un fuerte impacto en la enseñanza, por ello resultan indispensables instrumentos de evaluación del aprendizaje que sean válidos y confiables. La Facultad de Idiomas de la UABC no implementa una herramienta tecnológica para el desarrollo de exámenes estandarizados y su análisis psicométrico; por ello, este trabajo analizó la implementación de exámenes estandarizados referidos a un criterio, basados en la Teoría de Respuesta al Ítem. La metodología empleada requirió la integración de varios comités de especialistas para producir instrumentos que guiaron la construcción del examen, tales como la tabla de índices de relevancia curricular, la justificación de contenidos, etcétera. Se midió la calidad del examen utilizando un desarrollo innovador que cuantificó automáticamente los criterios psicométricos de validez y confiabilidad como el índice de dificultad de los ítems y el índice de discriminación. Los resultados permiten observar la calidad del examen identificando el rendimiento de cada ítem, así como el rendimiento y nivel de dominio de la muestra de alumnos participantes. Se concluye que el instrumento mostró utilidad tanto para la evaluación formativa como para la sumativa.

**(Palabras clave:** evaluación estandarizada, psicometría, metodología, calidad educativa, examen)

## SUMMARY

For any educational institution it is important to obtain information about the performance of their students, estimate the learning compared to the established goals. The evaluation has had a strong impact on teaching, which is why learning instruments that are valid and reliable are indispensable. The Faculty of Languages of the UABC does not implement a technological tool for the development of standardized tests and their psychometric analysis; Therefore, this paper analyzed the implementation of standardized tests referred to a criterion, based on the Item Response Theory. The used methodology required the integration of several committees of specialists to produce instruments that guided the construction of the test, such as the curricular relevance index table, the contents justification, etc. The test quality was evaluated using an innovating development that automatically quantified the psychometric criteria of validity and trustworthiness like the difficulty index of the items and the discrimination index. The results allow to observe the quality of the test identifying the yield of each item, as well as the yield and level of dominion of the participant students sample. It is concluded that the instrument showed utility for both formative and summative evaluation.

**(Key Words:** standardized evaluation, psychometric, methodology, quality, test.)

Dirección General de Bibliotecas de la UAQ

## AGRADECIMIENTOS

Quiero agradecer primeramente a Dios por darme la vida, la salud y la capacidad intelectual para poder desarrollar este trabajo, por darme lo necesario para poder avanzar durante estos 3 años hasta lograr concluir esta meta. Quiero agradecer también a mi esposa, quién estuvo apoyándome día a día a no desistir y dar lo mejor de mí, por compartir esas noches de desvelo para completar tareas, proyectos y artículos, su apoyo fue fundamental para poder llegar hasta este día. Gracias a ella tenía lo necesario para poder trabajar día a día, por sus palabras de ánimo, por sus sacrificios, por su paciencia, por su cariño, muchas gracias. Agradezco a mi director de tesis, quién desde un principio mostró ser el mejor guía académico que pude recibir, por enseñarme y asesorarme, por mostrar un compromiso con mi desarrollo más allá de lo normal, por enseñarme a confiar más en mi capacidad, por estar siempre disponible a atender mis dudas, por apoyarme de manera incondicional durante estos tres años. Agradezco a mi padre, a mi madre, a mi hermana, a mis tíos, a mi familia, quienes también me impulsaron a lograr más en mi carrera profesional. A mi madre que me enseñó desde pequeño a esforzarme por lograr la excelencia en todo cuanto hiciera, a luchar por obtener lo mejor, a dar lo mejor de mí. Gracias a mis compañeros de trabajo que me apoyaron en diversos aspectos durante mis 3 años de este doctorado, trámites administrativos, pruebas piloto, encuestas entre muchas otras actividades. Agradezco a todos y cada uno de aquellos que contribuyeron a que este día llegará, gracias al apoyo de todos yo pude lograr esta meta, este logro es de todos ellos.

## TABLA DE CONTENIDOS

<b>1. PROBLEMA DE INVESTIGACIÓN.....</b>	<b>14</b>
1.1 PLANTEAMIENTO DEL PROBLEMA .....	15
1.2 JUSTIFICACIÓN.....	18
1.3 OBJETIVOS.....	22
1.3.1 <i>Objetivo general</i> .....	22
1.3.2 <i>Objetivos específicos</i> .....	22
1.4 ALCANCES.....	23
1.4.1 <i>Limitaciones</i> .....	23
<b>2. ESTADO DEL ARTE .....</b>	<b>23</b>
2.1 EVALUACIÓN DEL APRENDIZAJE Y PRUEBAS ESTANDARIZADAS.....	25
2.2 USOS DE LAS PRUEBAS ESTANDARIZADAS .....	27
2.3 CONTRAPARTE DE LOS USOS DE LAS PRUEBAS ESTANDARIZADAS.....	29
2.4 TEORÍAS Y FUNDAMENTOS DE LAS PRUEBAS.....	31
2.5 PSICOMETRÍA.....	33
2.5.1 <i>Calidad de las pruebas</i> .....	33
2.5.2 <i>Análisis psicométrico y medición de la calidad</i> .....	34
2.6 TECNOLOGÍA, EVALUACIÓN Y PRUEBAS.....	38
2.6.1 <i>Software educativo</i> .....	39
2.6.2 <i>Evaluación del software educativo</i> .....	41
2.6.3 <i>Modelos de evaluación de software</i> .....	42
2.6.4 <i>Ingeniería y metodologías para el desarrollo de software educativo</i> .....	44
2.6.5 <i>Software psicométrico</i> .....	47
<b>3. REFERENTE TEÓRICOS.....</b>	<b>48</b>
3.1 EVALUACIÓN DEL APRENDIZAJE .....	48
3.1.1 <i>Evaluación Formativa</i> .....	49
3.1.2 <i>Evaluación Sumativa</i> .....	50
3.2 EVALUACIÓN ESTANDARIZADA .....	56
3.3 PSICOMETRÍA.....	58
3.3.1 <i>Teoría de Respuesta al Ítem</i> .....	60

3.4	DESARROLLO DE SOFTWARE EDUCATIVO .....	61
3.5	PLATAFORMAS EDUCATIVAS CON FUNCIONES EVALUATIVAS .....	69
3.5.1	<i>Plataformas para análisis psicométrico</i> .....	74
<b>4.</b>	<b>METODOLOGÍA</b> .....	<b>76</b>
4.1	UNIDAD DE ANÁLISIS .....	76
4.2	METODOLOGÍA PARA EL ANÁLISIS DEL ESTADO DEL ARTE.....	76
4.3	METODOLOGÍA DE LA INVESTIGACIÓN.....	77
4.4	METODOLOGÍA PARA EL DESARROLLO DEL EXAMEN .....	81
4.5	DESARROLLO DE LA TECNOLOGÍA PARA EL ANÁLISIS PSICOMÉTRICO .....	87
4.5	METODOLOGÍA PARA LA INTERVENCIÓN EDUCATIVA .....	94
<b>5.</b>	<b>PROCESO DE INTERVENCIÓN</b> .....	<b>98</b>
5.1	ANÁLISIS DE LAS DIMENSIONES DE GESTIÓN EDUCATIVA .....	98
5.2	PLAN DE GESTIÓN DE RIESGOS DE LA INTERVENCIÓN.....	100
5.2.1	<i>Etapas 1. Definición de la gestión</i> .....	100
5.2.2	<i>Etapas 2. Identificar los riesgos</i> .....	102
5.2.3	<i>Etapas 3. Análisis cualitativo de los riesgos</i> .....	104
5.2.4	<i>Etapas 4. Análisis cuantitativo de los riesgos</i> .....	105
5.3	RECURSOS DISPONIBLES .....	110
5.4	DISEÑO DE LA ESTRATEGIA DE INTERVENCIÓN .....	112
5.4.1	<i>Definición de objetivos y metas de intervención</i> .....	112
5.4.2	<i>Definición de metas</i> .....	113
5.5	PROYECCIÓN DE LA ESTRATEGIA .....	115
5.5.1	<i>Definición de indicadores</i> .....	115
5.5.2	<i>Diagnóstico de los indicadores</i> .....	120
5.5.3	<i>Análisis diagnóstico de las variables definidas</i> .....	123
5.5.4	<i>Acciones propuestas</i> .....	127
5.5.5	<i>Evaluación de las acciones</i> .....	128
5.5.6	<i>Selección de las acciones</i> .....	130
5.6	DESCRIPCIÓN ESPECÍFICA DE LA ESTRATEGIA .....	134
5.6.1	<i>Vinculación de las etapas en una cadena lógica de resultados</i> .....	136
5.6.2	<i>Descripción de las actividades que componen cada etapa</i> .....	136



5.7	INTEGRACIÓN DE LA INTERVENCIÓN CON LA EVALUACIÓN ACTUAL .....	139
5.7.1	<i>Evaluación formativa y evaluación sumativa</i> .....	140
5.7.2	<i>Retroalimentación</i> .....	141
5.8	SISTEMATIZACIÓN DEL PROCESO DE INTERVENCIÓN .....	144
5.9	EVALUACIÓN DE LA INTERVENCIÓN.....	148
<b>6.</b>	<b>RESULTADOS</b> .....	<b>150</b>
6.1	PRUEBA ESTANDARIZADA.....	150
6.1.1	<i>La retícula</i> .....	151
6.1.2	<i>Tabla de índices de relevancia curricular</i> .....	151
6.1.3	<i>Tabla de especificaciones del examen</i> .....	153
6.1.4	<i>Especificaciones de ítems</i> .....	154
6.2	PROPIEDADES PSICOMÉTRICAS .....	156
6.2.1	<i>Calibraciones entre ciclos de pilotaje</i> .....	159
6.3	EJECUCIÓN DEL PROCESO DE INTERVENCIÓN .....	160
6.3.1	<i>Componente evaluación general</i> .....	160
6.3.2	<i>Componente institucionalización de la intervención</i> .....	161
6.4	TECNOLOGÍA INNOVADORA SIEXAES.....	163
<b>7.</b>	<b>DISCUSIONES</b> .....	<b>169</b>
<b>8.</b>	<b>CONCLUSIONES</b> .....	<b>176</b>
<b>9.</b>	<b>REFERENCIAS</b> .....	<b>183</b>
	<b>ANEXO 1 ESPECIFICACIONES DE ÍTEMS</b> .....	<b>194</b>
	<b>ANEXO 2 EVALUACIÓN DE LA USABILIDAD</b> .....	<b>197</b>
	<b>ANEXO 3 ANÁLISIS DE INDICADORES</b> .....	<b>207</b>
	<b>ANEXO 4 ACCIONES ESTRATÉGICAS INTERVENTIVAS</b> .....	<b>216</b>
	<b>ANEXO 5 PLAN DE EVALUACIÓN DE LA INTERVENCIÓN</b> .....	<b>181</b>
	<b>ANEXO 6 PROYECCIÓN DE LA ESTRATEGIA</b> .....	<b>221</b>
	<b>ANEXO 7 DIAGRAMAS DE SUBPROCESOS DE SISTEMATIZACIÓN DE LA INTERVENCIÓN</b> .....	<b>263</b>

## **ANEXO 8 REPORTES TÉCNICOS GENERADOS POR EL SISTEMA SIEXAES268**

## **ANEXO 9 RETÍCULA DE LA ASIGNATURA MORFOLOGÍA DE LA SEGUNDA LENGUA .....273**

### **ÍNDICE DE TABLAS**

Tabla 1 Comparación de las metodologías tradicionales y ágiles .....	65
Tabla 2 Ejemplo de justificación de contenidos para el examen.....	83
Tabla 3 Ejemplo de especificación de un contenido temático.....	84
Tabla 4 Clasificación de impacto y ocurrencia .....	100
Tabla 5 Roles y responsabilidades en el plan de riesgos .....	100
Tabla 6 Riesgos identificados para el proyecto .....	102
Tabla 7 Niveles de riesgos.....	104
Tabla 8 Análisis cualitativo de los riesgos .....	104
Tabla 9 Análisis cuantitativo de riesgos .....	105
Tabla 10 Priorización de los riesgos.....	107
Tabla 11 Elementos asociados a los riesgos.....	108
Tabla 12. Indicadores que se van a trabajar en la intervención .....	115
Tabla 13. Diagnóstico de los indicadores asociados a los objetivos de intervención .....	120
Tabla 14. Análisis diagnóstico del estado actual de las variables definidas.....	124
Tabla 15 Evaluación de las acciones estratégicas .....	128
Tabla 16 Justificación de la elección de estrategias .....	130
Tabla 17 Estrategia #1: Valorar la calidad de la prueba.....	134

Tabla 18 Actividades de la estrategia 1, etapa 1.....	136
Tabla 19 Actividades de la estrategia 1, etapa 2.....	138
Tabla 20. Atributos valorados para obtener el IRC.....	152
Tabla 21 Ejemplo de la tabla de especificaciones del examen.....	153
Tabla 22 Estadísticas descriptivas del examen.....	156
Tabla 23 Análisis psicométrico de los ítems .....	157
Tabla 24 Propiedades psicométricas en el primer ciclo de prueba.....	159
Tabla 25 Propiedades psicométricas en el segundo ciclo de prueba.....	159

## ÍNDICE DE FIGURAS

Figura 1 Modelo para desarrollar software educativo.....	62
Figura 2 Fase 1. Diseño educativo.....	63
Figura 3 Pilares conceptuales .....	66
Figura 4 Elementos del diseño educativo .....	67
Figura 5 Modelo conceptual de espacio virtual de aprendizaje.....	70
Figura 6 Categorías, variables e indicadores de análisis en la investigación. ....	77
Figura 7. Información psicométrica recopilada con el software TAP .....	78
Figura 8. Instrumento para la evaluación de usabilidad .....	79
Figura 9. Instrumento de recopilación de información para calcular IRC. ....	80
Figura 10 Ejemplo de un segmento del formato de especificaciones de ítems .....	85
Figura 11 Diagrama de casos de uso general del sistema SIEXAES .....	89

Figura 12 Diagrama de secuencia del caso de uso consultar análisis psicométrico de un examen.....	90
Figura 13 Segmento del diagrama de clases que compone al sistema SIEXAES .....	91
Figura 14 Secuencia lógica de la creación de la prueba, el análisis y la presentación de resultados.....	92
Figura 15 Ambiente de diseño de interfaces en la plataforma de desarrollo de software. ....	93
Figura 16 Construcción de la función de generar archivo pdf a partir de un datagrid. ....	93
Figura 17 Árbol de problemas del proyecto de intervención. ....	112
Figura 18 Diagrama de flujo de las etapas dentro de la acción estratégica.....	136
Figura 19 Ejemplo del reporte técnico de resultados para el alumno.....	143
Figura 20 Diagrama del proceso de sistematización de la intervención.....	145
Figura 21 Ejemplo de la retícula de la materia de Morfología de la segunda lengua .....	151
Figura 22 Ejemplo de una especificación de ítems. ....	155
Figura 23 Gráfica del comportamiento del ítem 7.....	158
Figura 24 Interfaz de la sección de preguntas. ....	164
Figura 25 Interfaz del reporte resumido al concluir el examen.....	165
Figura 26 Interfaz de análisis de respuestas – distractores .....	165
Figura 27 Reporte técnico para el alumno.....	167
Figura 28 Reporte técnico para el docente. ....	168

## INTRODUCCIÓN

Diferentes propósitos educativos requieren diferentes pruebas y usos de estas. Cuando una prueba no guarda estrecha relación con sus propósitos, difícilmente pueden efectuarse inferencias válidas a partir de sus resultados. Desarrollar exámenes para la evaluación del aprendizaje es una tarea delicada, ya que evaluar con un instrumento mal diseñado puede tener un impacto negativo tanto para el sustentante como para el docente; incluso medir otros aspectos del conocimiento diferentes a los que se tenían pensados evaluar, va en detrimento de este. Lo anterior resalta la importancia de los aparatos metodológicos en el desarrollo o construcción de exámenes, puesto que a través de estos mecanismos se puede asegurar la calidad del instrumento, así como el aporte de estos para la obtención de información válida y confiable.

En las Instituciones de Educación Superior (IES) de los llamados países industrializados, es común que se utilicen evaluaciones diseñadas y validadas para propósitos de admisión. Para el caso específico de México, en 1994 se creó el Centro Nacional para la Evaluación de la Educación Superior (CENEVAL) con la finalidad de contar con un examen nacional de carácter indicativo, proporcionando información confiable y válida sobre los conocimientos y habilidades que adquieren las personas como beneficiarios de los distintos programas educativos, sirviendo el examen como un indicador para aquellas IES que participaron.

Actualmente existe el Examen de Competencias Básicas (EXCOBA), antes denominado EXHCOBA, el cual permite incluir elementos multimedia que enriquecen la experiencia evaluativa y ofrecen resultados automáticos; sin embargo, la estructura del EXCOBA está alineado al currículo nacional, por lo que evalúa competencias académicas básicas que se precisan en los planes de estudio de la educación obligatoria, presentando una propuesta innovadora de evaluación sobre dichas competencias escolares, pues se aleja del formato de opción múltiple y se acerca a formas más “auténticas o naturales” de evaluar los aprendizajes (Ferreya y Backhoff, 2016).

La Facultad de Idiomas de la Universidad Autónoma de Baja California (UABC) actualmente no cuenta con una herramienta tecnológica que permita construir exámenes estandarizados y que a su vez puedan valorarse mediante el análisis psicométrico. Si bien, existen softwares que permiten efectuar este tipo de análisis, estos requieren conocimientos previos por parte

del usuario, además que es necesario construir archivos fuente en formatos muy técnicos, lo que limita su uso para personal con experticia en estos campos del conocimiento. Lo anterior abre la posibilidad de innovar el proceso de la construcción de exámenes para la evaluación en los estudiantes, y a su vez la medición de su calidad desde una perspectiva psicométrica de forma más eficiente.

En los siguientes apartados se detalla, en primer lugar, cuál fue el método elegido para guiar el desarrollo del examen, lo que incluye la descripción de los instrumentos utilizados para construir y evaluar el examen. Se explica la interpretación numérica para los valores de calidad que se van a medir y lo que estos implican. Posteriormente en el apartado de resultados se muestra la valoración técnica del análisis psicométrico, que es la que determina la calidad y confiabilidad del examen desarrollado, como lo es el índice de dificultad y discriminación. Así mismo se presentan los datos obtenidos con tres de los principales instrumentos que guían la elaboración del examen. En el último apartado se concluye la importancia y relevancia de contar con instrumentos estandarizados de evaluación y las ventajas de contar con software automatizado para la aplicación, construcción y análisis psicométrico de los exámenes.

## **1. PROBLEMA DE INVESTIGACIÓN**

El trabajo de investigación aborda la temática de las evaluaciones del aprendizaje mediante la implementación de instrumentos de evaluación estandarizados. La evaluación del aprendizaje ha sido un tema de relevancia para las instituciones superiores, en el afán de mejorar la calidad educativa.

En este apartado se dan detalles de los principales problemas o deficiencias en cuanto a evaluación se refiere en la Facultad de Idiomas de la UABC, las cuales se desean abordar mediante la aplicación de pruebas estandarizadas del aprendizaje, con énfasis en el análisis psicométrico de las mismas. Por otro lado, también se explican las situaciones y circunstancias presentes en cuanto la cultura evaluativa en la institución académica para la cual se desarrolla el proceso de intervención educativa.

## 1.1 Planteamiento del problema

Las instituciones de educación superior en la búsqueda de mejorar la calidad educativa han implementado exámenes estandarizados en diferentes momentos de la vida del estudiante universitario, varios autores (Fernández, 2013; Fernández, Alcaraz y Sola, 2017; Márquez, 2014) expresan que se optan por este tipo de pruebas con la intención de contar con instrumentos de evaluación válidos y confiables que permitan formar estudiantes con los atributos y características necesarias para responder a las demandas sociales y empresariales sobre el tipo de egresado universitario que se desea.

Este tipo de exámenes estandarizados han sido implementados con fines distintos, por ejemplo, siendo de carácter normativo al filtrar o seleccionar solo a aquellos estudiantes con el mejor rendimiento, para procesos de admisión a la universidad o promoción de grado, y en algunos otros casos para pronosticar el desempeño académico futuro de los estudiantes (Hernández, Ramírez y Gamboa, 2018). Así mismo se han empleado con fines de evaluación formativa y sumativa al implementarse en pruebas ordinarias o departamentales al final de un semestre, o bien en los exámenes parciales realizados en cada unidad de aprendizaje dentro de una asignatura.

Con este tipo de iniciativas y procesos de evaluación estandarizada, las instituciones educativas buscan aumentar las probabilidades de éxito en los estudiantes, tanto durante todo su trayecto escolar, así como en cada una de las unidades que componen una asignatura en particular.

Lo anterior se convierte en un reto constante para todas las instituciones educativas, y con mayor razón para las de nivel superior, ya que es necesario poder identificar de manera acertada las capacidades, competencias, nivel de dominio, conocimientos y habilidades de los estudiantes a fin de poder adecuar los planes, programas y métodos educativos para mejorar el proceso de enseñanza y aprendizaje (Hernández, Ramírez y Gamboa, 2018). Esto sin duda expresa claramente la latente necesidad de que los métodos empleados en la construcción y diseño de pruebas realmente sean de calidad y con ello mejorar significativamente el proceso evaluativo (Tristan y Pedraza, 2017).

Con la creación del Centro Nacional para la Evaluación de la Educación Superior (CENEVAL) en 1994, y el Instituto Nacional para la Evaluación de la Educación (INEE) en el 2002, en México se ha observado el impulso de este tipo de pruebas estandarizadas, logrando así la creación del Examen Nacional de Ingreso a la Educación Superior, Examen General para el Egreso de la Licenciatura y el Examen Nacional de Ingreso al Posgrado, pruebas que por años han sido el instrumento para seleccionar, promocionar y obtener distinciones académicas en el sistema educativo del país (CENEVAL, 2017).

Este tipo de impulsos a nivel nacional propiciaron que instituciones como la Universidad Autónoma de Baja California (UABC) desarrollara instrumentos diseñados especialmente para responder a sus necesidades evaluativas, como lo fue el Examen de Habilidades y Conocimientos Básicos (EXHCOBA) utilizado para selección o admisión.

Y más actualmente la UABC ha buscado el desarrollo y aplicación de exámenes departamentales estandarizados y con ello coadyuvar al logro de una de las metas del Plan de Desarrollo Institucional alineada a la aplicación de exámenes departamentales y de trayecto para mejorar continuamente los niveles de aprendizaje de los alumnos.

Actualmente la Facultad de Idiomas no cuenta con una herramienta tecnológica que permita realizar exámenes estandarizados, así como tampoco se cuenta con un instrumento diseñado especialmente para efectuar análisis psicométricos con el cual determinar la calidad de los instrumentos de evaluación que se aplican. Lo anterior representa un área de oportunidad en cuanto a tecnología educativa con fines evaluativos se refiere.

Agregado a lo anterior es importante mencionar que los exámenes tanto del tipo formativo como los sumativos (ordinarios, departamentales, sumarios etc.) que se realizan para las diferentes asignaturas no siguen una metodología que pueda aportar confiabilidad y validez, no existe un análisis de resultados de los mismos para identificar áreas de oportunidad o dar alguna retroalimentación.

No hay actividades colegiadas para determinar la estructura de una prueba, en el caso particular de los exámenes de fin de semestre, conocidos como departamentales u ordinarios, lo que implica que la justificación de los contenidos temáticos representados en la prueba es



decidida únicamente con base al criterio determinado por el docente que imparte la asignatura.

Lo anterior pone en riesgo la correcta representatividad del universo de conocimientos determinado por el currículo de la asignatura en dicha prueba, al no emplear un proceso de jueceo más estricto y válido para que la prueba presente ítems de los contenidos temáticos que tengan un mayor impacto. Agregado a esto también la dificultad de la prueba podría variar entre docentes, significando así que un estudiante obtenga resultados diferentes en un examen de una misma materia. Esto representa un desacierto, ya que idealmente la materia debe ser evaluada en igualdad de posibilidades por cualquier alumno, independientemente del docente con quien la haya cursado.

La intervención fue realizada en la Facultad de Idiomas de la Universidad Autónoma de Baja California. Esta Facultad tiene presencia en todos los municipios del Estado. En su etapa inicial estas evaluaciones serán presentadas a gran escala cada semestre por una población estudiantil que oscila entre los 200 y 400 alumnos, se aplicarán para dos materias las cuales son las de mayor índice de relevancia curricular en el tronco común.

Se trabajaron con exámenes referidos a un criterio (Biggs, 2004; Heredia, 2009; Jornet y González, 2009), lo que nos permite estimar lo que un examinado puede hacer o no en un campo del conocimiento e identificar factores que afectan el desempeño en las pruebas. El universo de conocimientos está determinado por el currículo establecido en los planes de estudio y cartas descriptivas de cada una de las asignaturas. Los resultados de la evaluación se expresan en términos del grado con que la actuación de un individuo concuerda con los niveles de dominio esperados, y son independientes a los resultados de cualquier otro individuo. A diferencia de la evaluación referida a normas donde se emiten juicios sobre personas, la evaluación referida a criterios emite juicios sobre la actuación.

Agregado a lo anterior, un aspecto sobresaliente del instrumento a utilizar es la parte psicométrica del mismo, que permite no solo justificar los resultados obtenidos, sino que permite identificar las implicaciones prácticas y aplicadas que la medición conlleva (Barbero, 2010). Así mismo el análisis psicométrico permite determinar cuantitativamente que las

unidades (actuaciones de aprendizaje) sean correctas, y tengan un sustento sólido que avale dicha conclusión.

Por ello el objeto de estudio correspondiente a este trabajo de investigación trata el desarrollo de una prueba departamental estandarizada, así como la elaboración de una propuesta tecnológica sobre la cual se pueda implementar dicha prueba y realizar automáticamente los análisis psicométricos tradicionales basados en el modelo establecido por la teoría de respuesta al ítem. Lo anterior en el marco de un modelo de intervención educativa, con la finalidad de mejorar la calidad de la evaluación del aprendizaje en una institución de educación superior y contribuir a la creación de una cultura evaluativa con base a la utilización de procesos metodológicos rigurosos logrando así instrumentos de evaluación y procesos evaluativos válidos y confiables.

## **1.2 Justificación**

Para cualquier institución educativa es muy importante obtener información del rendimiento que tienen sus estudiantes, estimar el aprendizaje logrado y comparar dichos logros con las metas establecidas. En este punto, resulta indispensable contar con instrumentos de evaluación del aprendizaje que sean válidos y confiables; es decir, instrumentos que estén correctamente diseñados y que ofrezcan una seguridad sobre lo evaluado en términos de conocimientos y habilidades para los cuales se diseñó.

En sentido las pruebas estandarizadas son un instrumento de evaluación cuya característica sobresaliente es el exhaustivo control y método con el cual se realizan, y a este respecto Jornet (2017) destaca la relevancia de utilizar este tipo de pruebas ya que contribuyen a que las diferencias en relación a los resultados puedan imputarse a la diversidad del objeto o sujeto evaluado o a efectos debidos a una intervención sobre ellos, y no a posibles factores diferenciales derivados del método empleado con que se han diseñado, aplicado o interpretado los instrumentos.

Por lo anterior el proyecto de investigación se centra en el diseño, aplicación y análisis psicométricos de pruebas estandarizadas departamentales, este tipo de pruebas son clasificadas en la categoría de evaluaciones sumativas. Se le llaman así a este tipo de pruebas por que los resultados se utilizan para comprobar hasta qué punto los estudiantes han

aprendido bien lo que se supone que han aprendido, y por esta razón es que la evaluación se realiza después de concluir el episodio de enseñanza (Biggs, 2004).

Según Cruz y Quiñones (2012) el propósito más sobresaliente de este tipo de pruebas es tomar las decisiones pertinentes para asignar una calificación totalizadora a cada estudiante que refleje la proporción de objetivos logrados en el curso. Su función es explorar en forma equivalente el aprendizaje de los contenidos incluidos logrados en los resultados de forma individual. En contribución al aprendizaje equivalente de los contenidos es que se la prueba departamental también posee la característica de ser referida a criterios (Biggs, 2004; Heredia, 2009; Jornet y González, 2009) ya que con ello se delimita el universo de conocimientos de una manera sistematizada y con procedimientos que justifican los contenidos representados en la prueba.

Por medio de esta evaluación estandarizada departamental se trata de corroborar lo que ha sido alcanzado por el estudiante o el nivel de dominio logrado; situación que no es desconocida por el docente, ya que una vez alcanzada la evaluación sumaria, se cuentan con datos de todas las evaluaciones formativas realizadas durante el semestre y con ello se cuenta con un pronóstico de lo que se puede esperar de la evaluación sumaria. Emplear pruebas departamentales estandarizadas respaldará que es en el momento de la evaluación los resultados fueran inesperados, habría que desconfiar de la validez de las evaluaciones formativas o de la atención que se prestó a éstas para hacer los reajustes necesarios. Este tipo de conclusiones solo podrán efectuarse si la evaluación sumativa tiene un fundamento sólido establecido por una metodología rigurosa que permita comprobar la calidad de la prueba en términos psicométricos, es decir, que realmente evalúe aquello que se supone debe evaluar.

La Facultad de Idiomas carece de aparatos metodológicos con los cuales medir la confiabilidad y validez de los instrumentos de evaluación que aplica, por ello contar con una aplicación tecnológica que procese de manera automática los resultados de cada sustentante y generar los distintos reportes con los análisis psicométricos deseados, ayudaría a una toma de decisiones más oportuna. Lo que contribuye a procesos de retroalimentación más objetivos y respaldados por análisis cuantitativos que dan el sustento necesario para emitir juicios sobre el desempeño alcanzado por cada estudiante, y a su vez estimular o enriquecer los insumos para sugerir adecuaciones a las técnicas o estrategias de enseñanza aprendizaje

empleadas por los docentes a lo largo del semestre y por ende reflejado también en los procesos evaluativos de carácter formativo que implementa (Rosales, 2014).

Así mismo contar con un examen en una representación digital ayudaría al cumplimiento de metas del plan de desarrollo de la Universidad Autónoma de Baja California que apoyan la mejora continua del aprendizaje de los estudiantes.

Por otra parte, otro aspecto que estimula el desarrollo de este proyecto radica en el beneficio que aporta a la comunidad estudiantil este tipo de pruebas, ya que nuevamente por no contar con una metodología que contribuya a estimar la validez y confiabilidad de las pruebas, se presenta la deficiencia de consistencia entre evaluaciones de una misma asignatura impartida por más de un docente. Lo anterior da espacio a que existan variaciones de objetividad, dificultad y discriminación en las evaluaciones existentes para una misma asignatura, lo anterior como ya se mencionó en razón de que cada docente decide los criterios antes mencionados.

Agregado a lo anterior, una forma de zanjar estas variaciones de consistencia es aplicando el análisis psicométrico de la prueba que permite identificar los criterios de confiabilidad y validez con base en la teoría de respuesta al ítem. El objetivo sustancial de esta teoría es la elaboración de instrumentos de medición con propiedades invariantes entre poblaciones. Si dos individuos presentan idéntico nivel de rasgo medido ambos tendrán equivalente probabilidad de dar la misma respuesta, independientemente de la población de pertenencia (Attorresi, Lozzia, Abal, Galibert & Aguerri, 2009).

En este mismo sentido, los análisis psicométricos son procedimientos matemáticos que implican ciertos cálculos especializados y que requieren un mínimo de experticia, por esta razón el proyecto de investigación presenta una forma innovadora de acotar estos procedimientos al emplear un sistema que permite validar para cualquier examen de opción múltiple los criterios de validez y confiabilidad establecidos en la Teoría de Respuesta al Ítem (Hidalgo y French, 2016; Muñoz, 2010; Gómez, 2015) , los cuales son el índice de dificultad de la prueba, índice de dificultad de cada ítem, índice de discriminación y coeficiente de discriminación de cada ítem, y por último el análisis de respuestas de cada ítem con el cual se valoran las elecciones de cada una de las respuestas de cada ítem, esto

permite observar el comportamiento de los grupos altos y bajos al momento de elegir sus respuestas.

Además, la Facultad de Idiomas carece también de una cultura evaluativa basada en prácticas estandarizadas y con soporte metodológico como se ha mencionado anteriormente, por ello el proyecto de investigación busca lograr con dicha intervención la creación de una cultura evaluativa que estimule la aplicación de varias buenas prácticas educativas con TIC, ya que, por ejemplo, la propuesta tecnológica innovadora aporta una mejora al proceso educativo de las evaluaciones mediante exámenes departamentales, implementando procesos de evaluación sistematizados y automatizados, hablando claro del procesamiento de los datos, el análisis psicométrico y los reportes técnicos diseñados específicamente con la información relevante para cada uno de los actores participantes en el proceso, como lo son los directivos, coordinadores, docentes y alumnos. Lo anterior permite llevar la capacidad de retroalimentación más allá del alumno y el docente, permitiendo transitar a los espacios de administración educativa en la institución. González y Colás (en García-Valcárcel y Hernández, 2013, p. 30) menciona que “para que se pueda iniciar y desarrollar una práctica docente con TIC de valor pedagógico es preciso vincular la misma con la creación de una filosofía y cultura de trabajo en el centro basada en el uso de las TIC”.

La Facultad de Idiomas cuenta con plataformas educativas de carácter institucional con capacidades evaluativas como lo es Blackboard, Google Classroom, es importante mencionar que Moodle a partir de cierta versión ya incluye un apartado para analizar el índice de dificultad y el índice de discriminación de los ítems, sin embargo, no viene acompañado de ciertos atributos psicométricos ni de un aparato metodológico con el cual diseñar la prueba. Lo anterior muestra el vacío de conocimiento existente, ya que algunas plataformas educativas no están basadas en modelos psicométricos o no van acompañadas de metodologías para el diseño de pruebas de calidad. Es decir, que dichas plataformas no contemplan aquellos criterios de confiabilidad y validez de un examen, ni las mediciones que estos mismos conllevan.

Existe una falta de información detallada de los resultados de los exámenes, y por ende una toma de decisiones con menos elementos para la identificación y corrección de problemas en las asignaturas impartidas a través de estas plataformas. Aunado a esto, más del 75% de las

asignaturas que se imparten no emplean esas plataformas para realizar sus evaluaciones, por lo que sigue presente la ausencia de tecnología educativa que pueda ser llevada de una manera más práctica y sencilla para la realización de exámenes en formato digital, con capacidades de análisis psicométricos y la producción de reportes técnicos diseñados específicamente a las necesidades de información de los actores involucrados, logrando así contar con información relevante que apoye de manera significativa la toma de decisiones en los distintos niveles e implicados en el proceso de evaluación del aprendizaje.

De manera directa los beneficiados son los alumnos y el personal docente involucrados en la asignatura para la cual se desarrolla la prueba. Otros beneficiados directos son los responsables de las áreas de coordinación académica y directivos. De manera indirecta también se ven beneficiados todos los docentes que componen la planta académica de la institución, ya que, aunque el proceso de investigación no interviene todas las asignaturas que comprenden el plan de estudios, las estrategias y metodologías implementadas si pueden migrar a cualquier otra asignatura.

### **1.3 Objetivos**

#### ***1.3.1 Objetivo general***

Diseñar una propuesta innovadora de análisis psicométrico para los exámenes estandarizados de los aprendizajes en la Facultad de Idiomas de la Universidad Autónoma de Baja California.

#### ***1.3.2 Objetivos específicos***

- 1.-Analizar metodologías para el desarrollo de exámenes estandarizados.
- 2.-Analizar metodologías para el análisis psicométrico de exámenes.
- 3.- Examinar tecnologías o instrumentos con los cuales se realicen análisis psicométrico de exámenes.
- 4.- Diseñar una propuesta tecnológica para el análisis psicométrico de exámenes estandarizados.

## **1.4 Alcances**

- 1.- El proyecto de investigación comprende el desarrollo de una prueba estandarizada departamental para la asignatura de Morfología en la Segunda Lengua.
- 2.- La investigación comprende la aplicación de un modelo de intervención educativa, diseñando estrategias y acciones interventivas, así como la evaluación y justificación de las mismas.
- 3.- La prueba es de tipo formativo y de carácter criterial, por lo que solo comprende aquel universo de conocimientos definido en el currículo de la asignatura.
- 4.- La evaluación de la calidad técnica del instrumento solo comprende el análisis psicométrico de los índices de dificultad y discriminación, y el coeficiente de discriminación de los ítems que componen la prueba.
- 5.- El proceso de intervención comprende el diseño y producción de los instrumentos base sobre los cuales efectuar procesos de retroalimentación y la toma de decisiones.
- 6.- La investigación comprende la elaboración del reporte técnico sobre la calidad psicométrica del instrumento de evaluación.

### **1.4.1 Limitaciones**

- 1.- La cantidad de alumnos con los cuales hacer el pilotaje de la prueba y el primer análisis psicométrico está en razón de la cantidad de alumnos inscritos en las cuatro sedes que comprenden la Facultad de Idiomas.
- 2.- El proceso de intervención no incluye la definición de modelos o procedimientos para efectuar la retroalimentación posterior a la aplicación del instrumento de evaluación.
- 3.- El proyecto de investigación no contempla un análisis comparativo entre los resultados obtenidos entre grupos que aplican la prueba departamental al interior de la sede o unidad académica, ni tampoco entre sedes.

## **2. ESTADO DEL ARTE**

La evaluación estandarizada del aprendizaje o llamada también evaluación de gran escala, es una forma de evaluación sistematizada del aprendizaje, caracterizada por seguir un proceso riguroso con marcos referenciales teóricos y metodológicos con los cuales se miden rasgos

observables en la población objeto, estableciendo precisiones específicas, controles logísticos y administrativos. De igual manera se identifica por la sistematización de los instrumentos o técnicas con las que se recopila, analiza e interpreta la información de forma tal que se utilicen los mismos instrumentos durante todo el proceso.

Las pruebas estandarizadas han surgido como una posible respuesta o herramienta para la mejora de los procesos educativos en las instituciones, en particular del fin de contar con instrumentos de evaluación que sean válidos y confiables, aportando información con la cual las instituciones educativas puedan tomar decisiones y emprender acciones para mejorar la calidad de sus procesos, y con ello el aprendizaje de sus estudiantes. Es por lo anterior que se ha visto utilizar pruebas estandarizadas con el fin de instaurar parámetros de desempeño imprescindibles para establecer objetivos educativos, o como índices para predecir desempeño de estudiantes, entre otros usos.

Los retos que significa la evaluación de los aprendizajes en la educación superior, tanto como actividad que responde a las demandas y necesidades formativas del siglo XXI, así como mecanismo de análisis e intervención en el campo de la investigación educativa, hacen del tema un campo de estudio apremiante. De esta manera, en el presente estado del arte se da cuenta del desarrollo del conocimiento científico que se ha realizado en los últimos años sobre la evaluación estandarizada de los aprendizajes, en términos de aplicación de la psicometría como método de medición y aseguramiento de la calidad, así como la inclusión de la tecnología aplicada a los procesos educativos y de evaluación. En consecuencia, este estado del arte expone los hallazgos más relevantes en la temática, así como la demarcación de los alcances y vacíos que ha tenido la comunidad científica en los últimos 25 años.

La estructura de este apartado se relaciona con tres grupos de análisis que corresponden directamente con el enfoque de la evaluación estandarizada de los aprendizajes. El primero corresponde a las investigaciones que dan cuenta sobre las modalidades de evaluación, teorías y metodologías que se aplican bajo este enfoque. En el segundo se refieren las investigaciones que abordan específicamente los mecanismos de medición, evaluación y/o la calidad de las pruebas estandarizadas, haciendo énfasis en la intervención o uso de la psicometría. Por último, se analizan las investigaciones que relacionan la evaluación del aprendizaje con la evaluación estandarizada, así como el uso de tecnología aplicada a la



educación (software educativo con o sin funciones evaluativas). A guisa de cierre de este apartado, se realiza un análisis de los vacíos de conocimiento existentes en relación a la aplicación de tecnología en la elaboración y puesta en práctica de pruebas estandarizadas, haciendo un análisis particular en los aspectos involucrados para determinar la calidad psicométrica de las mismas.

La revisión de la literatura para este apartado se realizó aplicando una metodología que consistió en la definición de categorías, variables e indicadores diseñadas especialmente para los propósitos de este trabajo de investigación, tomando como referencia la producción científica de los últimos 25 años en lo relacionado a evaluación educativa, haciendo un énfasis especial en la evaluación estandarizada del aprendizaje (véase el apartado metodología en este documento).

## **2.1 Evaluación del aprendizaje y pruebas estandarizadas**

Las investigaciones que aquí se presentan se alinean al estudio del qué es y cómo se comprende la evaluación del aprendizaje, con un particular énfasis en las evaluaciones a gran escala o también conocidas como evaluaciones estandarizadas, en cuanto metodologías en el diseño o desarrollo de las mismas, y las teorías que fundamentan las técnicas o instrumentos con los cuales se elaboran este tipo de pruebas, cerrando con los usos y/o beneficios que se han logrado en la aplicación de estas.

Definir la evaluación del aprendizaje obliga la reflexión de las diferentes miradas que posicionan su objeto de estudio en este campo; esto con la finalidad de conceptualizarla en su justa medida, aunado al hecho que esta es una temática de interés actual para las instituciones de educación superior por la creciente exigencia sobre la calidad en la formación de las nuevas generaciones de profesionales.

La evaluación del aprendizaje puede clasificarse desde diferentes líneas y con base en una diversidad amplia de criterios; por ejemplo, el momento en que se evalúa (inicial o diagnóstica), por el actor que la realiza (autoevaluación, heteroevaluación, coevaluación, etcétera); por la forma en que esta se realiza (formativa, dinámica, estática, sumativa) o por el enfoque como lo es la evaluación autentica o alternativa centrada en los procesos (Izquierdo, 2008; Zúñiga, Solar, Lagos, Báez y Herrera, 2014).

En este sentido, la evaluación de los aprendizajes funge como base fundamental para la observación de los aciertos y errores en las estrategias empleadas para lograr la adquisición de conocimientos en los estudiantes y con ello tomar decisiones en pro de su formación educativa. Bogantes (2015) comenta que la evaluación del aprendizaje es el ejercicio educativo o formativo que dictamina qué, cómo, por qué y cuándo se deben de enseñar; es decir, se asocia la evaluación del aprendizaje con las decisiones que se hayan o tomen durante los procesos de enseñanza y aprendizaje.

En este tenor, la evaluación es un proceso continuo que se realiza prácticamente desde que se tiene conciencia; por ello, Mendivil (2012) asevera que esta no surge en y para el ámbito pedagógico, ya que su génesis se alinea a la vida cotidiana, puesto que en todo momento efectuamos una evaluación del comportamiento, la expresión, el rendimiento. Por otro lado; en el ámbito de la educación superior se observa una tendencia por describir la evaluación como el producto de una serie de interacciones entre diferentes tipos de personas en un espacio y tiempo determinados, interacción que particularmente ofrece un marco referencial sobre el cual los docentes orientan las prácticas de evaluación (Contreras, 2010).

A pesar de la relevancia que enmarca al campo, en la actualidad existe una confusión o mala interpretación del significado real de la evaluación del aprendizaje, que llegan a reducirla a la mera acción de calificar; al respecto Alcaraz (2015) explica que:

en la actualidad aún seguimos ... confundiendo algunas características de la evaluación con las de la calificación. De modo que, prácticas que dicen ser de evaluación tienen tras de sí muchas de las cualidades propias del acto de calificar. Decimos que evaluamos para comprobar si el alumnado está aprendiendo o no y olvidamos que la principal función de la evaluación no es tanto comprobar el aprendizaje como asegurar las condiciones para que se produzca dicho aprendizaje (pp. 210-211).

En general se observan definiciones de evaluación del aprendizaje que se pueden clasificar o resumir en una serie de fines particulares. Uno de ellos es la evaluación como mejora, que para ciertos autores se comprende por dos tipos de evaluación: la evaluación formativa y evaluación formadora (Umaña, A., Calvo, X. y Salas, N., 2017), la cual permite transformar la práctica del docente y el aprendizaje del estudiante; evaluación como rendimiento de cuentas del docente o de la institución, que se observa cuando los docentes demuestran que realizan correctamente su trabajo y que logran los objetivos educativos planteados por la

institución (este es uno de los usos que se le dan a las pruebas estandarizadas, también conocidas como pruebas de logros); la evaluación como rendimiento de cuentas del alumno, donde el estudiante demuestra, en la mayoría de los casos, la prueba de su aprendizaje de un determinado tiempo (Hidalgo y Murillo, 2017).

Por otro lado, la evaluación estandarizada de los aprendizajes centra sus acciones desde lo que explican Tristán y Pedraza (2017):

Las pruebas estandarizadas son los instrumentos de medición más utilizados en psicología, educación, ciencias de la salud y ciencias sociales, que cuentan con un amplio desarrollo técnico y metodológico con formas perfeccionadas para medir los rasgos observables o latentes, en la población focal específica y con un grado de precisión previamente establecido y controlado por procedimientos logísticos y administrativos igualmente objetivos. (p.18)

En relación a esta característica metodológica se observó también que la estandarización se entiende como un proceso de sistematización de todos aquellos elementos que están asociados a una acción de recogida e interpretación de información, de tal forma que se utilicen los mismos instrumentos o técnicas tanto para recopilar, análisis e interpretar la información (Jornet, 2017). Es importante notar que otros autores refieren como características particulares de este tipo de pruebas son los marcos de referencia teóricos y metodológicos rigurosos (Popham, 1999; Martínez, 2001; Gómez, 2004; Backhoff, 2018; Fernández, M., Alcaraz, N., & Sola, M., 2017), resaltando el hecho que se les asocia con fines de evaluación válidos y confiables. Las concepciones que se tiene de evaluación estandarizada se hace un especial énfasis en el elemento base de ser sistematizadas y que emplean métodos o instrumentos muy rigurosos para la recopilación y análisis de información. Se sustentan en análisis con marcos referenciales teóricos con los cuales se pueden efectuar mediciones que dan como resultado valoraciones cuantificables de atributos asociados a la calidad de la prueba, como son la validez y la confiabilidad.

## **2.2 Usos de las pruebas estandarizadas**

Un tema asociado directamente a las pruebas de evaluación del aprendizaje es el uso que se le da a estas. Uno de los propósitos o intenciones de evaluar es la obtención de información

específica con base en la cual se puedan efectuar acciones o decisiones educativas. En este sentido, el uso que se le otorgan a las pruebas estandarizadas normalmente se encuentran asociados a pruebas de gran escala o de objetivos, donde por citar un ejemplo se encuentran las pruebas de admisión a las universidades y aquellas para medir el desempeño de los estudiantes de un país. Inicialmente se revisa la percepción general de los usos de este tipo de pruebas, y posteriormente se analizan las diferencias o aspectos negativos que se observan sobre este tipo de pruebas.

En el caso particular de México este tipo de pruebas no han sido una práctica muy frecuente, la Secretaría de Educación Pública empieza a aplicarlas en 1972, estas se utilizaron para decidir la admisión de alumnos en secundaria (Martínez, 2009); sin embargo, la situación no tuvo otro avance sobresaliente hasta 1994, cuando se impulsó la realización de pruebas a gran escala en la educación básica mediante el proyecto denominado Estudio de Evaluación de la Educación Primaria.

Este año también fue importante para la evaluación en México, ya que con el ingreso a la Organización para la Cooperación y el Desarrollo Económicos (OCDE) se aumentaron los esfuerzos por integrarse a la vida económica y política internacional, lo cual incluía las evaluaciones educativas a gran escala. Esto produjo en consecuencia la creación del Centro Nacional para la Evaluación de la Educación Superior (CENEVAL) y con ello se extendió este tipo de evaluaciones a gran escala o estandarizadas en el país. En 1996 se realiza un trabajo de estándares curriculares produciendo evaluaciones en relación a ellos, lo que se conoció como Pruebas de Estándares Nacionales aplicadas por primera vez en 1998 (Martínez, 2001).

Como se mencionó al inicio de este trabajo, un variable de interés es observar cuales son los usos que se le dan a las pruebas estandarizadas. En este sentido sobre salen apreciaciones que asocian este tipo de pruebas a rendimiento escolar, a calidad educativa en las instituciones e incluso a dar razones de la educación de en un país (Gómez, 2004; Tiramonti, 2014).

Como se ha mencionado, una característica de este tipo de pruebas es el exhaustivo control y método con el cual se realizan, y a este respecto Jornet (2017) destaca la relevancia de utilizar este tipo de pruebas:

La estandarización ha estado defendida desde posiciones en las que se enfatiza la necesidad de plantear un mismo sistema de acercamiento a la realidad (sea en pruebas de rendimiento, escalas de actitudes o sistemas de evaluación de docentes, instituciones, programas...) con el fin de que las diferencias en relación a los resultados puedan imputarse a la diversidad del objeto o sujeto evaluado o a efectos debidos a una intervención sobre ellos, y no a posibles factores diferenciales derivados del modo en que han diseñado, aplicado o interpretado los instrumentos. (p. 6).

Por su parte, algunos autores consideran que la importancia de usar estas pruebas radica al hecho que valoran el aprendizaje de los estudiantes de una forma masiva y lo atribuyen a los efectos del sistema educativo al que pertenecen, con lo que se permite proporcionar información sobre las fortaleza o debilidades. Lo anterior se explica al decir que el sistema educativo produce aprendizajes en los estudiantes, y la aplicación de pruebas estandarizadas en la medición de los resultados de dichos aprendizajes, con lo cual se puede valorar la calidad del propio sistema educativo, y con ello señalar que cambios deben realizarse para mejorar.

Resulta interesante observar como el planteamiento principal de estos autores sobre la virtud de la prueba estandarizada es tratar de contar con un instrumento correctamente diseñado que minimice el error de medición, es decir que las diferencias en los resultados al ejecutar determinada prueba se deben a las características propias del individuo y no a un mal diseño o estructura de la prueba. Sin embargo, se esbozan también por su parte ciertos aspectos que tienen que ver con el uso de estas pruebas y que no son del todo favorables o acertados para evaluar el aprendizaje de los estudiantes, o más bien la calidad educativa de la institución.

### **2.3 Contraparte de los usos de las pruebas estandarizadas**

Como se mencionó anteriormente, hay autores que manifiestan las ventajas y usos de este tipo de pruebas (Jornet, 2017; Tiramonti, 2014), pero, así como han surgido usos y ventajas en la aplicación en este tipo de pruebas, también existe una contraparte que expresan otros investigadores sobre las desventajas de utilizar pruebas estandarizadas. En este sentido Popham (1999) haciendo referencia a las pruebas para medir la calidad educativa de una institución considera:

emplear pruebas estandarizadas de logros para averiguar la calidad educativa es como medir la temperatura con una cuchara. Las cucharas tienen la misión de medir cosas diferentes que el calor

o el frío. Las pruebas estandarizadas de logros tienen la misión de medir algo distinto que cuán buena o cuán mala es una escuela. Las pruebas estandarizadas de logros deberían usarse para hacer las interpretaciones comparativas que se supone deben suministrar. No deberían ser usadas para evaluar la calidad educativa. (p.4)

A esta línea de intelección se agregan otros autores que con el paso de los años han observado un uso desvirtuado de este tipo de pruebas, y como es que se ha llegado a conclusiones erróneas sobre la calidad educativa de una institución por utilizar resultados que midieron todo menos la calidad de la institución. Por ejemplo, Gómez (2004) explica que el conocimiento que se tiene sobre el rendimiento de los estudiantes se limita al desempeño obtenido en las pruebas estandarizadas, hay un desconocimiento de otras habilidades, actitudes y competencias que son implícitos a cualquier área de conocimiento. Y por esta razón las conclusiones que se emiten sobre la calidad de un sistema educativo son incompletas.

Siguiendo esta misma crítica, se agregan elementos como la mala práctica de preparar a los estudiantes exclusivamente para lograr un buen rendimiento en pruebas estandarizadas, esto por la misma presión administrativa por querer obtener un buen ranking y con ello acceder a ciertos apoyos o bien lograr un estatus social. De acuerdo a esta situación, se hace notar como los estudiantes son orientados, enseñados o educados para responder correctamente a estas pruebas, en lugar de enseñarlos a pensar por sí mismos y convertirse en aprendices creativos (Buckman, citado en Moreno, 2016).

La revisión ha arrojado otros resultados sobre el mal uso de las pruebas, las investigaciones ponen de manifiesto situaciones desfavorables que trascienden la cuestión técnica de la prueba, y con esto se hace referencia a la manipulación de la información por parte de las instituciones u organismos encargados de aplicar estas pruebas.

En relación a esto, se encuentra que ciertas instituciones manejan la información de manera secreta, cuando en realidad debería de ser difundida, desde luego tomando en cuenta controles de integridad y confidencialidad de la información de los participantes directos de la prueba, ya que compartir dichos resultados es clave para tomar acciones de mejora hacia el interior e incluso exterior de la institución. Sin embargo, este tipo de prácticas en ocasiones son practicadas por motivos políticos (Martínez, 2001).

Un punto común que expresan estos autores es que se suelen tomar decisiones con información o resultados de pruebas que no fueron diseñadas para ello. Además, se hace referencia que las instituciones manipulan los resultados de una manera hermética con lo que se entorpece la difusión de resultados, difusión que es elemental para lograr la mejora de la misma institución.

Es interesante como lo anterior indirectamente dice algo sobre la concepción de la calidad técnica de las pruebas, ya que, aunque se encuentran opiniones desfavorables sobre las pruebas estandarizadas, la mayoría de estas están enfocadas en el uso que se da a los resultados obtenidos con ellas, y no tanto a un mal diseño de la prueba, a los métodos o técnicas de recopilación y análisis de resultados.

#### **2.4 Teorías y fundamentos de las pruebas**

El diseño de una prueba debe ir acompañado de un sustento teórico que sirva de referencia para la aplicación de una o más estrategias pedagógicas y a su vez entender que influencia o corriente cognitiva a modelado la elaboración de una prueba, en este sentido es importante observar cuales han sido las principales teorías sobre las que gira la evaluación del aprendizaje y el diseño de pruebas.

Una primera aproximación hace referencia al positivismo del siglo XX, con los trabajos de de Mill en 1822 y Comte en 1842, Binet (citado en Tristán Y Pedraza, 2017) explica que:

La idea de base de las pruebas estandarizadas como instrumentos de medidas de objetos abstractos o rasgos latentes cuenta con una profunda influencia del positivismo del siglo XIX, que buscaba establecer con el mayor rigor metodológico posible una definición del objeto de estudio, por ejemplo, la inteligencia o el rendimiento escolar. (p. 18).

Por su parte Martínez (2001) en relación al principio de las teorías de medición (termino intrínseco a la evaluación), indica que surge en siglos más atrás. Él explica que las bases de la teoría de la medición fueron puestas por los trabajos de Laplace en 1796 sentando las bases para la teoría de probabilidades y el de Gauss en 1798 con los fundamentos de la teoría de números. El interés por aplicar estas teorías de medición en contextos educativos surge con mayor auge en países como Alemania, Inglaterra, los Estados Unidos y, en menor medida, Francia y las regiones francófonas de Suiza y Bélgica.

Más adelante en 1907 Binet construyó el primer test estandarizado de inteligencia, el cual fue desarrollado teniendo como referencia los trabajos de Galton, Wundt, Ebbinghaus, Pearson y Spearman, quienes a su vez sentaron las bases para la Antropometría y la Psicofísica (Aiken, citado en Escurra, 2011). Todos estos trabajos sentaron las bases de lo que más adelante sería conocida como la Teoría Clásica de los Test (TCT) siendo Gulliksen en 1950 quien lleva a cabo la síntesis canónica de este enfoque (Muñiz, 2010).

El avance en el campo de la medición y elaboración de pruebas gracias a la TCT continuó dando así el surgimiento de otras teorías como la Teoría de la Generalización (Cronbach, Gleser, Nanda y Rajaratnam, 1972) y la Teoría de Respuesta al Ítem (Rasch, 1960; Lord y Novick, 1968; Lord, 1980).

Ambas teorías son propuestas para mejorar aquellos aspectos susceptibles de error y ausencia de medición en la teoría clásica de los test. En el caso particular de la Teoría de Respuesta al Ítem (TRI) se busca un fundamento probabilístico al problema de medir constructos latentes (particularmente los no observables), en la cual se considera al ítem como la unidad básica de medición, con lo cual se puede pronosticar como responderá un sustentante a un ítem en particular (Cortada, 2004).

De Cortada (2004) también muestra una de las ventajas observadas de la TRI en relación a la Teoría Clásica de los Test, y tiene que ver con el sesgo, comenta:

Los tests han sido algunas veces criticados por considerarse que eran sesgados respecto a las minorías étnicas y una de las ventajas de la teoría de la respuesta al ítem es que proporciona un marco de referencia unificado para conceptualizar los sesgos a nivel del ítem. Esto se consigue hallando lo que se denomina el funcionamiento diferencial del ítem o DIF. Un ítem presenta DIF si los sujetos que tienen la misma aptitud, pero pertenecen a distintos grupos, no tienen la misma probabilidad de contestar bien el ítem. Por lo tanto, el DIF puede investigarse comparando los parámetros que describen las curvas características del ítem. (p. 108).

Actualmente se han desarrollado diversos trabajos basados en la TRI (Ferreya y Backhoff, 2016; Santelices y Valenzuela, 2015; Baladrón, Sánchez, Romeo, Curbelo, Villacampa y Jiménez, 2017), los cuáles muestran que esta teoría es aplicable en diferentes áreas del conocimiento.



## 2.5 Psicometría

### 2.5.1 Calidad de las pruebas

Es de suma importancia contar con instrumentos de evaluación que estén correctamente diseñados, con el fin de que los resultados que de estos se obtengan sean válidos y confiables, para a partir de ellos hacer inferencias, emitir juicios y una toma de decisiones acertada.

En la investigación una categoría de estudio es la Psicometría, para esta se han definido dos indicadores de interés particular asociados a atributos psicométricos, a saber, la validez y la confiabilidad de una prueba. De igual manera otro indicador asociado son los cálculos específicos que se realizan para determinar los atributos antes mencionados.

Haciendo una revisión inicial sobre la confiabilidad, se encuentra que este criterio se asocia con los errores de medición (Argibay, 2006), el autor explica que en toda medida el valor obtenido es compuesto por el valor verdadero y los errores en la medición, de forma tal que un instrumento será más confiable en razón de maximizar el valor verdadero. Al reducir el error de medición, se incrementa la similitud entre el valor verdadero y el obtenido con la medición, y en igual proporción la confiabilidad del instrumento.

En particular se señalan dos tipos de errores: los errores aleatorios que, dada su naturaleza, no hay una capacidad de control o predicción de los mismos, y por otro lado están los errores sistemáticos, que pueden ser controlados y ser sujetos de modificarse mediante alguna alteración en el mismo sistema. Para efectos de la calidad de una prueba, el error sobre el que hay interés es el error aleatorio (Argibay, 2006).

En términos académicos se ha encontrado que el concepto de confiabilidad se explica por ejemplo si en ausencia de cualquier cambio de manera permanente en una persona las calificaciones de una prueba varían en gran medida con el tiempo o en diferentes situaciones, es probable que la prueba no sea confiable, y por tal razón no pueda ser utilizada para explicar o predecir el comportamiento de los sustentantes (Aiken, citado en Árraga y Sánchez, 2012)

Como se observa, las investigaciones manifiestan la confiabilidad como un atributo elemental para considerar la calidad de una prueba. Sin embargo, la confiabilidad debe ir acompañada de otro atributo, la validez.

Una prueba puede ser confiable, pero si no es válida, los juicios emitidos a partir de los resultados serán erróneos o insuficientes para tomar decisiones.

Esto sucede con el uso que se tiene de las pruebas, en un apartado anterior Popham señaló que en ocasiones las pruebas estandarizadas eran similares a medir la temperatura con una cuchara, es decir que el instrumento no era válido para lo que se deseaba evaluar. Lo anterior significa que de nada sirve tener un instrumento confiable si este no es el adecuado para la medición que quiere realizarse.

Continuando con lo antes explicado, la conceptualización de validez es explicada por Aliaga (2007) de la siguiente manera:

como un juicio evaluativo global en que la evidencia empírica y los supuestos teóricos respaldan la suficiencia y lo apropiado de las interpretaciones y acciones en base a los puntajes de las pruebas, que son función no sólo de los ítems sino también de la forma de responder de las personas, así como del contexto de la evaluación. (p. 95).

En términos coloquiales, Hernández (citado en Árraga y Sánchez, 2012, p. 383) lo define como: “el grado en el que un instrumento realmente mide la variable que pretende medir”.

Para este atributo hay diferentes tipos de evidencia, la asociada con el contenido, la que se relaciona con el criterio y la referida al constructo. Sin embargo, Gregory (citado en Árraga y Sánchez, 2012) señala que algunos autores e investigadores en psicometría como Cronbach en 1988, Guion en 1980 y Messick en 1995, consideran a la validez de constructo como el elemento central para todos los tipos de evidencia de validez, dejando los restantes tipos solo como apoyo.

### ***2.5.2 Análisis psicométrico y medición de la calidad***

En apartados anteriores se analizó como surgieron las primeras teorías que sentaron las bases para el desarrollo de pruebas. En el caso particular de la Teoría de Respuesta el Ítem se han efectuado investigaciones para determinar una serie de atributos con los cuales se puede observar la validez y confiabilidad de una prueba. En el caso particular de la investigación una variable de estudio definida es la calidad de las pruebas mediante el uso de la psicometría. Para esta variable se definieron tres indicadores de interés especial, que a su vez son los

atributos con los que se observa la calidad técnica de una prueba basada en la TRI, estos son: el índice de dificultad del ítem, índice de discriminación y el coeficiente de discriminación.

El índice de dificultad del ítem se define como la proporción de una muestra o población que responde acertadamente un ítem o pregunta en una prueba (Medina, Ramírez y Miranda, 2019).

Croker y Algina (citado en Backhoff, Larrazolo y Rosas, 2000) mencionan que usualmente, a esta proporción se le denota con una  $p$ , la cual indica la dificultad del ítem. El cálculo de este atributo se realiza mediante la división del número de personas que contesto acertadamente el ítem entre el número de personas que en total contestaron el ítem.

La fórmula que expresa la operación anterior se representa de la siguiente manera:

$$p = \frac{A}{N}$$

En cuanto a los valores posibles de este atributo, estos van desde cero hasta uno, Wood explica (citado en Backhoff et al, 2000) “a mayor dificultad del ítem, menor será su índice” (p. 14). Lo anterior quiere decir que entre más cercano a uno se encuentre el valor de este índice, el ítem es más fácil de responder, y viceversa.

En relación al índice de discriminación de la prueba, la revisión encuentra que los autores han definido este atributo (Backhoff et al, 2000; Medina et al, 2019) con una analogía simple, quien haya obtenido una mejor puntuación en todo el examen deberá tener mayores probabilidades de contestar correctamente un ítem, así pues la discriminación es la cualidad que tiene un ítem para separar a los estudiantes con mejores puntuaciones de aquellos con menor puntuación final en la prueba.

Existen diferentes métodos para obtener la discriminación, uno de ellos consiste en separar a la población de los sustentantes en dos grupos, 50% con puntajes superiores a la media, y 50% con puntajes inferiores. El cálculo necesario para la obtención de este atributo está dado por la siguiente fórmula:

$$D_i = \frac{N_{sm} - N_{im}}{\frac{N}{2}}$$

Donde  $N_{sm}$  es el número de sustentantes con puntaje superior a la media que acertaron el ítem,  $N_{im}$  es el número de sustentantes con puntaje inferior a la media que acertaron el ítem y  $N$  es el número total de sustentantes que contestaron el ítem.

Otro método encontrado en la revisión de este apartado es tomando percentiles de la población, en lugar de considerar a todos los sustentantes, en este sentido Backhoff et al. (2000) solo considera al 54% de ellos, 27% de las puntuaciones más altas en el test y 27% de las puntuaciones más bajas en el test. Mientras que Medina et al. (2019) hace el cálculo con el 25% de la población tanto en grupos altos como bajos.

Para Backhoff et al. (2000) el cálculo necesario para la obtención de este atributo está dado por la siguiente formula:

$$D_i = \frac{GA_{aciertos} - GB_{aciertos}}{N_{grupomayor}}$$

Donde  $GA_{aciertos}$  es el número de aciertos en el ítem del 27% de los sustentantes con las puntuaciones más altas en el examen,  $GB_{aciertos}$  es el número de aciertos en el ítem del 27% de los sustentantes con las puntuaciones más bajas en el examen, y por último  $N_{grupomayor}$  es el número de sustentantes en el grupo más numeroso ya sea GA o GB.

Si bien existe una diferencia entre los autores referente a los percentiles al momento de calcular el índice, donde hay concordancia es en la ventaja que presenta emplear este método, ya que para ambos la ventaja de utilizar este método permite reducir la probabilidad de subestimar el nivel de discriminación de los ítems, precisamente por incluir solo a aquellos sustentantes con mayor consistencia en su rendimiento. Para este atributo el rango de valores es de  $[-1,1]$ .

Respecto a estos valores Ebel y Frisbie (citado en Backhoff et al, 2000) proponen una regla para clasificar la calidad del ítem en términos de índice de discriminación. Como mínimo este índice debe encontrarse por encima de .2 para considerarse regular, entre .3 y .39 se considera buena, y mayores a .39 son excelentes. Un índice por debajo de .2 significa que el ítem esta defectuoso, necesita revisión profunda o bien desecharse.

Además del índice de discriminación, también se ha investigado el coeficiente de discriminación (Pérez, Acuña y Arratia, 2008; Medina et al, 2019), el cual es conocido como el punto de correlación biserial ( $r_{pbis}$ ).

Este atributo se define como una medida de la consistencia de un ítem con toda la prueba en su conjunto, el cual refleja la correlación entre los puntajes de los sustentantes en un ítem en particular y sus puntajes en la prueba completa. Dicho atributo puede ser obtenido a partir del siguiente calculo:

$$r_{pbis} = \frac{\bar{X}_1 - \bar{X}_0}{S_x} * \sqrt{\frac{n_1 n_0}{n(n-1)}}$$

Esta fórmula es la que utiliza Medina et al (2019) en la investigación *Validez y confiabilidad de un test en línea sobre los fenómenos de reflexión y refracción del sonido*, así como la que utilizó Backhoff et al (2000) en su investigación *Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA)*.

Este coeficiente permite observar la probabilidad de que un ítem sea contestado correctamente por aquellos estudiantes con mayor puntuación en la prueba, esto se deduce al presentar una correlación positiva y que se encuentre cercana a uno, es decir que entre más cercano a uno sea la correlación, la probabilidad es más alta. En el caso contrario, si la correlación es negativa, significa que habrá una mayor tendencia de que los estudiantes con menor puntuación en la prueba acierten el ítem, lo que puede significar que el ítem es defectuoso. Por lo tanto, el rango de valores posibles para este atributo es de [-1,1].

Referente a atributos que permiten observar la confiabilidad de una prueba mediante una medición psicométrica están el Índice de Confiabilidad de Kuder-Richardson (KR20) y el Alfa de Cronbach, estos índices permiten medir la consistencia interna (Reidl, 2013) de instrumentos o pruebas, además de ser los procedimientos más comunes para dicho propósito. Para que el índice KR20 sea de calidad o aceptable, debe ser mayor a 0.70, mientras que para el Alfa de Cronbach debe ser de 0.80.

También otra forma de obtener una valoración de la confiabilidad de una prueba es mediante el método test retest, el cual consiste en aplicar la prueba en diferentes momentos a la misma muestra de sustentantes, para observar las fluctuaciones o variaciones en los resultados de estos. Robins (citado en Ezpeleta, De la Osa, Domenech, Navarro y Losilla, 1997) explican que: “La fiabilidad test-retest es una de las formas más rigurosas de medir la fiabilidad porque evalúa la estabilidad de la medida en el tiempo a pesar del cambio” (p. 530).

Una de los motivos por los cuales se opta por este método de evaluación de la confiabilidad de una prueba es la que comenta Serra y Peña (2006) al decir que normalmente se utiliza este método por la simplicidad y el bajo coste que implica.

La valoración de confiabilidad mediante el método test retest se hace con el cálculo del coeficiente de correlación intraclase (CII) conocido también como índice de concordancia (Mandeville, 2005), los valores aceptados para este coeficiente van de 0 a 1, donde entre más cercano a uno significa un mayor grado de acuerdo, según la escala de Landish y Koch (1977).

Como se ha analizado en esta revisión de las distintas investigaciones, la mayoría de los autores contemplan como elementos básicos de validez de las pruebas las características de dificultad y discriminación de los ítems, con sus respectivas variantes como se veía en el caso de Backhoff et al. (2000) y Medina et al. (2019). Se observa también una concordancia en relación a que los atributos de confiabilidad que tienen uso en la evaluación de la calidad de una prueba son el índice de consistencia interna conocido como KR20 y el Alfa de Cronbach, aunque también se utiliza el método del test retest mediante el cálculo del CII, aunque cabe destacar que la mayoría de las investigaciones que aplican este método están asociados a estudios de carácter médico o de pruebas clínicas con pacientes.

## **2.6 Tecnología, evaluación y pruebas**

En este siguiente grupo la revisión trata sobre aquellas investigaciones que abordan la inclusión de tic en procesos de evaluación y evaluación estandarizada del aprendizaje. De igual manera se señalan que teorías o metodologías se han desarrollado para diseñar software con fines educativos, y como se evalúa la calidad de estos. Por otra parte, también se analizan

que software existen, con la particularidad de tener como fin específico el análisis psicométrico.

### **2.6.1 *Software educativo***

Con la socialización del internet y el acceso a computadoras (Castells, 2010), la incursión de plataformas tecnológicas en contextos educativos no tardó mucho en surgir, ya que la tecnología permitió acceder a información de formas distintas a lo tradicional, facilitando así la consulta libre, así como generar representaciones virtuales o de contenido enriquecido de casi cualquier tema que se deseara. Todo esto encontró también un espacio de oportunidad para el ámbito educativo, con la creación de las primeras herramientas tecnológicas diseñadas específicamente con fines educativos.

En términos conceptuales varias investigaciones tienen similitudes (Vidal, Gómez y Ruíz, 2010; Cataldi, 2000; Careaga, citado en Almaguel, Álvarez y Pernía, 2016) en la definición de software educativo, agrandes rasgos se les ubica como programas computacionales que facilitan la enseñanza y el aprendizaje, algunos los señalan como facilitadores del aprendizaje, o que solo pueden recibir este termino de software educativo cuando el programa incorpora una intencionalidad pedagógica, que normalmente está orientado o apuntado a un objetivo de aprendizaje.

Como se observa, aunque las investigaciones antes mencionadas se encuentran registradas en un lapso de 15 años, concuerdan en que la característica principal es la inclusión de aspectos pedagógicos propios del proceso de enseñanza y aprendizaje mediados o representados por un programa o aplicación. Es interesante como incluso se asocia directamente con la calidad, al comentar que es una herramienta a la que se le atribuye mejora en la calidad del proceso de enseñanza – aprendizaje (Gómez, 1998).

A este respecto autores como Couturejuzón (2003) y Reyes, Fernández y Duarte (2015) comentan que el software educativo es un producto tecnológico diseñado para apoyar procesos educativos, el cual permite al docente y al estudiante alcanzar de una mejor manera las metas de aprendizaje establecidas. Además de ser un medio de amplia difusión para representar y desarrollar prácticamente cualquier contenido educativo, haciendo las veces de un libro, revista, video, dibujo, diagrama etc. con su propia representación y formato digital.

Lo que permite alcanzar una representación dinámica de todos los objetos, tales como figuras, diagramas o gráficas, así como de sonido para estimular o propiciar una mejor relación entre lo concreto y lo abstracto.

Vidal, Gómez y Ruiz (2010) también señalan que para que un software sea considerado educativo, este debe cumplir con un principio básico, el cual consiste en integrar de forma casi homogénea tres ciencias, la computación, la pedagógica y la ciencia en cuestión para la cual se desarrolla el software.

Como se puede observar, las investigaciones antes citadas consideran como punto de especial interés la capacidad del software para representar casi cualquier contenido en formato digital, lo que facilita la exposición al estudiante a diferentes contenidos que pueden ser enriquecidos con objetos multimedia, simulaciones, videos, entre otros. Y que a su vez se asocia este tipo de elementos como un aporte a la calidad del proceso de enseñanza.

Lo anterior se concluye, al considerar por ejemplo que este tipo de tecnologías facilitan el almacenamiento de información, el acceso a la misma, una mejor y más oportuna comunicación entre los diferentes actores que participan del proceso de enseñanza – aprendizaje, eliminando las barreras del espacio y el tiempo, permitiendo de esta manera a los estudiantes interactuar con objetos virtuales que tal vez en la realidad no pudieran acceder a ellos.

Si bien existe una variedad de software de características educativas, es de particular interés encontrar aquellos que utilicen estas tecnologías para evaluar el aprendizaje. A este respecto se encuentran investigaciones como la de López, Hernández y Farrán (2011) quienes hablan sobre una plataforma de evaluación automática con una metodología efectiva para la enseñanza/aprendizaje en programación de computadores. González (2006) habla sobre el desarrollo de un software educativo para el autoaprendizaje de bases de datos en asignaturas de nivel superior, con la capacidad de incluir de un módulo para evaluar aprendizaje. Estas investigaciones son ejemplos de software educativo con capacidad o funciones de evaluación dentro de los mismos.

Los hay también software con fines específicos de evaluación del aprendizaje, por ejemplo, Tirado, Backhoff y Larrazolo (2016) hablan de como los exámenes más acreditados



elaborados por el Educational Testing Service, como son el Test of English as a Foreign Language (TOEFL iBT) y el Graduate Record Examination (GRE) son presentados o ejecutados en plataformas digitales. Referente a pruebas a grande escala o estandarizadas, se tiene por ejemplo la evaluación del Programm for International Student Assessment (PISA), que también ha migrado su presentación a formatos digitales. Y en el caso particular de México, desde 1993, el Examen de Habilidades y Conocimientos Básicos (EXHCOBA) se ha aplicado en formato digital como instrumento de selección para los aspirantes a instituciones universitarias de educación media superior y superior.

Siguiendo con la línea de software evaluativo en México, se tiene el Examen de Conocimientos Básicos, el cual posee características más complejas que solo representar preguntas de opción múltiple, sustentado en tres soportes teóricos: la psicometría, teorías cognitivas y ciencias de la computación. Con lo cual se permite representar de manera digital reactivos estructurales constructivos, conocidos así porque la actividad a evaluar apela a un campo semántico del conocimiento, y no a un contenido puntual (Tirado et al, 2016).

El software educativo tiene fines educativos en cualquier elemento del proceso de enseñanza – aprendizaje, ya sea como herramienta para la representación de conceptos, acceso ágil al almacenamiento y consulta de información, aumentar la comunicación e interacción, minimizar costos, riesgos y peligros mediante simuladores, y desde luego aprovechar estos recursos para evaluaciones, de carácter ordinario o estandarizadas.

### **2.6.2 Evaluación del software educativo**

Este punto en particular ha sido investigado y puesto a discusión por varios autores, especialistas en educación, los cuales han generado como resultado una serie de normas, sugerencias o instrucciones con las cuales se puede evaluar un software en términos pedagógicos y técnicos (Martínez y Sauleda, 1992; Gómez, 1998; Marquès, 1998; Squires y McDougall, 1997; Navarro,1999; Galvis, 2000).

Es comprensible que en la medida que se ha difundido el uso de estas tecnologías en los contextos educativos, ha sido proporcional el interés por valorar realmente cual es el aporte significativo que tienen en relación al proceso de enseñanza aprendizaje. A este respecto Cataldi et al (citado en Reyes, Fernández y Duarte, 2015) dice: “La proliferación de estos

materiales de apoyo educativo lleva consigo la necesidad de evaluar su calidad pedagógica y su pertinencia con el entorno en el cual se van a utilizar” (p.187).

Es importante destacar que la evaluación de un software educativo depende de varios factores que determinan la misma, por ejemplo, el fin pedagógico deseado, características de los usuarios finales del software, infraestructura tecnológica (velocidad de internet, hardware e incluso otros softwares, por mencionar algunos) requerida para el correcto funcionamiento. Cova, Arrieta y Riveros (2008) en relación a lo anterior comentan que el software es valorado en razón de las funciones derivadas para el docente y de las interacciones que ocurran en el aula causadas directamente por el programa. Razón por la que en la evaluación deben incluirse las opiniones directas de los usuarios finales en el ambiente directo de uso, que en este caso se refiere al contexto de aprendizaje, además de las opiniones de los desarrolladores del programa (Cataldi, 2000). Cova, Arrieta y Riveros (2008) consideran de manera sintetizada que la evaluación del software educativo debe responder a dos interrogantes: “¿qué tipos de experiencias de aprendizaje puede suministrar o apoyar la aplicación? Y ¿con que enfoques de la enseñanza se identifica el software?” (p. 50).

Resulta notorio que los autores revisados consideran que un elemento crucial en la evaluación del software es la calidad pedagógica y la validez en el contexto final de aplicación, y que estos procesos de valoración deben considerar aspectos como el tipo de interacción e incluso las condiciones de infraestructura con las que se cuenta, y de esta forma determinar de una manera más acertada la calidad del software educativo.

### **2.6.3 Modelos de evaluación de software**

Se ha revisado como es que para ciertos autores la evaluación del software educativo debe responder a las experiencias de aprendizaje que aporta y el enfoque de la enseñanza con la que se sustenta. Para esto y en pro de normalizar los procesos de evaluación de software se han desarrollado diferentes modelos de evaluación del software. Aunque existen muchos modelos y siguen surgiendo (esto en razón de los avances tecnológicos en las ciencias computacionales, las nuevas capacidades de procesamiento, representación de imágenes, realidad aumentada, etc.) la revisión se centra solo en aquellos que han tenido mayor difusión tanto en la explicación de la evaluación como en la puesta en práctica.

A este respecto se mencionan los trabajos de Clarke, Pete y Naidoo (citado en Cova et al., 2008), que toman como eje central de evaluación tres dimensiones específicas: la pedagógica, la matriz de evaluación (con la cual se evalúan la conveniencia para el contexto de enseñanza, viabilidad, confiabilidad y validez), y la usabilidad del software (interface de interacción con el usuario).

También se encuentran las aportaciones de Poole (1999) que en su modelo establece lo que se conocen como Listas de control, estas listas poseen una serie de indicadores de calidad tanto de la parte educativa como de la parte técnica del software, estas listas de indicadores son desarrolladas considerando las necesidades individuales de la escuela y son especificados por docentes, desarrolladores de software y asesores expertos en el área.

Otra metodología es la planteada por Marqués (citado en Cova et al, 2008), en esta se presenta la Ficha de Catalogación y Evaluación Multimedia de los Programas Educativos, la cual:

recoge los rasgos fundamentales del software (título del material, temática, objetivos didácticos, contenidos tratados, educativo, destinatarios-edad, tipología, estrategia didáctica, función, mapa de navegación, valores que potencia, requisitos técnicos y otros) y las valoraciones sobre los aspectos funcionales (eficacia, facilidad de uso e instalación), técnicos y estéticos (entornos: audiovisual y multimedia, calidad de contenidos) y pedagógicos (capacidad de motivación, adecuación al usuario, recursos didácticos que utiliza, esfuerzo cognitivo que exigen sus actividades). (p.54).

Siguiendo con la revisión, se encuentra el modelo de evaluación de software educativo bajo enfoque sistémico de Díaz, Pérez, Mendoza y Grimán (2003). El modelo se basa en aplicar una serie de cuestionarios que permiten cuantificar la calidad en tres criterios principales: la usabilidad, la funcionalidad y la confiabilidad, que a su vez permite medir de forma general la calidad del programa en tres categorías: baja, intermedia o avanzada. Al igual que el modelo de Poole (1999), este modelo incluye la participación de especialistas en áreas de ciencias de la computación, además de docentes y estudiantes.

En una última revisión se encuentra un modelo propuesto por Straccia, Zanetti y Pollo (2019), este modelo toma como referencia los estándares de la Organización Internacional de Normalización (ISO por sus siglas en inglés) ISO/IEC 9126:2001, ISO/IEC 14598:1999 e ISO/IEC 25000:2005 actuales con los cuales se evalúa la calidad de un software y se agregan aquellos aspectos distintivos del software educativo, como lo son atributos para medir el

aprendizaje, la enseñanza, condiciones referidas a los destinatarios y contenido. Cada uno de estos atributos se mide con una serie de indicadores específicos.

Recopilando los aportes de los autores revisados en este apartado se puede notar que se tiene una fuerte influencia por incluir a profesionales en el área de las ciencias computacionales en los modelos de evaluación del software, lo cual tiene sentido puesto que hay muchos elementos programables que por su naturaleza de construcción pueden ir o no alineados a funciones pedagógicas que la necesidad educativa demande, razón por la que solo expertos en el área puede determinar si son factibles o no de incluir como requerimientos de funcionalidad o usabilidad en el software educativo.

También se observa que hay un común denominador en los autores al momento de clasificar la calidad del software mediante atributos observables, ya que los reducen a dos aspectos en concreto. El primero que tiene que ver con la parte educativa (llamada pedagógica en otros autores), y el segundo que comprende la parte tecnológica que (llamada técnica por otros autores) en algunos casos los autores la subdividen en 2 o más atributos de calidad.

#### ***2.6.4 Ingeniería y metodologías para el desarrollo de software educativo***

En relación a este punto, existen procesos propios de las ciencias de la ingeniería con las que se hacen aportes para el desarrollo de software educativo. Aquí se hace mención a unas cuantas propuestas para la mejora del desarrollo de software educativo producto de investigaciones correspondientes a la ingeniería de software y a las metodologías de desarrollo de software.

Con la incursión de la tecnología en los ambientes educativos, el desarrollo de software para estos fines representaba un reto. La percepción que se tenía del desarrollo de software educativo es captada por lo que dice Cataldi, Lage, Pessacq y García (1999) quienes explican:

Uno de los problemas más importantes con los que se enfrentan los ingenieros en software y los programadores en el momento de desarrollar un software de aplicación, es la falta de marcos teóricos comunes que puedan ser usados por todas las personas que participan en el desarrollo del proyecto informático para aplicaciones generales. El problema se agrava cuando el desarrollo corresponde al ámbito educativo debido a la total inexistencia de marcos teóricos interdisciplinarios entre las dos áreas de trabajo. (p. 185-186).

Se puede notar como, aunque existía poco desarrollo o propuestas para desarrollar software con fines educativos, ya se hacía notar la necesidad y se vislumbraba un posible crecimiento de este tipo de desarrollos. Las primeras aproximaciones o marcos de desarrollo de software educativo eran limitados, por ejemplo Cataldi et al (1999) proporcionan una explicación de los diferentes modelos o metodologías de desarrollo de software (señalados también como paradigmas de desarrollo) con la intención de que el diseñador del software empleara alguna de estas metodologías de acuerdo al mejor acoplamiento o facilidad de uso en relación al tipo de proyecto (software educativo) a realizar, los recursos disponibles etc. Se analiza como de acuerdo a la teoría educativa y el currículum, se deberá adaptar algunos de los paradigmas del ciclo de vida, discriminando en cada etapa las actividades a realizar con la documentación, las técnicas y herramientas a utilizar según la necesidad pedagógica en cuestión.

Lo anterior deja ver que, en su mayoría, con lo que se contaba en esos años era con una mediana adaptación de la ingeniería de software con la producción educativa, ya que solo se limitaba a dar una variedad de modelos o metodologías, y seleccionar aquella que se adecuara más a las necesidades del proyecto, pero no existía un marco referencial diseñado especialmente para desarrollar software de este giro.

Autores como Marqués (1995) proponían una metodología para desarrollar software educativo, pero esta se concentraba en el aparato pedagógico como principal fuente, en otras palabras, era adecuar el software a un modelo pedagógico, mientras que como se explicó anteriormente la ingeniería de software de esos años era lo contrario, adecuar una necesidad educativa a un modelo de desarrollo de software.

Sin embargo, con el pasar de los años, el avance de las ciencias computacionales y la llegada de internet (detonante en muchos otros aspectos tecnológicos y metodológicos) surgieron inevitablemente más investigaciones y trabajos en la búsqueda de crear un modelo o marco referencial que integrara un modelo para desarrollar software educativo de forma nativa.

Con el surgimiento de metodologías ágiles de desarrollo de software como lo son Programación Extrema (XP por sus siglas en inglés), CRYSTAL, SCRUM, surgieron propuestas metodológicas que integraban estos nuevos modelos en la ingeniería de software

con fines educativos. Así surgieron investigaciones como la de Orjuela y Rojas (2008), quienes proponían un modelo basado en siete etapas propias o de carácter general para cualquier desarrollo de software con base en una metodología ágil, pero a su vez cada una de estas etapas incluía ahora de forma nativa elementos propios del diseño educativo.

Así, por ejemplo, la etapa de planeación incluía un proceso de aspectos educativos, en la etapa de diseño se agregaban tres procesos más: diseño educativo, diseño de comunicación y diseño computacional; con los cuales se consideraba y observaban los elementos pedagógicos que el software requería en paralelo a las necesidades de usabilidad. Y lo mismo sucedía con las restantes cinco etapas del ciclo de desarrollo del software.

Con la creciente difusión de estas metodologías y la exposición social en masa de software gratuito en internet con fines educativos, el acceso a sistemas o plataformas para desarrollar software personal y no de grado industrial, surgieron más propuestas metodológicas para desarrollar software, pero ahora con un enfoque particular a las características particulares del contexto educativo para el cual se intervenía el desarrollo del software.

En apartados anteriores se observaba como autores que investigaron sobre evaluación de software educativo señalaban que un elemento importante era tener presente siempre las necesidades particulares de la institución, pero con los avances tecnológicos y metodológicos, las posibilidades fueron más allá permitiendo desarrollar software para situaciones o momentos particulares del proceso de enseñanza- aprendizaje, mediante el desarrollo de metodología adaptables a contextos y situaciones particulares. En este sentido se recopilan aportaciones como las de González (2006); García, Vite, Navarrate, García y Torres (2016); Madariaga, Riviero y Leyva (2016) y Esterkin y Pons (2017).

### 2.6.5 *Software psicométrico*

Con relación a esta temática se rescatan aquellas investigaciones sobre software con fines psicométricos o con capacidades psicométricas. Se entiende por capacidades psicométricas a aquel software que puede realizar por lo menos un análisis psicométrico básico de una prueba, tales como los revisados en apartados anteriores, calcular el índice de dificultad de la prueba, índice de dificultad del ítem, índice y coeficiente de discriminación del ítem, índice de consistencia interna KR20, Alfa de Cronbach entre otros.

Existen investigaciones que hacen uso de software especializado, tales como las de Cechova, Neubauer y Sedlacik, 2014; Thoe, Fook y Thah, 2008 en las que se utiliza el software ITEMAN de la compañía *Assessment Systems Corporation*. También existe el software Test Analysis Program (TAP) publicado por Brooks y Johanson (2003). Algunos otros trabajos de investigación (Backhoff, Larrazolo y Rosas, 2000) recurren al uso de software estadístico Statistical Package for the Social Sciences (SPSS) propiedad de IBM. También hay quienes (Marr, Gupchup & Anderson, 2016) implementan otro software conocido como R Studio, de distribución libre, con el cual se pueden efectuar análisis estadísticos.

Sin embargo, después de revisar las investigaciones o aportaciones sobre este campo, se hace notar la ausencia de software con capacidades de análisis psicométricos que estén integrados a su vez en software con fines educativos, sobre todo en aquellos que están diseñados con fines evaluativos específicamente. Esto señala que, primeramente, no todos los software poseen la capacidad de generar resultados en un formato o estructura base que sirva de insumo para software especializado para análisis psicométrico como es el caso de TAP y de ITEMAN, y en el caso de los software que se han revisado, no se integra en su diseño o funciones evaluativas la posibilidad de efectuar un análisis psicométrico en el mismo sistema, para esto recurren a un software de terceros, que en la mayoría de los casos que se analizaron necesitan un archivo fuente en un formato específico y particular.

### **3. REFERENTE TEÓRICOS**

#### **3.1 Evaluación del aprendizaje**

La evaluación puede definirse como aquel instrumento que permite verificar en términos de calidad que lo que se enseña, se enseña bien, pero no se limita solo a la interacción docente-alumno, sino que también aplica para contextos organizacionales, es decir a nivel de instituciones educativas, a las cuales se les mide la calidad de las acciones que realizan para ofrecer una educación de calidad.

Para Biggs (2004) el qué y cómo aprenderán los alumnos está influenciado en gran manera por la evaluación que se les realizará. Es común la creencia que los objetivos curriculares son el eje principal de la enseñanza, sobre todo desde la perspectiva del docente; sin embargo, hay autores como Ramsden (en Biggs, 2004) que aseveran que la perspectiva del estudiante dista de la percepción anterior, el alumno tiene la tendencia a aprender solo aquello que él considera será evaluado o bien representado en un examen. En este sentido, se expresa que la buena evaluación, consiste por lo tanto en asegurar que la evaluación este alineada al currículo.

La evaluación del aprendizaje puede clasificarse desde diferentes líneas y con base en una diversidad amplia de criterios; por ejemplo, el momento en que se evalúa (inicial o diagnóstica), por el actor que la realiza (autoevaluación, heteroevaluación, coevaluación, etcétera); por la forma en que esta se realiza (formativa, dinámica, estática, sumativa) o por el enfoque como lo es la evaluación autentica o alternativa centrada en los procesos (Izquierdo, 2008; Zúñiga, Solar, Lagos, Báez y Herrera, 2014).

Para efectos de este proyecto, la evaluación se aborda desde la clasificación particular de la forma en que esta se realiza, siendo principalmente la evaluación formativa y la evaluación sumativa descritas por autores como Rosales, 2014; Jaime Labarca, Catalina Figueroa, Bárbara Huidobro, Ana Cecilia Wright, Arnoldo Riquelme, Rodrigo Moreno 2014; De la Orden y Pimienta, 2016.



### **3.1.1 Evaluación Formativa**

La evaluación durante el proceso de aprendizaje o formativa es un concepto introducido en el año 1967 por M. Scriven (citado por Rosales, 2014) para referirse a los procedimientos utilizados por los profesores cuyo fin es adaptar su proceso didáctico a los progresos y necesidades de aprendizaje observados en sus alumnos. Esta evaluación se realiza durante el desarrollo del proceso de enseñanza aprendizaje para localizar las deficiencias cuando aún se está en posibilidad de corregirlas, es decir, hacer adecuaciones durante la ejecución del proceso que permitan encausar de mejor manera el logro deseado o planeado para el alumno, a saber, el dominio de conocimiento o aprendizaje esperado (Zuñiga et.al., 2014).

La evaluación formativa se centra en aquellas partes significativas del programa de estudio; propiciando una toma de decisiones a tiempo; la eficacia de éstas son el resultado de la riqueza de la información y el estímulo al trabajo en función del éxito. La evaluación formativa o continua no debe basarse únicamente en pruebas formales, sino que debe incluir la observación de la actividad y el análisis de tareas (Izquierdo, 2008). Esto significa que el proceso de la evaluación debe centrarse no solo en actividades específicas sino, en buena medida, en la actividad ordinaria del salón de clases, tales como: ejercicios, solución de problemas, trabajos, dibujos, ensayos, redacciones, lecturas, y esquemas, entre otros; con esto se permite recoger información no sólo sobre el resultado, sino también sobre el proceso mismo, permitiendo conocer mejor al alumno y con ello adecuar el trabajo pedagógico (Rosales, 2014).

Es así como la evaluación formativa tiene como fin regular el proceso de enseñanza-aprendizaje para posibilitar que los medios de formación respondan a las características de los estudiantes (De la Orden y Pimienta, 2016). Con este tipo de evaluación se busca identificar los puntos débiles en el aprendizaje, por encima del resultado obtenido con la acción de aprendizaje. Así mismo esta evaluación permite retroalimentar tanto al alumno como al docente acerca del desarrollo del proceso enseñanza aprendizaje, así como el grado de avance en el logro de los objetivos planteados en el curso.

Resulta claro también observar como esta evaluación se complementa con otro tipo de evaluación, con esto hacemos referencia a la evaluación de carácter sumativa, en la cual se

pueda constar qué fue lo que se aprendido durante todo el curso, y determinar hasta qué punto se aprendió bien. La anterior situación debe ser en cierta manera el reflejo de los resultados obtenidos con la evaluación formativa, es decir, debe existir una correlación significativa entre los resultados de ambas evaluaciones, de lo contrario se observaría entonces defectos en las evaluaciones, ya sean las formativas o la sumativa.

### ***3.1.2 Evaluación Sumativa***

La mayoría de los exámenes finales, de cierre de unidad, o departamentales como lo es el caso de este proyecto de investigación cumplen con la función de ser exámenes o evaluaciones de carácter sumativa. Se les llama así puesto que los resultados se utilizan para comprobar hasta qué punto los estudiantes han aprendido bien lo que se supone que han aprendido, y por esta razón es que se realiza después de concluir el episodio de enseñanza (Biggs, 2004). Así mismo este tipo de evaluaciones cumplen con funciones tales como selección, promover grados, expedir certificados entre algunos otras; por lo que normalmente se les considera de alto impacto, esto en razón del tipo de consecuencias que tienen dichos exámenes para el sustentante.

Dentro de esta clasificación de evaluaciones sumativas, existen a su vez dos modelos principales conocidos como el modelo de medida y el modelo de niveles. En el caso de del primero, este está formulado para captar características estables de los individuos, cuyo fin radica en poder efectuar comparaciones entre los miembros de una misma población, o bien con una norma de carácter general para dicha población. A este tipo de evaluación también se le conoce como evaluación referida a normas.

El modelo de niveles está diseñado para evaluar el logro alcanzado o la mejora del rendimiento académico en relación al aprendizaje adquirido. Esta evaluación también es conocida como evaluación referida a criterio (Biggs, 2004; Heredia, 2009; Jornet y González, 2009). En palabras de Cruz y Quiñones (2012) su propósito es tomar las decisiones pertinentes para asignar una calificación totalizadora a cada estudiante que refleje la proporción de objetivos logrados en el curso. Su función es explorar en forma equivalente el aprendizaje de los contenidos incluidos logrados en los resultados de forma individual.

Por medio de esta evaluación se trata de corroborar lo que ha sido alcanzado por el estudiante; situación que no representa nada nuevo para el docente y estudiante puesto que, al llegar a la evaluación sumaria, cuentan ya con suficientes datos obtenidos de las evaluaciones formativas que les harán vislumbrar lo que pueden esperar de la evaluación sumaria. Si en el momento de la evaluación los resultados fueran inesperados, habría que desconfiar de la validez de las evaluaciones formativas o de la atención que se prestó a éstas para hacer los reajustes necesarios. Lo anterior desde luego tendrá mayor validez si la evaluación sumativa tiene un fundamento sólido establecido por una metodología rigurosa que permita comprobar la calidad de la prueba en términos psicométricos, es decir, que realmente evalué aquello que se supone debe evaluar.

Un ejemplo de este tipo de evaluaciones sumativa son las pruebas departamentales (el proyecto de investigación parte del diseño y aplicación de una prueba de esta naturaleza), dichas pruebas están planeadas como un instrumento con el cual medir de manera directa lo que el estudiante ha aprendido y hasta qué punto (nivel de dominio) lo ha aprendido. De esta forma se puede obtener información que permita determinar a cierto nivel de detalle cuales actividades, estrategias o instrumentos evaluativos formativos fueron eficaces o bien deficientes como se mencionó anteriormente.

Por otro lado, como explica Biggs (2004), el estudiante tiene la tendencia a tomar como importante en su aprendizaje aquello que considere le será evaluado o que tendrá un impacto cuantificable en su historial académico, entiéndase lo anterior como una calificación parcial, semestral, la acreditación o no de un curso. Implementar una prueba referida a un criterio permite orientar de una manera objetiva que aquellos conocimientos que el currículo establezca estén debidamente representados en dicha prueba, que esta no sea producto de un juicio personal o apreciación única del docente, es decir, que el Maestro sea quien determine con base en su experiencia lo que considere debe venir representado o no en el examen, en términos de conocimiento.

Claro está, la función del docente en el aspecto concreto de la evaluación se encuentra dada, entre otras cosas, por la capacidad que se tiene de establecer hasta qué punto han aprendido con un grado de dominio adecuado aquello que se tenía previsto, pensado, planeado o deseado que tenían que aprender. Lo anterior nuevamente pone la atención en como

determinar qué es lo que se supone debe aprender el estudiante, la respuesta a lo anterior está dado por el universo de conocimientos que contempla la carta descriptiva o el currículo de la asignatura para la cual se desarrolla la evaluación.

Lo anterior expresa la intención directa de estimular al estudiante a lograr un aprendizaje más significativo y consciente, puesto que desarrollar un examen que refleje de manera eficiente el contenido de un currículo, estableciendo índices de relevancia curricular para cada uno de los temas incluidos, permite dirigir esfuerzos a aquellos temas que tienen un mayor impacto para el estudiante y que son esenciales para que pueda concretar conocimientos futuros. Y a su vez el estudiante recibe estimulación por dominar aquellos conocimientos que aun no siendo de alta relevancia curricular si tienen una representación en la prueba, aunque evidentemente en menor proporción.

Otro aspecto particular de la evaluación del aprendizaje mediante pruebas de carácter criterial son las siguientes:

- Los resultados de la evaluación se expresan en términos del grado con que la actuación de un individuo concuerda con los niveles de dominio esperados, y son independientes a los resultados de cualquier otro individuo.
- A diferencia de la evaluación referida a normas donde se emiten juicios sobre personas, la evaluación referida a criterios emite juicios sobre la actuación.

Analizar el aprendizaje desde una perspectiva cuantitativa según Biggs (2004) implica lo siguiente:

1. Las actuaciones de aprendizaje se unifican. Si la actuación que hay que evaluar deja de ser cuantitativa, esta se valora asignando categorías de manera subjetiva o bien arbitraria.
2. Estas unidades son objeto del error o bien del acierto en su definición, integración o concepción.
3. En caso que las unidades sean correctas, pueden ser tratadas de manera aditiva, es decir que la suma de cada una de ellas establece o crea el índice de aprendizaje, el cual es reflejado en una única escala.

4. Lo anterior resulta entonces que cada unidad que se acierte debe valer lo mismo que cualquier otra unidad, restando importancia a lo que esté bien, mientras haya un número suficiente de aciertos.

Sin embargo, uno de los aportes significativos de este trabajo de investigación radica en determinar cuantitativamente que las unidades (actuaciones de aprendizaje) sean correctas, y tengan un sustento sólido que avale dicha conclusión. Y, por otro lado, que se priorice y rescate la importancia de lo que se acierta bien, es decir, no solo se trata de contestar correctamente una pregunta en la prueba, sino saber que tanta relevancia curricular e importancia tiene para el estudiante responder correctamente a esa pregunta.

Esto último agrega peso a que no solo se trata de observar la calificación o índice obtenido al concluir la prueba, sino que más bien el número no refleja en esencia todos los matices de la actuación de aprendizaje. Por poner un ejemplo, un estudiante puede obtener una calificación final en un examen departamental de 85 sobre 100. En términos generales podríamos decir que el estudiante domina satisfactoriamente un determinado universo de conocimientos. Pero si pudiéramos observar cuanta relevancia curricular tienen los reactivos que no acertó, y se identifica que estos eran los de mayor relevancia curricular en la asignatura, que son conocimientos que se requieren para concretar otros aprendizajes en asignaturas siguientes. Es claro que entonces ese 85 ya no se puede observar de la misma manera, y la calificación oculta deficiencias de aprendizaje que trascienden o tienen un alto impacto.

Este tipo de detalles son muy significativos al momento de evaluar el aprendizaje de un estudiante y emitir un juicio al respecto. Se busca zanjar estas situaciones empleando metodologías que permitan diseñar pruebas las cuales se fundamenten o estructuren con base a criterios que permitan identificar casos como el ejemplo antes mencionado. Diseñar pruebas que tengan una correcta representación de conocimientos en relación a la carta descriptiva o currículo de la asignatura, así como una distribución de reactivos que permitan representar de manera efectiva los contenidos temáticos de mayor índice de relevancia curricular, con previo juicio y justificación de los mismos, evitando así la subjetividad o arbitrariedad.

Para lograr esto último, existen metodologías para el diseño de pruebas que forman o integran comités en la ejecución de diferentes etapas, con la característica particular de que estos comités se integran por docentes con características especiales como, por ejemplo, ser expertos en la asignatura, expertos en el plan curricular, expertos en la docencia entre otros. Esto con el fin de contar con un proceso de jueceo más objetivo y representativo de la realidad al contar con la opinión de diferentes expertos en la materia (Contreras, 2000; Contreras y Backhoff, 2004; Contreras y colaboradores, 2005).

El proyecto de investigación presentado en este trabajo se centra en el aspecto particular de la evaluación sumativa, bajo el modelo de niveles o evaluación referida al criterio. Se parte del diseño de una prueba departamental estandarizada, la cual mediante una aplicación innovadora permite efectuar además de la prueba un análisis psicométrico con el cual determinar la validez y confiabilidad de cada uno de sus ítems, así como de la prueba en su conjunto.

La evaluación de los aprendizajes funge como base fundamental para la observación de los aciertos y errores en las estrategias empleadas para lograr la adquisición de conocimientos en los estudiantes y con ello tomar decisiones en pro de su formación educativa. La evaluación del aprendizaje es el ejercicio educativo que da razones sobre el qué, cómo, por qué y cuándo se deben de enseñar; la evaluación del aprendizaje está estrechamente relacionada con las decisiones que se efectúan durante los procesos de enseñanza y aprendizaje (Bogantes, 2015).

En el ámbito de la educación superior la evaluación se explica como el producto de una serie de interacciones entre diferentes tipos de personas en un espacio y tiempo determinados, interacción que particularmente ofrece un marco referencial sobre el cual los docentes orientan las prácticas de evaluación (Contreras, 2010).

Ofrecer evaluación de calidad requiere de cumplir con 2 aspectos principalmente. En primer término, que técnicamente debe ser válida y confiable, y en segundo término que siempre debe estar orientada a la mejora (Murillo y Román, 2016). De manera general en una prueba diseñada para evaluar el aprendizaje su calidad puede ser observada mediante la

confiabilidad y validez que está presente, estos dos términos son muy importantes al momento de hablar de pruebas.

La confiabilidad de una prueba hace referencia a la consistencia de las puntuaciones obtenidas por los mismos individuos en diferentes ocasiones o con distintos conjuntos de ítems equivalentes (Arginay, citado en Zamora 2015). También puede definirse la confiabilidad como el grado con el que las diferencias individuales en las calificaciones de las pruebas, se atribuyen a errores aleatorios de la medición y el grado con el que se asignan a diferencias reales de las características o dominio en consideración (Muñiz, citado en Zamora, 2015).

Claro está, si bien una prueba puede ser confiable, no es suficiente si esta no es válida, los juicios emitidos a partir de los resultados serán erróneos o insuficientes para tomar decisiones. En relación a lo anterior, el concepto de validez se refiere al juicio evaluativo global en el que la evidencia empírica y el fundamento teórico respaldan lo apropiado de las interpretaciones y acciones con base a los puntajes obtenidos en la prueba. Y que, agregado a esto, no se limita solo a las preguntas de la prueba, sino también a cómo responden y al contexto de la evolución (Aliaga ,2007).

Evaluar el aprendizaje implica más que solo aplicar una prueba, sin embargo, para efectos de esta investigación el objeto de estudio se limita a evaluar aquellos atributos del aprendizaje en los estudiantes que pueden ser cubiertos con una prueba escrita, entendiéndose esta como un conjunto de ítems de opción múltiple que abarcan el contenido temático de una asignatura de acuerdo a lo que establece el currículo de la misma.

Bajo esta investigación el objetivo de evaluar el aprendizaje radica en determinar el grado de dominio que los estudiantes poseen de un determinado universo de conocimientos, o en su defecto la ausencia de dicho dominio. Para lograrlo se aplica una metodología de diseño de la prueba basada en la Teoría de Respuesta al Ítem y el análisis psicométrico de la misma. En el apartado de metodología se detalla específicamente cuales son los procedimientos que se realizan para determinar la validez y confiabilidad de la prueba.

### 3.2 Evaluación estandarizada

Las evaluaciones estandarizadas o también conocidas como pruebas a gran escala, son instrumentos de medición ampliamente utilizados en áreas como la educación, ciencias de la salud y ciencias sociales. Cuya principal característica radica en el amplio desarrollo técnico y metodológico con formas perfeccionadas para medir rasgos específicos en la población focal, estableciendo previamente un grado de precisión y control en los procedimientos logísticos y administrativos que rigen la prueba (Tristan y Pedraza, 2017).

Como parte de la estandarización también es característico el proceso de sistematización de todos aquellos elementos que están asociados a una acción de recogida e interpretación de información, de tal forma que se utilicen los mismos instrumentos o técnicas tanto para recopilar, analizar e interpretar la información (Jornet, 2017).

La evaluación estandarizada del aprendizaje es un elemento que actualmente tiene relevancia para la evaluación y desarrollo educativo, como menciona Cueto (2011):

En las últimas dos décadas el incremento de las evaluaciones estandarizadas de rendimiento escolar ... ha sido notable en América Latina ... la correlación entre el rendimiento y las posibilidades de acceder a la educación superior... y poder desempeñarse bien en otras áreas del currículo (p. 3).

Sin embargo, también se ha cuestionado si este tipo de exámenes pueden o no ser una herramienta que permita medir de manera transparente y real los logros de un determinado sistema educativo. Así mismo también se considera que un examen estandarizado solo puede medir una pequeña parte del desarrollo intelectual o la inteligencia, es decir, el aprendizaje se convierte en el producto de las interpretaciones que los propios estudiantes generan cada vez que se enfrentan a la información que les es presentada en diferentes contextos (Barrenechea, 2010).

En cuanto al uso de las evaluaciones estandarizadas habitualmente se encuentran las pruebas de gran escala o de objetivos, como lo son las pruebas de admisión a las universidades y aquellas para medir el desempeño de los estudiantes en los distintos niveles educativos.

En el caso particular de México este tipo de pruebas su aplicación se remonta a los inicios de los años 70, la Secretaría de Educación Pública empieza a aplicarlas en 1972 para decidir



la admisión de alumnos en secundaria (Martínez, 2009); pero no fue hasta más de dos décadas después que este tipo de pruebas tuvo mayor difusión, esto mediante la realización de pruebas a gran escala en la educación básica mediante el proyecto denominado Estudio de Evaluación de la Educación Primaria en 1994.

En ese mismo año se crea el Centro Nacional para la Evaluación de la Educación Superior (CENEVAL) y con ello se extendió este tipo de evaluaciones a gran escala o estandarizadas en el país. En 1998 se aplican por primera vez las Pruebas de Estándares Nacionales (Martínez, 2001).

La relevancia de utilizar este tipo de pruebas radica en utilizar un mismo sistema de acercamiento a la realidad, con el fin de las diferencias obtenidas con la medición sean en su mayoría atribuidas a la diversidad del sujeto evaluado o a intervenciones sobre ellos, y no a deficiencias en el diseño, aplicación o interpretación del instrumento con el que se le evalúa (Jornet, 2017).

Por lo anterior es importante entonces que un instrumento de evaluación pueda ofrecer datos válidos y confiables. Aquí es donde entra la Teoría de Respuesta al Ítem (TRI) al permitir obtener un número cuantificable de habilidades y conocimientos que posee un sustentante, número obtenido a partir de un análisis matemático que estima la relación funcional entre los valores de la variable que miden los ítems y la probabilidad de acertar estos (Muñiz, 2010, p. 64).

Las Tecnologías de la Información y Comunicación (TIC) ofrecen la posibilidad de innovar en ciertas prácticas o métodos que tradicionalmente se involucran en la enseñanza. La tecnología educativa debe innovarse constantemente, no solo para obtener un mayor mercado de ventas o crear una especie de moda que cambie de acuerdo a los cambios tecnológicos del hardware o software, sino más bien para cubrir las áreas de oportunidad o necesidad cambiante que se va creando conforme la sociedad va evolucionando.

Es relevante migrar las prácticas clásicas de un examen en papel a una evaluación más enriquecida a través de sistemas computacionales, y que permitan realizar análisis de resultados más ágiles.

El proyecto de investigación busca generar conocimiento sobre las mejoras y ventajas que representa contar con una aplicación tecnológica que permita la realización de exámenes estandarizados basados en la TRI y aquellos factores críticos que se deben considerar para lograr dicha mejora, factores como la elección de la plataforma de desarrollo, diseño visual, procesamiento y gestión de los datos entre otros. Se busca innovar el procesamiento de datos al tener la capacidad de analizar psicométricamente los resultados obtenidos y ofrecer de manera confiable y válida información que permita mejorar la toma de decisiones de los diferentes actores involucrados en el proceso.

Por mencionar a algunos actores y el impacto de la evaluación, se puede hablar del alumno, para quien el resultado del examen puede significar la promoción o no a un siguiente curso o semestre, mientras que para el docente puede significar inscribirse a cursos de estrategias de enseñanza, y en esferas administrativas puede significar remover a un docente de una asignatura.

Se quiere que la intervención produzca cambios significativos en la filosofía y cultura en la institución referente al uso de las TIC y que notablemente sea una mejora al proceso educativo. Producir datos cuantitativos que permitan justificar las diferentes decisiones sobre esta nueva forma evaluar dentro de la institución.

### **3.3 Psicometría**

Las pruebas o test en el campo educativo son una práctica habitual, puesto que estos tienen como propósito proporcionar información o evidencias específicas que permiten tomar decisiones o guiar intervenciones en los diferentes contextos en los que estos son aplicados. Las pruebas cumplen una importante función en relación a la evaluación y la intervención educativa, razón por la cual es sumamente importante conocer cómo funcionan, las propiedades y las condiciones necesarias para que las pruebas puedan ser utilizadas de manera adecuada (Meneses, Barrios, Lozano, Bonillo, Turbany, Cosculluela y Valer, 2014).

La psicometría es una rama de la psicología que se encarga del desarrollo de teorías, métodos y técnicas que dan soporte a los procesos de construcción y gestión de las pruebas. Uno de los objetivos más comúnmente señalados en la psicometría es proporcionar las garantías científicas necesarias para efectuar una medición objetiva y estandarizada de los fenómenos

psicológicos no observables a partir de una muestra de comportamientos (Meneses et. al., 2014). Lo anterior ha supuesto una contribución significativa a la teoría de la medición en otras disciplinas.

En relación a la revisión histórica de esta disciplina se considera que los trabajos de Laplace en 1796 precursor para la teoría de probabilidades y el de Gauss en 1798 con los fundamentos de la teoría de números fueron las bases de la teoría de la medición. Más adelante en 1907 Binet construyó el primer test estandarizado de inteligencia, el cual fue desarrollado teniendo como referencia los trabajos de Galton, Wundt, Ebbinghaus, Pearson y Spearman, quienes a su vez sentaron las bases para la Antropometría y la Psicofísica (Aiken, citado en Escurra, 2011).

También se señalan los trabajos efectuados por Thrustone en 1931 con el libro *The reability and validity of tests* en el cual se sistematizaba de manera general lo desarrollado hasta esa época en cuanto al desarrollo de pruebas. Todos estos trabajos sentaron las bases de lo que más adelante sería conocida como la Teoría Clásica de los Test (TCT) siendo Gulliksen en 1950 quién lleva a cabo la síntesis canónica de este enfoque (Muñiz, 2010).

La TCT propone un nuevo enfoque basado en el concepto de puntuación verdadera. Se parte de la puntuación empírica obtenida mediante los tests, y mediante un conjunto determinado de supuestos, el objetivo es separarla en dos partes fundamentales para valorar el error asociado al proceso de medida y, así, inferir el valor real que se pretende medir. Esta teoría estimuló el debate sobre la propia medida psicológica como el proceso de desarrollo de los tests durante las siguientes décadas (Meneses et. al., 2014).

El avance en el campo de la medición y elaboración de pruebas gracias a la TCT continuó dando así el surgimiento de otras teorías como la Teoría de la Generalización (Cronbach, Gleser, Nanda y Rajaratnam, 1972) y la Teoría de Respuesta al Ítem que en lo particular surgió como respuesta a las críticas de medición hechas a la TCT (Rasch, 1960; Lord y Novick, 1968; Lord, 1980).

### **3.3.1 Teoría de Respuesta al Ítem**

En el caso de esta teoría, esta surge en los años sesenta, al mismo tiempo que se generalizaba la idea de elaborar pruebas referidas a un criterio, que en el contexto educativo se refiere a evaluaciones cuyo objetivo es medir las destrezas de un individuo en un campo del conocimiento muy delimitado. La denominación TRI agrupa líneas de investigación psicométricas independientes iniciadas por Rasch (1960) y Birnbaum (1968). La convergencia de estas investigaciones es que establecen una relación entre el comportamiento de un sujeto frente a un ítem y el rasgo responsable de esta conducta lo que se denomina o conoce también como rasgo latente.

Para lograr esto, se recurre a funciones matemáticas que describen la probabilidad de dar una determinada respuesta al ítem para cada nivel del rasgo medido por este. El objetivo sustancial de esta teoría es la elaboración de instrumentos de medición con propiedades invariantes entre poblaciones. Si dos individuos presentan idéntico nivel de rasgo medido ambos tendrán equivalente probabilidad de dar la misma respuesta, independientemente de la población de pertenencia (Attorresi, Lozzia, Abal, Galibert & Aguerri, 2009).

En los ítems de habilidades la clave es la respuesta correcta y en los de personalidad es aquélla opción que indica la presencia de un nivel mayor de rasgo en el individuo. El gráfico de dicha función matemática se denomina Curva Característica del Ítem (CCI).

En la TRI se encuentran varios supuestos, uno de ellos consiste en asumir que la conducta de una persona ante un ítem sigue un modelo probabilístico con una forma determinada. Otro es que las respuestas a cualquier subconjunto de ítems no agregan ninguna información para el cálculo de probabilidad de respuesta a un ítem en particular. Lo anterior significa que las respuestas a distintos ítems son estadísticamente independientes (Attorresi et. al., 2009).

La TRI emplea una serie de cálculos estadísticos y matemáticos con los cuales se pueden establecer ciertas mediciones en cuanto a la calidad técnica de la prueba y a su vez en grado de acercamiento real con el que se mide determinado rasgo o habilidad en un sustentante, que para efectos de este trabajo de investigación significan dominio de conocimiento en una asignatura.

Entre los indicadores asociados a esta teoría se encuentran el índice de dificultad del ítem que permite establecer en una escala numérica en un rango de cero a uno cuán difícil es responder acertadamente un ítem. Entre más cercano se encuentre este valor el ítem es más sencillo de contestar. También se encuentra el índice de discriminación el cual indica en qué medida el ítem diferencia a los examinados con un nivel en el rasgo por encima o por debajo del parámetro de localización (Atorresi et. al., 2009).

Así mismo existe el coeficiente de discriminación que, a diferencia del anterior, es más preciso en el cálculo de la discriminación ya que contempla a todos y cada uno de los sustentantes, mientras que con el índice solo se considera al 54% de los sustentas, porcentaje distribuido en 27% de los estudiantes con más alto rendimiento y 27% de los sustentantes con menor rendimiento. cumple y el índice de confiabilidad. Este valor de discriminación se mide en un rango de menos uno a uno, donde entre más cercano a uno se encuentre el valor significa que el ítem discrimina mejor.

Otro atributo medible es el coeficiente de fiabilidad (conocido también como confiabilidad) el cual establece el grado de consistencia y estabilidad de las puntuaciones obtenidas a lo largo de continuos eventos de medición utilizando un mismo instrumento. Usualmente para el cálculo de este indicador se utilizan el Alfa de Cronbach y el coeficiente de Kuder-Richardson conocidos como KR20 y KR-21, los cuales miden la fiabilidad de un test con base en su grado de consistencia interna, indicando el grado en que los ítems de un test covarían. La consistencia interna se considera aceptable si se obtienen valores dentro del rango de .70 y .90.

### **3.4 Desarrollo de software educativo**

En la última década, la adopción de las tecnologías de la información y comunicación han afectado considerablemente las formas en las que se interactúa con el conocimiento. Utilizar software para mejorar las capacidades de los estudiantes conlleva inevitablemente que se realicen estudios para definir estándares o mejores prácticas en los procesos de desarrollo, esto con el fin de garantizar la calidad del producto.

Existen muchas metodologías de desarrollo de software que por su amplio uso han demostrado ser eficaces, sin embargo, desarrollar software con fines educativos implica

modificar ciertos procesos y agregar algunos otros. Por mencionar solo un elemento, las metodologías usuales o tradiciones de software se centran en los aspectos técnicos del producto, mientras que un software educativo tiene que considerar, aparte de este elemento que se menciona, el aspecto pedagógico, es decir ser construido tomando en cuenta aquellas estrategias didácticas en las cuales el software se ve involucrado.

Existen varias investigaciones en cuanto a metodologías de desarrollo de software educativo, como por ejemplo las de (Piñares, M., Miranda R., Rozo, F., y Lobo, M., 2009), (Arias, M., López, Á., y Honmy, J., 2015) y (Duarte, A., y Rojas, M., 2008). En general las metodologías desarrolladas en estas investigaciones muestran cómo se desarrollan cada una de las etapas del proceso de construir el software, y en particular todas incluyen una fase dedicada únicamente a los aspectos educativos.

Por ejemplo, hablando de la metodología propuesta por Piñares et al. (2009) que está basada en el enfoque de competencias, se integran cinco fases como se muestra en la figura 1.

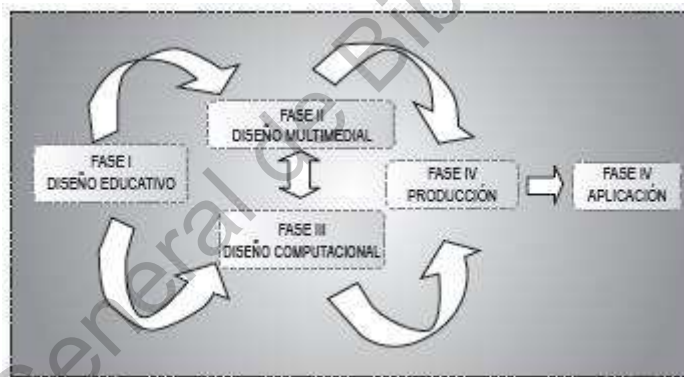


Figura 1. Modelo para desarrollar software educativo

Fuente: Piñares et al., 2009., p. 74.

La parte inicial de esta metodología se enfoca en el diseño educativo, esta etapa es de mucha importancia para el desarrollo del software, ya que como se aprecia en la figura 1, esta fase tiene repercusiones en las siguientes 2 fases. Dada la importancia de la fase, se requiere la ejecución de una serie de etapas que concluyan con el diseño adecuado de aprendizaje. Las etapas que componen a esta fase de diseño educativo se ilustran en la siguiente figura.



Figura 2. Fase 1. Diseño educativo.

Fuente: Piñares et al., 2009., p. 75.

A grandes rasgos esta fase identifica primeramente cual es la necesidad educativa que se quiere atender, para identificar las posibles causas de los problemas identificados y proponer alternativas de solución. Posteriormente en el diseño de fines educativos se especifica el tipo de conocimiento que será trabajado por los alumnos, que usualmente va identificado como objetivos. Seguido de esto, en el sistema de competencias se detalla de manera clara como se organizarán las competencias que se pretenden desarrollar con el software, esta parte abarca los objetivos de la competencia (el enunciado y sus elementos) y las normas (dimensión de la competencia que trata el contexto, recurso y evidencias de conocimientos).

Seguido de esta etapa se desprenden 2 más, por un lado, el diseño de contenidos que tiene que ver con la descripción de cada contenido que comprenda el problema estudiado, y por otro lado el diseño pedagógico, que trata el modelo pedagógico seleccionado como base para diseñar las actividades de aprendizaje que el software poseerá.

Por último, se concluye con la etapa del diseño de aprendizaje que tiene como finalidad dar a los estudiantes estrategias para mejorar el desempeño en áreas y contenidos específicos. Así se toman los elementos e indicadores seleccionados de la competencia a trabajar, se especifican las características del modelo pedagógico que el software va a incluir, de esta forma utilizando estos elementos se diseña la secuencia de aprendizaje para que los

estudiantes realicen las actividades solicitadas bajo las condiciones que el modelo pedagógico especifica.

Continuando con la siguiente fase de la metodología está el diseño Multimedial y el Computacional, en el caso del diseño Multimedial en esta fase se describen las etapas que permiten conocer la planeación de las actividades que se van a desarrollar, se utilizan representaciones gráficas y recursos multimedia que en conjunto mostraran el diseño de cada interfaz y la estructura que se implementará en el software.

Por su parte, el diseño Computacional comprende la descripción de los elementos que van a permitir que el sistema funcione con los procesos que se definieron en la fase del diseño educativo y el Multimedial, esta fase también utiliza representaciones de las interacciones que el usuario tendrá con el software y poder detallar el grado de apoyo al proceso de enseñanza aprendizaje.

La fase siguiente es la de Producción, en la cual se integran todos los productos obtenidos en las fases anteriores para que sean programados y así completar la construcción final del software. Normalmente este producto (software) es considerado un prototipo, el cual requiere de una serie de procesos de evaluación para determinar si realmente cumple con los objetivos propuestos.

Finalmente se concluye con la fase de implementación, en la cual el software se lleva al campo real de uso, es decir, se presenta a la población de estudio para lo utilicen. El fin de esto es poder comprobar con los usuarios finales si se cumplieron con los objetivos señalados en la primera fase de la metodología, a saber, las necesidades educativas identificadas.

Dentro de esta fase también se realiza el análisis de los resultados obtenidos al implementar el software, con este análisis se pretende identificar aquellas áreas que presentaron deficiencias o en las cuales no se lograron los propósitos definidos. De esta manera se pueden realizar ajustes para corregir los errores detectados y poder cumplir con los propósitos previstos.

Desde el punto de vista de Ingeniería de Software, esta metodología implementa elementos de desarrollo de software tradicional, como por ejemplo de la metodología en cascada o



espiral. Sin embargo, actualmente se utilizan otro tipo de metodologías de características más dinámicas o también conocidas como metodologías ágiles.

Las metodologías ágiles tienen como principal enfoque el factor humano, una estrecha colaboración con el cliente y la realización de desarrollos en periodos muy cortos que normalmente duran unas cuantas semanas. Además, se basan en ciertos valores bien definidos como por ejemplo desarrollar software funcional más que lograr buena documentación, colaboración con el cliente más que un buen contrato, agilidad de respuesta a cambios más bien que seguir un plan estricto (Duarte, A., y Rojas, M., 2008). A grandes rasgos una comparación entre este tipo de metodologías y las tradicionales serían las que se ilustran en la siguiente tabla:

**Tabla 1**

*Comparación de las metodologías*

METODOLOGÍAS ÁGILES	METODOLOGÍAS TRADICIONALES
Basadas en heurísticas provenientes de prácticas de producción de código.	Basadas en normas provenientes de estándares seguidos por el entorno de desarrollo.
Preparados para cambios durante el proyecto.	Cierta resistencia a los cambios.
Reglas de trabajo impuestas internamente (por el equipo.)	Reglas de trabajo impuestas externamente.
Proceso menos controlado, con pocos principios.	Proceso mucho más controlado, con numerosas políticas/normas.
Flexibilidad en los contratos debido a la respuesta a cambios.	Existe un contrato prefijado.
El cliente es parte del equipo de desarrollo.	El cliente interactúa con el equipo de desarrollo mediante reuniones en determinadas etapas del proceso.
Grupos pequeños (<10 integrantes) y trabajando en el mismo sitio.	Grupos grandes y posiblemente distribuidos trabajando en diferentes tareas.
Pocos roles.	Más roles
Menos énfasis en la arquitectura del software.	La arquitectura del software es esencial y se expresa mediante modelos.

Fuente: Duarte, A., y Rojas, M., 2008., p. 161.

Algunas de las metodologías ágiles más conocidas son Extreme Programming (XP), Crystal y SCRUM. El implementar este tipo de metodologías con propósitos educativos, al igual que la metodología analizada anteriormente, requiere de adecuaciones que permitan contemplar dentro de su marco de desarrollo los aspectos didácticos y pedagógicos del software. Un caso particular del uso de estas metodologías es la propuesta por Duarte, A., y Rojas, M. (2008), en la cual se explica la distribución de las fases del desarrollo y como la parte educativa se ve integrada en cada proceso.

Para desarrollar el producto de software en esta metodología se proponen 7 fases:

- Planeación
- Obtención de requerimientos
- Análisis
- Diseño
- Implementación
- Pruebas
- Evaluación

Partiendo de la fase de planeación, en esta se propone la integración del equipo de trabajo y las políticas que lo regularan, particularmente estas políticas deben estar fuertemente vinculadas a fortalecer los procesos de enseñanza-aprendizaje y la incorporación de las tecnologías en estos procesos. En esta fase es importante también como se dará la interacción entre los diferentes pilares del desarrollo, como se muestra en la figura 3.

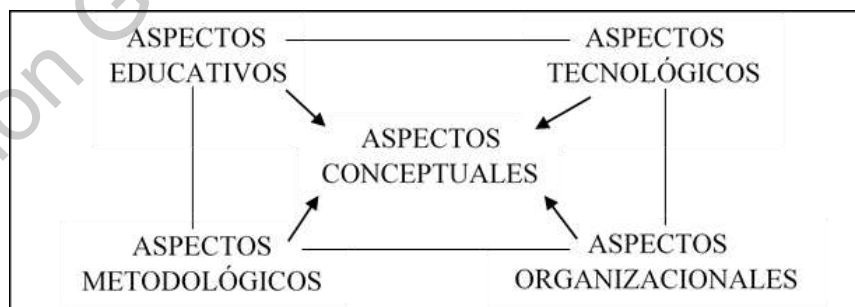


Figura 3. Pilares conceptuales.

Fuente: Adaptado de Duarte, A., y Rojas, M., 2008., p. 168.

Dentro de los aspectos educativos se debe considerar elementos como el modelo pedagógico que debe soportar el software, y como se articularan las actividades con el uso de estas tecnologías. Aquí también se especifican las estrategias didácticas a las cuales debe responder el software educativo. Con respecto a los aspectos tecnológicos se hace referencia a la descripción de software y hardware que soportan la implementación del software educativo. Los aspectos conceptuales son los contenidos que se van a transmitir con el software, en el pilar de aspectos metodológicos se describen todas las actividades y estrategias que acompañan el uso del software en la ejecución del proceso de enseñanza-aprendizaje. Por último, los aspectos organizacionales detallan aquellas actividades relacionadas con la gestión del proceso de desarrollo e implementación del software educativo.

Otra de las fases que también requiere de adecuaciones para ajustarse a un enfoque educativo de desarrollo es la de diseño. Aparte de especificar los elementos comunes de cualquier diseño de software, también se requiere que especifiquen criterios del diseño educativo. En este sentido se deben resolver las interrogantes sobre el alcance, contenido y el tratamiento que el software debe aportar. Se debe prestar especial atención al modelo pedagógico para diseñar las actividades interactivas con el software para que vayan de acuerdo a dicho modelo. En la siguiente figura se ilustra de forma general los factores a considerar dentro del diseño educativo:

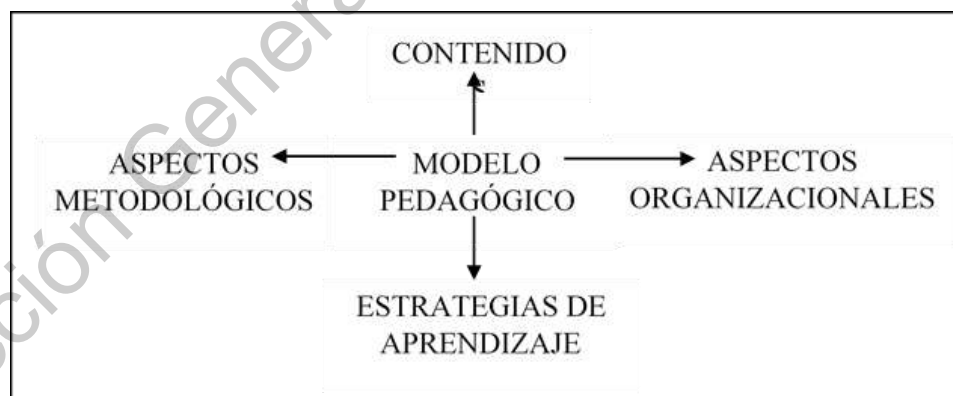


Figura 4. Elementos del diseño educativo

Fuente: Adaptado de Duarte, A., y Rojas, M., 2008., p. 168.

Otro elemento a considerar es el diseño computacional, el cual debe incluir la descripción de los usuarios con las funciones a las cuales tiene acceso, para posteriormente implementar

módulos de acuerdo al diseño más apropiado para que soporte los requerimientos funcionales, no funcionales y del propio modelo pedagógico. A grandes rasgos esta fase es la encargada de materializar los requerimientos definidos para el proyecto.

Una característica de esta metodología es que con cada ejecución del ciclo de desarrollo se obtiene como producto final un prototipo funcional del sistema, de manera que con cada nuevo ciclo se obtenga un producto con nuevas funcionalidades, esto permite que el software sea probado tanto por el cliente como por el equipo de desarrollo. Además, una ventaja de este tipo de metodología es que cada ciclo o iteración tiene una duración estimada de 30 días, ofreciendo productos tangibles y funcionales en periodos cortos de tiempo. Esta propuesta metodológica posee la característica de adaptarse a procesos de desarrollo de software educativo de diferentes niveles de complejidad, requerimientos y tamaño.

Otra metodología de reciente diseño es la propuesta por García, Vite, Navarrate, García y Torres (2016), la cual establece una serie de etapas donde se ven involucrados los aspectos de la ingeniería de software y el diseño instruccional. La metodología propuesta se compone de los siguientes elementos:

- Concepto o pre – producción:
  - En esta etapa se incluye la concepción del proyecto, agentes implicados, viabilidad y justificación. Engloba aspectos que dan razón de la importancia y beneficio de desarrollar el software.
- Análisis:
  - En esta etapa se elaboran las metas y objetivos de la aplicación, identificación de procesos, actividades y elementos. Una parte importante de esta etapa es que también se definen las estrategias y tareas de aprendizaje. Así como la especificación de los perfiles de los usuarios, los entornos de aprendizaje, análisis de contenidos, requerimientos técnicos para los entornos de aprendizaje, requerimientos de usuarios y requerimientos funcionales.
- Diseño
  - En esta etapa se elaboran los diseños lógicos de los aspectos psicopedagógicos, el diseño funcional de la operación educativa, los

esquemas de navegación, edición de interfaces y elementos de control con los cuales se establecen el diseño comunicacional.

- Desarrollo
  - Esta parte comprende la selección de herramientas de desarrollo, el manual técnico del software donde se especifican las sentencias, código que se realizan con la herramienta de diseño, y por último la edición de los elementos multimedia.
- Implementación
  - En esta etapa con base en el diseño pedagógico y físico, se generan las interfaces e incorporación de los elementos multimedia, obteniendo como resultado las pantallas que integran a el software educativo.
- Evaluación y validación del programa
  - Se realizan una prueba piloto del software, se evalúa la sintaxis, ortografía y contenidos generales del sistema para identificar errores, y por último se realiza una evaluación del entorno general, pedagógico y técnico/estético.
- Producción y elaboración de material complementario
  - Estas son las últimas dos etapas de la metodología y consisten primeramente en especificar un nombre representativo para el software, reproducirlo, distribuirlo entre otras. Así mismo comprende el desarrollo de los manuales de usuario y el manual de actividades didácticas.

### **3.5 Plataformas educativas con funciones evaluativas**

Una plataforma es entendida como aquel sistema informático que emplea software específico para desarrollar y ejecutar tareas educativas, entre ellas la de evaluar. Usualmente este tipo de plataformas están ubicadas en contextos web, también conocidas como e-learning, cuya característica principal es que toda la interacción y la comunicación entre los alumnos y los profesores se da a través de internet. Lo anterior puesto que se rompe la condicionante de trabajar exclusivamente en estaciones de trabajo locales, en las cuales el sistema o la plataforma se encuentre instalada.

Esto introduce un nuevo espacio de aprendizaje, el cual está asociado a la enseñanza no presencial, este nuevo espacio de aprendizaje se le conoce como el espacio virtual.

Naturalmente un espacio de enseñanza diferente al tradicional (presencial) requiere de nuevos diseños de aprendizaje, esto quiere decir una nueva planeación de las interacciones entre todos los elementos que se ven involucrados en las relaciones didácticas (Fernández-Pampillón, 2009).

La relación entre estos entornos virtuales y nuevos diseños de aprendizaje se ilustra en la siguiente figura:

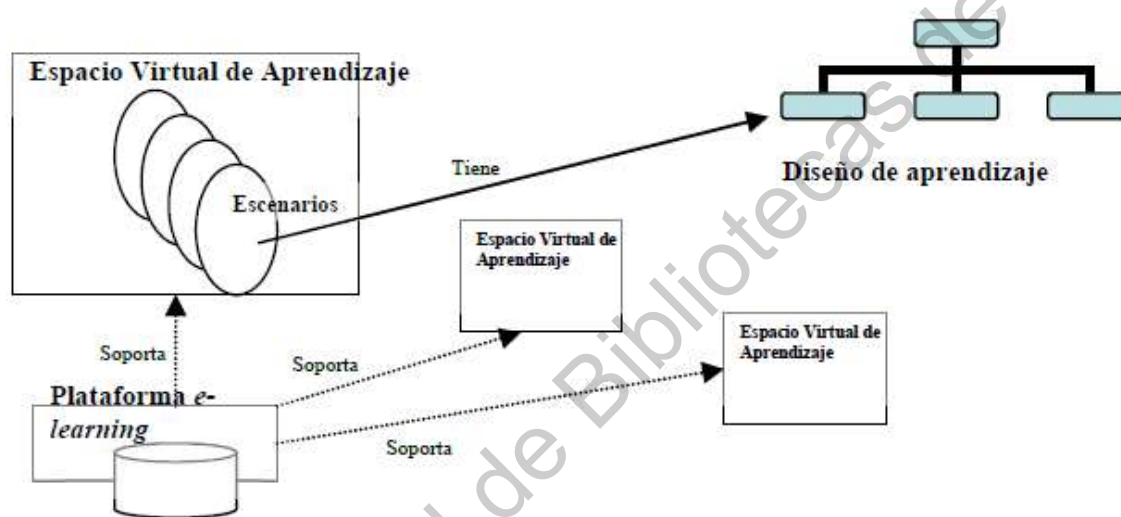


Figura 5. Modelo conceptual de espacio virtual de aprendizaje

Fuente: Fernández-Pampillón, 2009., p. 47.

Comúnmente las plataformas de e-learning se han utilizado con el propósito de continuar con la actualización profesional o capacitación de personal, ya que siempre son accesibles y el costo es muy bajo. Actualmente se han explotado estas plataformas para crear espacios educativos con los cuales ofertar diferentes modalidades, y con ello lograr llegar a un mayor número de personas. Y las IES no son la excepción, ya que incluso hay universidades que no cuentan con un campus físico, sino que son completamente virtuales.

Particularmente estas plataformas se pueden clasificar en 2 tipos: las generales y las específicas. Esta clasificación está dada por el tipo de enfoque pedagógico a utilizar, es decir, si pedagógicamente es neutro, o si por el contrario está enfocada al aprendizaje de una materia en particular, el desarrollo de una competencia o la realización de una función en concreto.

Para los casos donde se habla de enfoques neutros, los sistemas de software más comunes son los llamados *Learning Management Systems* (LMS), y como ejemplo de este tipo de sistemas están las plataformas de *Moodle*, *Blackboard*, *Google Classroom* y *e-Collage*.

Este tipo de sistemas permiten crear y administrar diferentes cursos virtuales, organizados en grupos para los cuales existe un usuario con el perfil de profesor y los restantes con roles de alumnos. Comúnmente estas plataformas vienen diseñadas con una estructura base, la cual puede personalizarse con una serie de funciones para adecuarse a las necesidades de aprendizaje que el curso demande (Fernández-Pampillón, 2009).

Un criterio en particular que debe considerarse de estas plataformas es la funcionalidad de evaluación, ya que todo proceso de enseñanza-aprendizaje requiere de ser evaluado para identificar el grado de logro de los estudiantes, cumplimiento de metas entre otras cosas.

La mayoría de los sistemas de LMS incluyen un módulo especial para la evaluación, este módulo incluye funciones con las cuales se pueden crear, editar y trabajar exámenes, desde especificar diferentes tipos de ítems como lo son preguntas abiertas, de selección múltiple, ordenamiento, verdadero o falso o incluso poder adjuntar otro tipo de archivos para completar el examen. Además, también permiten la calificación automática de los reactivos (para el caso de las preguntas que no sean abiertas, y que requieren de juicio directo del profesor), mostrar puntajes obtenidos, promedios generales entre otros datos.

Como se mencionó anteriormente, la idea básica de este tipo de plataformas es poder establecer cursos virtuales que sigan metodologías pedagógicas generales, por lo que las actividades y funciones que se disponen en estos sistemas son lo más genéricos posibles y con los cuales se puedan cumplir propósitos educativos de calidad pero sin llegar ser cursos más detallados o específicos, en los cuales se trabajen aspectos concretos del proceso enseñanza-aprendizaje como lo son el desarrollo de competencias específicas o habilidades.

Por ejemplo, la plataforma Moodle incluye un apartado en el cual se gestionan los aspectos evaluativos de un curso, permitiendo así calificar de manera automática actividades y que los resultados lleguen a través del correo electrónico a los estudiantes, evitando así la entrega de dichos resultados una clase después de la aplicación. Así mismo, se puede dar una retroalimentación a cada estudiante de acuerdo a los resultados obtenidos en los exámenes,

o proporcionar incluso exámenes para auto-estudio que no tienen influencia en las calificaciones generales del alumno.

Hablando de la parte visual o diseño del examen, *Moodle* permite trabajar con varios ordenamientos de preguntas o generarlos aleatoriamente, cambiar posiciones de preguntas arbitrariamente, determinar la cantidad de preguntas que se ven en cada página del examen. Incluso se puede elegir un modo de retroalimentación para el examen en el cual se puedan revelar las calificaciones o retroalimentación al estudiante mientras ejecuta el examen, darle oportunidad de resolver preguntas más de una vez (con menor o igual puntaje), o incluso poder volver a realizar todo el examen (Módulo de examen - MoodleDocs, 2016).

Otra plataforma que incluye un módulo para evaluación es Blackboard, este sistema web permite crear evaluaciones sencillas y de características generales, ya que al igual que Moodle básicamente se tienen que crear o registrar las preguntas que formarán parte del examen, esto significa redactar la pregunta, especificar las posibles respuestas y especificar la respuesta correcta.

Dentro de los distintos tipos de preguntas que Blackboard utiliza están las que son de tipo ensayo o preguntas abiertas, las cuales tienen que ser evaluadas directamente por el profesor, las de llenar datos faltantes, de relacionar, de opción múltiple, múltiple respuesta y las dicotómicas o verdadero /falso. También están las preguntas que se responden anexando un documento o archivo externo, preguntas del tipo escala o Likert, de ordenamiento y de respuesta corta. En general se presentan un buen número de posibles tipos de preguntas, las cuales responden en la mayoría de los casos a las necesidades de evaluación que se desean para los alumnos. Sin embargo, la parte del acomodo visual o diseño del examen es un poco restringida, debido a que se busca ser lo más genérico posible y no delimitar el examen a un grupo específico de aplicaciones o fines (Crear evaluaciones y encuestas – Ayuda de BlackBoard, 2018).

Sin embargo, para fines específicos tanto del proceso enseñanza – aprendizaje como los de evaluación, este tipo de plataformas pueden no cumplir con los recursos y funciones necesarios. Aquí entran aquellos sistemas especializados por ejemplo en el dominio o



aprendizaje de una competencia o materia en específico o incluso llegar a ser más específicos y concentrarse en el logro de una sola tarea.

Unos ejemplos de este tipo de plataformas serían las creadas para aprender idiomas, estas plataformas contienen elementos multimedia que anteriormente han sido probados en ambientes reales o de interacción presencial, material de lectura, audio y otro tipo de recursos que son los que normalmente se utilizan para aprender cualquier idioma. Básicamente los únicos cambios entre este tipo de plataformas con fines específicos es el tipo de ambiente (diseño) y el idioma para el que esta creado, por lo demás la mayoría de las actividades, ejercicios, contenidos, metodologías etc. son muy similares ya que están basadas, en la mayoría de los casos, en los mismos principios pedagógicos.

Ahora bien, cabe destacar que las plataformas genéricas antes mencionadas cumplen con la funcionalidad de crear exámenes, algunos con enfoque más normativo que criterial, pero ninguno de estas ofrece un análisis psicométrico de dichas aplicaciones, ofrecen estadísticos descriptivos de los resultados generales de los sustentantes, pero no hay un análisis detallado de la ejecución del examen, que ofrezca información con la cual se puedan emitir juicios del dominio de una determinada habilidad o conocimiento.

Si bien las plataformas especializadas se concentran en evaluar el dominio de cierta competencia, el tipo de exámenes que se realizan carecen de la evaluación psicométrica.

Existen otro tipo de plataformas con un poco más de funciones específicas, por ejemplo, sistemas *Content Management Systems* (CMSs), *Learning Content Management Systems* (LCMS), claro esta este tipo de plataformas están mayormente orientadas a la gestión de los cursos y no tanto a la funcionalidad específica de crear o aplicar exámenes.

Sin embargo, existen plataformas que están construidas específicamente para cumplir con la función de evaluar, tal es el caso del sistema EXCOBA, sistema cuyo único propósito es la aplicación de un examen que sirve como referencia para seleccionar a los estudiantes con la mayor probabilidad de éxito escolar al ingresar a diferentes programas educativos. Además, posee la cualidad de diagnosticar competencias básicas del sustentante y con ello detectar deficiencias de formación.

Este examen está alineado al currículo nacional por lo que evalúa aquellas competencias básicas que se precisan en los planes de estudio de la educación obligatoria (Ferreya y Backhoff, 2016). El tipo de ítems que se incluyen en el sistema responden a ciertos elementos en concreto, por ejemplo, la competencia que se evalúa, la estrategia de evaluación a seguir, entre otras cosas. Además, que para asegurar la confiabilidad y validez del instrumento se realiza un análisis psicométrico de cada uno de los ítems que se incluyen en el examen para determinar su calidad, basándose en criterios como el grado de discriminación, la dificultad y la correlación punto-biserial (función que determina la confiabilidad general del instrumento).

Este tipo de plataformas al ser diseñadas con un fin específico ofrecen información más relevante sobre la ejecución del examen y no solo obtener información normativa de la aplicación (saber el nivel que obtuvo el alumno en relación a todo el grupo que aplica el examen), sino de carácter criterial la cual detalla el grado de dominio de determinado conocimiento o habilidad que se desea conocer por parte del alumno.

### ***3.5.1 Plataformas para análisis psicométrico***

Efectuar un análisis psicométrico de un test de habilidad no es un proceso sencillo, ya que como se observó en apartados anteriores requiere de la aplicación de funciones y procesos de carácter matemático estadístico complejos. Por lo anterior en su mayoría, se recurre a software para realizar este tipo de operaciones complejas. Una situación particular con este tipo de software es que requieren de un archivo fuente el cual sirve como insumo para poder efectuar los análisis correspondientes, en algunos casos el archivo se realiza en formato de texto plano, siguiendo un formato específico como tabulaciones, separaciones de respuestas de cada ítem y por cada sustentante entre otros elementos.

Algunos otros permiten que los datos sean capturados en el mismo software, sin embargo, también requieren de un formato específico para la información, y especificar cierta información de carácter técnica o conceptos propios de la psicometría.

Lo anterior limita el uso de este tipo de sistema a usuarios con previa experticia en el manejo del software y de la psicometría. Por mencionar algunos de los softwares especializados que pueden realizar este tipo de análisis se encuentran ITEMAN de la compañía *Assessment*

*Systems Corporation* y Test Analysis Program (TAP) publicado por Brooks y Johanson (2003). También se puede mencionar el software estadístico Statistical Package for the Social Sciences (SPSS) propiedad de IBM, y en formato de distribución libre el software R Studio.

Algunos de estos softwares generar reportes en formatos distintos, pudiendo ser texto plano similar a un bloc de notas, o bien más robustos similares a un procesador de palabras el cual incluye gráficos descriptivos del comportamiento de cada uno de los ítems, estadísticos generales como desviación estándar, medias, promedios de cada atributo entre otros.

Como se mencionó en el apartado de Teoría de Respuesta al Ítem, los softwares antes mencionados permiten efectuar los cálculos para los atributos psicométricos de validez y confiabilidad como lo son el coeficiente de discriminación, índice de dificultad y la consistencia interna. Sin embargo, es importante recalcar que estos softwares son externos a los revisados anteriormente, los clasificados como plataformas educativas o softwares educativos. Estos últimos poseen elementos o funciones para elaborar y diseñar pruebas de habilidad en distintos formatos, como opción múltiple, respuesta construida, relacional, abiertas, entre otras; sin embargo, la capacidad funcional de análisis psicométricamente dichas pruebas no viene integrada a ellos, por lo que se recurre de manera externa a softwares especializados en psicometría con los mencionados anteriormente para lograr obtener estos análisis.

## **4. METODOLOGÍA**

### **4.1 Unidad de análisis**

La Facultad de Idiomas de la Universidad Autónoma de Baja California tiene presencia en 4 municipios del estado, por lo que para esta prueba se trabajó con una población de 394 alumnos provenientes de las cuatro sedes. Todos estos cursan el primer semestre del tronco común de la Licenciatura en Idiomas; de esta manera, el examen se desarrolló para la asignatura Morfología de la segunda lengua, la cual está seriada con Morfosintaxis, ambas ponderadas con la más alta relevancia curricular-en el nivel académico de los participantes. El género de la población fue homogéneo (50% para cada cual), de los cuales se eligieron a su vez un 33% de alumnos con bajo rendimiento (menores o iguales a 6.9), 33% de alumnos con rendimiento regular (entre 7 y 8.9) y 34% de alumno con alto rendimiento (mayores o iguales a 9).

### **4.2 Metodología para el análisis del estado del arte**

Para la elaboración del estado del arte de este trabajo de investigación fue necesario aplicar una metodología que guiara la revisión documental del mismo. Dicha revisión de la literatura se desarrolló tomando como líneas de intelección —que a su vez fungieron también como criterios e inclusión y exclusión— las categorías, variables e indicadores diseñadas ex profeso para esta investigación (ver Figura 6), las cuales dirigieron el análisis de contenido de los últimos 25 años de producción científico en el campo de la evaluación educativa, en lo general, y la evaluación estandarizada de los aprendizajes, en lo específico.

A partir del fichaje —entendido como la técnica de recolección, almacenamiento y organización de información— se logró operativizar los ejes de macro (categorías), meso (variables) y micro (indicadores) análisis con los que se recuperaron los puntos de encuentro, desencuentro, tensiones y vacíos de conocimiento en este campo de investigación educativa, y que dan sustento a la caracterización de las prácticas de la evaluación de los aprendizajes, las metodologías que se utilizan para una evaluación de calidad, los principales atributos con los que se puede medir la calidad de una prueba estandarizada, así como las teorías sobre las que recaen todas las prácticas, técnicas o instrumentos que se emplean en la consecución de la evaluación estandarizada del aprendizaje.

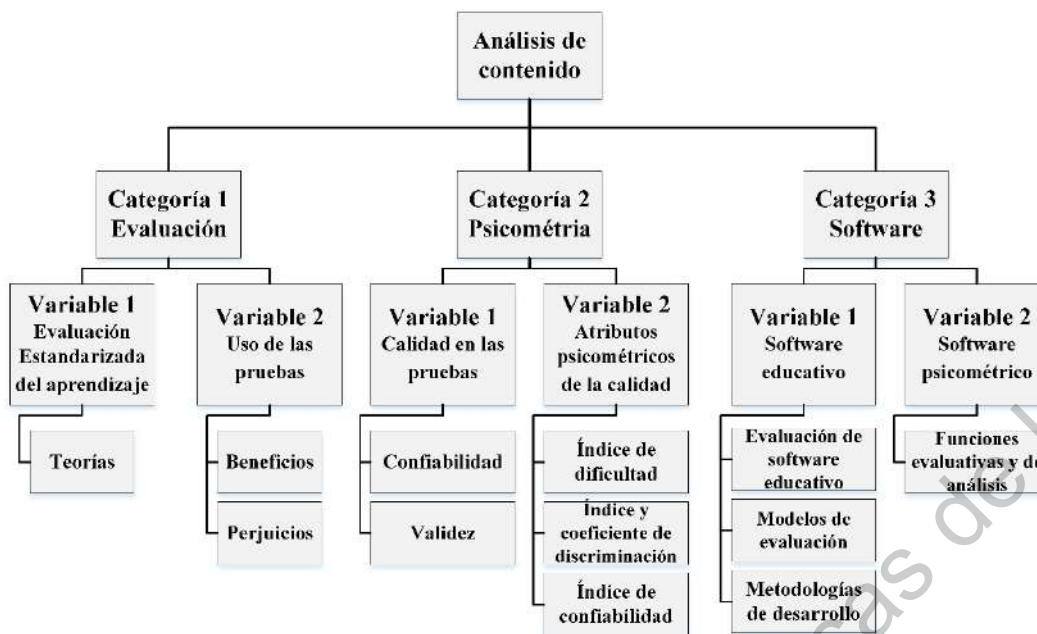


Figura 6. Categorías, variables e indicadores de análisis en la investigación.

Fuente: elaboración propia.

Con base en la estructura de análisis de investigación señalada en el esquema anterior se procedió a la revisión sistemática de cada uno de los elementos que componen el corpus de este apartado, efectuando así un análisis cualitativo de estos documentos enfocando los esfuerzos en encontrar elementos relacionados con los indicadores definidos.

### 4.3 Metodología de la investigación

El presente trabajo se sustentó desde el enfoque cuantitativo de investigación, de forma general este enfoque se caracteriza por realizar una constatación mediante la recolección de información cuantitativa orientada por conceptos empíricos medibles, derivados de los conceptos teóricos con los que se construyen las hipótesis conceptuales. Bajo este enfoque la medida y la cuantificación de los datos constituye el procedimiento empleado para alcanzar la objetividad en el proceso de conocimiento. Para ello se busca establecer promedios a partir del estudio de las características de un gran número de sujetos, los cuales usualmente son representativos (muestra) de todo el universo de estudio. De ahí se deducen leyes explicativas de los acontecimientos en términos de señalar relaciones de causalidad entre los acontecimientos estudiados. Las explicaciones proporcionadas se contrastan con la realidad factual de manera que su concordancia con ella define la veracidad y objetividad del

conocimiento obtenido (Monje, 2011). El proyecto se centra en aspectos observables susceptibles de cuantificación, y se utilizan diversos recursos estadísticos para el análisis de los datos. Por esta razón el enfoque cuantitativo contribuyó a la identificación y procesamiento de las variables asociadas a la investigación, como se puede observar en el apartado del proceso de intervención y que se detalla de manera completa en el anexo 3 de este trabajo. En este sentido la investigación emplea el software estadístico SPSS 25 (Wagner III, 2019), con el cual se realizan los análisis de estadísticos descriptivos de los diferentes instrumentos (cuestionarios, encuestas) que se utilizan para recopilar información.

Otro análisis y procedimiento estadístico que la investigación aplica son los asociados a la obtención de criterios (indicadores) de calidad pertenecientes al examen, estos son realizados con las plataformas de software ITEMAN y TAP, los resultados obtenidos con el procesamiento de estos sistemas son analizados en la última etapa de la metodología para el desarrollo del examen, con los cuales se mide la calidad psicométrica de la prueba.

Un ejemplo del instrumento de recopilación de información obtenido de la ejecución de la prueba y procesado por los softwares antes mencionados se puede observar en la siguiente figura:

Quick Item Analysis  
 \*\*\*\*\*

Item	Key	Number Correct	Item Diff	Disc. Index	# Correct in High Grp	# Correct in Low Grp	Point Biser	Adj PtBis
Item 08	( 2 )	29	0.76	0.35	9 (0.90)	6 (0.55)	0.38	0.33
Item 09	( 3 )	19	0.50	0.14	5 (0.50)	4 (0.36)	0.22	0.15
Item 10	( 3 )	9	0.24	0.31	4 (0.40)	1 (0.09)	0.40	0.35
Item 11	( 2 )	24	0.63	0.55	10 (1.00)	5 (0.45)	0.34	0.28
Item 12	( 1 )	25	0.66	0.44	8 (0.80)	4 (0.36)	0.30	0.24
Item 13	( 2 )	36	0.95	0.18	10 (1.00)	9 (0.82)	0.16	0.12
Item 14	( 2 )	21	0.55	0.35	8 (0.80)	5 (0.45)	0.24	0.17
Item 15	( 3 )	36	0.95	0.09	10 (1.00)	10 (0.91)	0.19	0.16
Item 16	( 3 )	34	0.89	0.27	10 (1.00)	8 (0.73)	0.26	0.22
Item 17	( 1 )	17	0.45	0.52	7 (0.70)	2 (0.18)	0.43	0.37

Figura 7. Información psicométrica recopilada con el software TAP.

Fuente: elaboración propia usando TAP versión 16.1.18.

Por otra parte, uno de los apartados del proceso metodológico para la construcción de la propuesta tecnológica para la ejecución y análisis psicométrico del examen es necesario la evaluación de la usabilidad, este proceso requiere de recopilar la opinión de personal directivo, de coordinación y docentes mediante una serie de entrevistas semi estructuradas, para ello se aplicó una metodología cualitativa con la cual se obtienen los discursos completos de los participantes para proceder luego a su interpretación. Un pequeño segmento de este instrumento se muestra en la siguiente figura (el instrumento completo se puede consultar en el anexo 9).

**Evaluación de la calidad y usabilidad del sistema SIEXAES (alumno)**

El propósito del siguiente cuestionario es ayudar a detectar problemas en el software especializado para la realización de exámenes estandarizados SIEXAES, y con ello mejorar la experiencia y calidad evaluativa que la Facultad de Idiomas te ofrece como estudiante.

Instrucciones:

Marque la casilla que mejor represente su experiencia

1. ¿Le parece adecuada la selección de contenidos destacados en la portada antes de iniciar el examen?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

2. ¿Es fácil identificar los elementos en los que se debe ingresar algún dato en la portada antes de iniciar el examen?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

Figura 8. Instrumento para la evaluación de usabilidad.

Fuente: elaboración propia.

Este instrumento se empleó para medir cuantitativamente los criterios de calidad de la usabilidad de la herramienta tecnológica innovadora, dichos criterios comprendieron cinco dimensiones, contenido, navegación, facilidad de aprendizaje, representación gráfica y utilidad.

Entre los métodos cuantitativos para la recopilación de la información, en el caso particular de la elaboración de la prueba estandarizada, se emplearon varios instrumentos, por ejemplo,

para la etapa 2 del diseño relacionada con el análisis del currículo se aplicó una encuesta a 3 docentes expertos en la asignatura, mediante los cuales se pudo obtener el índice de relevancia curricular. Este proceso de recolección y análisis cuantitativo se puede observar en la siguiente figura.

Contenido	Contribución al logro de la competencia de la unidad (20%)	Dosificación (cuantía de contenidos implícitos) 10%	Carga horaria (asignada o estimada para su aprendizaje) 10%	Relevancia disciplinaria (20%)	Cantidad de servicios que recibe	Proporción de servicios que recibe 20%	Cantidad de servicios que proporciona	Proporción de servicios que proporciona 20%	Índice de Relevancia Curricular (IRC)
1.1.1 Definition of the discipline	0.20	0.10	0.10	0.20	0.00	0.00	0.00	0.00	0.600
1.1.2 History	0.13	0.07	0.07	0.13	0.00	0.00	0.00	0.00	0.400
1.2.1 Free forms	0.20	0.10	0.07	0.13	0.00	0.00	1.00	0.03	0.533
1.2.2 Signs and morphemes	0.20	0.10	0.10	0.20	0.00	0.00	1.00	0.03	0.633
1.3.1 Morphemes and allomorphs	0.20	0.10	0.08	0.20	0.00	0.00	6.00	0.20	0.783
1.3.2 Free and bound morphemes	0.20	0.10	0.10	0.13	0.00	0.00	1.00	0.03	0.567

Figura 9. Instrumento de recopilación de información para calcular IRC.

Fuente: elaboración propia.

Mediante este instrumento se pudo recopilar la valoración que hizo cada uno de los tres jueces en relación a criterios definidos para establecer el IRC, se emplearon diferentes procedimientos estadísticos para establecer la consistencia entre la valoración que hicieron los jueces, tales como la correlación intra clase, la técnica de análisis de varianza (ANOVA) entre otros.

Así mismo en una de las etapas del proceso de intervención, para el análisis diagnóstico de la situación actual que presentaba cada indicador de intervención asociado a los objetivos interventivos se empleó una entrevista directa individual con cada uno de los docentes que componían la planta docente. Esta información se recopiló en una tabla para poder lograr una distribución uniforme de cada situación presente en cada indicador del logro de la intervención. Este instrumento puede ser consultado en el capítulo cinco, bajo la descripción del proceso de intervención.

El instrumento antes mencionado fue muy útil para la investigación, ya que nuevamente, bajo el enfoque cuantitativo, contar con un punto de partida en relación a cada uno de estos indicadores permitió establecer de manera clara el grado de avance y/o mejora de cada uno de ellos al concluir la intervención. Algunos de estos indicadores también fueron identificados gracias a la implementación de la técnica de árbol de problemas, identificado las causas y efectos del problema de investigación en cuestión.



#### 4.4 Metodología para el desarrollo del examen

Inicialmente se expresa que la evaluación del aprendizaje referida a un criterio se orienta a explorar el nivel de dominio que tiene el estudiante o el examinado respecto a un criterio definido, mostrando así las habilidades y conocimientos que este domina o en su caso las que no domina, pero en un campo específico del conocimiento (Contreras y colaboradores, 2005).

Bajo la idea anterior, el examen constituye por sí solo una muestra de varias tareas evaluativas que permiten ser una muestra representativa del dominio de un contenido en específico.

Este tipo de exámenes permiten describir de manera precisa y clara lo que se intenta medir. Un uso común de dichos exámenes o pruebas es en la evaluación de programas, para certificar competencias (Popham, 2000). La metodología que se utilizó para realizar este examen fue desarrollada en el Instituto de Investigación y Desarrollo Educativo (IIDE) (Contreras, 2000; Contreras y Backhoff, 2004; Contreras y colaboradores, 2005) con base en el modelo psicométrico que propuso originalmente Nitko (1994) para elaborar exámenes de gran escala de referencia criterial orientados por el currículum; es decir que el currículum es la base sobre la que se desarrolla el examen, y que así mismo todas aquellas decisiones sobre lo que se va a evaluar y la forma de evaluarlos están directamente orientados por lo que se establece en el currículum.

Dicha metodología contempla seis etapas del desarrollo del examen: a) definición del dominio de resultados que pretende el currículo, b) análisis del currículum, c) desarrollo de un plan de evaluación, d) producción y validación de ítems, e) análisis primario de resultados, y f) análisis secundario de resultados.

El trabajo que se realiza en el módulo del diseño del examen de Morfología de la Segunda Lengua abarca la definición del dominio de resultados que pretende el currículum, y el análisis del currículum, que corresponden *a las primeras dos etapas de la metodología*. En estas dos etapas los procedimientos fueron:

- Seleccionar al comité coordinador del examen (CCE)
- Capacitar al comité coordinador del examen

- Seleccionar al comité diseñador del examen (CDE)
- Capacitar al comité diseñador del examen
- Determinar el universo de contenido en la asignatura
- Analizar el currículum de la asignatura
- Determinar el universo de medida
- Justificar el universo de medida
- Diseñar la estructura general del examen

En referencia a los puntos anteriores, la integración de dichos comités es un proceso elemental, ya que estos se encargan de la ejecución de las diferentes actividades que se requieren para cada fase del desarrollo del examen. El primero es el comité diseñador, el cual se encarga del diseño general de la prueba, entre las acciones que se realizan está el análisis del currículo, elaboración de la retícula y el análisis de los índices de relevancia curricular.

El segundo comité es el de especificaciones, encargado de elaborar las especificaciones para la construcción de los ítems, detallando la estructura estándar que estos deben tener. El tercero es el comité de ítems, el cual se encarga de la producción de los ítems de acuerdo a las características señaladas en las especificaciones.

El cuarto es el comité de jueceo, quien realiza una valoración de diferentes aspectos, como la calidad técnica del ítem, la validez cognitiva, formato y edición entre otros. El último comité es el analítico informático, este realiza los procesos de evaluación psicométrica de los ítems mediante el procesamiento de la información con software especializado.

Otro detalle importante sobre la integración de estos comités es que se conforman por un docente experto en la asignatura, un docente experto en el currículo, un docente con años de experiencia en educación y un docente con amplia experiencia en evaluación, estos comités a su vez pueden preparar o capacitar a otros docentes para que participen en las actividades que se desempeñan.

La operación de tales procedimientos generó tres productos principales:

1) La retícula, que es la representación gráfica de los contenidos que el comité diseñador del examen hizo explícitos, tras analizar la documentación formal e informal que definía el

currículum, de igual forma en la retícula también se muestran las relaciones de servicios entre los contenidos identificados.

2) El documento de justificación del contenido que se juzgó como necesario evaluar en el universo de contenidos del examen, mismo que también es identificado en la retícula, un ejemplo es el mostrado en la tabla 2.

**Tabla 2**

*Ejemplo de justificación de contenidos para el examen*

Contenido a evaluar en el examen	Importancia asignada	Razones que justifican la decisión	Observaciones
1.1.1 Definition of the discipline	Muy importante  (IRC = 0.600)	<p>El contenido <b>1.1.1 Definition of the discipline</b> tiene un <b>IRC = 0.600</b>, de un máximo posible de 1.000. Aunque no proporciona ni recibe servicio de otros contenidos del programa, su importancia radica en que recibió los puntajes más altos en todos los criterios que integran el Índice de Relevancia Curricular (IRC): contribución al logro de la competencia de la unidad (Recibió 0.20 de un máximo posible de 0.20), cuantía de contenidos implícitos (Recibió 0.10 de un máximo posible de 0.10), carga horaria estimada para su aprendizaje (Recibió 0.10 de un máximo posible de 0.10) y relevancia disciplinaria (Recibió 0.20 de un máximo posible de 0.20).</p> <p>Como consecuencia de lo anterior, se elaborará una especificación para <b>producir 2 ítems</b> que permitan probar cabalmente su dominio por parte del examinado.</p>	<p>La especificación detallará la elaboración de dos ítems, uno que pruebe el dominio del concepto Morfología mediante su definición y otro que ponga a prueba la habilidad para identificar características de la morfología.</p>

Fuente: elaboración propia.

3) La tabla de especificaciones del examen, que presenta de una manera sintética las decisiones de estrategia evaluativa del examen. Un ejemplo de este proceso se ilustra en la siguiente tabla.

**Tabla 3***Ejemplo de especificación de un contenido temático*

Contenido	IRC	Especificaciones	Ítems	Foco del ítem	Tipo de ítem	Nivel taxonómico
1.1.1 Definition of the discipline	0.600	1	2	Probar el dominio del concepto Morfología mediante su definición.	OM	Comprender conceptos
				Probar el dominio del concepto Morfología mediante la identificación de sus características.	OM	Comprender conceptos

Fuente: elaboración propia.

Para la etapa tres de la metodología en lo que tiene que ver con el desarrollo del plan de evaluación del instrumento, se realizaron los siguientes procedimientos:

- Diseñar la estructura el examen
- Seleccionar al comité diseñador de especificaciones de ítems
- Capacitar al comité diseñador de especificaciones de ítems
- Elaborar especificaciones de ítems
- Seleccionar al comité elaborador de ítems
- Capacitar al comité elaborador de ítems

El procedimiento más importante derivado de esta etapa fue la de desarrollar especificaciones de ítems. La especificación de ítems es un proceso formal que describe, al responsable quien finalmente elaborará el ítem, las características que debe tener la tarea evaluativa.

En otras palabras, se habla de un retrato por escrito de un ítem, este retrato detalla las características que deben tener los reactivos y las respuestas a estos, de manera que pueda considerarse válida la tarea evaluativa.

Un ejemplo de este tipo de especificaciones se muestra en la figura 10 (se pueden consultar más especificaciones en el anexo 1). Entre los elementos que destacan en este proceso es la precisión de la dimensión del conocimiento y proceso cognitivo que el ítem evaluará, así como la posible información contextual requerida para la ejecución del mismo.

<b>Información contextual</b>	El contenido <i>Definition of the discipline</i> es el primer tema conceptual que el alumno aborda en la <i>Unidad 1. Basic Concepts</i> , en donde se aborda el concepto, así como las características de la morfología en la formación de palabras.							
1.1.1 Definition of the discipline	Por lo anterior, este contenido fue uno de los que obtuvo las puntuaciones más altas en los criterios de contribución al logro de la competencia de la asignatura, así como mayor puntuación con relación al índice de relevancia curricular.							
Part of the Linguistic that studies the admissible rules for the words formation for their delimitation, definition and classification	De esta manera se elaborarán 2 ítems: uno que pruebe el dominio del concepto de <i>Morphology</i> ; y otro que pruebe el dominio del concepto <i>Morphology</i> a través de la identificación de sus características.							
<b>Unidad 1. Basic concepts</b> <b>Competencia:</b> Identificar las unidades morfológicas mínimas implicadas en la formación de palabras en la lengua inglesa a través de la realización de ejercicios prácticos para crear conciencia de los mecanismos de construcción de significados; mostrando respeto, mostrando interés así como respeto hacia el profesor y sus compañeros al trabajar en el aula en equipos de trabajo o al realizar sus ejercicios de manera individual. <b>IRC:</b> 0.600 <b>Criterios:</b> • Contribución al logro de la competencia de la unidad: .20 • Dosificación: .10 • Carga horaria: .10 • Relevancia disciplinaria: .20 • N° servicios que recibe: 0 • N° servicios que proporciona: 0 <b>Recibe servicios de:</b> Ningún contenido.	<b>Información contextual o indicaciones para responder este ítem:</b> Ninguna.							
	<b>Información tabular, gráfica o textual a emplear en el ítem:</b> Ninguna.		Dimensión conocimiento	Dimensión proceso cognitivo				
	Conocimiento factual		Recordar	Comprender	Aplicar	Analizar	Evaluar	Crear
	Conocimiento conceptual			X				
	Conocimiento procedimental							
Conocimiento metacognitivo								
<b>Especificación de la base del ítem:</b> 1) Podrá presentar el término <i>Morphology</i> y solicitar que se identifique la definición de éste. Por ejemplo: <i>Which one is the definition of "Morphology"?</i> 2) Podrá presentar el término <i>Morphology</i> y solicitar que se identifique características propias de esta disciplina. Por ejemplo: <i>Which one are the characteristics of Morphology?</i>								
<b>Especificación de la respuesta correcta a.</b> 1) Será la definición de <i>Morphology</i> . Por ejemplo: <i>Science that studies word formation.</i> 2) Serán las características en cuestión. Por ejemplo: <i>delimited, classify and define words</i>								
<b>Especificación del distractor b.</b> 1) Se podrá presentar una definición incompleta del término, o que haga referencia a otro concepto. Por ejemplo: <i>Science that studies sentence formation.</i>								

Figura 10. Ejemplo de un segmento del formato de especificaciones de ítems.

Fuente: elaboración propia.

Para la etapa cuatro, referente a producir y validar ítems, e involucró la operación de los siguientes procedimientos:

- Elaborar ítems según las especificaciones
- Seleccionar al comité evaluador de ítems
- Capacitar al comité evaluador de ítems

La elaboración de un ítem es un trabajo técnico muy delicado y que requiere mucha atención, esto puesto que se busca mantener una estricta congruencia con la especificación que lo produce. Esto es altamente importante porque de esta estricta congruencia se obtienen evidencias relacionadas con la validez del contenido del examen, en cuyo caso contrario de no cumplirse se compromete de manera directa la calidad, confiabilidad y validez del examen.

Para la etapa cinco y seis de la metodología, se especificaron los análisis realizados a los ítems. El análisis del comportamiento de los ítems del examen de Morfología de la Segunda Lengua ante los examinados de la muestra se llevó a cabo mediante los procedimientos convencionales de la teoría de la respuesta al ítem como se explicó en el apartado de estado del arte y marco teórico respectivamente. Sin embargo, se detallan brevemente los análisis efectuados para cada ítem fueron:

- El índice de dificultad ( $p$ )
  - El cálculo de este atributo se realiza mediante la división del número de personas que contesto acertadamente el ítem entre el número de personas que en total contestaron el ítem. La fórmula que expresa la operación anterior se representa de la siguiente manera:
    - $$p = \frac{A}{N}$$
- El índice de discriminación mediante grupos contrastados (27% altos y 27% bajos)
  - El cálculo necesario para la obtención de este atributo está dado por la siguiente formula:
    - $$D_i = \frac{GA_{aciertos} - GB_{aciertos}}{N_{grupomayor}}$$
    - Donde  $GA_{aciertos}$  es el número de aciertos en el ítem del 27% de los sustentantes con las puntuaciones más altas en el examen,  $GB_{aciertos}$  es el número de aciertos en el ítem del 27% de los sustentantes con las puntuaciones más bajas en el examen, y por último  $N_{grupomayor}$  es el número de sustentantes en el grupo más numeroso ya sea GA o GB.
- La correlación del punto biserial ítem – total ( $r_{bis}$ )

- Es atributo se utiliza para saber si las personas “adecuadas” son las que obtienen las respuestas correctas, el poder predictivo del reactivo y cómo puede contribuir a las predicciones (Backhoff et. al., 2000).
- Dicho atributo puede ser obtenido a partir del siguiente calculo:

$$\blacksquare r_{pbis} = \frac{\bar{X}_1 - \bar{X}_0}{S_x} * \sqrt{\frac{n_1 n_0}{n(n-1)}}$$

Estos análisis fueron realizados mediante un programa de cómputo especializado ITEMAN de la compañía Assessment Systems Corporation, así como mediante el software TAP publicado por Brooks, G.P. & Johanson , G.A. (2003). Además de efectuar el análisis con estos softwares se utilizó el mismo software en el que se ejecutó el examen, el cual integra un módulo de evaluación psicométrico. Este software es la propuesta tecnológica innovadora de este proyecto de investigación, este software lleva por nombre Sistema de Exámenes Estandarizados (SIEXAES). Se analizaron los resultados de los tres softwares para observar variaciones entre las mediciones y con ello calibrar los cálculos de forma que se obtenga números exactos y reducir errores por operaciones con menor cantidad de decimales o bien redondeo de los mismos.

#### **4.5 Desarrollo de la tecnología para el análisis psicométrico**

Para cualquier desarrollo de software es muy importante contar con un análisis detallado de los requerimientos funcionales (RF) sobre los cuales se cimienta todo el desarrollo. Como menciona Arias (2005):

Un gran número de los proyectos de software fracasan por no realizar una adecuada definición, especificación, y administración de los requerimientos. Dentro de esa mala administración se pueden encontrar factores como la falta de participación del usuario, requerimientos incompletos y el mal manejo del cambio a los requerimientos (p. 1).

Por lo anterior, es necesario especificar de manera clara y detallada cuáles serán los requerimientos funcionales que el sistema SIEXAES debe contemplar, requerimientos sobre los cuales se efectuarán las siguientes etapas correspondientes al desarrollo de dicho sistema. Cabe destacar que existen requerimientos funcionales que tienen un mayor grado de importancia para el funcionamiento del sistema, o que requieren de un mayor grado de desarrollo, esto es en muchos casos por la complejidad de la función.

Por ejemplo, no requiere el mismo tiempo desarrollar una función como la de iniciar sesión en el sistema, en comparación con una función que requiere de procesar información más compleja o que afecta a un mayor número de entidades en la base de datos. Por ello es muy importante identificar aquellos requerimientos funcionales que tienen un alto impacto en el logro del propósito general del sistema.

Para lograr esto, es necesario contar con la aportación del personal docente de la Facultad de Idiomas, quienes en parte son usuarios finales del sistema, así como del personal administrativo que serán quienes interactuarán en mayor medida con el sistema. En base a sus aportaciones se empiezan a recopilar sus necesidades, que posteriormente son traducidos como requerimientos funcionales del sistema. Algunos requerimientos, como los de usabilidad también son contemplados en el desarrollo, pero estos se especifican en etapas posteriores en las cuales se empiezan a definir los aspectos visuales o estéticos del sistema.

Estos requerimientos son muy importantes ya que como se ha mencionado anteriormente, en ocasiones el bajo desempeño del sustentante es resultado de un mal diseño del instrumento de evaluación, refiriéndonos tanto a la redacción como al acomodo y tamaño de los elementos, los colores utilizados entre otros factores. Por lo que parte de este análisis de requerimientos va acompañado de la aplicación de buenas prácticas en el diseño de interfaces, elementos contemplados por la rama de la tecnología conocida como la interacción humano computadora (IHC).

La metodología de desarrollo de software aplicada fue Rational Unified Process (RUP), la cual utiliza el enfoque de la orientación a objetos en el diseño y emplea el lenguaje unificado de modelado (UML) para la representación de los procesos en ejecución. Las etapas del proceso de desarrollo son las siguientes:

1. Análisis de requerimientos

En esta etapa se elaboran los requerimientos funcionales y de usabilidad del sistema, lo que posteriormente se traduce en casos de uso del sistema. El análisis de requerimientos es un proceso muy robusto y elaborado ya que se desea tener la mayor cantidad posible de necesidades detectadas de forma clara y precisa para evitar cambios o errores durante las siguientes etapas de ejecución.



Por ello como se mencionó anteriormente en esta etapa la participación de los alumnos, personal docente y directivos fue muy importante, ya que con ello se obtuvo un banco de requerimientos funcionales lo suficientemente sólido para establecer las funciones y procesos cruciales que el software de evaluación debería poseer. Los casos de uso permiten modelar las funciones del sistema en escenarios particulares (usos del sistema) de forma que se pueden observar secuencias lógicas de interacción entre el usuario y el software en procesos específicos.

En esta etapa se recogen de igual forma todos los requerimientos asociados a la parte pedagógica de la aplicación, que en el caso particular del proyecto son exclusivamente de carácter evaluativo, es decir el software no incluye actividades o contenido didáctico, ni elementos de enseñanza aprendizaje. El sistema está diseñado específicamente para representar en formato digital el examen desarrollado mediante la metodología explicada en el apartado anterior. EL diagrama de caso de uso general del sistema SIEXAES está representado en la siguiente figura:

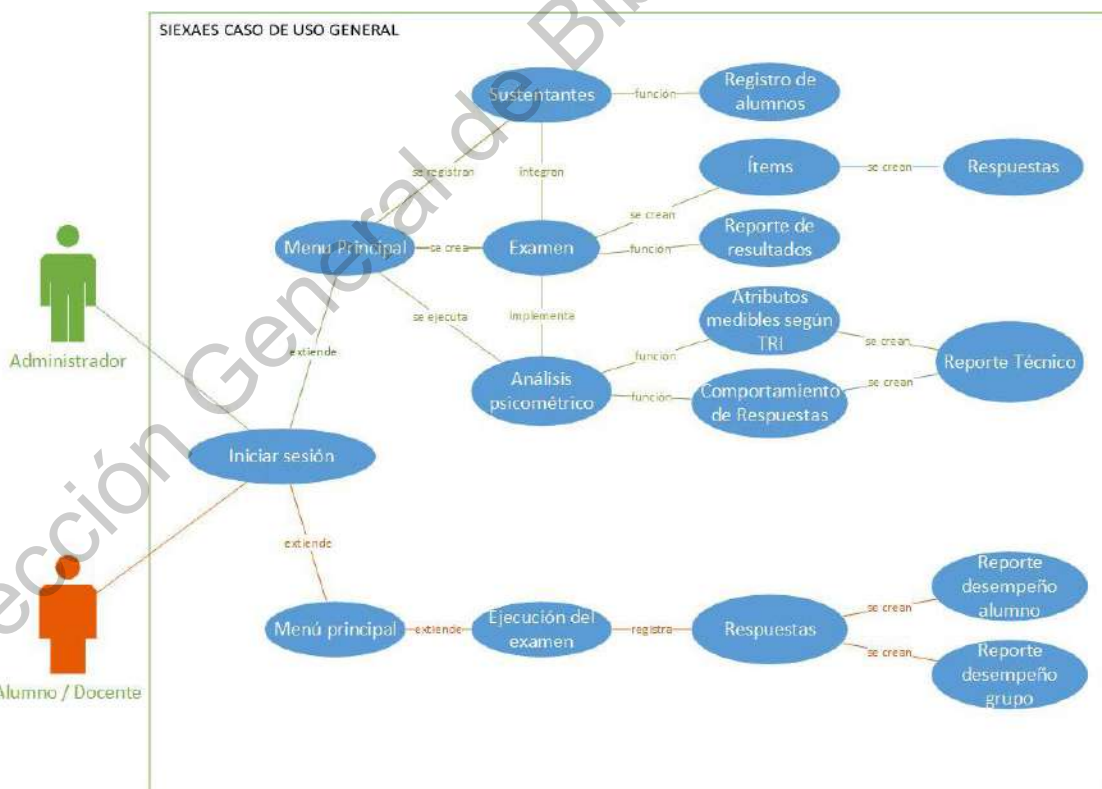


Figura 11. Diagrama de casos de uso general del sistema SIEXAES.

Fuente: elaboración propia.

## 2. Diseño funcional

En esta se elaboraron los diagramas de secuencia de todos y cada uno de los casos de uso identificados en la etapa anterior, cada requerimiento funcional puede tener uno o más casos de uso. Desarrollar diagramas de secuencia permite pre visualizar entidades de bases de datos, clases e interfaces que serán elaborados en etapas posteriores. Esto ahorra tiempo o facilita identificar elementos integradores o claves para la ejecución de los procesos.

Por ejemplo, en la siguiente figura, se ejemplifica solo un diagrama de secuencia de los desarrollados para la herramienta tecnológica, el diagrama comprende el proceso en el cual el administrador del sistema (en este caso es el personal responsable del área de evaluación en la Facultad) desea consultar los atributos psicométricos de la prueba.

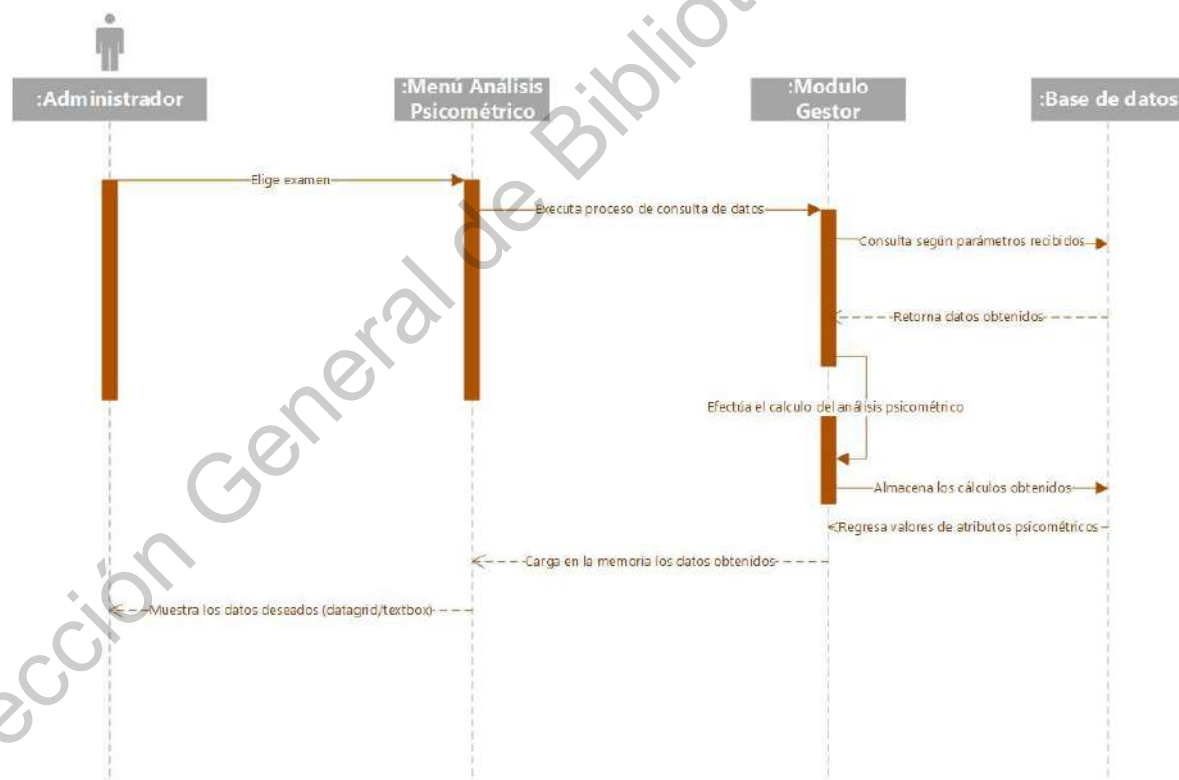


Figura 12. Diagrama de secuencia del caso de uso consultar análisis psicométrico de un examen.

Fuente: elaboración propia.

### 3. Diseño lógico

Esta etapa consiste en elaborar los diagramas de clases de todas y cada una de las funciones definidas para el ciclo de desarrollo, lo que implica definir objetos, atributos y métodos. También en esta etapa se elabora el diagrama entidad-relación y el diagrama relacional de la base de datos sobre la que estará operando la aplicación.

En la figura 13 se puede observar el diagrama de clases de algunas clases que componen al sistema SIEXAES, destacando la clase que posee la función que se encarga de efectuar el cálculo matemático para definir los cortes de los grupos bajos y altos, así como la operacionalización de los atributos del índice de discriminación, índice de dificultad y el coeficiente de discriminación.

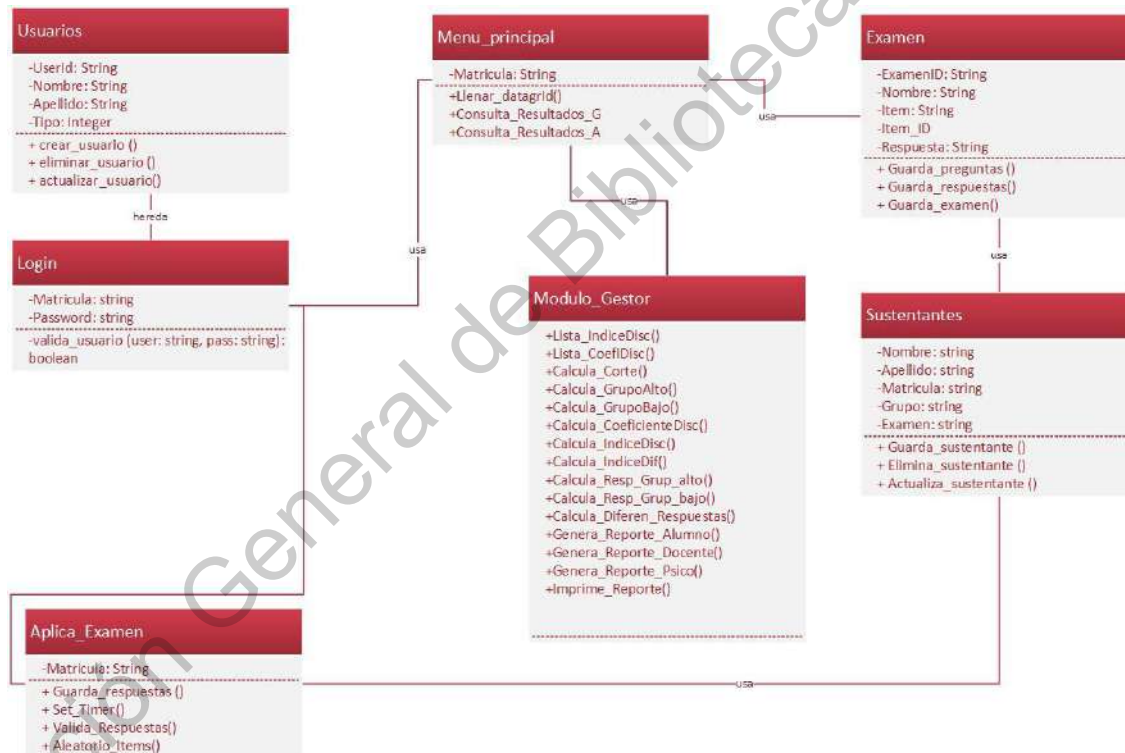


Figura 13. Segmento del diagrama de clases que compone al sistema SIEXAES.

Fuente: elaboración propia.

Un detalle interesante del diseño lógico del sistema, es que se pueden crear diferentes exámenes, y todos tendrán la capacidad de ser analizados psicométricamente, esto significa que sin que el docente formalmente aplique una metodología estandarizada para desarrollar la prueba y representarla en el sistema SIEXAES, podrá tener acceso a una valoración

psicométrica de la misma, obteniendo así los atributos de confiabilidad y validez definidos en la TRI, como lo son el índice de dificultad, índice y coeficiente de discriminación, y el comportamiento de las respuestas. El proceso lógico resumido de lo anteriormente descrito se puede observar en la siguiente figura 14.

Sin embargo, hay que considerar que el sistema solo tiene la capacidad de representar ítems o reactivos con el formato de opción múltiple. Lo anterior, como se ha mencionado anteriormente es uno de los aspectos innovadores de la tecnología desarrollada, ya que permite realizar en un mismo software la creación de la prueba, la ejecución y el análisis psicométrico de la misma.

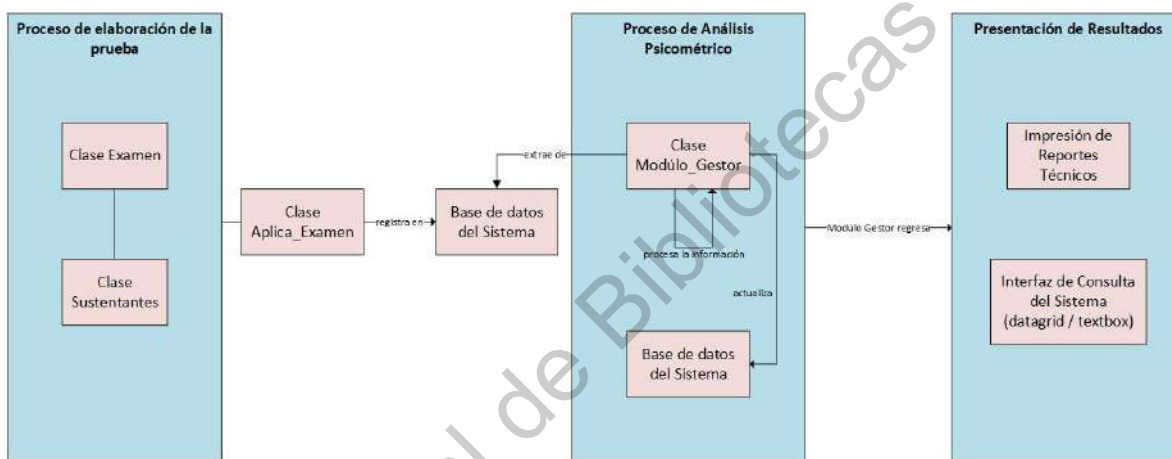


Figura 14. Secuencia lógica de la creación de la prueba, el análisis y la presentación de resultados.

Fuente: elaboración propia.

#### 4. Diseño físico

En esta parte del proceso se elaboran las interfaces del sistema, las cuales están guiadas por los requerimientos de usabilidad definidos en la primera etapa, además de los requerimientos asociados a las funciones, los cuales también son recogidos por etapas como el diseño funcional y el diseño lógico. Así mismo se elabora el diseño de la base de datos en un sistema gestor de bases de datos. Para el proyecto se empleó el metapaquete XAMPP versión 7.3.3, el cual incluye el sistema gestor de bases de datos MariaDB en su versión 10.1.38. Parte de lo elaborado en esta etapa se puede observar en la figura 15 referente al diseño de interfaces.



Figura 15. Ambiente de diseño de interfaces en la plataforma de desarrollo de software.

Fuente: elaboración propia.

## 5. Construcción

En esta etapa se programan todas las funciones del software de acuerdo a lo especificado por los diseños lógicos, tanto de las clases como la estructura de la base de datos. Para la construcción a nivel de codificación de la aplicación se utilizó la plataforma de desarrollo de Visual Studio Express 2017. En la figura 16 se muestra un pequeño ejemplo del proceso de codificación de los diseños lógicos.

```

1401     ** FUNCION PARA IMPRIMIR EL DATAGRID A PDF **
1402
1403
1404     1 referencia
1405     Private Function GetDataTable() As DataTable
1406         Dim dataTable As New DataTable("MyDataTable")
1407         'Create another DataColumn Name
1408         Dim dataColumn_1 As New DataColumn(DataGridView1.Columns(0).HeaderText.ToString(), GetType(String))
1409         dataTable.Columns.Add(dataColumn_1)
1410         Dim dataColumn_2 As New DataColumn(DataGridView1.Columns(1).HeaderText.ToString(), GetType(String))
1411         dataTable.Columns.Add(dataColumn_2)
1412         Dim dataColumn_3 As New DataColumn(DataGridView1.Columns(2).HeaderText.ToString(), GetType(String))
1413         dataTable.Columns.Add(dataColumn_3)
1414         Dim dataColumn_4 As New DataColumn(DataGridView1.Columns(3).HeaderText.ToString(), GetType(String))
1415         dataTable.Columns.Add(dataColumn_4)
1416         Dim dataColumn_5 As New DataColumn(DataGridView1.Columns(4).HeaderText.ToString(), GetType(String))
1417         dataTable.Columns.Add(dataColumn_5)
1418         'Now Add some row to newly created dataTable
1419         Dim dataRow As DataRow
1420         For t As Integer = 0 To DataGridView1.Rows.Count - 2
1421             dataRow = dataTable.NewRow()
1422         Next t
1423     End Function
  
```

Figura 16. Construcción de la función de generar archivo pdf a partir de un datagrid.

Fuente: elaboración propia.

## 6. Pruebas

En esta etapa se lleva el prototipo del software al ambiente real de uso, para lo cual participan alumnos y docentes. Una vez ejecutadas ciertas funciones del sistema se recogen las opiniones de los usuarios y las observaciones realizadas por personal del equipo de desarrollo. Se realiza un reporte de ejecución de las pruebas al software, dicho reporte se entrega a los responsables del diseño funcional para elaborar los ajustes necesarios y que sucesivamente estos se reproduzcan en las siguientes etapas del ciclo.

## 7. Evaluación de la usabilidad

Como parte de la metodología de desarrollo se aplica también la ingeniería de usabilidad, para ello se diseñaron dos instrumentos de recopilación de información con los cuales se miden los indicadores de:

- Contenido
- Navegación
- Facilidad de aprendizaje
- Representación gráfica
- Utilidad

Estos instrumentos están diseñados para recopilar la información de los estudiantes y el personal docente-administrativo, ya que estos son los usuarios finales del sistema; y a su vez, son los actores principales del proceso de intervención. Estos instrumentos pueden ser revisados en el anexo 2.

### **4.6 Metodología para la intervención educativa**

De manera general podemos mencionar que el concepto de intervención está asociado a las estrategias, técnicas, prácticas o acciones en educación, a escenarios prácticos que implican a sujetos con una problemática definida. La intervención educativa tiene como fin actuar de manera práctica en la solución de un problema educativo o bien mejorar la calidad educativa del o los estudiantes.

Para la realización de la intervención educativa que este proyecto de investigación implica se utilizó la metodología desarrollada por Castro y Castro (2013), esta metodología establece

un conjunto de etapas con las cuales se cubren de manera general la mayoría de las esferas involucradas en un proceso de intervención educativa. Este modelo se basa en un esquema en cascada que permite establecer una secuencia lineal de productos resultantes de la ejecución de tareas específicas, que a su vez funcionan como los insumos de las etapas subsecuentes. De esta forma, por ejemplo, para llegar a la etapa de diseño de las estrategias de intervención primero deben cubrirse dos etapas asociadas a los fundamentos y situación presente en el contexto de intervención, ya que las propuestas deben tener una base real de la situación actual en la institución donde se realizará el proceso de intervención, garantizando así la factibilidad de aplicación de las mismas. Este modelo se compone de nueve etapas, a continuación, se listan cada una de las fases y se describen las actividades principales que cada una de estas etapas contemplan.

#### Fase 1. Antecedentes y fundamentación

En este apartado se presenta el sustento teórico, de modo que el quehacer pedagógico se fundamente en bases científica sólidas. Este fundamento teórico debe proporcionar los constructos teóricos que sustentarán toda acción educativa, de modo que permita comprender las razones teóricas que explican los cambios que se presenten en el alumno, el salón de clases, la escuela etc. Para efectos de este proyecto en este apartado se tratan los aspectos teóricos de la evaluación del aprendizaje, evaluación estandarizada del aprendizaje, la psicometría y el uso de software especializado para análisis psicométricos. Así mismo se analizaron los antecedentes que la institución educativa en la cual se está interviniendo presenta sobre el uso de pruebas estandarizadas, herramientas tecnológicas asociadas a este tipo de pruebas, entre otros.

#### Fase 2. Análisis de las dimensiones de la gestión educativa

Fundamentalmente, se busca con este análisis hacer conscientes de las decisiones tanto de gestión, como del currículum, así como las posibilidades de abrirse a dimensiones vinculadas directamente con el aula y la investigación de la acción. Principalmente se habla de la dimensión de gestión educativa, dimensión pedagógica, dimensión curricular y dimensión investigativa.

### Fase 3. Diagnóstico situacional

Esta etapa trata la información referida al espacio específico en el que se ubica el objeto o situación de intervención. Particularmente, se refiere a la evaluación del contexto institucional, el que se realiza en función de identificar las necesidades y problemas globales de la escuela, que sean atingentes y justifiquen la intervención.

### Fase 4. Formulación de la propuesta

Como dice Castro y Castro (2013) un problema de la escuela o del aula es una situación educativa insatisfactoria que debe ser cambiada, resultando en un campo de acción sobre el cual es necesario actuar y, como tal, constituye un desafío una reflexión, creatividad y voluntad de los que participan directamente en ella. Por lo tanto, se debe seleccionar racionalmente el problema al que se le quiere encontrar solución desde el currículum y la didáctica.

### Fase 5. Fines y propósitos del proyecto

Esta etapa es la que se define de manera concreta cual es la finalidad de la intervención, así como el objetivo general y específicos del proyecto. El objetivo general expresa la idea global del estado futuro que se pretende alcanzar al final del periodo de duración del plan de intervención. El objetivo específico es la situación o estado deseable de alcanzar como resultado de la eliminación de uno de los factores que causan el problema principal; es decir, cada objetivo específico apunta a la superación de cada una de las causas más importantes que determinan el problema principal. Esta etapa del proceso está definida en el capítulo 5, bajo las secciones 5.1 y 5.2.

### Fase 6. Estrategias y acciones

Esta fase permite visualizar las posibilidades de concreción del proyecto, las estrategias tienen un carácter multidimensional, los cambios que se desean lograr dependen de varios factores, los cuales deben ser enfrentados a través de las líneas de acción. Esta puede combinar acciones orientadas a acciones con los alumnos, formación o preparación de los docentes para llevar adelante el proyecto, relación con la comunidad y entorno social, adquisición de recursos, modificación de asuntos organizativos, entre



otros. Todos estos aspectos se analizan en esta etapa del plan. Para Cruz (2017) esta fase es descrita como la fase de implementación, ya que es en la cual se define formalmente al plan de gestión respecto a las acciones que este va a contemplar. Esta fase es una de las más complejas y robustas del proceso de intervención, ya que implican a detalle cada una de las tareas y actividades que se van a realizar con cada una de las estrategias y acciones de intervención definidas, las cuales fueron previamente validadas. Esta fase del proyecto está representada en todo el capítulo 5 del trabajo de investigación.

#### Fase 7. Recursos

En este apartado se determinan y establecen los medios y recursos que se requieren para llevar a la práctica el plan de gestión, se incluyen aspectos como el recurso humano, financiero, infraestructura, material, externo entre otros. Se definen los recursos existentes como los necesarios.

#### Fase 8. Evaluación

La evaluación es una de las fases más importantes en el plan de gestión, ya gracias a esta se dispone de información acerca de la implantación y ejecución del plan lo que, al mismo tiempo, permite la toma de decisiones adecuadas y las posibilidades de proyección en términos de consolidación (Castro y Castro, 2013). Esta etapa se analiza más a detalle en el apartado 5.6 Evaluación de la Intervención.

#### Fase 9. Impacto

Aunque si bien las fases de estrategias y evaluación del plan de intervención son imprescindibles, es necesario también proveer de información acerca del impacto que la intervención pudo haber generado más allá de lo establecido en los objetivos. De forma tal que puedan observarse impactos en áreas internas (áreas que no han sido, específicamente, intervenidas en el contexto de la institución escolar) o áreas externas como por ejemplo la comunidad en la que está inmersa la institución educativa.

## 5. PROCESO DE INTERVENCIÓN

### 5.1 Análisis de las dimensiones de gestión educativa

La intervención implica la definición de una serie de dimensiones que engloban los distintos aspectos interventivos que deben contemplarse para asegurar la calidad del proceso. A continuación, se describe cada una de ellas.

#### *Dimensión gestión educativa*

##### a) Gestión estratégica

El plan de gestión de intervención está alineado directamente a la misión y visión de la institución para la cual se realiza. El proyecto contribuye con la misión al aportar a “Formar integralmente profesionales de calidad en los campos de enseñanza de lenguas y sus culturas, traducción, interpretación y áreas afines a nivel de licenciatura y posgrado”, ya que la intervención pretende aumentar la calidad de la evaluación del aprendizaje en los estudiantes, lo que definitivamente abona precisamente a formar estudiantes con niveles de dominio más concretos, observables y demostrables. En relación a la integración del proyecto de intervención con la visión de la institución, la intervención abona directamente con lo siguiente: “La Facultad de Idiomas de la Universidad Autónoma de Baja California en 2019 continúa siendo la institución líder, innovadora y de vanguardia reconocida en el ámbito nacional e internacional”, específicamente con la parte de ser una institución innovadora, ya que el proyecto de intervención genera como producto una herramienta tecnológica que integra un proceso innovador que permite construir, representar, ejecutar exámenes, y a la vez la capacidad de realizar análisis psicométricos automáticos.

##### b) Gestión curricular

La intervención tiene como fundamento de desarrollo el currículo de una de las asignaturas que tienen el mayor índice de relevancia curricular del tronco común de las carreras de Enseñanza de Lenguas y Traducción. El currículo representa al universo de conocimientos que se desea evaluar, y por ende las pruebas deben reflejar dicho universo. Algunos contenidos temáticos tienen mayor representación en una prueba en relación a otros, esto está determinado por el índice de relevancia curricular, dicho índice es un número obtenido a

partir de la valoración de dicho contenido temático en relación a la cantidad de conocimiento que aporta para concretar otros conocimientos, que pueden estar dentro de la misma unidad o no, incluso que pueden afectar el aprendizaje de temas de otras asignaturas que no necesariamente sean del mismo semestre.

#### *Dimensión pedagógica*

El proyecto de intervención presenta una forma innovadora de evaluar el aprendizaje de los estudiantes, esto mediante análisis psicométricos generados por computadora de manera automática con cada ejecución que se realiza del examen. Esto permite estar evaluando constantemente la calidad del examen, y con ellos calibrar de una manera más oportuna el examen, para que este cuente con los elementos más adecuados para evaluar el dominio de conocimiento en cuestión. Esto indudablemente fomenta también una cultura evaluativa diferente hacia el interior de la institución educativa, ya que contar con instrumentos estandarizados de aprendizaje válidos y confiables requiere que los docentes empleen este tipo de instrumentos como criterios de promoción de grado, trayectoria de alumnos entre otros.

Así mismo los resultados son útiles para observar unidades temáticas donde hay bajo nivel de dominio por parte de los estudiantes, lo que fomenta un análisis interno hacia las estrategias de aprendizaje empleadas por el docente y que no cumplieron con la función de apoyar a el logro de un aprendizaje significativo. Lo anterior resulta entonces en la promoción de cursos de capacitación en estrategias de enseñanza – aprendizaje para el docente, diseño de nuevos materiales, recursos entre otros.

#### *Dimensión investigativa*

El proyecto de intervención fomenta ciertos procesos investigativos, por ejemplo, el análisis de trayectoria académica de los estudiantes, análisis de deficiencias en el currículo de las asignaturas, estudios comparativos de niveles de dominio en el aprendizaje de asignaturas con mayor índice de reprobación, estudios comparativos de desempeño de los estudiantes aplicando evaluación estandarizada en relación a evaluaciones ordinarias o que no siguen alguna metodología para su desarrollo.

## 5.2 Plan de gestión de riesgos de la intervención

El modelo de riesgos que se implementa es el definido por el PMBOK, a continuación, se presenta el plan de riesgos y la definición de cada uno de los elementos que el modelo implica.

### 5.2.1 Etapa 1. Definición de la gestión

El plan establece las siguientes clasificaciones de impacto y ocurrencia de los riesgos:

**Tabla 4**

*Clasificación de impacto y ocurrencia*

Probabilidad de ocurrencia	Grado de impacto
Baja	Bajo
Media	Medio
Alta	Alto
Continua	Critico

Fuente: elaboración propia.

Los roles y responsabilidades definidas se muestran en la siguiente tabla:

**Tabla 5**

*Roles y responsabilidades en el plan de riesgos*

Rol	Responsabilidad
Director de proyecto	Se encarga de la identificación, priorización y seguimiento de riesgos, proponer acciones para afrontar los riesgos identificados y la autorización de acciones propuestas para mitigar los riesgos.
Supervisor	Se encarga del seguimiento de riesgos y proponer acciones para afrontar los riesgos.
Equipo de trabajo	Se encarga de la identificación de riesgos.

Fuente: elaboración propia.

Los formatos definidos para la identificación de riesgos, así como para las respuestas a dichos riesgos son los siguientes:

*Formato #1. Identificación de riesgos primario*

Numero	Riesgo identificado	Características

*Formato #2. Análisis de riesgos cualitativo*

Numero	Riesgo identificado	Probabilidad de ocurrencia	Impacto

*Formato #3. Análisis de riesgos cuantitativo*

Numero	Riesgo identificado	Probabilidad de ocurrencia	Impacto	Prioridad

*Formato#4. Respuestas de riesgos*

# Riesgo	Afecta	Calidad	Tiempo	Alcance
<b>Descripción</b>				
<b>Causa</b>				
<b>Ocurrencia</b>		<b>Impacto</b>		
<b>Estrategia</b>				
<b>Acciones a tomar</b>				
<b>Plan de contingencia</b>				

### 5.2.2 Etapa 2. Identificar los riesgos

Los riesgos que identificados para el proyecto son los mostrados en la siguiente tabla:

**Tabla 6**

*Riesgos identificados para el proyecto*

Numero	Riesgo identificado	Características
1	Falla en el equipo tecnológico con el cual se desarrolla el software del examen	Este tipo de riesgo está asociado a las posibles fallas técnicas que todo equipo de cómputo puede presentar, y que en ocasiones es ajeno al uso propio del equipo.
2	Falla en el equipo de cómputo en el cual se aplicaran los exámenes	Este riesgo se refiere a que puede ser el caso que durante la ejecución de la prueba en el laboratorio algún equipo presente alguna falla, o pueda presentarse una falla en el suministro de energía.
3	Resistencia a la intervención	El riesgo implica la resistencia de los docentes a querer cambiar su metodología evaluativa, compartir información, ser observados etc.
4	Criterios de calidad de la prueba (validez y confiabilidad) deficientes	Se refiere a los valores establecidos de calidad en las mediciones psicométricas, el índice de dificultad, índice de discriminación, coeficiente de discriminación.
5	Capacitación inadecuada	Para utilizar el desarrollo tecnológico se requiere que el docente esté capacitado en la operación del mismo.
6	Reportes técnicos deficientes	Si los reportes técnicos con los resultados de la ejecución no son claros y entendibles se complica la toma de decisiones.

<b>7</b>	Número de docentes participantes	Para diferentes etapas en la metodología del desarrollo de la prueba se requiere la participación de diferentes docentes en diferentes comités.
<b>8</b>	Falta de disponibilidad para efectuar pruebas en el ambiente final de prueba	Se requiere hacer prueba piloto de la prueba y del software para valorar atributos de calidad en el software y la mecánica de la aplicación.
<b>9</b>	Aumento del tiempo asignado para cada actividad	Pudiera ser el caso que algunas etapas del desarrollo requieran mayor tiempo del previsto, o que surgen contratiempo ajenos a los docentes que participan en los comités, reuniones académicas , fechas de reuniones etc.
<b>10</b>	Impacto de la prueba	Consecuencias de los resultados para los alumnos, docente y/o procesos académicos.
<b>11</b>	Cantidad de alumnos evaluados	Si los grupos son más grandes que la capacidad tecnológica que poseen los laboratorios de cómputo.
<b>12</b>	Cambios al currículo de la asignatura	El examen se desarrolla con base en el currículo o la carta descriptiva de la asignatura, si este presenta cambios por actualización o cambio de plan de estudios, el examen pudiera ser obsoleto.

Fuente: elaboración propia.

### 5.2.3 Etapa 3. Análisis cualitativo de los riesgos

Para los riesgos identificados se contempla la siguiente determinación del nivel del riesgo:

**Tabla 7**

*Niveles de riesgos*

Nivel de riesgo	Rango de valores
Alto	Mayor al 70%
Medio	Entre el 30% y el 70%
Bajo	Menor del 30%

Fuente: elaboración propia.

El análisis cualitativo de los riesgos es el siguiente:

**Tabla 8**

*Análisis cualitativo de los riesgos*

Numero	Riesgo identificado	Probabilidad de ocurrencia	Impacto
1	Falla en el equipo tecnológico con el cual se desarrolla el software del examen	baja	alto
2	Falla en el equipo de cómputo en el cual se aplicaran los exámenes	baja	alto
3	Resistencia a la intervención	media	alto
4	Criterios de calidad de la prueba (validez y confiabilidad) deficientes	media	alto
5	Capacitación inadecuada	baja	medio



6	Reportes técnicos deficientes	baja	alto
7	Número de docentes participantes	media	alto
8	Falta de disponibilidad para efectuar pruebas en el ambiente final de prueba	media	medio
9	Aumento del tiempo asignado para cada actividad	media	medio
10	Impacto de la prueba	baja	medio
11	Cantidad de alumnos evaluados	media	medio
12	Cambios al currículo de la asignatura	baja	crítico

Fuente: elaboración propia.

#### 5.2.4 Etapa 4. Análisis cuantitativo de los riesgos

Referente a este apartado del plan de riesgos de la intervención, para cada uno de los riesgos se codifico en una escala de 0 a 1 los valores de probabilidad de ocurrencia, y el impacto. Y con base a los números anteriores se obtuvo el indicador de prioridad, también representado en una escala de 0 a 1.

**Tabla 9**

*Análisis cuantitativo de riesgos*

Numero	Riesgo identificado	Probabilidad de ocurrencia	Impacto	Prioridad
1	Falla en el equipo tecnológico con el cual se desarrolla el software del examen	.3	.9	.27

<b>2</b>	Falla en el equipo de cómputo en el cual se aplicaran los exámenes	.3	.7	.21
<b>3</b>	Resistencia a la intervención	.6	.7	.42
<b>4</b>	Criterios de calidad de la prueba (validez y confiabilidad) deficientes	.7	.9	.63
<b>5</b>	Capacitación inadecuada	.4	.5	.20
<b>6</b>	Reportes técnicos deficientes	.3	.8	.24
<b>7</b>	Número de docentes participantes	.4	.7	.28
<b>8</b>	Falta de disponibilidad para efectuar pruebas en el ambiente final de prueba	.4	.5	.20
<b>9</b>	Aumento del tiempo asignado para cada actividad	.4	.4	.16
<b>10</b>	Impacto de la prueba	.3	.5	.15
<b>11</b>	Cantidad de alumnos evaluados	.5	.4	.20
<b>12</b>	Cambios al currículo de la asignatura	.3	.9	.27

Fuente: elaboración propia.

El siguiente paso en el análisis cuantitativo es la priorización de los riesgos, valor obtenido en razón de la probabilidad de ocurrencia por el impacto que le riesgo produce. Los riesgos ordenados quedan de la siguiente manera y son representados en la tabla 10.

**Tabla 10***Priorización de los riesgos*

Numero	Riesgo identificado	Probabilidad de ocurrencia	Impacto	Prioridad
1	Criterios de calidad de la prueba (validez y confiabilidad) deficientes	.7	.9	.63
2	Resistencia a la intervención	.6	.7	.42
3	Número de docentes participantes	.4	.7	.28
4	Falla en el equipo tecnológico con el cual se desarrolla el software del examen	.3	.9	.27
5	Cambios al currículo de la asignatura	.3	.9	.27
6	Reportes técnicos deficientes	.3	.8	.24
7	Falla en el equipo de cómputo en el cual se aplicaran los exámenes	.3	.7	.21
8	Capacitación inadecuada	.4	.5	.20
9	Falta de disponibilidad para efectuar pruebas en el ambiente final de prueba	.4	.5	.20
10	Cantidad de alumnos evaluados	.5	.4	.20
11	Aumento del tiempo asignado para cada actividad	.4	.4	.16
12	Impacto de la prueba	.3	.5	.15

Fuente: elaboración propia.

El análisis cuantitativo de los riesgos ha arrojado valoraciones sobre elementos que están asociados a los riesgos, con los cuales se pueden anticipar aquellos riesgos con mayor probabilidad de ocurrir. Los datos obtenidos son los siguientes:

**Tabla 11**

*Elementos asociados a los riesgos*

<b>Elementos asociados a riesgos</b>	<b>Contexto presente</b>
<b>Profesores no capacitados</b>	Solo el 4.76% (1 docente) de los profesores que componen la planta académica en la institución tiene la capacitación para elaboración de pruebas estandarizadas y análisis psicométrico. Sin embargo el 14.28% (tres docentes) tienen cursos y/o talleres enfocados específicamente a la evaluación educativa.
<b>Tiempo para desarrollar pruebas</b>	El 76.19% de los docentes consideran que es muy tardado hacer una prueba estandarizada, razón por la cual deciden hacer pruebas con otro tipo de métodos.
<b>Resistencia a estandarizar criterios</b>	El 57.14% de los docentes considera que es complicado estandarizar criterios de evaluación, consideran que requiere análisis extensos para encontrar aquellos medios o atributos que permitan lograr medir los mismos aspectos del conocimiento.
<b>Auto apreciación de expertos</b>	El 100% de los docentes comenta que tiene la capacidad de desarrollar un examen para evaluar cualquier conocimiento asociado a la asignatura que imparte.
<b>Libertad de catedra</b>	El 52.38% considera que estandarizar una prueba es una forma de limitar su libertad de catedra lo que implica la forma y/o instrumentos con los que evalúa el aprendizaje de sus estudiantes.
<b>Cultura evaluativa</b>	El 61.90% considera que su clase ha funcionado con los instrumentos y estrategias que han aplicado, por lo que no

	consideran relevante querer implementar algunos otros, pero casi el 40% considerarían analizar otras alternativas si hay buenos fundamentos.
<b>Objetivos de las pruebas</b>	El 85.71% de los docentes consideran como objetivo final del examen el observar la calificación numérica que estos obtienen, dejando por un lado la identificación de los temas en los que menor desempeño hubo, o características comunes en el grupo en cuanto ciertos dominios de conocimiento.
<b>Resistencia a ser observado</b>	El 85.71% de los docentes considera que se deben estipular bien los usos de la prueba y garantizar la privacidad de la información para poder participar en una prueba estandarizada, ya que se tiene la creencia de que ese tipo de pruebas son para evidenciar malas prácticas o bajo desempeño.
<b>Estímulo a evaluar la calidad de las pruebas</b>	El 76.19% de los docentes consideran que sus instrumentos son de calidad, aunque no los sometan a ningún tipo de juicio o revisión por pares, ni ningún tipo de análisis. El restante 23.81 considera que le gustaría saber cómo poder medir la calidad de sus exámenes.
<b>Resistencia a compartir estrategias</b>	El 80.95% de los docentes consideran que el material que desarrollan no debe ser compartido libremente ya que requirió de un esfuerzo propio, sin embargo es importante destacar que el 90.47 estaría dispuesto a trabajar en un trabajo donde el producto sea la integración de las actividades de cada uno, solo el 9.53% mantuvo la postura de no trabajar en proyectos de esa naturaleza.
<b>Impacto de la prueba</b>	El 81.94% de los alumnos considera que es importante saber la calidad de la prueba, y qué tanto valor tiene un examen en

	<p>relación a la calificación final del semestre. El 91.66% manifestó sentirse más seguro de hacer una prueba en la que tengan la seguridad del contenido específico que se evaluara. En el caso de los docentes el 95.23% expresó que es importante saber cuál será el uso del resultado de las pruebas, si tienen consecuencias para ellos y de qué tipo, o si son formativas. El 100% considero elemental saber para que servirá la prueba en relación a su deseo de participar o no en un proceso de estandarización.</p>
<p><b>Operatividad</b></p>	<p>Analizando las características de la infraestructura tecnológica (software y hardware) y de red, se cuenta con dos laboratorios con capacidad para 28 y 24 usuarios respectivamente. Actualmente ningún grupo supera los 28 estudiantes, sin embargo, a lo largo de la historia se han presentado grupos con hasta 28 alumnos, lo que condiciona la capacidad de los laboratorios para efectuar un solo examen para todo el grupo. Esto es importante cuidar y señalar para efectos de que las condiciones del examen sean las mismas para todos, lo que incluye el espacio donde se realiza la prueba, el horario, el día entre otros aspectos.</p>

Fuente: elaboración propia.

### 5.3 Recursos disponibles

Dentro del análisis de riesgos se deben identificar los distintos tipos de recursos con los que se cuenta para la realización de la intervención, algunos de los cuales pueden ser considerados en su ausencia como riesgos, o bien para determinar con que elementos se pueden contar para mitigar el efecto de algún riesgo. Sin embargo, de forma general independientemente al análisis de riesgos, el proceso de intervención requiere la definición de los recursos para el proyecto. La intervención contempla son los siguientes:

- Tecnológicos
  - Servidor de aplicaciones dedicado para albergar la aplicación.
  - Equipo de cómputo y software para la construcción de la aplicación.
  - Conectividad a la red interna de la institución.
- Humano
  - Docentes expertos en la asignatura para la cual se construye el examen.
  - Docentes expertos en el currículo.
  - Docentes expertos en la enseñanza.
  - Docentes expertos y con amplia experiencia en el plan de estudios.
  - Diseñador y programador de la herramienta tecnológica.
  - Responsable de infraestructura tecnológica y de red en la institución educativa.
- Infraestructura
  - Laboratorios de cómputo pertenecientes a la institución educativa (para la realización de las pruebas).
  - Cubículo designado para el desarrollo y construcción de la herramienta tecnológica.
  - Cubículos reservados para trabajar la construcción del examen.
- Institucionales
  - Asignación de carga horaria al personal docente que participa en la construcción del examen.
  - Asignación de responsable de evaluación.

## 5.4 Diseño de la estrategia de intervención

Como primer paso en el diseño de la estrategia en la intervención educativa a efectuar con este proyecto de investigación, se define mediante el instrumento grafico del árbol de problemas la situación actual presente en la Facultad de Idiomas, institución sobre la cual se va intervenir. Un segmento de dicho instrumento se representa en la siguiente figura:

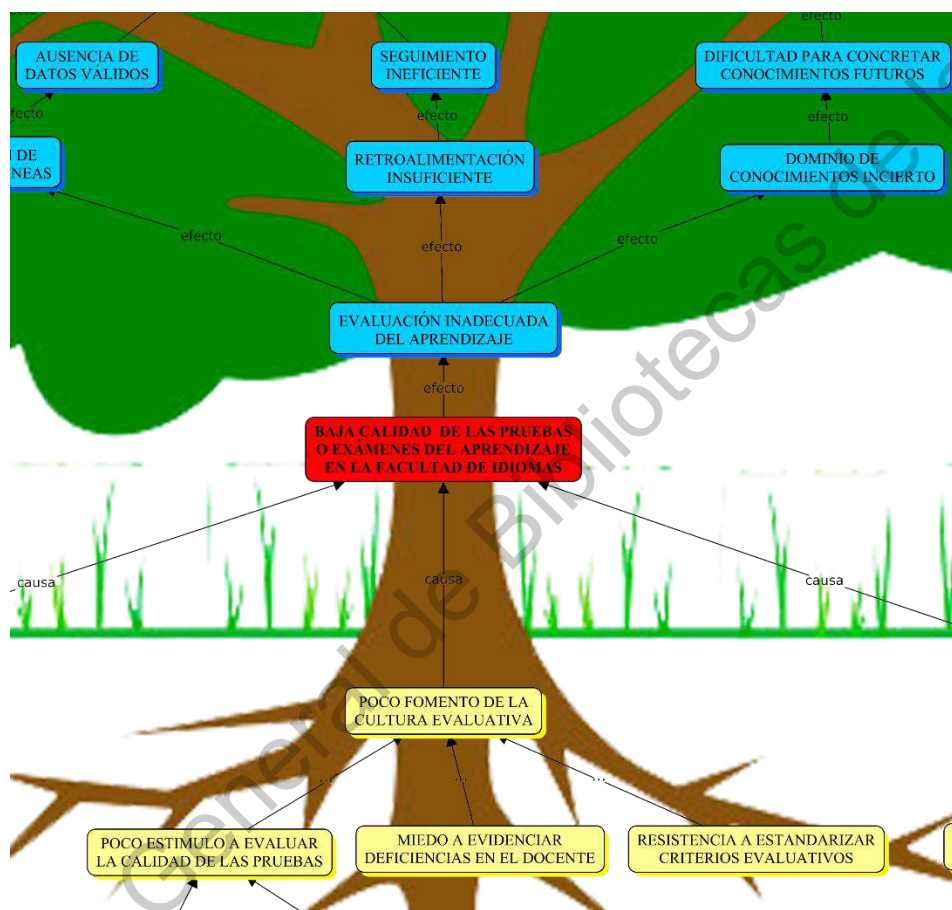


Figura 17. Árbol de problemas del proyecto de intervención.

Fuente: elaboración propia.

### 5.4.1 Definición de objetivos y metas de intervención

El siguiente paso en el diseño de la estrategia es establecer en términos de la intervención, los objetivos y metas a lograr.

#### Objetivos

1.- Aumentar el índice de confiabilidad y validez de las pruebas que se aplican.



- 2.- Establecer un proceso de seguimiento oportuno de resultados en las pruebas para establecer métodos de retroalimentación efectivos para los estudiantes.
- 3.- Estandarizar el diseño de pruebas para tener un control más efectivo de la calidad de los instrumentos de evaluación que se utilizan.
- 4.- Eliminar las inconsistencias en la evaluación de docentes que imparten una misma asignatura.
- 5.- Reducir los factores externos que condicionan el desempeño de las pruebas (áreas, condiciones físicas, horarios, días, eventos, entre otros).
- 6.- Establecer mecanismos metodológicos que permitan evaluar de manera confiable y válida el dominio de conocimiento por parte de los estudiantes.
- 7.- Diseñar reportes técnicos que permitan analizar los atributos de calidad de las pruebas (índice de discriminación, índice de dificultad, índice de correlación de distractores) y con ellos aportar datos válidos para tomas de decisiones oportunas y eficaces.
- 8.- Establecer un programa de capacitación en evaluación estandarizada del aprendizaje a los docentes que conforman la institución educativa.
- 9.- Integrar tecnología en la aplicación y ejecución de las pruebas estandarizadas.

#### **5.4.2 Definición de metas**

- 1.- Los índices de confiabilidad (Índice de dificultad de los ítems, Índice de dificultad de la prueba, Índice de relevancia curricular) y validez (índice de discriminación, correlación ítem-total) en las pruebas cumplen con los valores de calidad definidos por la Teoría de Respuesta al Ítem (TRI).
- 2.- Existe un programa institucional de seguimiento para los resultados obtenidos en las pruebas, cumpliendo con un mínimo de una retroalimentación para cada estudiante en relación al desempeño mostrado en las pruebas.
- 3.- Se cuenta con pruebas estandarizadas para las asignaturas de mayor índice de relevancia curricular.

4.- Se implementa un programa institucional de capacitación para el diseño de pruebas estandarizadas y evaluación de calidad de las mismas.

5.- Las inconsistencias de evaluación entre docentes que imparten una misma materia son erradicadas, al contar con instrumentos que pueden medir el mismo dominio de conocimiento, aunque en su estructura empleen diferentes ítems.

6.- Se aplica una metodología para el diseño de pruebas estandarizadas que contemplen los criterios de calidad definidos por la TRI.

7.- En la institución se cuenta con un grupo colegiado del aprendizaje encargado de gestionar los aspectos evaluativos en la institución, expertos en los análisis psicométricos, diseño de pruebas, capacitación entre otros.

8.- Se cuenta con tres reportes técnicos, diseñados con información clara, precisa y relevante de cada uno de los aspectos medidos en la prueba, para el alumno, el docente, personal administrativo y responsables de evaluación (comité colegiado del aprendizaje).

9.- El examen se realiza utilizando software computarizado que automatiza la ejecución, evaluación y entrega de resultados.

## 5.5 Proyección de la estrategia

### 5.5.1 Definición de indicadores

En esta etapa se definen aquellos indicadores del logro de los objetivos de intervención propuestos, así como la codificación de las variables asociadas a estos. Para cada indicador se deben definir ciertos atributos, como la dimensión cualitativa y cuantitativa que poseen, y para cada variable el cálculo del índice de influencia y dependencia. Los detalles de estos últimos procesos se encuentran en el anexo 3.

**Tabla 12**

*Indicadores que se van a trabajar en la intervención*

Objetivo de intervención	Indicador	Dimensión cualitativa	Dimensión cuantitativa
Aumentar el índice de confiabilidad y validez de las pruebas que se aplican.	Índice de dificultad del ítem Índice de dificultad del examen Índice de relevancia curricular	La dificultad de un ítem se entiende como la proporción de personas que responden correctamente un reactivo de una prueba. Entre mayor sea esta proporción, menor será su dificultad.  El índice de relevancia curricular es aquel valor que expresa la importancia de un contenido en comparación con otros contenidos, enmarcados dentro el currículo de una asignatura. Estos contenidos con alto IR son clave para concretar otros conocimientos dentro de una misma asignatura o incluso para cursos posteriores.	Se desea obtener un índice de dificultad en el examen aproximado entre .5 y .65.  Los ítems deben mantener un índice de dificultad en mayor proporción con valores superiores a .4, sin embargo, deben estar presentes ítems dentro del rango de .2 a .4 para discriminar aún más los grupos altos.  El índice de relevancia curricular se tomará como clave a aquellos que estén superiores o iguales a .7.
	Índice de discriminación  Coeficiente de discriminación	La discriminación se explica de la siguiente manera: si la prueba y un ítem miden la misma habilidad o competencia, se espera que quien tuvo	Se desea obtener un índice de discriminación como mínimo de .2.

	(correlación punto biserial)	<p>una puntuación alta en toda la prueba deberá tener altas probabilidades de contestar correctamente el ítem. Y se espera lo mismo para el caso contrario, es decir aquellos que tuvieron puntuaciones bajas. Así, un buen ítem debe discriminar entre aquellos que obtuvieron buenas calificaciones en la prueba y aquellos que obtuvieron bajas calificaciones.</p> <p>La correlación se utiliza para saber si las personas “adecuadas” son las que obtienen las respuestas correctas, qué tanto poder predictivo tiene el reactivo, y cómo puede contribuir a las predicciones.</p>	Así mismo se desea obtener un valor de rpbis de .2.
Establecer un proceso de seguimiento oportuno de resultados en las pruebas para establecer métodos de retroalimentación efectivos para los estudiantes.	<p>Reporte con resultados de la ejecución de cada estudiante.</p> <p>Reporte con las observaciones y/o recomendaciones (retroalimentación) hechas al estudiante</p>	Es necesario contar con un proceso que permita observar el seguimiento que se les da a los estudiantes de acuerdo al desempeño mostrado en las pruebas. Para ello es necesario un reporte que describa los resultados obtenidos por los estudiantes en la prueba y así mismo uno que deje evidencia de las recomendaciones o sugerencias sobre el aprendizaje que el estudiante demostró en la prueba.	<p>Se desea por lo menos tener un reporte general de resultados obtenidos por los estudiantes, clasificados por asignatura y por maestro cuando aplique.</p> <p>Se desea que como mínimo se de una retroalimentación al 100% de los estudiantes, para ello debe evidenciarse en un reporte.</p>

<p>Eliminar las inconsistencias en la evaluación de docentes que imparten una misma asignatura.</p>	<p>Contenidos temáticos representados en la prueba</p>	<p>Una misma asignatura impartida por más de un docente debe ofrecer la misma probabilidad de acreditar un examen para los estudiantes independientemente de que se curse con docentes diferentes, por ello las pruebas deben evaluar el mismo universo de conocimientos y medir el mismo dominio.</p>	<p>Se desea identificar en un instrumento grafico los contenidos temáticos que se deben tratar en la asignatura y clasificarlos por nivel de relevancia curricular para que estos tengan representatividad en la prueba. En la prueba deben estar representados el 100% de contenidos con un IR mayor o igual a .7.</p>
<p>Reducir los factores externos que condicionan el desempeño de las pruebas (áreas, condiciones físicas, horarios, días, eventos, entre otros).</p>	<p>Calidad de los espacios de aplicación</p>	<p>En el desempeño de una prueba ademas de influir aspectos internos hay factores externos que pueden condicionar la ejecución de la prueba, tales como la hora del día en que se aplica, la ventilación del lugar, distracciones externas, duración de la prueba, exámenes previos o posteriores, clima entre muchos otros, por ello es importante</p>	<p>Se desea utilizar los mismos horarios y el mismo lugar para la aplicación de estos exámenes.  Considerando que el día y la hora definidos no coincidan previamente o posteriormente con exámenes, actividades cultural o deportivas entre otros.</p>
<p>Estandarizar el diseño de pruebas para tener un control más efectivo de la calidad de los instrumentos de evaluación que se utilizan.</p>	<p>Tipo de ítems</p>	<p>Existen números estudios que comprueban la efectividad de cierta representación de ítems (opción múltiple, escalas de valor, relacionales entre otros) con los que se puede medir el dominio de conocimiento en los estudiantes, por lo que para poder evaluar de manera más eficiente las respuestas de los estudiantes se debe estandarizar el tipo de ítems que se utilizan en las pruebas.</p>	<p>Se desea tener una representación de un 90% de ítems de selección en la prueba, dando un margen de solo 10% para preguntas abiertas.</p>

<p>Establecer mecanismos metodológicos que permitan evaluar de manera confiable y valida el dominio de conocimiento por parte de los estudiantes.</p>	<p>Metodología para el diseño de pruebas estandarizadas con análisis psicométrico</p>	<p>Para estandarizar la evaluación del aprendizaje es necesario contar con una metodología que guie el proceso de diseño de las pruebas que se quieran realizar en la institución, y que junto a ello establezcan mecanismos psicométricos para evaluar la calidad de las pruebas. Para ello se utilizará una metodología para desarrollo de exámenes referidos a un criterio (el criterio es el currículo de la asignatura) y el análisis psicométrico se hará en base a la TRI.</p>	<p>Se desea que exista un producto por cada una de las fases que se establecen en la metodología de desarrollo de la prueba para poder hacer valoraciones en periodos más cortos y no una sola al final del todo el proceso.</p> <p>Se desea cumplir con los criterios de calidad psicométrica establecidos por la TRI (índice de dificultad, índice de discriminación, coeficiente de discriminación, correlación ítem-total)</p>
<p>Diseñar reportes técnicos que permitan analizar los atributos de calidad de las pruebas (índice de discriminación, índice de dificultad, índice de correlación de distractores) y con ellos aportar datos válidos para tomas de decisiones oportunas y eficaces.</p>	<p>Número de reportes</p>	<p>Es necesario que se cuente con un reporte técnico diseñado específicamente para describir los resultados psicométricos de la ejecución de la prueba, para poder observar aquellos indicadores de cada ítem y con ello poder hacer las correcciones necesarias a los distractores, a las respuestas correctas entre otros.</p> <p>De igual forma se necesita un reporte técnico que incluya información clara y específica sobre aquellos datos que son relevantes para el alumno (unidad temática con menos dominio, temas específicos con mejor o</p>	<p>Reporte técnico para el alumno con los resultados de la ejecución (promedio, puntaje por temas, por unidad, clasificación del resultado por áreas de mayor y menor dominio entre otros datos).</p> <p>Reporte técnico para el docente con los resultados de la ejecución (promedio grupal, promedio individual, promedios de puntaje por temas, por unidad, clasificación del resultado por áreas de mayor y menor dominio entre otros datos).</p>

		<p>menor dominio etc.), más allá de solo presentar una calificación.</p> <p>Lo mismo sucede con el reporte técnico para el docente, se necesita más que solo promedios grupales e individuales, se requiere conocimiento sobre unidades temáticas que menor grado de dominio y con ello establecer estrategias para reforzar dichos conocimientos, ya que en algunos casos ciertas unidades poseen temas de mayor relevancia curricular y por ende deben ser atendidos más estrictamente.</p>	<p>Reporte técnico para áreas directivas o de coordinación con los resultados de la ejecución (comparativos entre grupos de criterios como promedios, alumnos aprobados, reprobados, distribución de calificaciones entre otros).</p> <p>Reporte técnico con la evaluación psicométrica de la prueba.</p>
<p>Establecer un programa de capacitación en evaluación estandarizada del aprendizaje a los docentes que conforman la institución educativa.</p>	<p>Curso de capacitación y número de docentes participantes</p>	<p>Programa institucional de capacitación docente en el diseño de pruebas estandarizadas del aprendizaje.</p> <p>Para lograr crear una cultura evaluativa en la institución es necesario que participen el mayor número de docentes posible en dichos cursos.</p>	<p>Mínimo un curso de capacitación por semestre.</p> <p>Mínimo de un 80% de docentes participantes en dicho curso.</p>

Fuente: elaboración propia.

### 5.5.2 Diagnóstico de los indicadores

Es necesario identificar el estado actual que presentan los indicadores definidos para cada objetivo de intervención, cabe mencionar que el diagnóstico mostrado a continuación es efectuado en la Facultad de Idiomas de la Universidad Autónoma de Baja California, la cual tiene presencia en los cinco municipios que conforman al estado, sin embargo, la intervención específicamente fue realizada en la unidad académica de la ciudad de Tecate. La siguiente tabla detalla el estado que se tenía para cada uno de los indicadores definidos en la intervención.

**Tabla 13**

*Diagnóstico de los indicadores asociados a los objetivos de intervención*

Objetivo de intervención	Indicador	Estado actual
<p>Aumentar el índice de confiabilidad y validez de las pruebas que se aplican.</p>	<p>Índice de dificultad del ítem Índice de dificultad del examen Índice de relevancia curricular</p>	<p>De las 35 asignaturas obligatorias en el programa educativo de Lic. En Enseñanza de Lenguas, solo 2 cuentan con un estudio psicométrico, por lo que solo el 5.71% de estas asignaturas cuentan con una definición de los indicadores antes mencionados. Estas asignaturas son impartidas en 1ero y 2do semestre, pertenecientes al tronco común, dichas asignaturas son Morfología en el segundo Idiomas y Morfosintaxis en la segunda lengua respectivamente.</p> <p>En el caso de la asignatura de Morfología en el segundo Idioma el índice de relevancia curricular (IRC) de los 36 contenidos a considerar (incluidas practicas) se señalaron 10 contenidos con el mayor IRC.</p>
	<p>Índice de discriminación Coeficiente de discriminación (correlación punto biserial)</p>	



<p>Establecer un proceso de seguimiento oportuno de resultados en las pruebas para establecer métodos de retroalimentación efectivos para los estudiantes.</p>	<p>Reporte con resultados de la ejecución de cada estudiante.</p> <p>Reporte con las observaciones y/o recomendaciones (retroalimentación) hechas al estudiante</p>	<p>De las 35 asignaturas obligatorias en el programa educativo, solo 2 cuentan con un reporte técnico de la ejecución de la prueba por parte del alumno, estas dos asignaturas representan el 5.71% del total de asignaturas en el programa.</p> <p>Ninguna asignatura cuenta con un reporte que especifique la retroalimentación que se le dio al alumno, a cuantos se atendió, cuantas veces se ha atendido a cada alumno o progreso en las áreas oportunidad.</p>
<p>Eliminar las inconsistencias en la evaluación de docentes que imparten una misma asignatura.</p>	<p>Contenidos temáticos representados en la prueba</p>	<p>Solamente el 5.71% (dos asignaturas) contemplan un análisis curricular con el cual se tengan identificados la representatividad de los contenidos temáticos en ítems de la prueba.</p>
<p>Reducir los factores externos que condicionan el desempeño de las pruebas</p>	<p>Calidad de los espacios de aplicación</p>	<p>Solo para la aplicación de dos exámenes se toman en cuenta los criterios de tiempo, espacio y momento para ser realizados, sin embargo, no existe un programa</p>

<p>(áreas, condiciones físicas, horarios, días, eventos, entre otros).</p>		<p>institucional que estandarice los criterios o buenas prácticas a considerar al momento de aplicar una prueba.</p>
<p>Estandarizar el diseño de pruebas para tener un control más efectivo de la calidad de los instrumentos de evaluación que se utilizan.</p>	<p>Tipo de ítems</p>	<p>El 5.71% de las asignaturas de todo el programa educativo aplican una misma estructura para la prueba, utilizan ítems de opción múltiple. El restante 94.29% de las asignaturas aplican exámenes con diversos tipos de ítems, que pueden variar incluso en el tiempo aunque se trate del mismo contenido temático a evaluar.</p>
<p>Establecer mecanismos metodológicos que permitan evaluar de manera confiable y válida el dominio de conocimiento por parte de los estudiantes.</p>	<p>Metodología para el diseño de pruebas estandarizadas con análisis psicométrico</p>	<p>El 94.29% de las asignaturas impartidas en el plan de estudios no siguen una metodología para el diseño de las pruebas, estas se construyen con base en la experiencia y criterio de cada docente, lo que significativamente pone de manifiesto que las pruebas pueden poseer elementos determinados solamente por juicios personales.</p>
<p>Diseñar reportes técnicos que permitan analizar los atributos de calidad de las pruebas (índice</p>	<p>Número de reportes</p>	<p>Para el 5.71% de las asignaturas se cuenta con tres reportes técnicos, uno dirigido al alumno, otro al docente y uno más para las áreas de coordinación o directivos.</p>

<p>de discriminación, índice de dificultad, índice de correlación de distractores) y con ellos aportar datos válidos para tomas de decisiones oportunas y eficaces.</p>		<p>El restante 94.29% no cuentan con ningún tipo de reporte técnico o general que señale el rendimiento del alumno en la prueba.</p>
<p>Establecer un programa de capacitación en evaluación estandarizada del aprendizaje a los docentes que conforman la institución educativa.</p>	<p>Curso de capacitación y número de docentes participantes</p>	<p>No se cuenta con ningún programa institucional que capacite a los docentes en evaluación estandarizada del aprendizaje. Solo el 4.76% (1 docente) de los profesores que componen la planta académica en la institución tiene la capacitación en el diseño de pruebas.</p>

Fuente: elaboración propia.

### 5.5.3 Análisis diagnóstico de las variables definidas

Para el análisis de la situación presente en la institución donde se realizó la intervención se realizaron una serie de encuestas, la mayoría están dirigidas a los 21 docentes que conforman la planta académica (algunos de estos docentes a su vez desempeñan puestos de coordinación o dirección) puesto que muchas de estas variables están asociadas a prácticas o actividades desempeñadas por docentes, sin embargo también hay algunas que están asociadas a aspectos de infraestructura, alumnos (se tomó una muestra de 72 alumnos de los 8 semestres que contempla carrera, la población total de alumnos en este ciclo escolar es de 141) y a procesos.

**Tabla 14***Análisis diagnóstico del estado actual de las variables definidas*

<b>Variable</b>	<b>Estado actual</b>
<b>Profesores no capacitados</b>	Solo el 4.76% (1 docente) de los profesores que componen la planta académica en la institución tiene la capacitación para elaboración de pruebas estandarizadas y análisis psicométrico. Sin embargo el 14.28% (tres docentes) tienen cursos y/o talleres enfocados específicamente a la evaluación educativa.
<b>Tiempo para desarrollar pruebas</b>	El 76.19% de los docentes consideran que es muy tardado hacer una prueba estandarizada, razón por la cual deciden hacer pruebas con otro tipo de métodos.
<b>Resistencia a estandarizar criterios</b>	El 57.14% de los docentes considera que es complicado estandarizar criterios de evaluación, consideran que requiere análisis extensos para encontrar aquellos medios o atributos que permitan lograr medir los mismos aspectos del conocimiento.
<b>Auto apreciación de expertos</b>	El 100% de los docentes comenta que tiene la capacidad de desarrollar un examen para evaluar cualquier conocimiento asociado a la asignatura que imparte.
<b>Libertad de catedra</b>	El 52.38% considera que estandarizar una prueba es una forma de limitar su libertad de catedra lo que implica la forma y/o instrumentos con los que evalúa el aprendizaje de sus estudiantes.
<b>Ausencia de tecnología</b>	Solo para el 5.7% de las asignaturas se posee un examen estandarizado en un formato digital, el cual es aplicado mediante un software diseñado a la medida. Sin embargo, el 48 de las asignaturas cuentan con la posibilidad de ser impartidas de manera semi-presencial por lo que se apoyan en plataformas como Blackboard, desde la cual los exámenes se realizan de forma digital. Sin embargo, para que un docente pueda dar una asignatura en esta modalidad tiene que haber cubierto un mínimo de 4 cursos de capacitación en el uso de estas herramientas y diseño de cursos.

<b>Ausencia de análisis psicométricos</b>	Solo para el 5.7% de las asignaturas se realiza el análisis psicométrico de las pruebas que se aplican.
<b>Cultura evaluativa</b>	El 61.90 considera que su clase ha funcionado con los instrumentos y estrategias que han aplicado, por lo que no consideran relevante querer implementar algunos otros, pero casi el 40% considerarían analizar otras alternativas si hay buenos fundamentos.
<b>Objetivos de las pruebas</b>	El 85.71% de los docentes consideran como objetivo final del examen el observar la calificación numérica que estos obtienen, dejando por un lado la identificación de los temas en los que menor desempeño hubo, o características comunes en el grupo en cuanto ciertos dominios de conocimiento.
<b>Métodos estandarizados ausentes</b>	El 95.23% de la planta académica desarrolla sus pruebas siguiendo métodos diseñados por ellos mismos, o consultando material en internet.
<b>Resistencia a ser observado</b>	El 85.71% de los docentes considera que se deben estipular bien los usos de la prueba y garantizar la privacidad de la información para poder participar en una prueba estandarizada, ya que se tiene la creencia de que ese tipo de pruebas son para evidenciar malas prácticas o bajo desempeño.
<b>Estímulo a evaluar la calidad de las pruebas</b>	El 76.19% de los docentes consideran que sus instrumentos son de calidad, aunque no los sometan a ningún tipo de juicio o revisión por pares, ni ningún tipo de análisis. El restante 23.81 considera que le gustaría saber cómo poder medir la calidad de sus exámenes.
<b>Resistencia a compartir estrategias</b>	El 80.95% de los docentes consideran que el material que desarrollan no debe ser compartido libremente ya que requirió de un esfuerzo propio, sin embargo es importante destacar que el 90.47 estaría dispuesto a trabajar en un trabajo donde el producto sea la integración de las actividades de cada uno, solo el 9.53% mantuvo la postura de no trabajar en proyectos de esa naturaleza.

<p><b>Impacto de la prueba</b></p>	<p>El 81.94% de los alumnos considera que es importante saber la calidad de la prueba, y qué tanto valor tiene un examen en relación a la calificación final del semestre. El 91.66% manifestó sentirse más seguro de hacer una prueba en la que tengan la seguridad del contenido específico que se evaluara. En el caso de los docentes el 95.23% expresó que es importante saber cuál será el uso del resultado de las pruebas, si tienen consecuencias para ellos y de qué tipo, o si son formativas. El 100% considero elemental saber para que servirá la prueba en relación a su deseo de participar o no en un proceso de estandarización.</p>
<p><b>Operatividad</b></p>	<p>Se realizó un estudio de las características de la infraestructura tecnológica (software y hardware) y de red, se cuenta con dos laboratorios con capacidad para 28 y 24 usuarios respectivamente. Actualmente ningún grupo supera los 28 estudiantes, sin embargo, a lo largo de la historia se han presentado grupos con hasta 28 alumnos, lo que condiciona la capacidad de los laboratorios para efectuar un solo examen para todo el grupo.</p> <p>Esto es importante cuidar y señalar para efectos de que las condiciones del examen sean las mismas para todos, lo que incluye el espacio donde se realiza la prueba, el horario, el día entre otros aspectos.</p> <p>En el siguiente apartado de recursos se dan más detalles de esta variable y los aspectos a considerar.</p>
<p><b>Viabilidad</b></p>	<p>En este sentido actualmente y con los recursos con los que se dispone es posible una puesta a prueba y operación de la propuesta de intervención, sin embargo, hay que afinar ciertos detalles para mejorar los porcentajes de aceptación y participación por parte del personal docente principalmente, en cuanto a la capacidad de respuesta tecnológica se cuenta para cubrir el 100% de los estudiantes y las asignaturas que se imparten si se desearan aplicar exámenes estandarizados. Sin embargo actualmente solo se cuenta con dos exámenes estandarizados los cuales serán la muestra para poner en practica la intervención.</p>

Fuente: elaboración propia.

#### 5.5.4 *Acciones propuestas*

En este apartado se registran aquellas acciones estratégicas que permitirán alcanzar los objetivos y metas planteadas en etapas anteriores. Para cada estrategia se tienen que definir ciertos atributos y características. Para acción se recopilan los siguientes datos:

- Nombre de la acción estratégica
- Descripción general
- Actores involucrados
- Participación esperada de los actores
- Plazos de realización
- Objetivo que atiende
- Meta con la cual se relaciona
- Productos y resultados esperados

Además de detallar la información antes mencionada, también se hace un análisis de cada una de las acciones en relación a los siguientes aspectos:

- Obstáculos a los que se enfrentaría
- Resistencias que pueden presentarse entre los actores involucrados
- Recursos requeridos (infraestructura, equipos, materiales, etc.)
- Recursos humanos (personal de apoyo que se requiere para el desarrollo de la acción)
- Políticas y normas institucionales existentes vinculadas a la idea o que se requieren para desarrollarla.
- Valores institucionales que son importantes de recuperar, promover o fortalecer para el correcto desarrollo de la acción estratégica
- Consideraciones adicionales

Puesto que este apartado es de los más extensos las acciones son listadas y descritas en el anexo 4.

### 5.5.5 Evaluación de las acciones

En este apartado se efectúa un análisis detallado sobre las acciones definidas anteriormente, la intención de este proceso es evaluar cuales acciones son más factibles, viables y pertinentes para proyectar la estrategia de intervención. El puntaje asignado para cada criterio está en una escala de 3 a 1, donde 3 corresponde a un nivel alto, 2 a un nivel suficiente y 1 a un nivel insuficiente. El análisis resultante se muestra en la siguiente tabla:

**Tabla 15**

*Evaluación de las acciones estratégicas*

	Factibilidad	Viabilidad	Pertinencia	Congruencia teórico metodológica	Aceptación de los involucrados.	TOTAL
1. Valorar la calidad de la prueba.	3	3	3	3	2	14
2. Estandarizar métodos de construcción y/o diseño de pruebas	3	3	3	3	2	14
3. Diseñar reportes técnicos de la ejecución de la prueba	3	3	3	3	3	15
4. Crear un grupo colegiado de evaluación del aprendizaje (GCEA)	2	1	2	1	1	7
5. Diseñar el proceso de	2	2	2	1	2	9



seguimiento a los resultados de las pruebas.						
6. Formar docentes con habilidades y competencias en evaluación del aprendizaje.	3	3	3	2	2	13
7. Integrar tecnología en la aplicación de la prueba.	3	3	3	3	3	15
8. Estandarizar las condiciones de ejecución de la prueba.	3	3	3	2	2	13

Fuente: elaboración propia.

Dirección General de Bibliotecas de la UAQ

### 5.5.6 Selección de las acciones

Después de realizar la evaluación de las estrategias se han elegido las siguientes estrategias, y se detalla la justificación de la elección.

**Tabla 16**

*Justificación de la elección de estrategias*

	<b>Elección</b>	<b>Justificación</b>
<b>1. Valorar la calidad de la prueba</b>	SI	Esta estrategia si puede ser implementada en la intervención, ya que se cuenta con una persona capacitada en el análisis psicométrico de los criterios de calidad para una prueba estandarizada con base en la TRI. Se cuenta con el personal docente necesario para integrar cada uno de los comités que se requieren para desarrollar la prueba, y en este caso, específicamente para la Producción y Validación de Ítems (Etapa cuatro del ciclo de desarrollo de la prueba).
<b>2. Estandarizar métodos de construcción y/o diseño de pruebas</b>	SI	Esta estrategia para esta intervención sería aplicada específicamente para los exámenes ordinarios o departamentales de las asignaturas. Puesto que estos exámenes evalúan todo el contenido de la asignatura, es decir todo el currículo, situación que empata directamente con las características de la metodología seleccionada para desarrollar pruebas estandarizadas (exámenes referidos a un criterio, por criterio se refiere al currículo de la asignatura). Posteriormente se podría migrar a exámenes parciales o de cortes de unidades temáticas por mencionar algunos.
<b>3. Diseñar reportes</b>	SI	Al contar con un experto en evaluación estandarizada del aprendizaje y evaluación psicométrica, este personal

<p><b>técnicos de la ejecución de la prueba</b></p>		<p>puede encargarse de diseñar los reportes y de incluir la información técnica correspondiente a los resultados de la prueba. Aunque de igual forma, se cuenta con personal docente que puede integrar e comité diseñador de la prueba, lo que incluye participar en el diseño y elaboración de los reportes técnicos para los diferentes usuarios.</p>
<p><b>4. Crear un grupo colegiado de evaluación del aprendizaje (GCEA)</b></p>	<p>NO</p>	<p>Esta estrategia para este proyecto de intervención se encuentra fuera del alcance, ya que requiere de ciertos procesos y gestiones administrativas propias de la institución educativa. Situación que es ajena a los objetivos primordiales de la intervención, que es implementar pruebas estandarizadas referidas a un criterio para la evaluación del aprendizaje integrando tecnología. Si bien contar con un grupo facilitaría el desarrollo de más pruebas en menor tiempo, encargarse de actividades de capacitación, talleres entre otros, por el momento no es viable. Además, el trabajo de investigación no trata sobre creación de comités o grupos de trabajo, pero pudiera ser contemplado para una futura intervención.</p>
<p><b>5. Diseñar el proceso de seguimiento a los resultados de las pruebas.</b></p>	<p>NO</p>	<p>Aunque si bien este es un complemento clave para poder obtener el mayor beneficio de contar con pruebas de calidad, el trabajo de investigación no contempla información teórica o referentes metodológicos para el diseño de procesos para el seguimiento de trayectoria estudiantil. Sin embargo, este también sería un apartado contemplado para una futura intervención, la cual se enfoque especialmente en el diseño y evaluación de un modelo específico para el seguimiento del rendimiento</p>

		<p>escolar, un esquema de toma de decisiones entre otros aspectos.</p> <p>La intervención alcanza solamente a unas primeras aproximaciones de tomas de decisiones con los resultados obtenidos en las pruebas y representados en los reportes técnicos, pero no tan complejo como un proceso de trayectoria.</p>
<p><b>6. Formar docentes con habilidades y competencias en evaluación del aprendizaje.</b></p>	SI	<p>Esta estrategia viene implícita en la aplicación de la metodología para el desarrollo de pruebas estandarizadas, ya que para cada comité que se integra en cada una de las etapas del desarrollo de la prueba hay un proceso de capacitación por parte del Responsable de Evaluación del Aprendizaje, con lo cual los docentes adquieren habilidades para desempeñar ciertas tareas, como hacer un análisis curricular, establecer índices de relevancia curricular, justificar contenidos, elaborar especificaciones de ítems, producir ítems entre otros.</p>
<p><b>7. Integrar tecnología en la aplicación de la prueba.</b></p>	SI	<p>La institución cuenta con el recurso tecnológico y de infraestructura para poder desarrollar un software especialmente para construir y aplicar exámenes. También se cuenta con el personal capacitado para diseñar y programar en diferentes lenguajes. Se cuenta con un espacio reservado y dedicado para desarrollo y administración tecnológica en la unidad académica.</p> <p>La unidad académica posee dos laboratorios de computo con capacidad de hasta 28 personas, con equipo de última generación, por lo cual hay una garantía de que el equipo que actualmente se utiliza en la institución posee los</p>

		requisitos de software y hardware necesarios para ejecutar el sistema de exámenes.
<b>8. Estandarizar las condiciones de ejecución de la prueba.</b>	SI	<p>Existen espacios suficientes para atender la población estudiantil, considerando que las pruebas en esta primera intervención son las departamentales u ordinarias, se cuenta con una mayor disponibilidad de horario para el uso de los laboratorios, actualmente ningún grupo supera los 28 alumnos, motivo por el cual todos los grupos realizarían la prueba en las mismas condiciones. Los laboratorios cuentan con clima controlado, iluminación recientemente mejorada, y se encuentran ubicados de forma que no hay distracciones visuales inmediatas.</p> <p>Referente a interferencias sonoras del exterior, también se encuentran en espacios alejados de plazas o áreas deportivas, por lo que en dado caso de presentarse actividades en alguna de estas áreas no afectarían durante la ejecución de la prueba.</p>

Fuente: elaboración propia.

El siguiente paso en el proceso de la proyección de las acciones estratégicas es detallar cada una de las etapas que las componen. Posteriormente cada una de estas etapas tienen que ser descritas específicamente en actividades, y para cada una de estas actividades se describen elementos como el tiempo, participantes, estrategia de seguimiento y los recursos necesarios para desarrollar cada una de ellas.

## 5.6 Descripción específica de la estrategia

A continuación, se ejemplifican para la primera estrategia los aspectos antes mencionados, las restantes estrategias pueden ser consultadas en el anexo 6.

**Tabla 17**

*Estrategia #1: Valorar la calidad de la prueba*

ETAPA	OBJETIVO(S)	INDICADORES DE LOGRO
<b>1. Diseñar especificaciones de ítems</b>	<p>Una especificación de ítems expresa de manera clara cuales son los elementos que debe incluir el ítem en la prueba, es el manual con las instrucciones para construir el ítem.</p> <p>Esta etapa se encarga de desarrollarla el comité de especificaciones de ítems, y es supervisado por el responsable de evaluación colegiada del aprendizaje (ECA).</p>	<p>Especificación de ítem(s) de cada uno de los contenidos temáticos que se juzgaron como pertinentes de ser incluidos en la prueba.</p>
<b>2. Producir ítems de acuerdo a las especificaciones</b>	<p>Se construyen los ítems de acuerdo a las especificaciones definidas para ellos, el ítem debe reflejar todos los elementos que la especificación detalla.</p> <p>Esta etapa se encarga de desarrollarla el comité de producción de ítems, y es supervisado por el responsable de evaluación colegiada del aprendizaje.</p>	<p>Ítems (la cantidad se define en la etapa de justificación de contenidos) producidos por cada contenido temático que la prueba incluye.</p>
<b>3. Pilotaje de los ítems con una muestra de la población</b>	<p>Se deben de probar en el ambiente final de ejecución los ítems producidos, se construye con los ítems una prueba piloto que es aplicada por una muestra de la población estudiantil.</p>	<p>Cantidad de alumnos participantes en el pilotaje.</p> <p>Reporte de incidencias o</p>

	Esta etapa es llevada a cabo por el responsable de ECA, y el responsable de pruebas del software que se utiliza para la aplicación del examen.	ejecución de la prueba en el sistema.  Base de datos con la información recabada.
<b>4.- Recopilación de los datos (para análisis con el software especializado)</b>	Los datos de la aplicación del examen deben ser vaciados en un archivo plano y formateado con una estructura específica de acuerdo al software especializado con el que se van a analizar los datos producidos (respuestas de cada sustentante)  Esta etapa es llevada a cabo por el responsable de ECA, y el responsable de pruebas del software que se utiliza para la aplicación de la prueba.	Archivos fuente formateados para cada software especializado (ITEMAN, TAP)
<b>5.- Análisis psicométrico de los resultados</b>	Después de ejecutar las pruebas de validez y confiabilidad a cada uno de los ítems se determina la calidad de cada ítem y de la prueba en general.  Esta etapa es llevada a cabo por el responsable de ECA, y el comité de evaluación de ítems.	Índice de dificultad del ítem y el examen  Índice de discriminación  Coeficiente de discriminación
<b>6.- Reporte técnico del análisis</b>	Los resultados obtenidos con el análisis psicométrico se integran en un reporte técnico que explica cada valor obtenido y se emiten criterio de juicio sobre la calidad de cada ítem.  Esta etapa es realizada por el responsable de ECA.	Reporte técnico con los resultados obtenidos, incluyendo el juicio de cada ítem.

Fuente: elaboración propia.

### 5.6.1 Vinculación de las etapas en una cadena lógica de resultados

Esta actividad consiste en elaborar un diagrama de flujo que permita observar como cada etapa de la acción estratégica se relacionan entre sí, y producen una secuencia lógica de insumos y productos entre las mismas.

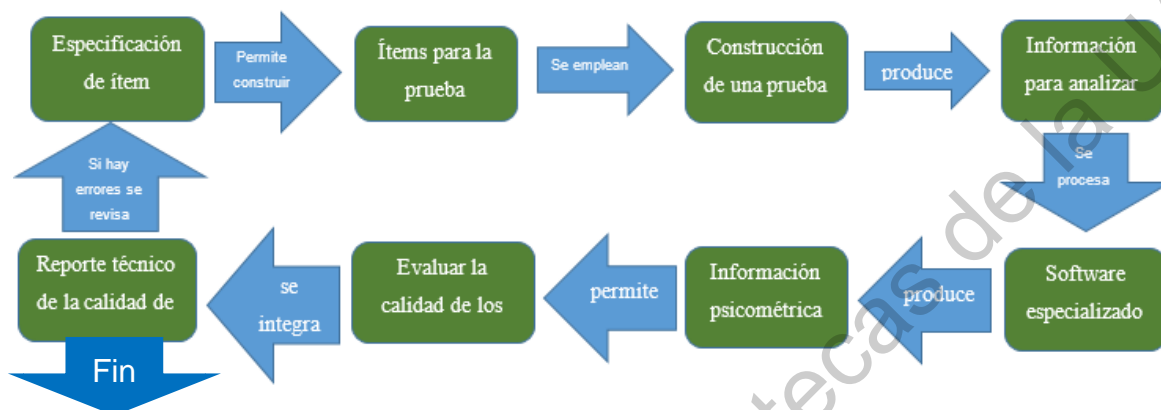


Figura 18. Diagrama de flujo de las etapas dentro de la acción estratégica.

Fuente: elaboración propia.

### 5.6.2 Descripción de las actividades que componen cada etapa

Este proceso implica describir para cada una de las etapas que componen a cada acción estratégica todas las acciones necesarias que deben realizarse para lograr el objetivo previsto, señalando ciertos criterios de cumplimiento como los el tiempo y los recursos necesarios.

**Tabla 18**

*Actividades de la estrategia 1, etapa 1.*

ETAPA 1: Diseñar especificaciones de ítems				
<b>OBJETIVO(S):</b>	Contar con las guías de diseño para la producción o construcción de ítems de algún contenido temático que va a ser incluido en la prueba			
<b>INDICADOR:</b>	Especificación de ítem(s) de cada uno de los contenidos temáticos que se juzgaron como pertinentes de ser incluidos en la prueba.			
Actividades	<b>Tiempo</b>	<b>Participantes</b>	<b>Estrategias de seguimiento</b>	<b>Recurso</b>
<b>1.- Consultar el documento de justificación de contenidos para</b>	1-2 hrs, ya que además de consultarla hay que producir un	Responsable de ECA: se encarga de revisar el documento de	Lista de cotejo con la cantidad de ítems por	Documento con la justificación de contenidos (producto



<b>saber cuántos ítems deben ser producidos por cada contenido temático.</b>	nuevo producto el cual solo incluya contenido temático y cantidad de ítems requeridos	justificación de contenidos para especificar la cantidad de ítems que se producirá con cada especificación (Requiere conocimientos en la elaboración de justificación de contenidos).	contenido temático.	generado con la etapa anterior de la metodología de desarrollo de la prueba)
<b>2. Capacitación del comité elaborador</b>	2 hrs	Responsable de ECA: capacita al comité encargado de construir las especificaciones de ítems (requiere amplia experiencia y experticia en el diseño de especificaciones de ítems).	Lista de asistencia de participantes.  Minuta con las observaciones y/o comentarios de la capacitación.  Evaluación de calidad de la capacitación	Espacio para dar la capacitación, proyector, ordenador
<b>3.- Elaborar la especificación del ítem por cada contenido temático.</b>	Para esta actividad el tiempo estimado es de 15 días (considerando cantidad de participantes y los tiempos de disponibilidad).	El comité de especificaciones de ítems: se encarga de construir las especificaciones de ítems. (para esta función se requiere de un dominio amplio y experto de la asignatura para la cual se desarrolla la prueba).	Modelo Canvas (por hacer / haciendo / terminado) de cada una de las especificaciones a desarrollar.  Lista de cotejo con las especificaciones ya revisadas y evaluadas por el responsable de ECA.	Estaciones de trabajo con el software procesador de palabras.  Plantilla y/o formato de especificación.  Espacio físico para que el comité trabaje en la elaboración.

Fuente: elaboración propia.

**Tabla 19***Actividades de la estrategia 1, etapa 2.*

ETAPA 2: Producción de ítems				
<b>OBJETIVO(S):</b>	Se construyen los ítems de acuerdo a las especificaciones definidas para ellos, el ítem debe reflejar todos los elementos que la especificación detalla.			
<b>INDICADORE(S):</b>	Ítems (la cantidad se define en la etapa de justificación de contenidos) producidos por cada contenido temático que la prueba incluye.			
Actividades	Tiempo	Participantes	Estrategias de seguimiento	Recurso
<b>1.- Revisión de las especificaciones de ítems</b>	1 hr	Responsable de ECA: presenta al comité de producción de ítems las especificaciones diseñadas (se requiere amplia experiencia en el diseño y uso de especificaciones de ítems).	Reporte con los comentarios y/u observaciones realizadas a las especificaciones.	Especificaciones de ítems  Proyector  Ordenador  Lugar (cubículo, laboratorio)
<b>2.- Capacitación del comité elaborador</b>	2 hrs	Responsable de ECA: capacita al comité encargado de construir las especificaciones de ítems (se requiere amplia experiencia y experticia en la producción de ítems con base en especificaciones de ítems).	Lista de asistencia de los participantes.  Minuta de las observaciones y/o comentarios de los participantes.  Evaluación de calidad de la capacitación.	Espacio para dar la capacitación, proyector, ordenador
<b>3.- Producir ítems de acuerdo a las especificaciones</b>	Para esta actividad y la cantidad de ítems que se tienen previstos, el tiempo estimado es de	Responsable de ECA: se encarga de revisar y monitorear la producción de ítems (se requiere amplia experiencia y	Modelo Canvas (por hacer / haciendo / terminado) de cada una de los ítems por desarrollar.	Estaciones de trabajo con el software procesador de palabras.

	10 días (considerando cantidad de participantes y los tiempos de disponibilidad).	<p>experticia en la producción de ítems).</p> <p>El comité de producción de ítems: desarrolla los ítems que cada contenido temático requiere (para desempeñar esta función se requiere de un dominio amplio y experto de la asignatura para la cual se desarrolla la prueba).</p>	<p>Lista de cotejo con las especificaciones ya revisadas y evaluadas (estructura y forma) por el responsable de ECA.</p>	<p>Especificación del ítem.</p> <p>Espacio físico para que el comité trabaje en la producción de ítems.</p>
--	---	---	--	---

Fuente: elaboración propia.

### 5.7 Integración de la intervención con la evaluación actual

A nivel institucional las pruebas departamentales se han definido como de carácter obligatorio para todos los estudiantes. Como ya se mencionó, este proyecto de investigación trata el diseño y aplicación de una prueba estandarizada departamental para la asignatura de Morfología de la Segunda Lengua, asignatura cuya relevancia curricular es una de las dos más altas para el tronco común de lenguas en la Facultad de Idiomas de la UABC.

Estas pruebas departamentales tienen un peso que oscila entre el 10% y 15% de la calificación total que el estudiante obtiene al final de semestre, es decir, lo acumulado con las evaluaciones formativas que el estudiante realiza durante todo el semestre tienen un valor del 85%, y el restante se obtiene a partir del resultado de la prueba departamental. El docente decide cuanto porcentaje asignarle a la evaluación, sin embargo, debe encontrarse entre los límites antes mencionados.

Por otra parte, como parte de la socialización de este tipo de pruebas, al inicio de cada semestre se muestra a cada uno de los grupos que participan en la asignatura en cuestión la estructura y formato del examen, se tienen una versión con ítems muestra de la prueba, con

el fin de que los alumnos se familiaricen con el formato de la prueba y las características de la misma.

Por otro lado, en el caso de los docentes se tiene una plática especial con aquellos que vayan a impartir la materia para explicarles la estructura general del examen, pero de manera detallada, mencionando aspectos como el universo de medida, cuantos ítems hay por cada contenido temático, se muestran las tablas de especificaciones de ítems, la retícula y algunos ítems muestra (los instrumentos antes mencionados pueden ser observados a detalle en el apartado de resultados).

Así mismo se les comunica cual es el propósito de la prueba y el impacto que tiene, así como el fin o uso que tienen los resultados recabados con la aplicación de la prueba. Esto se realiza con la intención de estimular el uso adecuado de la prueba, reduciendo o restringiendo que se utilice con fines distintos a los que pretende.

#### **5.7.1 Evaluación formativa y evaluación sumativa**

Como se ha mencionado, el proceso de intervención implica la aplicación de exámenes estandarizados departamentales. Los análisis psicométricos realizados a la prueba departamental de morfología de la segunda lengua permiten evidenciar la calidad del instrumento, lo que ofrece una seguridad al momento de tomar decisiones sobre el nivel de dominio que los estudiantes reflejan en la prueba. Es un hecho conocido que la actividad evaluativa no se limita a un solo instrumento o a una sola acción, ya que la importancia de evaluar se ve completa al tomar acciones correctivas con base a los resultados obtenidos tanto en las actividades evaluativas diarias frente a clase, evaluaciones parciales y evaluaciones semestrales como en el caso de los exámenes departamentales.

Dicho de otra manera, la evaluación debe complementarse entre las que son de tipo formativo y las que son de tipo sumativa, que en el caso de esta investigación se habla de una evaluación de carácter más formativo. Sin embargo, no hay duda de que un instrumento de esta naturaleza permite apreciar de una manera válida y confiable el nivel de dominio de un estudiante al momento de ejecutar la prueba, claro está todo esto dentro del marco de lo establecido en el currículo de la asignatura.

Como se ha expresado anteriormente por medio de este tipo de evaluación se trata de comprobar el nivel de dominio alcanzado en forma individual por el estudiante; lo anterior desde luego no debe significar un conocimiento totalmente nuevo para los implicados hablando del docente y el alumno, puesto que previo a la realización de la prueba departamental existe ya un sustento del nivel de dominio alcanzado por cada alumno como resultado de todas las actividades evaluativas de carácter formativo que se han realizado durante el semestre. Esto último es muy importante para el proceso de intervención, ya que permite pronosticar un posible resultado de la evaluación sumativa en relación a lo logrado con las formativas, significando así que ambos resultados tendrían que mantener una similitud. Si lo antes mencionado no se presentara, se hablaría entonces de una posible deficiencia en la calidad del procedimientos o instrumentos evaluativos empleados de manera formativa, situación altamente probable considerando el hecho de que este tipo de evaluaciones (formativas) en la mayoría de los casos no emplean una metodología que de sustento a la calidad de a mismas.

### **5.7.2 Retroalimentación**

Hablando particularmente de la enseñanza universitaria, algunos autores como Ferguson, Tillema, Leenknecht & Segers (en Canabal y Margalef, 2017) explican las percepciones de los estudiantes en relación a lo que se considera una retroalimentación de calidad de acuerdo a las formas, los detalles, los tipos y los tiempos, y el hecho sobresaliente de que los estudiantes valoran positivamente una retroalimentación significativa orientada a su aprendizaje.

En este sentido la intervención como parte de la metodología del diseño de la prueba implementa una serie de reportes técnicos detallados, los cuales están diseñados precisamente con fines de retroalimentación tanto para el docente como para el alumno. De esta manera al finalizar el curso, si el alumno no recibió una retroalimentación considerable durante sus procesos evaluativos formativos a lo largo del semestre, podrá obtener una retroalimentación detallada a nivel de todo el universo de conocimientos implicados en la asignatura (pueden consultarse los ejemplos de estos reportes técnicos en el anexo 8).

Estos reportes técnicos están diseñados con un nivel de abstracción tal que permiten al estudiante ubicar su nivel de dominio a nivel de unidad de aprendizaje, tema, sub tema, pico tema y así respectivamente. Además, se incluyen detalles específicos sobre cuántos aciertos de los temas de mayor relevancia curricular obtuvo por unidad, y desde luego a nivel global en toda la prueba.

Lo anterior permite al estudiante conocer de manera más específica un panorama sobre su logro académico, el aprendizaje alcanzado y superar la limitante final de solo observar una nota o calificación en la prueba.

Hablando en el caso del docente, los reportes técnicos le permiten observar el comportamiento del grupo a nivel de unidades de aprendizaje, temas, subtemas, pico temas etc., además se destacan aquellos de mayor relevancia curricular. Con esta información el docente puede encausar o reflexionar sobre las estrategias aplicadas durante el proceso de enseñanza y aprendizaje, así como durante la evaluación formativa realizada.

Puesto que la prueba departamental tiene un sustento metodológico y psicométrico que demuestra la validez y confiabilidad de la prueba, el docente está en condiciones seguras de poder observar deficiencias entre la evaluación sumativa que representa la prueba estandarizada departamental y las evaluaciones formativas que implemento en el curso. Por lo que la retroalimentación no solo se limita al alumno, sino que, por su parte, el docente también resulta beneficiado.

A continuación, se muestra un ejemplo de uno de los reportes técnicos que sirven como instrumento de retroalimentación para el alumno:

## REPORTE DE RESULTADOS

Alumno: n#n#n#



El examen de morfología aplicado contenía 63 ítems. Tu puntaje obtenido fue de:

47 ítems acertados

Obteniendo un promedio (en escala de 0 a 100) general de:

74.60

El promedio general del grupo fue de:

83.78

### Informe detallado por unidad de aprendizaje

	UNIDAD DE APRENDIZAJE 1	Temas de mayor relevancia curricular (IRC)	Temas dominados de mayor IRC	Temas no dominados de mayor IRC
Distribución de aciertos	Ítems acertados: 10/13	Ítems acertados: 2/3	1.3.1	
	Calificación en la unidad: 76.92 /100	Porcentaje de dominio: 66.66%	Morphemes and allomorphs 1.3.3 Roots, affixes	1.2.2 Signs and morphemes
	Nivel de dominio: Satisfactorio			

Figura 19. Ejemplo del reporte técnico de resultados para el alumno.

Fuente: elaboración propia.

Como se puede observar en la figura anterior, el reporte técnico ofrece información significativa para el alumno, como lo es los temas de mayor relevancia curricular por unidad de aprendizaje que el alumno domina, cabe destacar que este reporte se enfoca en el tipo de retroalimentación centrada en la tarea, brindando así información sobre los logros, aciertos, errores etc. Este es un aporte significativo a la retroalimentación que el alumno obtiene de manera automática con la finalización del proceso de evaluación de la prueba estandarizada, destacando la importancia y el beneficio que la herramienta tecnológica innovadora ofrece para la mejora de la retroalimentación. De esta manera el alumno obtiene de manera directa información que él puede interpretar, generando así también un estímulo a la retroalimentación auto dirigida, permitiendo al estudiante reflexionar sobre su aprendizaje en aspectos concretos, sin necesidad que esta venga directamente del docente.

Aunado a lo anterior, puede complementarse la retroalimentación al estudiante con una retroalimentación enfocada en la propia persona, sin embargo, esto ya implica una participación directa del docente en la cual destaque del estudiante aspectos como el esfuerzo y compromiso manifestado con el proceso de aprendizaje.

### **5.8 Sistematización del proceso de intervención**

Para lograr representar de una manera más clara el proceso de intervención se optó por elaborar un diagrama de flujo que representa gráficamente las diferentes etapas del proceso (ver figura 20), y como se ven involucradas ciertas acciones, tales como el plan de gestión de riesgos, la proyección de las estrategias interventivas y por último la evaluación de la intervención (de este proceso se dan más detalles en el siguiente apartado). Sin embargo, este proceso es muy extenso, y para lograr dar una claridad y detalle más específico de cada proceso se realizaron 8 diagramas correspondientes a los subprocesos más relevantes del proceso general de intervención. Algunos sub procesos y sus respectivos diagramas pueden ser revisados en el anexo 7. Así mismo se presenta una descripción de las etapas que componen este proceso de automatización de la intervención.



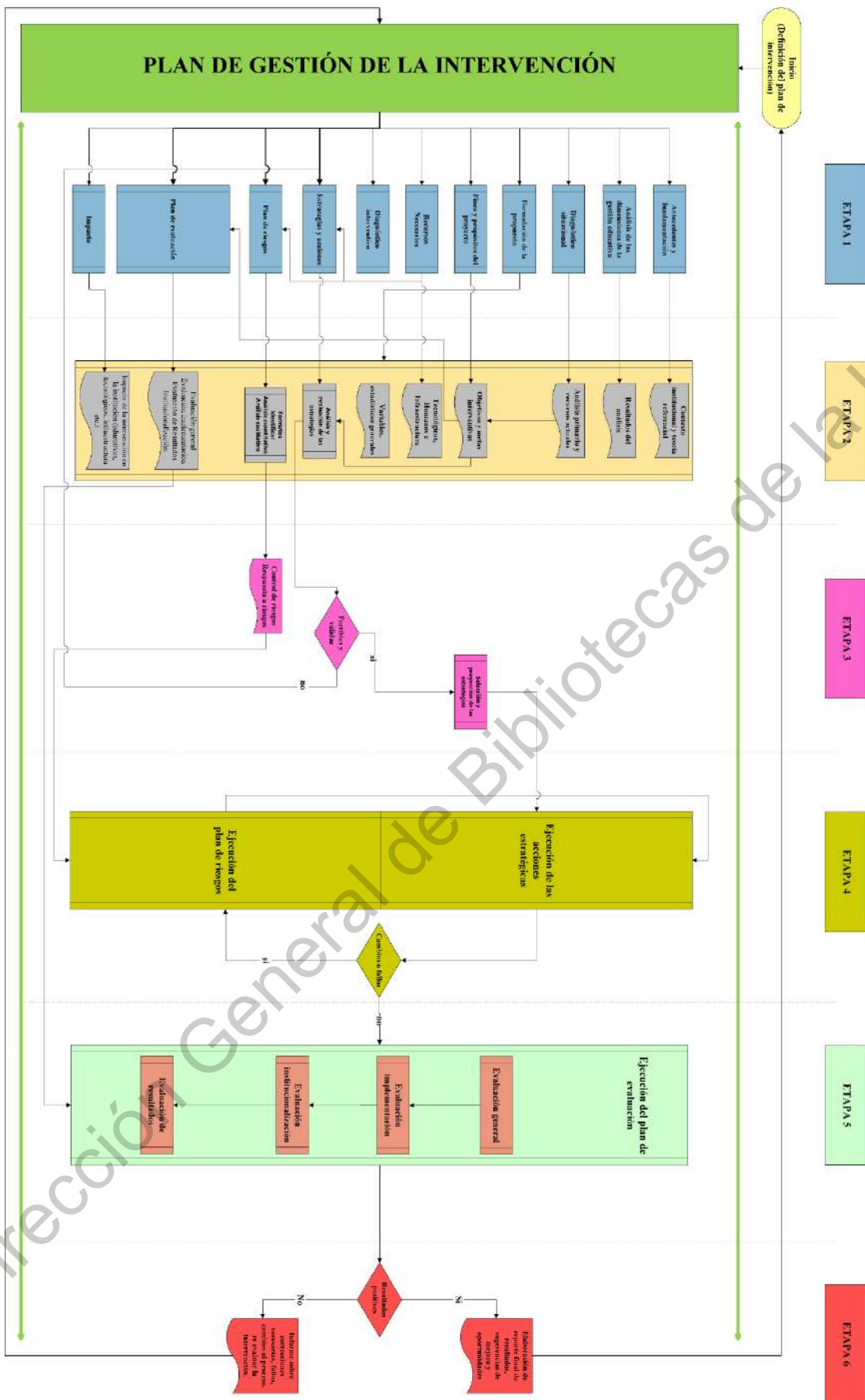


Figura 20. Diagrama del proceso de sistematización de la intervención.

Fuente: elaboración propia.

## **Etapa 1**

En esta se integran todas las fases que componen al Plan de gestión de Intervención del proyecto doctoral, cada fase está representada en el diagrama como un subproceso, ya que cada fase requiere de la ejecución de ciertas tareas y actividades, algunas son más complejas y requieren a su vez de un mayor número de subprocesos. Tal es el caso de la fase de *Estrategias y Acciones*, *Plan de Evaluación* y *Plan de Gestión de Riesgos*. Estas fases tienen a su vez su propio diagrama de sistematización para detallar lo que sucede dentro de cada fase.

## **Etapa 2**

Esta consiste en la producción de los primeros resultados de la ejecución de las fases del plan de gestión de la intervención, en su mayoría los resultados se traducen en documentos recopilatorios de información y/o su análisis. Sin embargo, en esta etapa destacan dos subprocesos muy importantes: análisis y evaluación de estrategias de intervención, y el diseño, identificación, análisis cuantitativo y cualitativo de los riesgos presentes en la intervención. Estos subprocesos también tiene su propia representación en un diagrama de sistematización, ya que dada su complejidad es necesario clarificar de una manera más abstracta lo que se realiza.

## **Etapa 3**

Esta etapa es crucial para la intervención, ya que es el momento en el que a partir del resultado del sub proceso de *análisis y evaluación de las estrategias* se valora la factibilidad y validez de dichas estrategias, y se eligen solo aquellas que cumplen con estos criterios, en caso contrario se regresa el proceso a la fase de *Estrategias y Acciones* para volver a analizar y proponer acciones estratégicas para la intervención. Además de la selección de las acciones estratégicas, se ejecuta también la proyección de las mismas. Estas últimas, junto al documento con las respuestas y control de riesgos, son el producto final de esta etapa, que sirve de insumo para la siguiente etapa.

#### **Etapa 4**

En esta etapa sucede el grueso de la intervención, ya que es el momento en que se ejecutan todas las estrategias previamente analizadas y evaluadas. Además, también es donde entra en juego el sub proceso de *Plan de Gestión de Riesgos* ya que en caso de presentarse algún fallo o error en la ejecución de las acciones estratégicas se recurre a este plan de riesgos y se procede a actuar según sea el caso. Para aquellas acciones estratégicas que no presentan errores, se turnan a la siguiente etapa, en la cual ahora se ejecuta la evaluación de ellas.

#### **Etapa 5**

Esta etapa al igual que la anterior es una de las más robustas, ya que para cada una de las acciones estratégicas que se han efectuado en el proceso de intervención se realiza una evaluación sistemática. Esta etapa es un súper proceso que se a su vez se compone de cuatro sub procesos asociados a la evaluación. En total se evalúan cuatro criterios que permiten observar la calidad y el logro de la intervención, cada criterio corresponde a una evaluación independiente. De esta forma se realiza una evaluación general de la intervención, una evaluación del proceso de implementación, una evaluación específica de los resultados de la intervención en relación a las metas establecidas en una de las fases definidas en la primera etapa del diagrama, y por último se evalúa el grado que la intervención se ha institucionalizado. Cada evaluación tiene su propio diagrama de sistematización, y a su vez se cuenta con el diagrama general del súper proceso de *Ejecución del Plan de Evaluación*.

#### **Etapa 6**

Esta es la última etapa del proceso de sistematización, el cual consiste en la valoración final de los resultados obtenidos de manera integradora de la ejecución del proceso de evaluación de la intervención. Este documento en el caso de obtener resultados positivos recopila los puntos fuertes de cada proceso de evaluación y a su vez en marca las áreas de mejora y/o sugerencias para obtener resultados aún más satisfactorios. Este documento puede servir como base para replantear un nuevo proceso de intervención, a manera de implementar un ciclo de mejora continua. En el caso de que los resultados sean negativos este documento recopila los elementos más críticos de fracaso de la intervención y puntualiza los indicadores que tuvieron mayor incidencia en los errores de la ejecución de las acciones estratégicas,

sirviendo así, como insumo para replantear los procesos particulares de fallo o bien todo el proceso interventivo.

## **5.9 Evaluación de la intervención**

Para evaluar el proceso de intervención del proyecto doctoral, se aplicó el modelo propuesto por Pizarro y Uribe (2015) el cual emplea cuatro componentes principales y una serie de criterios que a su vez son evaluables de acuerdo a ciertos indicadores. La descripción e interpretación de los criterios e indicadores que el plan de evaluación aplicó son los siguientes:

### ***Fase. Componente 1: Evaluación general***

#### **Criterio: Pertinencia**

En la evaluación de la intervención educativa, uno de los principales aspectos a considerar es la estrecha relación que guarda el proyecto de intervención con las necesidades y expectativas de los usuarios, actores o beneficiados. Por ello los indicadores que se evaluaron de este criterio son los siguientes:

- Necesidades de la población y expectativas de la intervención.
- Recursos con los que se cuenta.
- Viabilidad de la intervención.

### ***Fase. Componente 2: Evaluación del proceso de implementación***

#### **Criterio: Coherencia**

El criterio de coherencia, operativamente, se trata del grado de consistencia de la propuesta de intervención y la aplicación en el contexto real. Los indicadores que permiten evaluar este grado de consistencia son los siguientes:

- Metodología implementada
- Operatividad

- Capacitación

### ***Fase. Componente 3: Evaluación de los resultados***

#### **Criterio: Eficacia**

Este criterio contempla la valoración de los resultados obtenidos con la implementación de la propuesta de intervención. Formalmente este criterio se define como la distancia entre los objetivos planeados inicialmente en el diseño y los logros obtenidos, la relación entre lo asignado y lo alcanzado. Los indicadores que evalúan el cumplimiento de este criterio son:

- Logros
- Efectos generados
- Toma de decisiones (indicador agregado al plan)

### ***Fase. Componente 4: Institucionalización de la intervención***

#### **Criterio: Relevancia**

Pérez (citado en Pizarro y Uribe, 2015) incorpora en la evaluación de programas educativos el componente de institucionalización como el establecimiento de un continuo entre evaluación-mejora evaluación; que en el plan de evaluación se refiere al componente de institucionalización. El criterio de relevancia se entiende como la importancia que se le otorga a un programa para ser implementado y cubrir las necesidades individuales y/o sociales. Los indicadores que permiten medir este criterio son:

- Políticas
- Filosofía
- Adopción general del proceso

Es así como con los cuatro componentes antes descrito se evalúan los aspectos más relevantes de la intervención educativa que el proyecto de investigación debe cubrir para ser considerada exitosa, una descripción de todo el plan de evaluación y sus resultados se pueden consultar en el anexo 5.

## **6. RESULTADOS**

La ejecución del plan de intervención produjo una serie de resultados en diferentes aspectos, en este capítulo se analizan los resultados más sobresalientes de las etapas clave en el proceso de intervención.

Por ejemplo, se destacan las propiedades técnicas del examen estandarizado, y los resultados obtenidos en los atributos de calidad definidos para el análisis psicométrico, además de señalar algunos productos obtenidos de la etapa del desarrollo de la prueba. Por otro lado, se muestran también los resultados obtenidos de la ejecución del plan de evaluación de la intervención, destacando los valores obtenidos en los indicadores de cumplimiento de las metas propuestas para la intervención.

Otro aspecto importante fueron los resultados obtenidos en cuanto a la tecnología innovadora para la realización del examen y la integración del análisis psicométrico en la misma, observando las interfaces y los reportes técnicos generados automáticamente por el sistema SIEXAES. A continuación, se describen de manera más detallada los resultados antes mencionados, así como algunos otros de relevancia consideración.

### **6.1 Prueba estandarizada**

Uno de los propósitos de este trabajo de investigación era desarrollar exámenes estandarizados, para lo cual como se mencionó anteriormente se implementó una metodología basada en un modelo psicométrico para elaborar exámenes de gran escala de referencia criterial orientados por el currículum.

Es precisamente que producto del desarrollo de esta prueba se obtuvieron una serie de productos que contribuyen de manera directa a la mejora de la comprensión del universo de conocimientos que se pretende medir con la prueba.

### 6.1.1 La retícula

Uno de los productos resultantes de aplicar la primera etapa de la metodología y que tiene amplia utilidad para los docentes fue la retícula, este instrumento permitió apreciar de manera completa cómo los contenidos temáticos de la asignatura se relacionaban entre sí, mediante una serie de conectores textuales y flechas que permitían señalar como los temas y subtemas interactuaban para concretar conocimientos de temas que se encuentran en otras unidades. También permite identificar aquellos contenidos que tenían un mayor impacto para concretar otros conocimientos, esto se observaba por tener una mayor cantidad de conexiones con otros temas. Un segmento de esta retícula se ilustra en la figura 21.

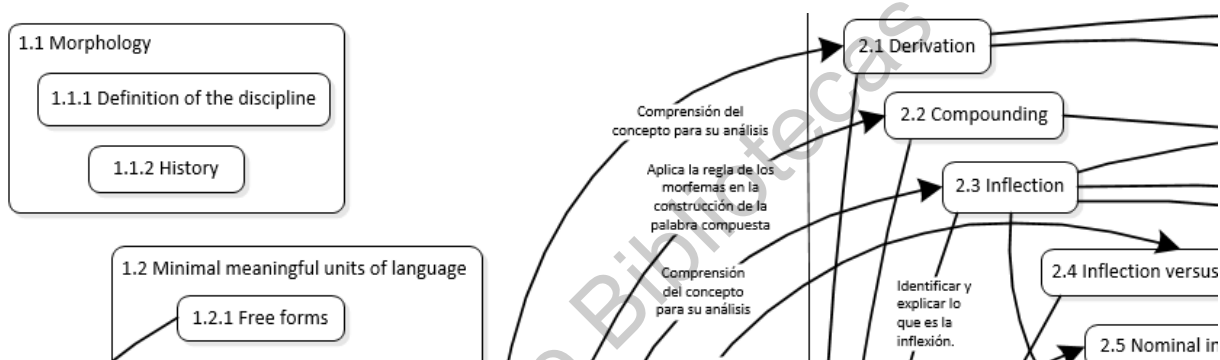


Figura 21. Ejemplo de la retícula de la materia de Morfología de la segunda lengua.

Fuente: elaboración propia.

La figura 21 ejemplifica un segmento de lo que toda la retícula es (para revisarla completa consultar el anexo 9), en esta se puede observar mediante las flechas cómo los contenidos temáticos de la primera unidad tienen relación con los de la unidad 2, y cómo algunos tienen un mayor número de relaciones con temas de unidades anteriores o posteriores. Este tipo de contenidos son los que tienen un mayor impacto en el aprendizaje del estudiante.

### 6.1.2 Tabla de índices de relevancia curricular

Este producto es cuantitativamente significativo, ya que permite obtener el valor de relevancia curricular que tiene cada uno de los contenidos temáticos que componen a todo el universo de conocimiento que será evaluado. Estos índices se representan en una tabla, que, ordenándose de mayor a menor, permite rápidamente clasificar cuales temas son más

importantes en relación a la competencia general del curso, así como para concretar otros conocimientos.

El producto correspondiente a los resultados del Índice de Relevancia Curricular (IRC) de los contenidos temáticos se muestra en la tabla 20, en la cual se pueden apreciar los diferentes atributos valorados para obtener este índice. Es importante mencionar que esta valoración la realizaron tres jueces, que de manera anónima evaluaron estos indicadores, utilizando una escala numérica definida, posteriormente se compararon las valoraciones hechas por cada juez, y mediante un proceso estadístico se analiza la consistencia interna de las valoraciones dadas por cada juez. En la tabla siguiente se muestra un ejemplo de la tabla de IRC con algunos de los contenidos temáticos de la asignatura.

**Tabla 20.**

*Atributos valorados para obtener el IRC.*

Contenido	Contribución al logro de la competencia de la unidad (20%)	Dosificación (cuantía de contenidos implícitos) 10%	Carga horaria (asignada o estimada para su aprendizaje) 10%	Relevancia disciplinaria (20%)	Índice de Relevancia Curricular (IRC)
1.1.1 Definition of the discipline	0.20	0.10	0.10	0.20	0.600
1.1.2 History	0.13	0.07	0.07	0.13	0.400
1.2.1 Free forms	0.20	0.10	0.07	0.13	0.533
1.2.2 Signs and morphemes	0.20	0.10	0.10	0.20	0.633

Fuente : elaboración propia.

La tabla anterior indica algunos de los criterios que se valoraron para obtener el IRC; se valoró cada uno de estos atributos para todo el universo temático de la asignatura, se observó cómo algunos temas tenían una relevancia tan baja que si estos se omitían prácticamente no tenían ninguna afectación en el aprendizaje del estudiante. Esto resulta interesante puesto que centra la atención en el diseño del currículo de la asignatura, ya que según los resultados



mencionados incluye contenido temático que contribuye muy poco a lo que el estudiante debe conocer y saber hacer.

Dado que el propósito de estos exámenes es evaluar el dominio de un determinado conocimiento que los estudiantes poseen, se utilizan una mayor cantidad de ítems de contenidos temáticos que posean un mayor IRC, ya que este tipo de temas requieren la integración de una mayor cantidad de conocimientos y por ende implica un mayor grado de dominio. Estos fueron algunos de los resultados obtenidos en las primeras etapas de la metodología, sin embargo, para las siguientes etapas los resultados fueron más técnicos, ya que implicaron la valoración de mayor peso en el desarrollo del examen, y la medición de la calidad a través del análisis psicométrico.

### 6.1.3 *Tabla de especificaciones del examen*

Este producto resultado de la elaboración de la prueba, es un importante instrumento, ya que de una manera sintetizada permite plasmar las decisiones estratégicas en relación a la cuestión evaluativa del examen. Mediante esta tabla se representa de una forma organizada cuantos ítems o reactivos deben elaborarse para cada tema en cuestión, así como definir cuál es el enfoque o foco que se pretende lograr con el reactivo, incluyendo también cual es el nivel taxonómico que implica dicho ítem. Esta tabla incluye todos los temas que tendrán una representación en la prueba, se puede ver un ejemplo de la misma en la tabla 21.

**Tabla 21**

*Ejemplo de la tabla de especificaciones del examen*

Contenido	IRC	Especificaciones	Ítems	Foco del ítem	Tipo de ítem	Nivel taxonómico
<b>2.4 Inflection versus derivation</b>	0.558	1	1	Probar la habilidad para distinguir la inflexión de la derivación	Opción Múltiple	Comprender conceptos

Fuente: elaboración propia.

Como se observa en la tabla anterior, esta sirve como una guía que permite saber cuántas especificaciones de ítems (instrumento que detalla la forma en que se elabora un reactivo) se

requieren, así como especificar claramente cuál es el objetivo que se busca con el reactivo, que tipo de reactivo es (opción múltiple, relacionar conceptos, ordenar, completar etc.) y por ultimo a que nivel taxonómico se llega con el mismo.

#### **6.1.4 Especificaciones de ítems**

En el desarrollo de la prueba, fue muy importante contar con el instrumento de la especificación de ítems. Este producto resultado de la etapa cuatro es el insumo principal para poder construir los ítems que formaran parte de la prueba. La importancia de estas especificaciones radica en que son la guía exacta para que se pueda construir el ítem de la manera correcta y estandarizada, la especificación provee todos los insumos necesarios para quien elaborará el ítem, desde instrucciones para la redacción, dimensiones del conocimiento y cognitivos que el ítem contempla, así como ítems muestra con las opciones correctas y los distractores. Este producto es indispensable para la elaboración de la prueba, ya que, dado que la prueba es estandarizada, los criterios e instrucciones para la construcción de cada ítem debe ser la misma, de forma tal que si existen errores en la construcción del ítem se deban en mayor proporción a la ejecución de quien elabora el ítem, y no por deficiencias en el instrumento que guía dicha construcción.

Cada uno de los ítems definidos en la tabla de especificaciones de la prueba, tiene a su vez una especificación de ítem, por lo que se generaron las especificaciones de ítems para todos y cada uno de los temas, obteniendo así una guía para la elaboración de ítems de cada uno de los contenidos temáticos que comprenden el universo de medida que el examen comprende, que como se ha dicho anteriormente, está determinado por el currículo de la asignatura. Lo anterior significa que se pueden construir cuantos ítems se deseen, permitiendo así contar con más de una versión de la prueba, sin embargo, independientemente de la versión, el examen evalúa de la misma forma y en las mismas condiciones el dominio de conocimiento que posee el estudiante. Lo anterior es un aporte significativo para los estudiantes, ya que no hay un examen versión fácil o difícil, aunque haya  $n$  versiones de la prueba, todas son igual de válidas y confiables para el estudiante.

En la siguiente figura se ilustra la especificación de uno de los contenidos temáticos de la prueba.

### Especificación de ítems de la prueba de Morfología del Segundo Idioma

<p><b>Información contextual</b></p> <p>1.1.2 History</p>	<p>El contenido <i>History</i> es el segundo tema conceptual que el alumno aborda en la <i>Unidad 1. Basic concepts</i>, en donde se abordan aspectos históricos de la Morfología como disciplina, tales como principales personajes, teorías o corrientes.</p> <p>Por lo anterior, este contenido obtuvo puntuaciones medias en los criterios de contribución al logro de la competencia de la asignatura, y una puntuación baja con relación al índice de relevancia curricular.</p> <p>Se elaborará un ítem que ponga a prueba si el alumno recuerda los aspectos fundamentales de la historia de la disciplina tales como personajes, teorías y corrientes lingüísticas importantes.</p>																																									
<p>As a discipline that records and interprets past events and the continuation of these happening in succession that lead to a present of the discipline.</p>	<p><b>Información contextual o indicaciones para responder este ítem:</b> Ninguna.</p>																																									
<p><b>Unidad 1. Basic concepts</b> <b>Competencia:</b> Identificar las unidades morfológicas mínimas implicadas en la formación de palabras en la lengua inglesa a través de la realización de ejercicios prácticos para crear conciencia de los mecanismos de construcción de significados; mostrando respeto, mostrando interés así como respeto hacia el profesor y sus compañeros al trabajar en el aula en equipos de trabajo o al realizar sus ejercicios de manera individual. <b>IRC:</b> 0.400</p> <p><b>Criterios:</b></p> <ul style="list-style-type: none"> <li>• Contribución al logro de la competencia de la unidad: .13</li> <li>• Dosisificación: .07</li> <li>• Carga horaria: .07</li> <li>• Relevancia disciplinaria: .13</li> <li>• N° servicios que recibe: 0</li> <li>• N° servicios que proporciona: 0</li> </ul> <p><b>Recibe servicios de:</b> Ningún contenido.</p> <p><b>Proporciona servicios a:</b> Ningún contenido.</p>	<p><b>Información tabular, gráfica o textual a emplear en el ítem:</b> Ninguna.</p> <table border="1" data-bbox="917 441 1339 630"> <thead> <tr> <th rowspan="2">Dimensión</th> <th colspan="6">Dimensión proceso cognitivo</th> </tr> <tr> <th>Recordar</th> <th>Comprender</th> <th>Aplicar</th> <th>Analizar</th> <th>Evaluar</th> <th>Crear</th> </tr> </thead> <tbody> <tr> <td>Conocimiento factual</td> <td>X</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Conocimiento conceptual</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Conocimiento procedimental</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Conocimiento metacognitivo</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Dimensión	Dimensión proceso cognitivo						Recordar	Comprender	Aplicar	Analizar	Evaluar	Crear	Conocimiento factual	X						Conocimiento conceptual							Conocimiento procedimental							Conocimiento metacognitivo						
Dimensión	Dimensión proceso cognitivo																																									
	Recordar	Comprender	Aplicar	Analizar	Evaluar	Crear																																				
Conocimiento factual	X																																									
Conocimiento conceptual																																										
Conocimiento procedimental																																										
Conocimiento metacognitivo																																										
	<p><b>Especificación de la base del ítem:</b></p> <ol style="list-style-type: none"> <li>1) Podrán presentarse datos característicos de un personaje referente en la historia de la morfología, y solicitar al examinado que identifique el personaje histórico del que se trata. Por ejemplo: <i>Who was known as the father of the "Structural linguistic" of the 20th century?</i></li> <li>2) También se podrían presentar datos característicos de un determinado periodo, corriente histórica o teoría de la disciplina, y solicitar al examinado que identifique el periodo, corriente histórica o teoría del que se trata. Por ejemplo: <i>Which current emerges as a reaction to the 20th century comparative linguistics research?</i></li> </ol>																																									
	<p><b>Especificación de la respuesta correcta a.</b></p> <ol style="list-style-type: none"> <li>1) Será el nombre del personaje referido. Por ejemplo: <i>Ferdinand de Saussure</i></li> <li>2) Será la corriente histórica, periodo o teoría de la que se habla. Por ejemplo: <i>Linguistic structuralism</i></li> </ol>																																									
	<p><b>Especificación del distractor b.</b></p> <ol style="list-style-type: none"> <li>1) Se podrá presentar a un personaje que haya realizado trabajos similares. Por ejemplo: <i>Charles S. Peirce</i> (que fue del mismo periodo que Ferdinand y colaboró en teorías del estructuralismo).</li> <li>2) Se podrá presentar una corriente histórica, periodo o teoría diferente pero afín al que se habla. Por ejemplo: <i>Linguistic generativism</i>.</li> </ol>																																									
	<p><b>Especificación del distractor c.</b></p> <ol style="list-style-type: none"> <li>1) Se podrá presentar a un personaje que haya realizado trabajos similares. Por ejemplo: <i>Noam Chomsky</i>.</li> <li>2) Se podrá presentar una corriente histórica, periodo o teoría diferente pero afín a la que se habla. Por ejemplo: <i>Linguistic functionalism</i>.</li> </ol> <p>En todos los casos, el distractor c no repetirá un personaje, teoría, corriente histórica o periodo ya especificado en la opción b.</p>																																									

<b>Ítem muestra</b>
Who was known as the father of the "Structural linguistic" of the 20th century?
a. Ferdinand de Saussure
b. Charles S. Peirce
c. Noam Chomsky

Figura 22. Ejemplo de una especificación de ítems.

Fuente: elaboración propia.

Como se observa en la figura anterior, la especificación también dio detalles contextuales, tales como la relevancia curricular, una introducción al contenido temático para poder contextualizar a quien elaborara el ítem de la manera más exacta posible de forma tal que cumpla con todos los requisitos de elaboración del ítem. Cuidando particularmente que el ítem realmente evalué aquello que se pretende evaluar, por ello la importancia de que esta guía este bien diseñada, ya que, si hay un defecto técnico en la misma, provocara que todos los ítems construidos a partir de ella, tendrán deficiencias.

## 6.2 Propiedades psicométricas

El análisis psicométrico resultante de la aplicación del pilotaje mostró a nivel de ítem los diferentes criterios de calidad que se deseaban medir. Fueron un total de 63 ítems los que se aplicaron para el examen de Morfología de la segunda lengua, para cada ítem se emitió un juicio para determinar si cada uno de estos se aceptaban o no, y las razones de dicho dictamen. Las estadísticas descriptivas generales del examen se muestran en la tabla 19.

**Tabla 22**

*Estadísticas descriptivas del examen*

Ítems	Media de aciertos	Desviación Estándar	Puntaje Mínimo	Puntaje Máximo	Media del valor P	Media de Rpbis	Alpha
63	40.211	7.242	27	61	0.638	0.212	0.796

Fuente : elaboración propia.

Entre los datos a destacar de la tabla 19 se encuentra el valor medio de  $p$  o el índice de dificultad del examen, este valor se ubicó en .638 lo que indica que el examen es de dificultad media con una leve tendencia a ser más fácil que a ser regular. También se destaca el hecho de que en promedio se aciertan 40 de las 63 preguntas, por otro lado, también se logró una puntuación muy cercana a la perfecta obteniendo 61 aciertos.

Ahora bien, detallando los resultados de algunos ítems se encuentran los mostrados en la tabla 20. En esta tabla se observa por ejemplo al ítem siete cuya dificultad es regular, es decir que el ítem no es fácil ni difícil, aunque se inclina muy levemente a ser difícil de responder, lo anterior se concluye puesto que el valor  $p$  del ítem es de .486, como se mencionó en el apartado anterior, entre más cercano a 1 se encuentre este valor indica que el ítem o pregunta es más fácil de responder.

**Tabla 23***Análisis psicométrico de los ítems*

ANÁLISIS	IT7	IT25	IT28	IT48
Dificultad (p)	Ítem regular con p=.486	Ítem fácil con p=.789	Ítem muy difícil, con una p=.289	Ítem muy difícil con .21
Discriminación	Discriminación muy buena con .51	Discriminación regular con .26	La discriminación es buena con .33	La discriminación es negativa con .16
rpBis	Muy buena con .40	Buena con .264	Es baja con .132	Es negativa con .271
Distractor	Los distractores funcionaron bien se obtuvieron diferencias de -.264 y -.355.	Los distractores funcionaron bien se obtuvieron diferencias de -.173 y -.091.	Los distractores funcionaron bien se obtuvieron diferencias de -.064 y -.264.	Los distractores fueron elegidos , pero en igual proporción por el grupo alto y bajo

Fuente: elaboración propia.

El poder discriminatorio del ítem siete es excelente con un valor obtenido de .51, además, el contestarlo correctamente tiende a guardar relación con los buenos puntajes en todo el examen (rpBis de .40).

El siguiente factor que se observa en el ítem siete son los distractores o respuestas no correctas, estos fueron elegidos en todos los casos y con mayor proporción por los grupos bajos (alumnos con bajo rendimiento, promedios menores o iguales a 6.9) o los que tuvieron menor nota al final del examen. Para saber si un ítem discrimina bien o no, los valores obtenidos de *rbis* para la respuesta correcta deben ser mayores a .2, las respuestas incorrectas o distractores deben ser negativas o pudieran ser positivas, pero con valores menores al valor de la respuesta correcta. Agregado a esto, también debe considerar que todos los distractores deben ser elegidos.

El ítem 7 es el ítem ideal que se busca tener un examen, ya que cumple correctamente con los estándares de calidad, para todos los criterios de valoración antes mencionados. La figura 23 ilustra el comportamiento del ítem.

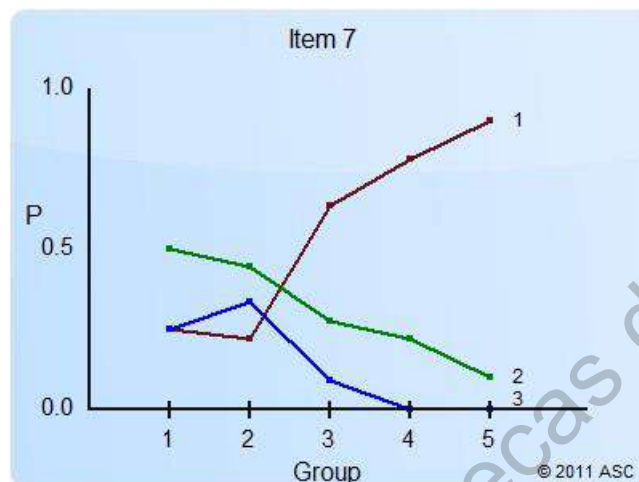


Figura 23. Gráfica del comportamiento del ítem 7.

Fuente: elaboración propia utilizando ITEMAN 4.3.0.3.

Como se puede observar en la figura 23, las tres respuestas posibles se representan con una línea de color verde, una marrón y una azul. El eje  $x$  u horizontal representa a los grupos bajo, medio y alto, mientras que el eje  $y$  o vertical representa a la población. La línea marrón corresponde a la respuesta correcta, y como se puede ver en la figura fue elegida en mayor proporción por el grupo alto que estaría ubicado entre la categoría 4 y 5. Y fue menor su elección por el grupo bajo que se encuentra la categoría 1 y 2.

Los resultados antes mostrados se obtuvieron a partir del análisis psicométrico realizado con el software SIEXAES, dicho software generó automáticamente los reportes técnicos con los estadísticos descriptivos del examen en general como ya se mostró anteriormente, y los criterios de confiabilidad y validez que se establecieron como indicadores de calidad, el índice de dificultad de cada ítem, el poder discriminatorio, la correlación ítem –total y el funcionamiento de cada distractor. Para el caso de las gráficas estas se obtuvieron a partir del análisis efectuado con el software ITEMAN.

### 6.2.1 Calibraciones entre ciclos de pilotaje

Otro de los aspectos interesantes de la puesta en práctica del pilotaje de la prueba, fue observar como los indicadores de calidad psicométrica de la prueba variaban entre una primera ejecución y la segunda, aunque la muestra de participantes vario en un 7.8%, las características de la población siguieron siendo las mismas.

En un primer pilotaje los resultados psicométricos obtenidos fueron los mostrados en la siguiente tabla:

**Tabla 24**

*Propiedades psicométricas en el primer ciclo de prueba*

Media de aciertos	Puntaje Mínimo	Puntaje Máximo	Media del valor P	Media del valor D	Media de Rpbis	Alpha
40.211	27	61	0.638	.272	0.212	.796

Fuente: elaboración propia.

Como resultado de este análisis psicométrico se identificaron, un total de 13 ítems que presentan valores de calidad inferiores a los deseados. Sin embargo, después de analizar a detalle las posibles causas se determinó que los ítems debían mantener su representación en la prueba, pero se les debían hacer adecuaciones a la redacción y mejorar la calidad de los distractores, así como aumentar la dificultad de algunos ítems que eran demasiado fáciles y que por ende no discriminan correctamente. Una vez realizado los ajustes antes mencionados, el análisis psicométrico arrojó lo siguientes resultados:

**Tabla 25**

*Propiedades psicométricas en el segundo ciclo de prueba*

Media de aciertos	Puntaje Mínimo	Puntaje Máximo	Media del valor P	Media del valor D	Media de Rpbis	Alpha
43.263	34	60	0.647	.331	0.322	.791

Fuente: elaboración propia.

De la comparación de ambos ciclos de ejecución se observan mejoras en los indicadores promedio de los atributos que definen la validez y la confiabilidad de la prueba. Por ejemplo, destaca el hecho de que el valor de  $r_{pbis}$  o el coeficiente de discriminación subió de ubicarse en un rango de regular a bueno con .322 estando muy cerca de ser un valor excelente. Así mismo el índice de discriminación o valor  $D$  subió a .331, indicando ahora una calidad buena. La media de aciertos también subió un poco, pudiendo ser a causa de los ajustes de esos 13 ítems que presentaban defectos técnicos en el primer pilotaje, y que para esta segunda ejecución fueron corregidos. Claro está, este instrumento conforme sean más las aplicaciones que se vayan realizando podrá seguir calibrándose y obtener mejores indicadores, sin embargo, hasta este momento los resultados obtenidos dan muestra de que el instrumento es válido y confiable, y de calidad sobresaliente.

### **6.3 Ejecución del proceso de intervención**

Como parte del proceso de intervención, la metodología integró la puesta en acción de diferentes planes de gestión que cubrían a su vez cada una de las etapas o fases que componen el proceso general de la intervención educativa.

Entre ellos destaca el plan de evaluación de la intervención, ya que este plan contempla la valoración y medición del logro de los indicadores de éxito para la intervención. Este plan de evaluación se encuentra dividido en cuatro fases-componentes, y a su vez cada fase posee indicadores de cumplimiento.

#### **6.3.1 Componente evaluación general**

Para el componente 1 que trata la evaluación general del proceso de intervención, el criterio de cumplimiento es la pertinencia, en este criterio se analizaron las necesidades o expectativas de la institución educativa, y se asociaron a objetivos interventivos, con el fin de cubrir dichas necesidades. Una necesidad sobresaliente era un bajo nivel de confiabilidad y validez en las pruebas, sin embargo, posterior al proceso de intervención, se obtuvo un instrumento de evaluación que cumplía con valores muy cercanos a los excelentes para los criterios psicométricos de confiabilidad y validez.



Otro resultado interesante a evaluar de esta fase fue que en la institución no se contaba con reportes técnicos estandarizados que aportaran información significativa sobre el desempeño de los estudiantes a nivel individual y grupal, pero como objetivo interventivo para atender esta necesidad se desarrollaron 3 reportes técnicos que aportaron información sobre la calidad técnica de la prueba, nivel de dominio de los estudiantes dosificado a nivel de subtemas, así como información sobre todo el universo de conocimientos evaluados generando datos específicos a nivel de grupo en cada uno de los temas representados en la prueba, destacando los de mayor relevancia curricular, análisis comparativo de unidades o temas con mayor dominio y aquellos deficientes.

### **6.3.2 *Componente institucionalización de la intervención***

Para este componente uno de los criterios fue la relevancia, y para medir este criterio se utilizó el indicador de políticas, destacando que a nivel institucional se desarrollaron políticas que normalizaron por ejemplo aspectos como las condiciones de ejecución de las pruebas estandarizadas, la metodología a seguir para el desarrollo futuro de instrumentos de evaluación estandarizados y por último una propuesta de cursos de capacitación continuos para que los docentes reciban preparación en el desarrollo de instrumentos de evaluación basados en modelos psicométricos y estandarizados. Es importante mencionar, que estas políticas propuestas, que se desarrollaron al interior de la Facultad de Idiomas como producto del proceso de intervención, pudieron ser vinculadas a políticas y normas ya establecidas a nivel institucional.

Otro de los indicadores para valorar la relevancia de la intervención fue el de la filosofía. Es bien sabido que el propósito de una intervención busca entre otras cosas provocar un cambio en la filosofía de los procesos que se realizan en la institución, en el caso particular de este proyecto aquellos asociados a la evaluación del aprendizaje. En este sentido uno de los objetivos de la intervención fue crear una cultura evaluativa donde toda acción evaluativa sea precedida de un fundamento teórico y metodológico que garantice la calidad del instrumento evaluativo.

A este respecto el indicador filosofía arrojó datos sobresalientes, por ejemplo, antes de la intervención menos del 15% de los docentes evaluaba la calidad técnica de los instrumentos

de evaluación que aplicaba. La meta de la intervención era lograr que el 50% de los docentes cambiaran esa filosofía, y optaran por interesarse en analizar la calidad de sus pruebas, que observaran las ventajas y el impacto que tenía esto en sus procesos de evaluación. Sin embargo, después de evaluar este indicador se obtuvo que el 76.19% de la población docente manifestó el interés por realizar la evaluación de calidad técnica de sus instrumentos.

Una situación similar se observó en el interés que los docentes manifestaban sobre realizar un análisis más profundo de los datos generados por cada evaluación. La meta de la intervención era lograr que el 50% de la planta docente analizara más que solo la calificación numérica de la prueba, efectuando análisis más detallados y concretos del desempeño del estudiante. Después de la intervención se observó que el 80.95% de los docentes que participaron en la capacitación de los diferentes comités de diseño de la prueba y en el uso de los reportes técnicos manifestaron un interés por efectuar análisis más profundos de los resultados obtenidos en las pruebas.

En cuando el indicador de adopción general del proceso también hubo resultados sobresalientes, por citar un ejemplo, antes de la intervención el 52.38% de los docentes consideraba que estandarizar una prueba era una forma de limitar su libertad de cátedra. Sin embargo, después de la intervención la evaluación arrojó que el 76.19% acepta las pruebas estandarizadas como método válido para evaluar en su práctica docente.

Así mismo destaca el hecho de que inicialmente el 61.90% de los docentes consideraba que sus clases habían funcionado con los instrumentos y estrategias que han aplicado, por lo que no consideraban relevante querer implementar algunos otros. Posterior a la intervención y con instrucción recibida en los cursos de capacitación, el 85.71% de los docentes considera necesario aplicar otros instrumentos de evaluación y/o poder comparar resultados con los que actualmente aplica en el desarrollo de sus clases.

La intervención buscaba lograr hacer un cambio notorio en la filosofía evaluativa en la institución, además de los aspectos antes mencionados, también era de interés proporcionar instrumentos a los docentes para que obtuvieran información más relevante y especializada del aprendizaje y desempeño de los estudiantes. Es decir, ir más allá de solo observar la calificación en una prueba. Inicialmente el 85.71% de los docentes consideraban como

objetivo final del examen el observar la calificación numérica que estos producían. Después del proceso interventivo este número se redujo solo al 19.04%, lo que significa que más del 80% de los docentes buscan generar información más relevante sobre el desempeño de sus estudiantes que solo la nota observada en la prueba.

Más detalles de los resultados obtenidos en los restantes indicadores del plan de evaluación pueden ser revisados en el anexo 5.

#### **6.4 Tecnología innovadora SIEXAES**

Uno de los resultados sobresalientes del proyecto de intervención fue desarrollar una propuesta tecnológica que permitiera construir y ejecutar una prueba, y que de manera automatizada procesara los análisis psicométricos para determinar la calidad de la prueba a nivel de reactivos y de respuestas. Es precisamente en esto último, que radica la innovación de la propuesta tecnológica, de integrar en un mismo sistema la capacidad para ejecutar la prueba y realizar el análisis psicométrico de manera automática, sin necesidad de otro software especializado. Por ello la metodología que se utilizó para desarrollar dicha tecnología fue la de Rational Unified Process o conocida como RUP.

Dentro de las diferentes etapas de desarrollo de la tecnología una fase importante era la que correspondía a la usabilidad del software, ya que dado el impacto que tiene la prueba en los estudiantes se deseaba que el software cumpliera con los requisitos de usabilidad y de interfaz necesarios, para reducir el porcentaje de error que pudiera afectar a la prueba en razón de factores asociados directamente al software.

Para ello se realizaron una serie de encuestas directas con los usuarios finales del sistema, destacando claramente los alumnos en primera instancia, y posteriormente los docentes. Entre ellos se destaca una de las preguntas que evaluaba la característica de ser intuitivo, obteniendo que un 93.7% considero que los propiedades gráficas y elementos incluidos favorecían a los usuarios poder interactuar con el sistema de manera intuitiva y natural. También un aspecto importante era saber cómo se sentía la carga grafica en el software, específicamente en aspectos como el tipo de colores, tamaño de letras, tipo de letras entre otros elementos.

En este indicador, considerando al total de la muestra de alumnos y docentes, se obtuvo un 91.3% de aceptación. El restante 8.7% considero que los colores pudieran variar un poco, sin embargo, los colores incluidos en el sistema estaban condiciones por la institución, puesto que son los que la universidad establece. Mientras que algunos otros expresaron la capacidad de aumentar el tamaño de la fuente, ya que algunos poseen capacidades visuales distintas. Esta fue una función que se incluyó en los requerimientos funcionales a integrar en uno de los subsecuentes ciclos de desarrollo y mejora del software.

En las siguientes figuras se pueden observar detalles del diseño gráfico que presentan algunas de las interfaces del sistema SIEXAES.

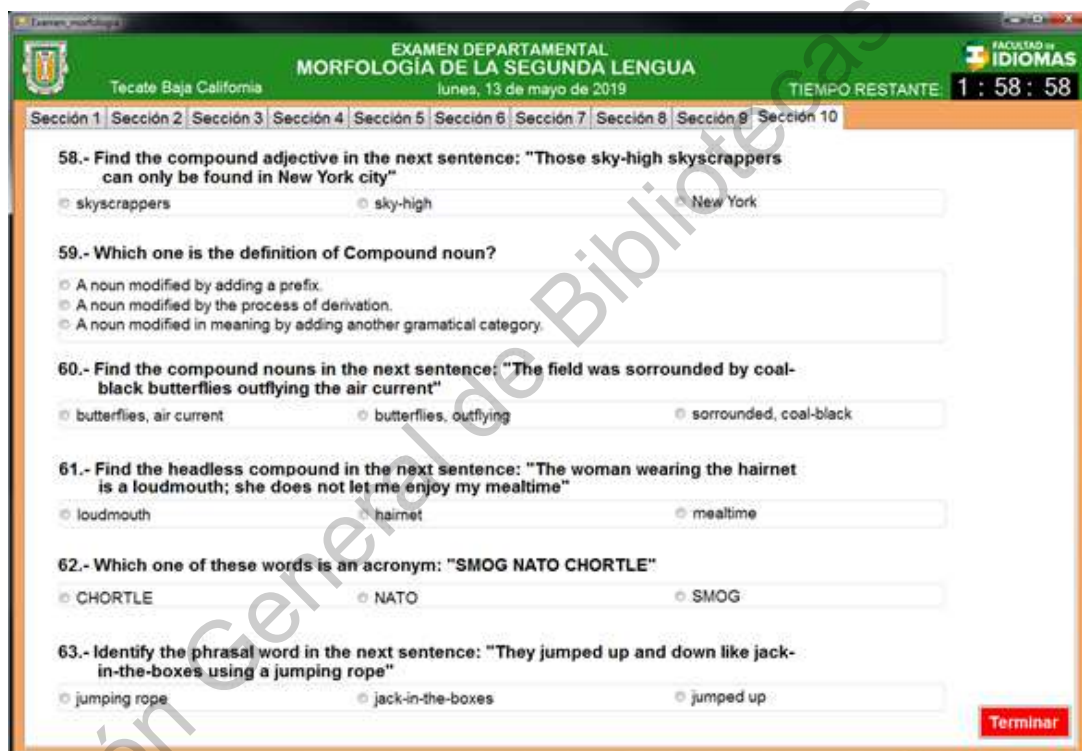


Figura 24. Interfaz de la sección de preguntas.

Fuente: elaboración propia.

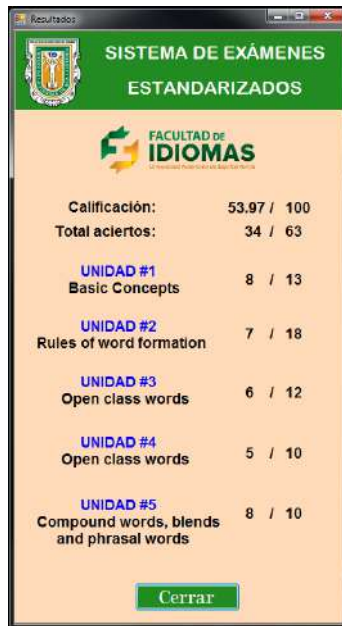


Figura 25. Interfaz del reporte resumido al concluir el examen.

Fuente: elaboración propia.

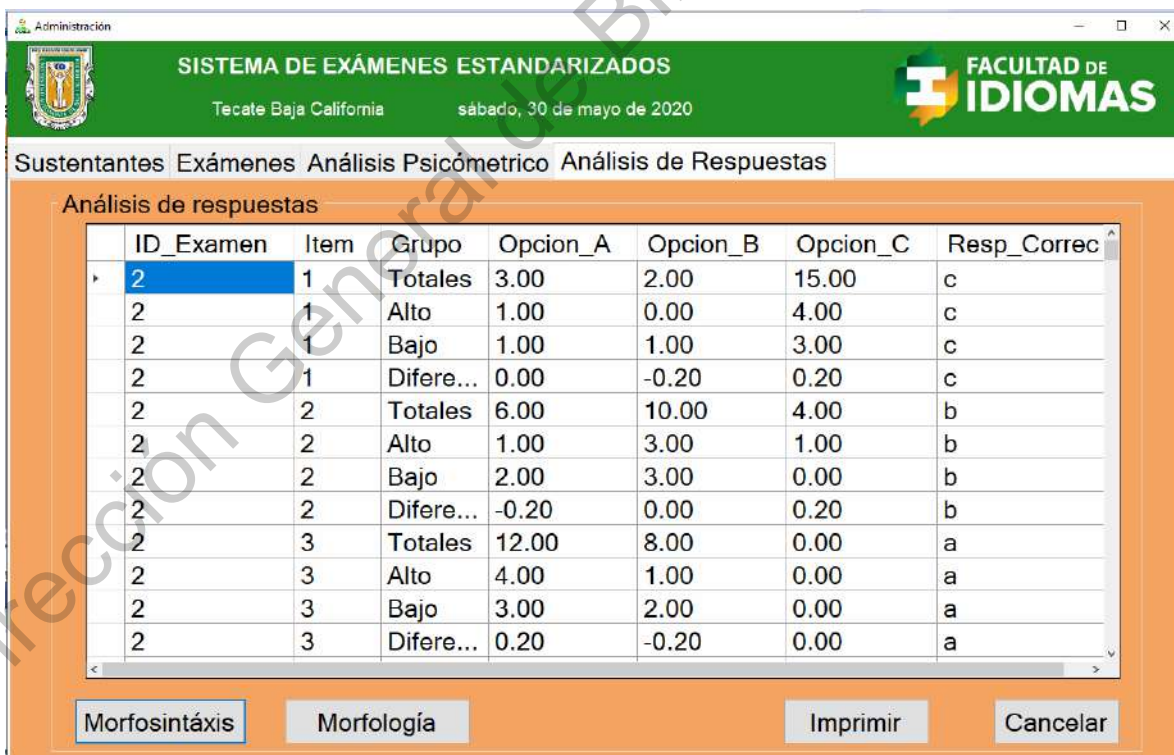


Figura 26. Interfaz de análisis de respuestas – distractores.

Fuente: elaboración propia.

Las figuras anteriores son ejemplos de las interfaces que componen al sistema, en la figura 24 se muestra la ventana principal de la prueba, esa interfaz es en la que el alumno contesta el examen, cuenta con un reloj digital que permite al alumno tener presente el tiempo que dispone para la realización de la prueba, y como se determinó en la elaboración de la prueba todos los reactivos son de opción múltiple, permitiendo al alumno cambiar sus respuestas cuantas veces lo desee.

La figura 25 es el reporte resumido que se genera automáticamente cuando el alumno termina la prueba, solo contempla información en primer nivel, es decir, calificación en escala numérica y calificación en número de aciertos, incluyendo aciertos por cada una de las unidades aprendizaje.

La figura 26 es la interfaz desde la cual el administrador del software puede consultar directamente los atributos psicométricos obtenidos en la prueba. En esta figura se aprecia la distribución de elecciones de los sustentantes entre grupos altos y bajos en cada una de las preguntas de la prueba, identificar la respuesta correcta, cantidad de sustentantes que han elegido cada respuesta entre otros aspectos.

Claro está, otra parte importante de la propuesta tecnológica asociado a un objetivo de intervención era la capacidad de ofrecer reportes técnicos con un grado de detalle que permitiera una toma de decisiones más acertada, y trascender solamente los casos numéricos. En la figura 27 se muestra un fragmento del reporte técnico completo que el alumno obtiene una vez concluido el examen, mediante el cual puede hacer autoanálisis más detallado sobre su nivel de dominio en temas específicos de toda la asignatura, así como determinar realmente cuan efectiva es la calificación numérica que alcanzo en la prueba.

## REPORTE DE RESULTADOS

Alumno: n#n#n#



El examen de morfología aplicado contenía 63 ítems. Tu puntaje obtenido fue de:

47 ítems acertados

Obteniendo un promedio (en escala de 0 a 100) general de:

74.60

El promedio general del grupo fue de:

83.78

### Informe detallado por unidad de aprendizaje

	UNIDAD DE APRENDIZAJE 1	Temas de mayor relevancia curricular (IRC)	Temas dominados de mayor IRC	Temas no dominados de mayor IRC
Distribución de aciertos	Ítems acertados: 10/13	Ítems acertados: 2/3	1.3.1	
	Calificación en la unidad: 76.92 /100	Porcentaje de dominio: 66.66%	Morphemes and allomorphs 1.3.3 Roots, affixes	1.2.2 Signs and morphemes
	Nivel de dominio: Satisfactorio			

Figura 27. Reporte técnico para el alumno.

Fuente: elaboración propia.

En la figura 27 se puede notar que el alumno puede determinar su posición en relación al promedio general del grupo en la prueba. También se observa que hay una estratificación detallada de la información asociada al dominio resultante obtenido por el alumno, mostrando por unidad de aprendizaje cuantos ítems acertó, la escala numérica de la calificación obtenida a nivel de unidad, así como una representación textual del nivel de dominio pudiendo ser: insuficiente, suficiente, buena y sobresaliente. Además, se destacan los temas de mayor relevancia curricular de dicha unidad, y se señala cuantos ítems de estos acertó, una escala porcentual del nivel de dominio y por último cuáles temas exactamente domino y cuáles no.

En este mismo sentido, en la figura 28 se muestra un segmento del reporte técnico que el sistema SIEXAES genera para el docente como resultado de la aplicación de la prueba.

El examen de morfología aplicado contenía 63 ítems. El promedio de aciertos obtenido por el grupo fue de (promedio de aciertos / aciertos posibles):

42.37 / 63
------------

El promedio (en una escala de 0 a 100) ponderado del grupo fue de:

67.25
-------

A continuación, se presenta el informe detallado:

Alumno	Aciertos Totales	Promedio Ponderado
Ejemplo 1	40	63.49
Ejemplo 2	54	85.71
Ejemplo n	32	50.79

**Unidad(es) con mejor rendimiento (promedio de aciertos / aciertos posibles):**

Unidad 1: 9.68 / 13	Unidad 3: 9.42 / 12
---------------------	---------------------

**Unidad(es) con rendimiento insuficiente:**

Unidad 5: 5.53 / 10
---------------------

**Contenido Temático con mayor insuficiencia (aciertos / aciertos posibles):**

Unidad 1	1.3.2 History	15/36
Unidad 2	2.6.2 Aspect	9/18
Unidad 5	5.1 Compounds versus phrases	1/18
Unidad 5	5.2 Compound Verbs	20/36
Unidad 5	5.5 Headed and headless compounds	9/18

**Comportamiento grupal en los temas de mayor Índice de relevancia curricular**

Contenido de mayor IRC	Promedio aciertos
2.5 Nominal inflection	n/2
2.3 Inflection	n/3
1.3.1 Morphemes and allomorphs	n/2
3.1 Nouns	n/3

**Comportamiento grupal en todo el universo de medida (promedios):**

Contenido temático	Promedio
1.1.1 Definition of the discipline	0.00
1.1.2 History	0.00
1.2.1 Free forms	0.00
1.2.2 Signs and morphemes	0.00
1.3.1 Morphemes and allomorphs	0.00
1.3.2 Free and bound morphemes	0.00
1.3.3 Roots, affixes	0.00
2.1 Derivation	0.00
2.2 Compounding	0.00
2.3 Inflection	0.00
2.4 Inflection versus derivation	0.00
2.5 Nominal inflection	0.00
2.6.1 Tense	0.00
2.6.2 Aspect	0.00
2.6.3 Mood	0.00
2.7 Other word formation	0.00
3.1 Nouns	0.00
3.2 Adjectives	0.00
3.3 Adverbs	0.00
3.4 Verbs	0.00
4.1 Pronouns	0.00
4.2 Prepositions	0.00
4.3 Conjunctions	0.00
4.4 Interjections	0.00
5.1 Compounds versus phrases	0.00
5.2 Compound verbs	0.00
5.3 Compound adjectives	0.00
5.4 Compound nouns	0.00
5.5 Headed and headless compounds	0.00
5.6 Blends and acronyms	0.00
5.7 Phrasal words	0.00

Figura 28. Reporte técnico para el docente.

Fuente: elaboración propia.

Como se puede apreciar en la figura 28, el reporte técnico para el docente incluye una mayor cantidad de atributos observables de evaluación, que van desde aspectos generales como el promedio general del grupo, unidades con el mejor y peor desempeño, contenido a nivel de subtemas y pico temas que tuvieron el más bajo número de aciertos. También se destaca el comportamiento a nivel grupal en los temas de mayor relevancia curricular para la asignatura, permitiendo al docente comprender más efectivamente el dominio del grupo en los temas más importantes para la concreción de otros conocimientos y a su vez que contribuyen más al logro de la competencia general del curso.



## 7. DISCUSIONES

Después de la aplicación de la metodología para la construcción del examen estandarizado, y la realización del análisis psicométrico, se pudieron identificar claramente como ciertos reactivos en el examen presentaban indicadores de calidad aceptables, mientras que algunos otros aun necesitaban ser corregidos. Sin embargo, el proceso en sí de diseñar y construir el examen aportó información suficiente para poder determinar si el instrumento era válido y confiable, pero no solo analizando el examen, sino también la herramienta tecnológica que se utilizó para la construcción, aplicación y posterior análisis del examen. Y es precisamente este último punto el cual va apuntado a el logro de uno de los objetivos de la investigación, que es el diseño de una propuesta tecnológica para el análisis psicométrico de exámenes estandarizados.

La intención de la herramienta propuesta es eliminar las dificultades que representa para un docente efectuar un análisis psicométrico de un examen, ya que el software que comúnmente se utiliza con estos fines requiere de cierta experticia por parte de los usuarios, tanto del uso del software como de la interpretación de los resultados, ya que los reportes que generan son muy técnicos.

Por otro lado, otro objetivo de la investigación fue analizar metodologías para desarrollar exámenes estandarizados referidos a un criterio, en primera instancia la evaluación del aprendizaje referida a un criterio se orienta a explorar el nivel de dominio que tiene el estudiante o el examinado respecto a un criterio definido, mostrando así las habilidades y conocimientos que este domina o en su caso las que no domina, pero en un campo específico del conocimiento (Contreras, Encinas y De Las Fuentes, 2005).

Pues bien, después de aplicar cada una de las fases propuestas en la metodología, se obtuvieron una serie de productos que permitieron conocer de mejor manera el universo de conocimientos que se tenía que evaluar en el examen, así como identificar de una manera gráfica y cuantificable la importancia de cada uno de los contenidos temáticos que la asignatura específica. También el integrar comités con la participación de diferentes docentes, y de las distintas unidades académicas permitió obtener juicios más objetivos, con

lo que se enriquecieron las etapas de justificación de contenidos, la elaboración de especificaciones de ítems, la evaluación de los ítems entre otras etapas.

Sin embargo, aún existen elementos que deben ser mejorados, por ejemplo, se observó que de los 63 ítems que integran el examen, existen deficiencias en aproximadamente 12 de ellos, de los cuales 8 particularmente en el índice de discriminación, por lo que se deben analizar las respuestas que dichos ítems contemplan. De igual forma aún deben producirse más ítems para el examen y con ello generar diferentes versiones del examen.

También es importante dar seguimiento a las observaciones que durante el desarrollo de la tabla con los índices de relevancia curricular y la retícula se hicieron sobre aquellos contenidos temáticos que aportan muy poco al aprendizaje de los estudiantes, ya que este tipo de situación implican directamente la calidad del diseño del currículo de la asignatura. Así mismo, dar continuidad a las sugerencias y valoraciones que los alumnos y maestros participantes le hicieron al software SIEXAES, para poder mejorar la calidad de la herramienta y cumplir de mejor manera los criterios de usabilidad evaluados, tales como el tamaño de los objetos, los colores, botones de navegación entre otros elementos.

Retomando un poco el aspecto educativo de este trabajo, se optó por utilizar exámenes referidos a un criterio, ya que la intención principal del examen es explorar el nivel de dominio que tiene el estudiante sobre un universo de conocimientos claramente delimitado, es decir, poder medir las habilidades y conocimientos que este domina o en su caso las que no domina, pero en un campo específico del conocimiento. Como se ha mencionado el criterio para guiar el desarrollo del examen es el currículo de la asignatura en cuestión, dicho currículo especifica los contenidos temáticos a evaluar y la secuencia de los mismos, todas aquellas decisiones sobre lo que se va a evaluar y la forma de evaluarlos están directamente orientados por lo que se establece en el currículum. Por lo anterior se tiene la ventaja de poder describir de manera precisa y clara lo que se intenta medir.

Mediante este tipo de exámenes se obtienen puntuaciones individuales asociadas únicamente al desempeño del estudiante, y no en función del desempeño del grupo al estudiante pertenece, es decir a las puntuaciones que sus compañeros de salón de clase obtienen, como se supone la evaluación normativa. Además, que obtener puntuaciones particulares a cada

alumno ofrece una posibilidad de establecer estrategias individuales de mejora, situación que con exámenes normativos no permite.

La evaluación criterial tiene como punto de mayor importancia la representatividad del elemento respecto al universo de medida, entendida en el aspecto educativo como congruencia entre el ítem y el objetivo, es así como el estudiante puede tener la certeza de que lo que se le está evaluando es realmente lo que aprendió y no otro aspecto del conocimiento ajeno a su aprendizaje, mientras que para el docente representa la certeza que el puntaje obtenido en el examen refleja exactamente el aprendizaje del alumno en relación a lo enseñado en clase y de acuerdo a los contenidos temáticos determinados en el currículo.

Tomando como referencia el comentario anterior, se recogen algunas de las experiencias de los participantes en la aplicación del examen.

Estudiante 1: “me gusto el formato de opción múltiple, facilita responder el examen”.

Estudiante 2: “las preguntas fueron claras y sin tantos rodeos, como esas que son capciosas”.

Estudiante 3: “siento que el examen realmente puso a prueba lo que sabía de la materia, creo que abarco todos los temas vistos”.

Docente 1: “un examen que sigue un método cuantificable para determinar si una pregunta en un examen está bien hecha o no, me da la tranquilidad de saber que si un alumno mío responde correctamente a dicha pregunta es por que realmente sabe lo que se le está preguntando, y de cierta manera me dice que mis estrategias de aprendizaje han ayudado a que él pueda contestar correctamente”.

Docente 2: “la puntuación final del examen me hace saber si en general enseñe bien o no lo que se supone que se debe ver en la materia, pensando en que la clase se apega a lo que dice la carta descriptiva de la materia”.

Docente 3: “en un principio era escéptico de la calidad del examen, pero después de observar que cada tema se evaluó de acuerdo a la experiencia de más de 1 docente, que se justificó cada pregunta, que no fueron los mismos maestros los que hicieron todo el examen sino que participaron diferentes comités durante todo el proceso, y que después de eso se aplicaron

cálculos para demostrar lo que debía venir y lo que no en el examen, me hizo darme cuenta que la calificación que un alumno obtuviera en el examen en realidad era la que merecía tener, que ese número significaba realmente lo que sabía de la materia”.

Con los comentarios anteriores se pone de manifiesto la importancia de contar con exámenes estandarizados, ya que sin importar si existen “n” cantidad de versiones de un examen, todas y cada una de las diferentes versiones evalúan de igual forma el aprendizaje del estudiante, no hay exámenes más difíciles o más fáciles. De esta forma el docente como el alumno tienen la seguridad de que el instrumento es válido y confiable.

Por otro lado, en relación a la revisión sistemática que se hizo del estado del arte de este trabajo de investigación se hicieron notar ciertas similitudes, así como diferencias entre los autores, según el grupo temático que se abordó. Por ejemplo, en cuanto a la evaluación del aprendizaje, se hizo ver que en este campo se ha avanzado mucho y que ha sido un tema de constante investigación en las últimas décadas, la existencia de una influencia Europea sobre las metodologías o técnicas que competen a esta área, así como las teorías sobre las que se sustentan.

El interés por la evaluación del aprendizaje en el nivel superior es más latente, ya que hay una presión constante de la sociedad hacia las instituciones de educación superior por formar profesionales integrales, cuya capacidad de aprender a aprender sea permanente. Sin embargo, siguen existiendo espacios de confusión hacia dentro de las instituciones, ya que algunos autores referían que se sigue asociando el término evaluar con el de calificar.

Por otra parte, fue relevante observar en cuanto a las pruebas estandarizadas como un grupo de autores defienden este tipo de pruebas por su calidad técnica, por la utilización de métodos científicos, marcos de referencia teóricos y metodológicos rigurosos, mientras que algunos otros, aunque no las rechazan por completo si hacen notar o señalar puntualmente cuáles son sus defectos. Aunque en realidad, todos estos autores que señalaban las desventajas o la contraparte de utilizar pruebas estandarizadas hacían alusión a que estas pruebas se utilizaban para otros fines distintos para los cuales habían sido diseñadas, es decir, en si no hay una mala concepción con el diseño y la calidad técnica de las pruebas, sino más bien que normalmente se les ha utilizado para medir aspectos educativos que no son adecuados mucho

menos acertados, y que desgraciadamente en ocasiones los resultados de estas pruebas y sus malos usos tienen un alto impacto en las instituciones educativas.

Referente a las teorías y fundamentos de la prueba se observó la tendencia moderna de aplicar por ejemplo la TRI por encima de la TCT, también se notó como la mayoría de los autores concuerdan en que los atributos de validez y confiabilidad de la prueba pueden ser medidos con análisis de ciertos atributos psicométricos, como los índices de dificultad y de discriminación, pero resaltar que se considera más efectivo el coeficiente de discriminación. Y que en cuanto a la consistencia interna la mayoría de los estudios consideran de calidad los valores que se obtienen a partir del cálculo del KR20 y el Alfa de Cronbach.

En el último apartado, tecnología, evaluación y pruebas, se nota la tendencia en los autores por resaltar la importancia que ha tenido el desarrollo tecnológico en la mejora de la calidad de la educación en general, y en particular de las capacidades para mejorar el proceso de enseñanza-aprendizaje. Se dio cuenta del surgimiento de los softwares educativos y como estos en sus primeros años no tenían un marco de referencia para ser diseñados y construidos, tuvieron que pasar un par de años y un nuevo salto tecnológico para facilitar estos procesos, que a su vez también fue impulsado por los mismos cambios en esquemas y modelos educativos, que cada vez más hacían inclusión de tecnología en sus procesos; de esta manera, la presente investigación da cuentas de la necesidad de repensar los software educativos como una posibilidad para medrar, por un lado, la calidad de la educación superior a partir de sus procesos evaluativos, por otro, desde una perspectiva teórica y conceptual sobre la evaluación estandarizada de los aprendizajes, este trabajo responde a los límites encontrados en el estado del arte sobre la materia

Por último, y el espacio vacío más notorio después de hacer esta extensa revisión, fue la ausencia de software educativo con capacidades de análisis psicométrico integradas, en las investigaciones revisadas se encuentran estos dos elementos de forma separada, es decir, por un lado, se encuentra el software educativo que puede o no contar con capacidades evaluativas (esto dependiente del propósito del software, algunos son solo de carácter informativo, comunicativo etc.), y por otro el software diseñado específicamente para realizar análisis psicométrico.

Sin embargo, se recalca nuevamente, que no existe la fusión de los mismos, probablemente como se ha abordado en la sección de medición de la calidad de la prueba, los cálculos que se efectúan son un tanto complejos, y al considerar que se trata de pruebas estandarizadas (o de gran escala) se tenga la creencia que incluir la capacidad psicométrica en un software que permita representar digitalmente una prueba lo haga muy robusto o pueda restarle rendimiento. Sin embargo, con los avances tecnológicos en cuanto a plataformas para desarrollo de software, gestores de bases de datos y hardware de alto rendimiento, lo anterior no debería ser una limitante.

Es precisamente por lo anterior que cobra mayor relevancia esta investigación, ya que, según la revisión de la literatura, se identificaron investigaciones que abordan el análisis psicométrico, pero para ello emplean software de terceros especializado para dicha tarea (Cechova, Neubauer y Sedlacik, 2014; Thoe, Fook y Thah, 2008; Marr, Gupchup & Anderson, 2016), situación que destaca el vacío de conocimiento en cuanto tecnología educativa que integre las capacidades de análisis psicométrico con la funcionalidad de ejecutar y construir una prueba, todo en el mismo sistema.

Y agregado a esto otro atributo de innovación en la tecnología del sistema SIEXAES son los reportes técnicos especializados para el docente y el alumno; que aprovechan los atributos psicométricos y la información generada por ciertos instrumentos producto de la metodología para el desarrollo de la prueba estandarizada, para ofrecer información detallada y de relevancia para los actores involucrados, permitiendo así un conocimiento más cercano a la realidad del nivel de dominio en los estudiantes, y por su parte al docente, un panorama del aprendizaje de sus alumnos.

Dicha innovación fue alcanzada integrando procesos de la metodología para el desarrollo de la prueba con procesos de la metodología para la construcción de la propuesta tecnológica, ya que se habían identificado ciertos atributos e información de importancia para los actores principales del proceso de intervención, los cuales fueron traducidos en elementos multimedia y funciones automatizadas del software SIEXAES para lograr obtener información objetiva, especializada, válida y confiable.

Resultando lo anterior en términos de su calidad y relevancia para el desarrollo de los procesos de enseñanza y aprendizaje, en el cumplimiento de los objetivos (y en algunos casos superándolos) para los cuales se planteó este proyecto.

Dirección General de Bibliotecas de la UAQ

## 8. CONCLUSIONES

A lo largo de este proyecto de investigación pudieron observarse como los criterios o indicadores de calidad en el logro de la intervención iban cumpliéndose conforme se avanzaba en el ciclo de desarrollo del proyecto, hasta el grado que, en algunos casos, las metas esperadas fueron superadas hasta en un 25% de los deseado. Por mencionar un ejemplo, en el caso particular de las propiedades psicométricas, una primera ejecución del ciclo de evaluación produjo resultados que se clasificaban dentro del rango de regulares, sin embargo, posterior a la segunda ejecución del ciclo, los resultados alcanzaron casi los indicadores de excelente.

Lo anterior da pauta a como el proyecto fue evolucionando con la conclusión de cada etapa una de las etapas que las múltiples metodologías aplicadas en el proyecto, sobre todo aquellas que abordaban las estrategias interventivas con las cuales se deseaba resolver el problema de investigación, a saber, mejorar la calidad en la evaluación de los aprendizajes, esto mediante la utilización de pruebas estandarizadas mediadas por una tecnología innovadora que ejecutara análisis psicométricos de forma automatizada.

La evaluación de los aprendizajes siempre será un tema de polémica en las esferas educativas, siempre se encuentra presente la opinión en contra o a favor del uso de pruebas estandarizadas. Por un lado, expresando la ventaja de contar con instrumentos con un sustento teórico metodológico muy robusto que permite observar la validez y la confiabilidad del instrumento, aportando así una seguridad a la toma de decisiones. Sin embargo, se encuentra también la contraparte la cual expresa que la evaluación no puede seguir una única línea o interpretación cuantitativa del aprendizaje, situación que una prueba, aun cuando esta haga uso de las bondades tecnológicas, no podrá cubrir totalmente o de forma natural como pudiera lograrse mediante un conjunto más variado de tareas evaluativas.

Está claro que la evaluación no debe concluirse estrictamente solo con la escala numérica o distribución porcentual que un examen escrito o digital presenta, sin embargo, no se puede ignorar que este tipo de pruebas departamentales cumplen un propósito formativo que desde la perspectiva del autor de esta investigación debe ser parte fundamental de la evaluación general del curso. Considero que este tipo de pruebas pueden aportar significativamente a la mejora de la calidad del aprendizaje de los estudiantes, por la sencilla razón de que al contar



con un respaldo metodológico que comprende análisis cuantitativos y cualitativos por parte de un grupo heterogéneo de expertos en la materia, expertos en el currículo, expertos en docencia y expertos en el plan curricular de la carrera, los resultados obtenidos a partir de la prueba sugieren una mayor confiabilidad y validez. Situación que, al compararse con las pruebas formativas que el docente aplica durante el semestre, permite hacer una comparación significativa y objetiva del nivel de dominio, o aprendizaje alcanzado por el alumno, por lo menos al momento de la ejecución de la prueba.

Esto abre una serie de oportunidades y ventajas, por ejemplo, el docente puede tener la seguridad de que, si hay una variación en los resultados obtenidos en la prueba departamental y las obtenidas durante el semestre, lo más probable es que esta se atribuya a las condiciones durante las que se ejecutaron las tareas evaluativas, así como a la calidad misma de dichas evaluaciones. Permitiendo al docente hacer una reflexión fundamentada sobre su práctica y sus estrategias. Por otra parte, el alumno puede confirmar de manera confiable si lo que aprendió durante el semestre realmente fue significativo, ya que de ser así la prueba departamental tendría que ser un reflejo de lo alcanzado durante este periodo.

Así mismo el alumno puede tener la seguridad de que la prueba, aun cuando existan diferentes versiones, evalúa de la misma manera el nivel de dominio adquirido, por lo que no deben existir diferencias, por lo menos atribuidas al instrumento. Pero aquí entra en juego la situación con el docente, en muchos casos los alumnos optan por llevar un curso o unidad de aprendizaje con determinado docente, puesto que “es más fácil pasar, o los exámenes son más fáciles”, lo anterior, empleado estas pruebas departamentales ofrecen la posibilidad que el estudiante independientemente del docente con quien curse la asignatura, al final tenga las mismas probabilidades de acreditarlo.

Claro está, esto también tiene una contraparte negativa, porque si el docente ofrece una baja calidad en la instrucción y las estrategias de enseñanza-aprendizaje, en consecuencia, el alumno tendrá menos probabilidades de acreditar la prueba departamental. Por ello, la prueba no puede ser el instrumento definitivo para promover o no el curso. Lo que sí es un hecho, es que la prueba si puede contribuir a que las áreas administrativas puedan hacer observaciones o comparaciones entre los docentes que imparten la asignatura, ya que todos los docentes que impartan esa materia trabajan sobre el mismo universo de conocimientos

definido en el currículo, conocen la estructura y formato de la prueba departamental, incluso asignan el porcentaje de valor que la prueba tendrá. Lo anterior entonces reduce las posibles causas de diferencias entre los resultados obtenidos durante el semestre y los resultados obtenidos con esta prueba sumativa.

Considero que uno de los aportes más significativos de este trabajo de investigación es la capacidad integrada que tiene la herramienta tecnológica innovadora SIEXAES, ya que un docente con muy poca o casi nula experiencia en análisis psicométricos puede obtener un reporte técnico detallado de la calidad psicométrica de la prueba. Obteniendo información diseñada y presentada de manera que sea comprensible, además, los reportes de resultados para el docente y el alumno presentan información con un nivel de abstracción que permite observar situaciones particulares del aprendizaje a nivel de alumno y a nivel de grupo. Permitiendo así hacer conclusiones y análisis acertados sobre el estado del aprendizaje que manifiesta el estudiante, con un mayor grado de acercamiento a la realidad.

Así mismo estos reportes contribuyen a un proceso de evaluación más significativo y objetivo, ya que ofrece una serie de datos con los cuales la retroalimentación trasciende la escala numérica representada en la prueba. Permitiendo así al alumno conocer con detalle cómo fue su desempeño a nivel general, por unidad de aprendizaje, por tema y destacando aquellos temas de mayor impacto o relevancia curricular.

Y, aunque no es el propósito de la prueba, también puede comparar su desempeño en relación a los demás sujetos pertenecientes a su grupo. Por su parte el docente puede observar el comportamiento del grupo con un grado de abstracción que le permite identificar distintos niveles de información, tales como promedios generales en la asignatura, promedios por unidad de aprendizaje, por contenido temático, y desde luego en especial los temas de mayor relevancia curricular. Lo anterior le ofrece al docente la posibilidad de observar a detalle que aspectos del universo de conocimientos no se cubrieron adecuadamente, y poder efectuar ajustes a su práctica docente, y como se dijo anteriormente, el docente tiene un grado de certeza más alto de que las deficiencias son en mayor proporción a factores externos, y no atribuidos a la prueba departamental.

Contribuye también a realizar autoevaluaciones sobre la calidad de sus instrumentos, identificar áreas de oportunidad en los reactivos que emplea en sus pruebas. Y lo anterior se potencializa al emplear una metodología para el desarrollo de la prueba. Como se mencionó anteriormente, el propósito de esta intervención era crear una cultura evaluativa enfocada en el uso de pruebas estandarizadas empleando metodologías para el desarrollo de las mismas, ya que esto permite al docente poder visualizar un panorama más completo del comportamiento de los estudiantes, al poder identificar aquellos temas de mayor relevancia curricular, cuántos de estos domina el grupo o en su defecto cuales no entre otros elementos. De esta forma el docente no solo percibe un valor numérico como única referencia en el aprovechamiento – rendimiento de un grupo, sino que por el contrario realmente puede dosificar aspectos más concretos del universo de conocimientos que está evaluando.

Sin embargo, aún existen aspectos pendientes por mejorar e investigar, tales como la inclusión de ítems que permitan medir de una forma más natural o cercana a la realidad el universo de conocimientos, permitir observar niveles cognitivos más complejos. Lo anterior implicaría utilizar ítems del tipo de respuesta construida, de relación o pareo, completar, entre algunos otros. Emplear nuevos formatos de ítems requerirá de un proceso de investigación que permita comprobar si hay una mejora en el nivel de dominio de los estudiantes, al interactuar con ítems que permitan evidenciar de forma más amplia y natural el conocimiento adquirido. Incluso se podrían hacer comparativas entre los resultados obtenidos con los tipos de ítems que se emplean en las pruebas actualmente y los obtenidos empleado este otro tipo de ítems.

También sería útil efectuar un análisis del comportamiento estudiantil en estas pruebas considerando ahora un periodo de acción más amplio, que implique a un mayor número de generaciones, y dar un seguimiento al rendimiento alcanzado por los estudiantes en aquellas materias que guardan relación directa con los conocimientos adquiridos en la asignatura para la cual se desarrolló la prueba estandarizada de esta investigación. Obteniendo así información que permita validar si un buen desempeño en la prueba estandarizada tiene relación con mejor rendimiento en otras asignaturas, o en su defecto lo contrario.

Así mismo, es importante mencionar la definición de una estructura lógica que permita establecer al momento de la creación de la prueba el marco de referencia sobre el universo

de medida que compone la prueba, es decir la identificación de aquellas unidades, temas, subtemas etc. que posean el mayor índice de relevancia curricular en relación a los ítems que componen la prueba. De forma tal que se pueda saber que ítem corresponde a que contenido temático, y si el mismo es representativo de alguno de aquellos contenidos que son más relevantes para la concreción de la competencia final del curso, o como se ha dicho anteriormente, de mayor relevancia curricular.

El impacto positivo que logró la intervención en la institución educativa es notorio, ya que, por citar un ejemplo, los docentes ahora analizan con más detenimiento el comportamiento de su grupo, contrastando las evaluaciones formativas realizadas en el transcurso del semestre y la evaluación sumativa realizada con la evaluación estandarizada. Además, ha impulsado a mejorar ciertas estrategias de aprendizaje e implementar algunas otras, esto con el fin de mejorar el aprendizaje en los estudiantes, particularmente en aquellos temas que, gracias a la metodología de desarrollo de la prueba y los reportes técnicos generados por el sistema, son de mayor relevancia curricular para la concreción de la competencia del curso y a la vez que son aquellos que tienen influencia en la adquisición de competencias futuras presentes en otras asignaturas.

También ha facilitado la comparativa entre el desempeño de grupos que llevan la misma asignatura, permitiendo observar a nivel de temas y subtemas el rendimiento del grupo y poder analizar en sesiones plenarias y de retroalimentación las diferentes estrategias aplicadas entre un docente y otro, logrando enriquecer los beneficios didácticos para los docentes que imparten dicha asignatura. Siendo también positivo el hecho de que los docentes ven la prueba estandarizada como un ejercicio de autoevaluación a su práctica, permitiendo identificar el comportamiento de todo el grupo en el universo completo de conocimientos que la carta descriptiva de la asignatura comprende.

Creo que cualquier docente estaría interesado en tener a un clic de distancia el poder evaluar la calidad psicométrica de sus instrumentos de evaluación, y poder obtener reportes detallados, robustos y completos de toda la información correspondiente a la prueba.

Sin embargo, para lograr el éxito de este tipo de pruebas no solo depende de la capacidad de funciones y procesamiento que la herramienta SIEXAES posea, sino que definitivamente

tiene que ir acompañada por el marco metodológico que permite elaborar la prueba. Ya que, por muchas funciones y capacidades de análisis que posea la herramienta tecnológica, si el docente construye una prueba sin sustento metodológico y teórico, los resultados solo le confirmarían que la prueba no está bien hecha, y por ende la toma de decisiones se interfiere por completo.

Considero que este trabajo de investigación empleó el modelo de intervención adecuado para la necesidad y problemática que existía, ya que se pudieron observar desde diferentes perspectivas como todos los elementos que formaban parte del proceso se alineaban a dichas problemáticas identificadas. La etapa de evaluación del proceso de intervención arrojó resultados positivos sobre lo implementado, mostrando un impacto favorable en la institución. Lo anterior se pudo concluir con las variaciones obtenidas en las apreciaciones de los docentes en relación al proceso de intervención, se veía como al principio algunos docentes consideraban que estas pruebas no eran útiles o se encontraban reacios a replantear la calidad de sus instrumentos de evaluación, y al final de la intervención el porcentaje de aceptación subió más de un 45%.

Uno de los propósitos de este proyecto, y que de hecho fue un indicador de éxito en la intervención era la creación o mejorar la cultura evaluativa al interior de la institución educativa. Y desde mi perspectiva se logró, ya que existen una serie de trabajos futuros con los cuales se podrá establecer un marco metodológico completo para la evaluación del aprendizaje en la institución, haciendo uso de la herramienta tecnológica para no solo evaluaciones sumativas, sino también para aquellas evaluaciones formativas que puedan ser representadas de manera digital. Así mismo se tiene pensado trabajar en un programa de capacitación continua sobre el uso de los reportes que genera el sistema SIEXAES para establecer procesos de retroalimentación más objetivos.

Por otra parte, se tiene pensando la conformación de un equipo de evaluación colegiada del aprendizaje, el cual se encargue a nivel institución de la elaboración de programas de capacitación continuos sobre el desarrollo de instrumentos de evaluación empleando la metodología vista en este trabajo de investigación, así como de buenas prácticas en la elaboración de diferentes instrumentos de evaluación formativa. Logrando con lo anterior

una continua mejora en la cultura evaluativa de la institución, y en general en la calidad del aprendizaje.

La puesta en marcha de este proyecto me permitió comprobar la importancia de evaluar bien, desde la concepción del proceso, la ejecución y la conclusión del mismo, la cual dista de solo presentar una nota al final de la prueba. Pude comprobar lo importante que es contar con instrumentos válidos y confiables sobre los cuales tomar decisiones de alto impacto, que trascienden a los alumnos, y que pueden llegar hasta las esferas de gestión o dirección de la institución. Considero que la evaluación es un elemento de la educación que muchas ocasiones se subestima y en algunos otros se cree se domina muy bien, pero vista a la luz de metodologías con años de aplicación y con sustentos teóricos significativos demuestran que la tarea de evaluar no es algo sencillo, pero que con una correcta integración de tecnología pueden efficientizar muchos procesos. Contribuyendo así a la calidad del aprendizaje en los estudiantes, tarea que conforme pasan los años se va haciendo más complicada para los docentes, ya que existen muchos medios y fuentes al alcance de los alumnos, por lo que constantemente el docente debe replantear su práctica y sus estrategias de enseñanza-aprendizaje.

Un mal instrumento de evaluación no queda solo en el mesabanco, afecta a todos los actores involucrados en el proceso de evaluación, por ello la importancia de contar con instrumentos de evaluación válidos y confiables. Una buena evaluación es aquella que mide lo que debe medir, que con sus resultados se puedan tomar decisiones efectivas en pro de la mejora de la calidad del aprendizaje de los estudiantes. Hay que recordar que la única manera de saber cómo nos encontramos y a donde vamos es evaluando.

Por ello, la responsabilidad de nosotros como educadores no solo se debe limitar a la calidad de nuestra enseñanza, sino también a la calidad con la que evaluamos el aprendizaje de nuestros estudiantes, solo así podremos contribuir de manera significativa a la formación y desarrollo de todos aquellos seres humanos que crucen por nuestras aulas.

## 9. REFERENCIAS

- Aigeneren, Miguel (1999), "Análisis de contenido: una introducción", en: <https://webcache.googleusercontent.com/search?q=cache:GyfFmKMDIwQJ:https://revistas.udea.edu.co/index.php/ceo/article/download/1550/1207+&cd=1&hl=es&ct=clnk&gl=mx&client=safari> (consulta: 1 abril de 2020).
- Alcaraz Salarirche, Noelia (2015), "Evaluación versus calificación", *Aula de Encuentro*, vol. II, núm. 17, pp. 209-236.
- Aliaga Tovar Jaime (2006), "Psicometría: test psicométricos, confiabilidad y validez", en: Quintana Alberto, *Psicología: Tópicos de Actualidad*. Lima: UNMSM. pp. 85-108.
- Almaguel Guerra, Adrian; Álvarez Mora, Denis; y Pernía Nieves, Luis Alberto (2016), "Software educativo para el trabajo con matrices", *Revista Digital: Matemática, Educación e Internet*, vol. XVI, núm. 2. pp. 1-12.
- Argibay, Juan Carlos (2006), "Técnicas psicométricas. Cuestiones de validez y confiabilidad", *Subjetividad y procesos cognitivos*, vol. VIII, pp. 15-33.
- Árraga Barrios, Marisela Virginia y Sánchez Villarroel, Marhilde (2012), "Validez y confiabilidad de la Escala de Felicidad de Lima en adultos mayores venezolanos", *Universitas Psychologica*, vol. XXI, núm. 2, pp. 381-393.
- Attorresi, H. F., Lozzia, G. S., Abal, F. J. P., Galibert, M. S., & Aguerri, M. E. (2009). Teoría de Respuesta al Ítem. Conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina de Clínica Psicológica*, 18(2), 179-188.
- Backhoff Escudero, Eduardo (2018), "Evaluación estandarizada de logro educativo: contribuciones y retos", *Revista Digital Universitaria*, vol. XIX, núm. 6, pp. 1-14.
- Backhoff Escudero, Eduardo; Larrazolo Reyna, Norma; y Rosas Morales, Martín (2000), "Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA)", *REDIE. Revista Electrónica de Investigación Educativa*, vol. II, núm. 1, pp.11-28.
- Backhoff, E. y Tirado, F. (1992). Desarrollo del Examen de Habilidades y Conocimientos Básicos. *Revista de la Educación Superior*, 83, 95-117.

- Baghaei, P., y Carstensen, C. (2013). Fitting the Mixed Rasch Model to a Reading Comprehension Test: Identifying Reader Types. *Practical Assessment, Research & Evaluation*, 18(5).
- Baladrón, Jaime; Sánchez Lasheras, Fernando; Romeo Ladrero, José; Curbelo, José; Villacampa Menéndez, Paloma; y Jiménez Fonseca, Paula (2018), “Evolución de los parámetros dificultad y discriminación en el ejercicio de examen MIR. Análisis de las convocatorias de 2009 a 2017”, *FEM: Revista de la Fundación Educación Médica*, vol. XXI, núm 4, pp. 181-193.
- Barbero, M. (2010). *Psicometría*. Madrid: Sanz y Torres.
- Barrenechea, I. (2010). *Evaluaciones Estandarizadas: Seis Reflexiones Críticas*.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee’s ability. En F. Lord & M. Novick (Eds.). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison Wesley.
- Bogantes Pessoa, Jenny (2015), “Estrategias para la evaluación en educación a distancia: un análisis de las opciones empleadas en el programa de educación general básica de la UNED”, *Innovaciones educativas*, vol. XVII, núm. 22, pp. 15-25.
- Brooks, Gordon y Johanson, George. (2003), “TAP: Test Analysis Program”, *Applied Psychological Measurement*, vol. XXVII, núm. 4, pp. 303–304.
- Cabero, J. (2010). Los retos de la integración de las TICs en los procesos educativos. Límites y posibilidades. *Perspectiva Educativa, Formación de Profesores*, 49 (1), 32-61.
- Canabal, C., y Margalef, L. (2017). La retroalimentación: la clave para una evaluación orientada al aprendizaje. *Profesorado. Revista de Currículum y Formación de Profesorado*, 21(2),149-170.
- Cataldi, Zulma (2000), *Una metodología para el diseño, desarrollo y evaluación de software educativo*, Tesis de Magister, Buenos Aires, Universidad Nacional de la Plata (Argentina).
- Cataldi, Zulma; Lage, Fernando; Pessacq, Raúl y García, Ramón (1999), “Ingeniería de software educativo”, ponencia presentada en el congreso “V Congreso Internacional



de ingeniería informática”, Buenos aires, en: <http://laboratorios.fi.uba.ar/lsi/c-icie99-ingenieriasoftwareeducativo.pdf>.

Cechova, Ivana; Neubauer, Jiri y Sedlacik, Marek (2014), “Computer-adaptive testing: item analysis and statistics for effective testing”, ponencia presentada en “European Conference on e-Learning”, Copenhagen, 30-31 de octubre de 2014.

Contreras Pérez, Gloria (2010), “Diagnóstico de dificultades de la evaluación del aprendizaje en la universidad: un caso particular en Chile”, *Educación y Educadores*, vol. XIII, núm. 2, pp. 219-238.

Contreras, L. (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. Ensenada, B. C.: IIDE. UABC. Disponible en: <http://eduweb.ens.uabc.mx/egresados/Tesis/indicetesis.htm>

Contreras, L. y Backhoff, E. (2004). Metodología para elaborar exámenes criteriosales alineados al currículo. En: Castañeda, S. (Ed.), Educación aprendizaje y cognición, teoría en la práctica. México: Manual Moderno. ISBN 970 729 088 9.

Contreras, L., Encinas, J. y De Las Fuentes, M. (2005). Evaluación Colegiada del Aprendizaje en la Universidad Autónoma de Baja California: el caso del examen de Matemáticas I de la Facultad de Ingeniería Mexicali. En: Memoria del VIII Congreso Nacional de Investigación Educativa. México: COMIE.

Cortada De Kohan, Nuria (2004), “Teoría de respuesta al ítem: supuestos básicos”, *Revista Evaluar*, vol. IV, núm. 1, pp. 95-110.

Couturejuzón González, Lourdes (2003), “Cumplimiento de los principios didácticos en la utilización de un software educativo para la educación superior”, *Educación Médica Superior*, vol XVII, núm. 1, pp. 53-57.

Cova, Ángela; Arrieta, Xiomara y Riveros, Victor (2008), “Análisis y comparación de diversos modelos de evaluación de software educativo”, *Enl@ ce: Revista Venezolana de Información, Tecnología y Conocimiento*, vol. V, núm. 3, pp. 45-67.

Cronbach, Lee; Gleser, Goldine; Nanda, Hariender y Rajaratnam, Nageswari (1972), *The dependability of behavioral measurements: theory of generalizability for scores and profiles*, New York, John Wiley and Sons.

- Cruz Núñez, Fabiola, y Quiñones Urquijo, Abel (2012). Importancia de la evaluación y autoevaluación en el rendimiento académico. *Zona Próxima*, núm. 16, pp. 96-104.
- Cueto, S. (2011). La evaluación estandarizada de la educación ciudadana en América Latina: Estado de la cuestión y algunas propuestas., PREAL, p. 3.
- De la Orden Oz, Arturo y Pimienta Prieto, Julio Herminio (2016), Instrumento para determinar los tipos de evaluación utilizados por los profesores universitarios. *Revista electrónica de investigación educativa*, vol. XVII, núm. 2, pp. 40-52.
- Díaz, Gabriela; Pérez, María; Mendoza, Luis y Grimán, Anna (2003), “Calidad Sistémica del Software Educativo”, ponencia presentada en el “Congreso Internacional Edutec' 2003: Gestión de las Tecnologías de la Información y la Comunicación en los diferentes ámbitos educativos”, Caracas, Universidad Central de Venezuela, 24 al 28 de noviembre de 2003.
- Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas, 18, 1-27.
- Esterkin, Viviana y Pons, Claudia (2017), “Evaluación de calidad en el desarrollo de software dirigido por modelos”, *Ingeniare. Revista chilena de ingeniería*, vol. XXV, núm. 3, pp. 449-463.
- Ezpeleta, Lourdes; De La Osa, Nuria; Domenech, José María; Navarro, José y Losilla, Josep (1997), “Fiabilidad test-retest de la adaptación española de la Diagnostic Interview for Children and Adolescents (DICA-R)”, *Psicothema*, vol. IX, núm. 3, pp. 529-539.
- Fernández-Pampillón, Ana. (2009). Las plataformas e-learning para la enseñanza y el aprendizaje universitario en Internet. En *Las plataformas de aprendizaje. Del mito a la realidad*. Biblioteca Nueva, Madrid, pp. 45-73. ISBN 978-84-9742-944-3
- Fernández Navas, Manuel; Alcaraz Salarirche, Noelia, y Sola Fernández, Miguel (2017), “Evaluación y pruebas estandarizadas: Una reflexión sobre el sentido, utilidad y efectos de estas pruebas en el campo educativo”, *Revista Iberoamericana de Evaluación Educativa*, vol. X, núm. 1, pp. 51-67.
- Ferreira, María y Backhoff Escudero, Eduardo (2016), “Validez del Generador Automático de Ítems del Examen de Competencias Básicas (Excoba)”, *RELIEVE-Revista Electrónica de Investigación y Evaluación Educativa*, vol. XXII, núm. 1, pp. 1-16.

Galvis Panqueva, Alvaro (2000), *Ingeniería de software educativo*, Colombia, Ediciones Uniandes.

García Sánchez, Eduardo; Vite Chávez, Osbaldo; Navarrate Sánchez, Miguel Ángel; García Sánchez, Miguel Ángel y Torres Cosío, Verónica (2016), “Metodología para el desarrollo de software multimedia educativo MEDESME”, *CPU-e. Revista de Investigación Educativa*, núm. 23, pp. 216-226.

García, A., Aguilera, M., Pérez, M. y Muñoz, G., & Instituto Nacional para la Evaluación de la Educación (México). (2011). *Evaluación de los aprendizajes en el aula: Opiniones y prácticas de docentes de primaria en México* (Primera ed.). México D.F: Instituto Nacional para la Evaluación de la Educación.

García-Valcárcel, A. y Hernández, A. (2013). *Recursos tecnológicos para la enseñanza e innovación educativa*. Madrid: Síntesis, p. 30.

Gómez del Castillo, María Teresa (1998), “Un ejemplo de evaluación de software educativo multimedia”, ponencia presentada en el “Congreso Internacional EDUTEC 97: Creación de materiales para la innovación educativa con nuevas tecnologías”, Málaga, 27-29 de octubre de 1997.

Gómez Yepes, Ricardo León (2004), “Calidad educativa: más que resultados en pruebas estandarizadas”, *Revista educación y pedagogía*, vol. XVI, núm. 38, pp. 75-89.

González Maya, Yuliana Lourdes (2006), “*Diseño e implementación de un software educativo para el autoaprendizaje del diseño de bases de datos relacionales*”, Tesis de licenciatura, Ciudad de México, Universidad Nacional Autónoma de México (México).

Hambleton, R. y Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer.

Heredia Manrique, Alfonso. (2009). “Dos formas diferenciadas de evaluación didáctica: evaluación normativa para seleccionar a los alumnos y evaluación criterial para el dominio del conocimiento básico”, *Bordon: Revista de pedagogía*, vol. LX1, núm. 4, pp. 39-48.

- Hidalgo Farran, Nina y Murillo Torrecilla, Francisco Javier (2017), “Las concepciones sobre el proceso de evaluación del aprendizaje de los estudiantes”, *REICE: Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, vol. XV, núm. 1, pp. 107-128.
- Izquierdo, Beatriz (2008), “De la evaluación clásica a la evaluación pluralista. Criterios para clasificar los distintos tipos de evaluación”, *EMPIRIA. Revista de Metodología de las Ciencias Sociales*, núm. 16, pp. 115-134.
- Labarca, J., Figueroa, C., Huidobro, B., Wright, A. C., Riquelme, A., & Moreno, R. (2014). Utilidad de la evaluación formativa en cursos clínicos integrativos en estudiantes de pregrado de medicina. *Revista médica de Chile*, vol. CXLII, núm. 9, pp. 1193-1199.
- Jornet Meliá, Jesús Miguel (2017), “Evaluación estandarizada”, *Revista iberoamericana de evaluación educativa*, vol. X, núm. 1, pp. 5-8.
- Jornet Meliá, Jesús Miguel & José González Such (2009), “Evaluación criterial: determinación de estándares de interpretación (EE) para pruebas de rendimiento educativo”, *Estudios sobre Educación*, núm. 16, pp. 103-123.
- Landish, Richard, y Koch Gary (1977), “The measurement of observer agreement for categorical data”, *Biometrics*, vol. XXXIII, núm. 1, pp. 159-174.
- López Noguero, Fernando (2002), “El análisis de contenido como método de investigación”, *Revista de Educación*, núm. 4, pp. 167-179.
- López Reguera, Jorge; Hernández Rivas, Cecilia y Farran Leiva, Yussef (2011), “Una plataforma de evaluación automática con una metodología efectiva para la enseñanza/aprendizaje en programación de computadores”, *Ingeniare. Revista Chilena de Ingeniería*, vol. XIX, núm. 2, pp. 265-277.
- Lord, Frederic (1980), *Applications of item response theory to practical testing problems*, New York, Hillsdale: Erlbaum Associates.
- Lord, Frederic y Novick, Melvin (1968), *Statistical theories of mental test scores*, New York, Addison-Wesley.
- Madariaga Fernández, Carlos Jesús; Rivero Peña, Yasnalla y Leya Téllez, Arquimedes René. (2016), “Propuesta metodológica para desarrollo de software educativo en la Universidad de Holguín”, *Ciencias Holguín*, vol. XXII, núm. 4, pp.1-17.

- Mandeville, Peter. (2005), "El coeficiente de correlación intraclase (ICC)", *Ciencia UANL*, vol. VIII, núm. 3, pp. 414-416.
- Marquès Graells, Pere (1998), "La evaluación de programas didácticos", *Comunicación y Pedagogía*, núm. 149, pp. 53-58.
- Marr, Lisa; Gupchup, Gireesh, y Anderson, Joe (2012), "An evaluation of the psychometric properties of the Purdue Pharmacist Directive Guidance Scale using SPSS and R software packages", *Research in Social and Administrative Pharmacy*, vol. VIII, núm. 2, pp. 166-171.
- Martínez Rizo, Felipe (2001), "Evaluación educativa y pruebas estandarizadas. Elementos para enriquecer el debate", *Revista de la educación superior*, vol. XXX, núm. 120, pp. 1-12.
- Martínez Rizo, Felipe (2009), "Evaluación formativa en aula y evaluación a gran escala: hacia un sistema más equilibrado", *Revista electrónica de investigación educativa*, vol. XI, núm. 2, pp. 1-18.
- Martínez Ruíz, María y Sauleda Parés, Narciso (1992), "La evaluación de software educativo en el escenario de la evolución de los paradigmas educativos", *Enseñanza y Teaching: Revista Interuniversitaria de Didáctica*, vol. X, pp. 161-174.
- Martínez, F. (2013). Las Pruebas Estandarizadas y el Diseño de la Política Educativa en México. Recuperado de: <http://estepais.com/site/2013/las-pruebas-estandarizadas-y-el-diseno-de-la-politica-educativa-en-mexico>
- Martínez, M. (1995). *Psicometría: Teoría de los Tests Psicológicos y Educativos*. Madrid: Síntesis.
- Mascheroni, M., Greiner, C., Dapozo, G., y Estayno, M. (2013). Ingeniería de Usabilidad. Una Propuesta Tecnológica para Contribuir a la Evaluación de la Usabilidad del Software. *Revista Latinoamericana de Ingeniería de Software*, 1(4), 125-134.
- Medina Paredes, Jhonny; Ramírez Díaz, Mario Humberto y Miranda, Isaías (2019), "Validez y confiabilidad de un test en línea sobre los fenómenos de reflexión y refracción del sonido", *Apertura: Revista de Innovación Educativa*, vol. XI, núm. 2, pp. 104-121.

- Mendivil Zuñiga, Tulio Nel (2012), "Sistema de evaluación del aprendizaje en los estudiantes de educación superior en la región caribe colombiana", *Dimensión empresarial*, vol. X, núm. 1, pp. 100-107.
- Meneses, J., Barrios, M., Lozano, L. M., Bonillo, A., Turbany, J., Cosculluela, A., y Valer, S. (2014). *Psicometría*. Editorial UOC.
- Mneimneh, Z., Heeringa, S., Tourangeau, R., y Elliott, M. (2014). Bringing Psychometrics and Survey Methodology: Can Mixed Rasch Models Identify Socially Desirable Reporting Behavior? *Journal of Survey Statistics and Methodology*, 2(3), p. 257-282.
- Monje, C. (2011). Metodología de la investigación cuantitativa y cualitativa. Guía didáctica. *Neiva, Colombia: Universidad Surcolombiana*.
- Moreno Olivos, Tiburcio (2016), "Las pruebas estandarizadas en la escuela contemporánea, ¿llave o cerrojo para la mejora de la educación?", *Temas de Educación*, vol. XXII, núm. 1, pp. 83-96.
- Muñiz, J. (1997). Introducción a la teoría de respuesta a los ítems. Madrid: Ediciones Pirámide.
- Muñiz, J. (2010). Las teorías de los tests: teoría clásica y teoría de respuesta a los ítems. *Papeles del Psicólogo*, 31 (1), p. 57-66.
- Muñiz, José. (2010), "Las teorías de los tests: teoría clásica y teoría de respuesta a los ítems", *Papeles del Psicólogo: Revista del Colegio Oficial de Psicólogos*, vol. XXXI, núm. 1, pp. 57-66.
- Murillo, F., y Román, M. (2016). La evaluación educativa como derecho humano. *Revista Iberoamericana de evaluación educativa*, 1(1), p. 1-5.
- Navarro, Eulalia (1999), "Análisis de productos multimedia educativos", *Comunicación y Pedagogía*, núm. 157, pp. 37-40.
- Nitko, A. (1994). A Model for Development Curriculum-Driven Criterion- Referenced and Norm-Referenced Examination for Certification and Selection of Students. Documento presentado en la Conference of Education, Evaluation and Assessment for the Association Studies of Educational Evaluation in Sudafrica (ASEESA). Sudáfrica.

- Olivar, A., y Daza, A. (2007). Las tecnologías de la información y comunicación (TIC) y su impacto en la educación del siglo XXI. *Negotium*, 3 (7), p. 21-46.
- Orjuela Duarte, Ailin y Rojas, Mauricio (2008), “Las metodologías de desarrollo ágil como una oportunidad para la ingeniería del software educativo”, *Revista Avances en Sistemas e Informática*, vol. V, núm. 2, pp. 59-171.
- Ortega, F. (2017). Principios e implicaciones del Nuevo Modelo Educativo. *Revista Latinoamericana de Estudios Educativos (México)*, XLVII (1), 43-62.
- Ortiz, G., Díaz, P., Llanos, O., Pérez, S., y González, K. (2015). Dificultad y discriminación de los ítems del examen de Metodología de la Investigación y Estadística. *Revista Educación Médica del Centro*, 7(2), 19-35.
- Pérez Tapia, Julio Hernán; Acuña Aguilar, Nelson y Arratia Cuela, Elvis Ruben (2008), “Nivel de dificultad y poder de discriminación del tercer y quinto examen parcial de la cátedra de cito-histología 2007 de la carrera de medicina de la UMSA”, *Cuadernos Hospital de Clínicas*, vol. LIII, núm. 2, pp.16-22.
- Poole, Bernand (1999), *Tecnología Educativa: educar para la sociedad de la comunicación y del conocimiento*, España, Mc Graw Hill. 2da. Edición.
- Popham, James (1999), “Why standardized test don't measure educationaly quality (Programa de Promoción de la Reforma Educativa en América Latina y el Caribe, Grupo de Trabajo sobre Estándares y Evaluación, Trans)”, *Educational Leadership*, vol. LVI, núm. 6, pp. 2-11.
- Popham, W. (2000). *Modern educational measurement*. Boston: Allyn y Bacon.
- Rasch, Georg (1960), *Probabilistic models for some intelligence and attainment tests* Denmark, Copenhague Denmarks Paedagogiske Institut.
- Reidl Martínez, Lucy María (2013), “Confiabilidad en la medición”, *Investigación en educación médica*, vol. II, núm. 6, pp. 107-111.
- Reyes Caballero, Faustino; Fernández Morales, Flavio Humberto y Duarte, Julio Enrique (2015), “Herramienta para la selección de software educativo aplicable al área de tecnología en educación básica”, *Entramado*, vol. XI, núm. 1, pp. 186-193.

- Rosales Mejía, M. (2014). Proceso evaluativo: evaluación sumativa, evaluación formativa y Asesment su impacto en la educación actual. En *Congreso Iberoamericano de Ciencia, Tecnología, Innovación y Educación*, vol. IV, pp. 662.
- Santelices Etchegaray, María Verónica y Valenzuela Rojas, Francisco (2015), “Importancia de las características del profesor y de la escuela en la calidad docente: Una aproximación desde la Teoría de Respuesta del Ítem”, *Estudios pedagógicos*, vol. XLI, núm. 2, pp. 233-254.
- Serra Mayoral, Anna y Peña Casanova, Jordi (2006), “Fiabilidad test-retest e interevaluador del Test Barcelona”, *Neurología*, vol. XXI, núm. 6, pp. 277-281.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, p. 72-101.
- Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *American Journal of Psychology*, 18, p. 161-169.
- Spearman, C. (1913). Correlations of sums and differences. *British Journal of Psychology*, 5, p. 417-426.
- Squires, David y McDougall, Ann (1997). *Cómo elegir y utilizar software educativo: guía para el profesorado*, Madrid, Ediciones Morata.
- Straccia, Luciano; Zanetti, Paula y Pollo Cattaneo, María Florencia (2019), “Definición de un estándar para la evaluación de calidad de software educativo”, ponencia presentada en el “XXV Congreso Argentino de Ciencias de la Computación”, Córdoba, 14 al 18 de octubre de 2019.
- Thoe, Ng Khar; Fook, Fong Soon; Thah, Soon Seng. (2009), “Use of ICT tool for Item Analysis of a Science Performance Test”, *Journal of Educational Technology*, vol. IX, núm.1, pp 5-15.
- Thurstone, L. (1925). A method of scaling psychological and educational tests. *The Journal of Educational Psychology*, 16, p. 433-451.
- Tirado, F., Backhoff, E., Larrazolo, N., y Rosas, M. (1997). Validez predictiva del Examen de Habilidades y Conocimientos Básicos (EXCOBA). *Revista Mexicana de Investigación Educativa*, 2 (3), p. 67-84.



- Tirado, Felipe, Backhoff Escudero, Eduardo y Larrazolo Reyna, Norma (2016), “La revolución digital y la evaluación: un nuevo paradigma”, *Perfiles educativos*, vol. XXXVIII, núm. 152, pp. 182-201.
- Tiramonti, Guillermina (2014), “Las pruebas PISA en América Latina: resultados en contexto”, *Avances en Supervisión Educativa*, núm. 20, pp. 1-24.
- Tristán López, Agustín y Pedraza Corpus, Nancy Yahibé. (2017), “La objetividad en las pruebas estandarizadas”, *Revista Iberoamericana de evaluación educativa*. vol. X, núm. 1, pp. 11-31.
- Ulloa, Cesar (2015), “Análisis de contenido”, en: [https://cpb-us-w2.wpmucdn.com/blogs.udla.edu.ec/dist/b/364/files/2015/08/Cesar-Ulloa-An%C3%A1lisis-de-contenido\\_jun\\_15-1-11lzctz.pdf](https://cpb-us-w2.wpmucdn.com/blogs.udla.edu.ec/dist/b/364/files/2015/08/Cesar-Ulloa-An%C3%A1lisis-de-contenido_jun_15-1-11lzctz.pdf) (consulta: 1 abril de 2020).
- Umaña Mata, Ana Cristina; Calvo Cruz, Xinia y Salas Quirós, Natalia (2017), “Evaluar para aprender: estado actual de catorce asignaturas en la universidad estatal a distancia de Costa Rica”, *Revista Electrónica Calidad en la Educación Superior*, vol. VIII, núm. 2, pp. 24-61.
- Velasco, A. (2003) Ética en las ciencias sociales y humanas. En Ibarra y Olivé (Eds.) *Cuestiones éticas en ciencia y tecnología en el siglo XXI*. Madrid, OEI, p. 245- 276.
- Vidal Ledo, María; Gómez Martínez, Freddy y Ruiz Piedra, Alina (2010), “Software educativos”, *Educación Médica Superior*, vol. XXIV, núm. 1, pp. 97-110.
- Wagner III, W. E. (2019). *Using IBM® SPSS® statistics for research methods and social science statistics*. Sage Publications.
- Zamora, J. (2015). Análisis de la confiabilidad de los resultados de la prueba de diagnóstico matemática en la Universidad Nacional de Costa Rica utilizando el modelo de Rasch. *Actualidades en Psicología*, 29(119), 153-165.
- Zúñiga, María; Solar, María Inés; Lagos, Jorge; Báez, Mario y Herrera, Ricardo (2014), “Evaluación de los aprendizajes: un acercamiento en educación superior”, en: CINDA-Centro Interuniversitario de Desarrollo. *Evaluación del aprendizaje en innovaciones curriculares de la educación superior*, Santiago de Chile, Ediciones e Impresiones Copygraph, pp. 15-38.

## **ANEXO 1 Especificaciones de Ítems**

Dirección General de Bibliotecas de la UAQ

## Especificación de ítems de la prueba de Morfología del Segundo Idioma

<p><b>Información contextual</b></p> <p><b>1.1.1 Definition of the discipline</b></p> <p>Part of the Linguistic that studies the admissible rules for the words formation for their delimitation, definition and classification</p> <p><b>Unidad 1. Basic concepts</b>  <b>Competencia:</b> Identificar las unidades morfológicas mínimas implicadas en la formación de palabras en la lengua inglesa a través de la realización de ejercicios prácticos para crear conciencia de los mecanismos de construcción de significados; mostrando respeto, mostrando interés así como respeto hacia el profesor y sus compañeros al trabajar en el aula en equipos de trabajo o al realizar sus ejercicios de manera individual.</p> <p><b>IRC:</b> 0.600</p> <p><b>Criterios:</b></p> <ul style="list-style-type: none"> <li>• Contribución al logro de la competencia de la unidad: .20</li> <li>• Dosificación: .10</li> <li>• Carga horaria: .10</li> <li>• Relevancia disciplinaria: .20</li> <li>• Nº servicios que recibe: 0</li> <li>• Nº servicios que proporciona: 0</li> </ul> <p><b>Recibe servicios de:</b> Ningún contenido.</p> <p><b>Proporciona servicios a:</b> Ningún contenido.</p>	<p>El contenido <i>Definition of the discipline</i> es el primer tema conceptual que el alumno aborda en la <i>Unidad 1. Basic Concepts</i>, en donde se aborda el concepto, así como las características de la morfología en la formación de palabras.</p> <p>Por lo anterior, este contenido fue uno de los que obtuvo las puntuaciones más altas en los criterios de contribución al logro de la competencia de la asignatura, así como mayor puntuación con relación al índice de relevancia curricular.</p> <p>De esta manera se elaborarán 2 ítems: uno que pruebe el dominio del concepto de <i>Morphology</i>; y otro que pruebe el dominio del concepto <i>Morphology</i> a través de la identificación de sus características.</p> <p><b>Información contextual o indicaciones para responder este ítem:</b> Ninguna.</p> <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 10px;"> <tr> <th rowspan="2" style="width: 30%;">Información tabular, gráfica o textual a emplear en el ítem: Ninguna.</th> <th style="width: 10%;">Dimensión conocimiento</th> <th colspan="5" style="text-align: center;">Dimensión proceso cognitivo</th> </tr> <tr> <th></th> <th style="text-align: center;">Recordar</th> <th style="text-align: center;">Comprender</th> <th style="text-align: center;">Aplicar</th> <th style="text-align: center;">Analizar</th> <th style="text-align: center;">Evaluar</th> <th style="text-align: center;">Crear</th> </tr> <tr> <td>Conocimiento factual</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Conocimiento conceptual</td> <td></td> <td></td> <td style="text-align: center;">X</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Conocimiento procedimental</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Conocimiento metacognitivo</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table> <p><b>Especificación de la base del ítem:</b></p> <ol style="list-style-type: none"> <li>1) Podrá presentar el término <i>Morphology</i> y solicitar que se identifique la definición de éste. Por ejemplo: <i>Which one is the definition of "Morphology"?</i></li> <li>2) Podrá presentar el término <i>Morphology</i> y solicitar que se identifique características propias de esta disciplina. Por ejemplo: <i>Which one are the characteristics of Morphology?</i></li> </ol> <p><b>Especificación de la respuesta correcta a.</b></p> <ol style="list-style-type: none"> <li>1) Será la definición de <i>Morphology</i>. Por ejemplo: <i>Science that studies word formation.</i></li> <li>2) Serán las características en cuestión. Por ejemplo: <i>delimited, classify and define words</i></li> </ol> <p><b>Especificación del distractor b.</b></p> <ol style="list-style-type: none"> <li>1) Se podrá presentar una definición incompleta del término, o que haga referencia a otro concepto. Por ejemplo: <i>Science that studies sentence formation.</i></li> <li>2) Se podrán presentar características correctas e incorrectas. Por ejemplo: <i>combine, describe and form words.</i></li> </ol> <p><b>Especificación del distractor c.</b></p> <ol style="list-style-type: none"> <li>1) Se podrá presentar una definición incompleta del término, o que haga referencia a otro concepto. Por ejemplo: <i>Science that studies word meaning.</i></li> <li>2) Se podrán presentar características correctas e incorrectas. Por ejemplo: <i>construct, relate and define words.</i></li> </ol> <p>En todos los casos, el distractor c. no repetirá una característica ya incluida en alguna de las modalidades de la opción b.</p>	Información tabular, gráfica o textual a emplear en el ítem: Ninguna.	Dimensión conocimiento	Dimensión proceso cognitivo						Recordar	Comprender	Aplicar	Analizar	Evaluar	Crear	Conocimiento factual								Conocimiento conceptual			X					Conocimiento procedimental								Conocimiento metacognitivo							
Información tabular, gráfica o textual a emplear en el ítem: Ninguna.	Dimensión conocimiento		Dimensión proceso cognitivo																																												
		Recordar	Comprender	Aplicar	Analizar	Evaluar	Crear																																								
Conocimiento factual																																															
Conocimiento conceptual			X																																												
Conocimiento procedimental																																															
Conocimiento metacognitivo																																															

<b>Ítem muestra</b>	
	<i>Which one is the definition of "Morphology"?</i>
a.	<i>Science that studies word formation</i>
b.	<i>Science that studies sentence formation</i>
c.	<i>Science that studies word meaning</i>

<b>Ítem muestra</b>	
	<i>Which ones are the characteristics of Morphology?</i>
a.	<i>delimit, classify and define words</i>
b.	<i>combine, describe and form words</i>
c.	<i>construct, relate and define words</i>

## Formato para la especificación de ítems de la prueba de Morfología del Segundo Idioma

<b>Información contextual</b>	<p>El contenido de <i>Inflection</i> es el tercer tema conceptual que el alumno aborda en la Unidad 2. <i>Rules of word formation</i>, en donde se aborda el concepto, así como la estructura y características de la inflexión de palabras. Este contenido da servicio a la práctica de la unidad 4 para el logro de la competencia en lo que se refiere identificar los mecanismos de formación de palabras propios de la lengua inglesa, para comprender mejor los principios de la formación del léxico.</p> <p>Para el aprendizaje del contenido de <i>Inflection</i> es importante que el estudiante haya comprendido previamente los conceptos de <i>Morphemes and allomorphs</i>, vistos en la Unidad 1. <i>Basic concepts</i>. A su vez, el aprendizaje de este contenido es insumo para la parte práctica de la Unidad 2, ya que deberá identificar y explicar lo que es la inflexión. Además, el dominio de <i>Inflection</i> es necesario para el aprendizaje del contenido <i>Tense</i> (de la Unidad 2), ya que debe saber identificar los cambios que sufre un verbo. También, este contenido proporciona servicio a otros dos contenidos de la unidad siguiente: <i>Nouns</i> y <i>Adjectives</i>. Por lo anterior, este contenido fue uno de los que obtuvo las puntuaciones más altas en los criterios de contribución al logro de la competencia de la asignatura, así como mayor puntuación con relación al índice de relevancia curricular.</p> <p>De esta manera se elaborarán 3 ítems: uno que pruebe el dominio del concepto de <i>Inflection</i>, otro donde el alumno pueda identificar el tipo de palabras que pueden sufrir inflexión, y un tercero que pruebe el reconocimiento de palabras que sufren inflexión en un texto real.</p>							
2.3 Inflection								
<p>In grammar, inflection is the modification of a word to express different grammatical categories</p>								
<p><b>Unidad 2. Rules of word formation</b>  <b>Competencia:</b> Identificar los mecanismos de formación de palabras propios de la lengua inglesa en ejemplos reales de uso, para comprender mejor los principios de la formación del léxico mostrando iniciativa en el desarrollo de las clases.</p> <p>IRC: 0.792</p> <p><b>Criterios:</b></p> <ul style="list-style-type: none"> <li>• Contribución al logro de la competencia de la unidad: .20</li> <li>• Dosificación: .10</li> <li>• Carga horaria: .10</li> <li>• Relevancia disciplinaria: .20</li> <li>• N° servicios que recibe: 1</li> <li>• N° servicios que proporciona: 5</li> </ul>								
<p><b>Recibe servicios de:</b> 1.3.1 Morphemes and allomorphs</p> <p><b>Proporciona servicios a:</b> 2.6.1 Tense 3.1 Nouns 3.2 Adjectives 3.4 Verbs</p> <p style="text-align: center;">Práctica 2</p> <p>P2 Elaborar un resumen en el que plasmen las características de los diferentes mecanismos de formación de palabras, proveyendo ejemplos reales de uso que faciliten el uso posterior como texto de referencia, de acuerdo a la normatividad de la APA.</p>								
<b>Información contextual o indicaciones para responder este ítem:</b>								
Ninguna.								
<b>Información tabular, gráfica o textual a emplear en el ítem:</b>								
Ninguna.								
		Dimensión conocimiento	Dimensión proceso cognitivo					
			Recordar	Comprender	Aplicar	Analizar	Evaluar	Crear
		Conocimiento factual						
		Conocimiento conceptual		X				
		Conocimiento procedimental						
		Conocimiento metacognitivo						
<b>Especificación de la base del ítem:</b>								
<p>1) Podrá presentar el término y solicitar que se identifique la definición de este. Por ejemplo: <i>Which one is the definition of Inflection.</i></p> <p>2) Podrá presentar categorías gramaticales y solicitar se identifique aquellas que pueden sufrir <i>inflexión</i>. Por ejemplo: <i>Which of the next categories can be inflected.</i></p> <p>3) Podrá presentar una oración y solicitar que se identifiquen las palabras que tienen <i>inflexión</i>. Por ejemplo: <i>Which words are inflected in the next phrase: There were four rows of seats on the economics class.</i></p>								
<b>Especificación de la respuesta correcta a.</b>								
<p>1) Será la definición de <i>Inflection</i>. Por ejemplo: <i>Adding a suffix to a root.</i></p> <p>2) Serán las categorías gramaticales correctas. Por ejemplo: <i>Verbs and Adjectives.</i></p> <p>3) Serán las palabras que sufren <i>inflexion</i>. Por ejemplo: <i>Rows and seats.</i></p>								
<b>Especificación del distractor b.</b>								
<p>1) Se podrá presentar una definición incompleta del término o que haga referencia a otro concepto. Por ejemplo: <i>Adding a prefix to a Root.</i></p> <p>2) Se podrán presentar categorías gramaticales correctas junto con otras que no lo son. Por ejemplo: <i>Verbs and Adverbs.</i></p> <p>3) Se podrá presentar palabras correctas junto con otras que no lo son. Por ejemplo: <i>Seats and economics.</i></p>								
<b>Especificación del distractor c.</b>								
<p>1) Se podrá presentar una definición incompleta del término o que haga referencia a otro concepto. Por ejemplo: <i>Adding an infix to a root.</i></p> <p>2) Se podrán presentar categorías gramaticales correctas junto con otras que no lo son. Por ejemplo: <i>Prepositions and Conjunctions.</i></p> <p>4) Se podrá presentar palabras correctas junto con otras que no lo son. Por ejemplo: <i>Rows and class.</i></p>								
En todos los casos, el distractor c no repetirá una característica ya incluida en alguna de las modalidades de la opción b.								

## **ANEXO 2 Evaluación de la usabilidad**

Dirección General de Bibliotecas de la UAQ

## Evaluación de la calidad y usabilidad del sistema SIEXAES (alumno)

El propósito del siguiente cuestionario es ayudar a detectar problemas en el software especializado para la realización de exámenes estandarizados SIEXAES, y con ello mejorar la experiencia y calidad evaluativa que la Facultad de Idiomas te ofrece como estudiante.

Instrucciones: marque la casilla que mejor represente su experiencia

1. ¿Le parece adecuada la selección de contenidos destacados en la portada antes de iniciar el examen?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

2. ¿Es fácil identificar los elementos en los que se debe ingresar algún dato en la portada antes de iniciar el examen?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

3. ¿Puede ver en la portada y en las demás ventanas, la forma en que se navega?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

4. ¿Existen elementos dentro de las ventanas, que le permitan saber exactamente dónde se encuentra y como volver a otra ventana o menú?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

5. ¿Le pareció adecuada la forma en que se muestran las preguntas?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

6. ¿Le pareció adecuada la forma en que se eligen las respuestas?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

7. ¿Le pareció adecuada la transición entre las diferentes secciones de preguntas que componen el examen?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

8. ¿Le pareció que la organización de los contenidos o funciones facilita recordarlos?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

9. ¿Los colores utilizados en las diferentes ventanas y objetos fueron adecuados?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

10. ¿El tamaño de los objetos en las ventanas fueron adecuados?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

11. ¿Le pareció adecuada la señalización para identificar preguntas contestadas?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

12. ¿Considera que gráficamente el software está equilibrado?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

13. ¿Tras un primer uso, le queda claro cuál es el objetivo, contenidos y servicios que ofrece?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

14. ¿Puede listar los servicios que el software le ofrece?

--

15. ¿Qué es lo que más te gusta del software?



16. ¿Qué sugieres como mejora?

Dirección General de Bibliotecas de la UAQ

## Evaluación de la calidad y usabilidad del sistema SIEXAES (docente – directivo)

El propósito del siguiente cuestionario es ayudar a detectar problemas en el software especializado para la realización de exámenes estandarizados SIEXAES, y con ello mejorar el proceso de evaluación del conocimiento en los estudiantes. De la misma manera se busca ofrecer información confiable y válida para una toma de decisiones más oportuna y conocer más detalladamente la situación que los estudiantes a nivel individual y colectivo presentan en la Facultad de Idiomas.

Instrucciones: marque la casilla que mejor represente su experiencia

1. ¿Le parece adecuada la selección de contenidos destacados en la portada o ventana de inicio?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

2. ¿Es fácil identificar los elementos en los que se debe ingresar algún dato en la portada o pantalla de inicio?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

3. ¿Puede ver en la portada y en las demás ventanas, la forma en que se navega?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

4. ¿Existen elementos dentro de las ventanas, que le permitan saber exactamente dónde se encuentra y como volver a otra ventana o menú?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

5. ¿Le pareció adecuada la forma en que se muestran los diferentes menús del sistema?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

6. ¿Le pareció intuitiva la interacción o navegación entre las diferentes ventanas del software?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

7. ¿Le pareció que la organización de los contenidos o funciones facilita recordarlos?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

8. ¿Le pareció adecuada la fluidez entre los diferentes menús o secciones del software?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

9. ¿El tamaño de los objetos en las ventanas fueron adecuados?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

10. ¿Los colores utilizados en las diferentes ventanas y objetos fueron adecuados?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

11. ¿Considera que gráficamente el software está equilibrado?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

12. ¿Tras un primer uso, le queda claro cuál es el objetivo, contenidos y servicios que ofrece?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

13. ¿Puede listar los servicios que el software le ofrece?

--

Dirección General de Bibliotecas de la UAQ

14. ¿Cree que los contenidos y servicios que se ofrecen en el software son de utilidad para desempeñar mejor sus funciones?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

15. ¿Considera que los reportes técnicos contienen información relevante o importante para su función?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

16. ¿Considera que la información mostrada en los reportes técnicos es clara y entendible?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

17. ¿Considera que el software es seguro en el manejo de la información?

Totalmente en desacuerdo	En desacuerdo	Regular	De acuerdo	Totalmente de acuerdo

18. ¿Qué es lo que más te gusta del software?

--

19. ¿Qué sugieres como mejora?

--

## Indicadores de usabilidad

La siguiente tabla organiza los diferentes indicadores de usabilidad que son asociados a las distintas preguntas incluidas en el test.

Para el cuestionario del alumno

Indicador	Preguntas
Contenido	1,2
Navegación	3,4,7
Facilidad de aprendizaje	8
Representación gráfica	5,6,9,10,11,12
Utilidad	13,14,15,16

Para el cuestionario del docente y directivo

Indicador	Preguntas
Contenido	1,2,14,15,16
Navegación	3,4,5,8
Facilidad de aprendizaje	6,7
Representación gráfica	9,10,11
Utilidad	12,13,17,18,19

## **ANEXO 3 Análisis de indicadores**

Dirección General de Bibliotecas de la UAQ

Tabla con los indicadores que se van a trabajar

Definición de indicadores				
Objetivo de intervención	Indicador	Dimensión cualitativa	Dimensión cuantitativa	Temporalidad
Aumentar el índice de confiabilidad y validez de las pruebas que se aplican.	Índice de dificultad del ítem Índice de dificultad del examen Índice de relevancia curricular	La dificultad de un ítem se entiende como la proporción de personas que responden correctamente un reactivo de una prueba. Entre mayor sea esta proporción, menor será su dificultad. El índice de relevancia curricular es aquel valor que expresa la importancia de un contenido en comparación con otros contenidos, enmarcados dentro el currículo de una asignatura. Estos contenidos con alto IR son clave para concretar otros conocimientos dentro de una misma asignatura o incluso para cursos posteriores.	Se desea obtener un índice de dificultad en el examen aproximado entre .5 y .65. Los ítems deben mantener un índice de dificultad en mayor proporción con valores superiores a .4, sin embargo deben estar presentes ítems dentro del rango de .2 a .4 para discriminar aún más los grupos altos. El índice de relevancia curricular se tomara como clave a aquellos que estén superiores o iguales a .7.	4 meses
	Índice de discriminación Coeficiente de discriminación (correlación punto biserial)	La discriminación se explica de la siguiente manera: si la prueba y un ítem miden la misma habilidad o competencia, se espera que quien tuvo una puntuación alta en toda la prueba deberá tener altas probabilidades de contestar correctamente el ítem. Y se espera lo mismo para el caso contrario, es decir aquellos que tuvieron puntuaciones bajas. Así, un buen ítem debe discriminar entre aquellos que obtuvieron buenas	Se desea obtener un índice de discriminación como mínimo de .2. Así mismo se desea obtener un valor de rpbis de .2.	



		<p>calificaciones en la prueba y aquellos que obtuvieron bajas calificaciones.</p> <p>La correlación se utiliza para saber si las personas “adecuadas” son las que obtienen las respuestas correctas, qué tanto poder predictivo tiene el reactivo, y cómo puede contribuir a las predicciones.</p>		
<p>Establecer un proceso de seguimiento oportuno de resultados en las pruebas para establecer métodos de retroalimentación efectivos para los estudiantes.</p>	<p>Reporte con resultados de la ejecución de cada estudiante.</p> <p>Reporte con las observaciones y/o recomendaciones (retroalimentación) hechas al estudiante</p>	<p>Es necesario contar con un proceso que permita observar el seguimiento que se les da a los estudiantes de acuerdo al desempeño mostrado en las pruebas. Para ello es necesario un reporte que describa los resultados obtenidos por los estudiantes en la prueba y así mismo uno que deje evidencia de las recomendaciones o sugerencias sobre el aprendizaje que el estudiante demostró en la prueba.</p>	<p>Se desea por lo menos tener un reporte general de resultados obtenidos por los estudiantes, clasificados por asignatura y por maestro cuando aplique.</p> <p>Se desea que como mínimo se de una retroalimentación al 100% de los estudiantes, para ello debe evidenciarse en un reporte.</p>	<p>Como máximo 5 días después de la aplicación de la prueba.</p>
<p>Eliminar las inconsistencias en la evaluación de docentes que imparten una misma asignatura.</p>	<p>Contenidos temáticos representados en la prueba</p>	<p>Una misma asignatura impartida por más de un docente debe ofrecer la misma probabilidad de acreditar un examen para los estudiantes independientemente de que se curse con docentes diferentes, por ello las pruebas deben evaluar el mismo universo de conocimientos y medir el mismo dominio.</p>	<p>Se desea identificar en un instrumento grafico los contenidos temáticos que se deben tratar en la asignatura y clasificarlos por nivel de relevancia curricular para que estos tengan representatividad en la prueba. En la prueba deben estar representados el 100% de contenidos con un IR mayor o igual a .7.</p>	<p>Inicio del siguiente semestre después de la intervención</p>
<p>Reducir los factores externos que condicionan el</p>	<p>Calidad de los espacios de aplicación</p>	<p>En el desempeño de una prueba ademar de influir aspectos</p>	<p>Se desea utilizar los mismos horarios y el mismo lugar para</p>	<p>Antes de cada evaluación y después de la misma</p>

desempeño de las pruebas (áreas, condiciones físicas, horarios, días, eventos, entre otros).		internos hay factores externos que pueden condicionar la ejecución de la prueba, tales como la hora del día en que se aplica, la ventilación del lugar, distracciones externas, duración de la prueba, exámenes previos o posteriores, clima entre muchos otros, por ello es importante	la aplicación de estos exámenes. Considerando que el día y la hora definidos no coincidan previamente o posteriormente con exámenes, actividades culturales o deportivas entre otros.	para analizar posibles mejoras.
Estandarizar el diseño de pruebas para tener un control más efectivo de la calidad de los instrumentos de evaluación que se utilizan.	Tipo de ítems	Existen números estudios que comprueban la efectividad de cierta representación de ítems (opción múltiple, escalas de valor, relacionales entre otros) con los que se puede medir el dominio de conocimiento en los estudiantes, por lo que para poder evaluar de manera más eficiente las respuestas de los estudiantes se debe estandarizar el tipo de ítems que se utilizan en las pruebas.	Se desea tener una representación de un 90% de ítems de selección en la prueba, dando un margen de solo 10% para preguntas abiertas.	Antes de cada evaluación.
Establecer mecanismos metodológicos que permitan evaluar de manera confiable y válida el dominio de conocimiento por parte de los estudiantes.	Metodología para el diseño de pruebas estandarizadas con análisis psicométrico	Para estandarizar la evaluación del aprendizaje es necesario contar con una metodología que guíe el proceso de diseño de las pruebas que se quieran realizar en la institución, y que junto a ello establezcan mecanismos psicométricos para evaluar la calidad de las pruebas. Para ello se utilizara una metodología para desarrollo de exámenes referidos a un criterio (el criterio es el currículo de la asignatura) y el análisis psicométrico se hará en base a la TRI.	Se desea que exista un producto por cada una de las fases que se establecen en la metodología de desarrollo de la prueba para poder hacer valoraciones en periodos más cortos y no una sola al final del todo el proceso.  Se desea cumplir con los criterios de calidad psicométrica establecidos por la TRI (índice de dificultad, índice de discriminación, coeficiente de discriminación, correlación ítem-total)	Antes de la intervención

<p>Diseñar reportes técnicos que permitan analizar los atributos de calidad de las pruebas (índice de discriminación, índice de dificultad, índice de correlación de distractores) y con ellos aportar datos válidos para tomas de decisiones oportunas y eficaces.</p>	<p>Número de reportes</p>	<p>Es necesario que se cuente con un reporte técnico diseñado específicamente para describir los resultados psicométricos de la ejecución de la prueba, para poder observar aquellos indicadores de cada ítem y con ello poder hacer las correcciones necesarias a los distractores, a las respuestas correctas entre otros.</p> <p>De igual forma se necesita un reporte técnico que incluya información clara y específica sobre aquellos datos que son relevantes para el alumno (unidad temática con menos dominio, temas específicos con mejor o menor dominio etc), más allá de solo presentar una calificación.</p> <p>Lo mismo sucede con el reporte técnico para el docente, se necesita más que solo promedios grupales e individuales, se requiere conocimiento sobre unidades temáticas que menor grado de dominio y con ello establecer estrategias para reforzar dichos conocimientos, ya que en algunos casos ciertas unidades poseen temas de mayor relevancia curricular y por ende deben ser atendidos más estrictamente.</p>	<p>Reporte técnico para el alumno con los resultados de la ejecución (promedio, puntaje por temas, por unidad, clasificación del resultado por áreas de mayor y menor dominio entre otros datos).</p> <p>Reporte técnico para el docente con los resultados de la ejecución (promedio grupal, promedio individual, promedios de puntaje por temas, por unidad, clasificación del resultado por áreas de mayor y menor dominio entre otros datos).</p> <p>Reporte técnico para áreas directivas o de coordinación con los resultados de la ejecución (comparativos entre grupos de criterios como promedios, alumnos aprobados, reprobados, distribución de calificaciones entre otros).</p> <p>Reporte técnico con la evaluación psicométrica de la prueba.</p>	<p>Se elaboran antes de la ejecución de la prueba, y son entregados después de la evaluación en un periodo no mayor a días.</p>
---	---------------------------	---	---	---

Establecer un programa de capacitación en evaluación estandarizada del aprendizaje a los docentes que conforman la institución educativa.	Curso de capacitación y número de docentes participantes	Programa institucional de capacitación docente en el diseño de pruebas estandarizadas del aprendizaje.  Para lograr crear una cultura evaluativa en la institución es necesario que participen el mayor número de docentes posible en dichos cursos.	Mínimo un curso de capacitación por semestre.  Mínimo de un 80% de docentes participantes en dicho curso.	Antes de la ejecución de las pruebas. Cada semestre posterior a la intervención.
---	--	--	---	---

### Análisis relacional de las variables

a) Identificación de las variables. Con base en las causas ubicadas en el árbol de problemas, identifique y defina las variables. Use la tabla de apoyo.

Variable	Definición
<b>Profesores no capacitados</b>	Se observa que la mayor parte de la población docente no cuenta con alguna capacitación en el diseño de pruebas estandarizadas, análisis psicométrico o que conozcan del tema.
<b>Tiempo para desarrollar pruebas</b>	La mayoría de los docentes consideran que la construcción de un examen estandarizado demanda mucho tiempo, el cual no disponen.
<b>Resistencia a estandarizar criterios</b>	Se considera que cada docente es libre de evaluar el aprendizaje de sus estudiantes de acuerdo a los criterios que el considere sin necesidad de apegarse a los de otro o a tener una única forma de evaluar.
<b>Auto apreciación de expertos</b>	La mayoría de los docentes se consideran competentes como para diseñar sus propios instrumentos de evaluación a los cuales les atribuyen las características de ser de calidad.
<b>Libertad de cátedra</b>	Se considera una restricción a su práctica docente el querer implementar un estándar para evaluar el aprendizaje.
<b>Ausencia de tecnología</b>	La mayoría de las pruebas que se realizan en la institución no cuentan con alguna tecnología que facilite la aplicación y análisis de las mismas.
<b>Ausencia de análisis psicométricos</b>	Las pruebas no son evaluadas en cuanto a su calidad psicométrica, por lo que condiciona la validez y confiabilidad de los datos que arrojan.
<b>Cultura evaluativa</b>	La mayoría de los docentes no presentan una buena disposición a variar sus instrumentos, estrategias o criterios de evaluación.
<b>Objetivos de las pruebas</b>	En general los docentes se concentran solo en la escala lineal de las pruebas con lo cual emiten el juicio sobre el nivel de dominio que un estudiante posee de determinado conocimiento.
<b>Métodos estandarizados ausentes</b>	Se observa que cada docente construye las pruebas de acuerdo a los criterios y juicios que el mismo se forma, sin seguir alguna metodología estandarizada.

<b>Resistencia a ser observado</b>	La mayoría de los docentes no gustan de emplear pruebas estandarizadas por que se tiene el temor a ser comparado en cuanto a desempeño con los demás docentes, o que se observen los resultados que obtienen sus estudiantes.
<b>Estímulo a evaluar la calidad de las pruebas</b>	La mayoría de los docentes no se preocupa por saber si los instrumentos de evaluación que aplica son de calidad, es decir válidos y confiables. Debido al poco tiempo disponible para esta actividad se confía más en la experiencia que el docente tiene para diseñar la prueba.
<b>Resistencia a compartir estrategias</b>	Un número considerable de docentes son celosos de las estrategias y/o el conocimiento que poseen sobre buenas prácticas en la enseñanza-aprendizaje, motivo por el cual es difícil hacer trabajos colaborativos.
<b>Impacto de la prueba</b>	Para estudiantes y docentes las consecuencias que tienen los resultados de una prueba condiciona el entusiasmo por participar en la misma. En ocasiones se asocia evaluar con conceptos negativos.
<b>Operatividad</b>	Es necesario saber si se cuentan con los recursos de infraestructura tecnológica necesarios para que la operación continua de aplicar las pruebas utilizando tecnología es sustentable.
<b>Viabilidad</b>	Es necesario saber la realidad de aplicar determinada metodología para el diseño de un examen en relación a las características del lugar donde se desea aplicar, lo mismo sucede con la metodología para el análisis psicométrico.

- b) Matriz de análisis estructural. Analice la relación de dependencia e influencia entre las variables causales identificadas (procure identificar entre 15 y 20 variables). Elabore la matriz de análisis estructural y calcule los índices de influencia y dependencia.

Variable	Causa 1	Causa 2	Causa 3	Causa 4	Causa 5	Causa 6	Causa 7	Causa 8	Causa 9	Causa 10	Causa 11	Causa 12	Causa 13	Causa 14	Causa 15	Causa 16	Suma de influencias
<b>Causa 1</b>	/	2	2	3	3	3	3	2	2	3	2	2	0	1	3	3	34
<b>Causa 2</b>	2	/	0	2	2	0	0	2	3	3	0	2	1	1	1	2	21
<b>Causa 3</b>	2	2	/	3	3	1	3	1	2	3	3	2	1	2	0	1	29
<b>Causa 4</b>	3	2	3	/	3	2	2	2	3	2	3	2	1	0	3	3	34
<b>Causa 5</b>	3	2	2	2	/	3	2	1	3	3	2	2	3	3	3	3	37
<b>Causa 6</b>	3	3	0	0	0	/	3	1	0	1	0	2	0	0	3	3	19
<b>Causa 7</b>	1	2	2	2	0	0	/	0	2	2	0	3	0	3	3	3	23
<b>Causa 8</b>	3	2	3	2	1	1	2	/	3	3	2	3	2	2	2	2	33
<b>Causa 9</b>	2	3	2	0	1	2	1	2	/	2	3	3	0	3	2	3	29
<b>Causa 10</b>	3	3	2	2	2	0	2	1	1	/	0	2	0	2	2	3	25
<b>Causa 11</b>	1	0	3	2	3	0	2	2	0	3	/	3	1	0	2	3	25
<b>Causa 12</b>	2	2	2	2	0	0	3	2	0	2	0	/	0	2	0	0	17
<b>Causa 13</b>	0	0	3	3	3	0	0	2	0	0	3	0	/	0	3	3	20

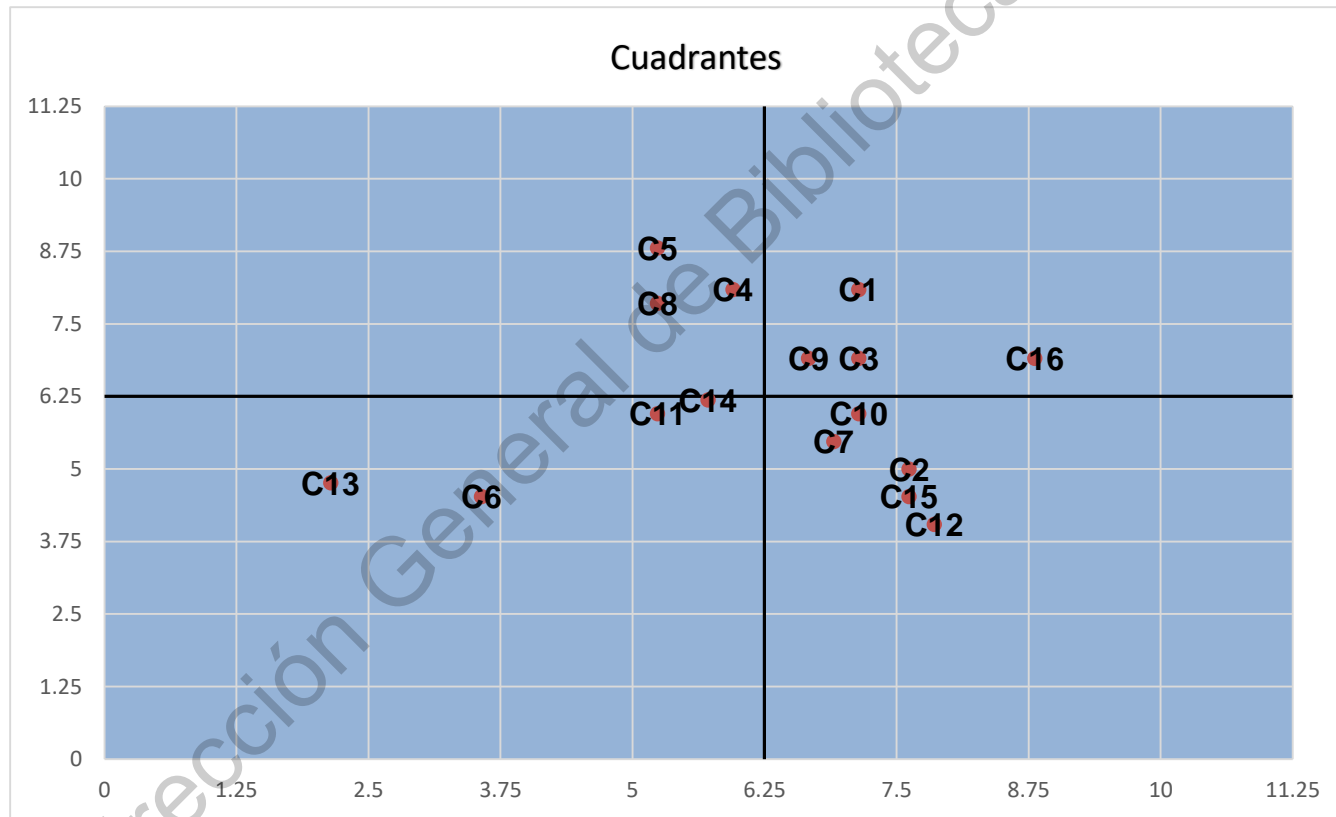
<b>Causa 14</b>	2	3	2	1	1	0	3	0	3	0	3	3	0	/	3	2	26
<b>Causa 15</b>	0	3	2	1	0	0	0	2	3	0	0	2	0	3	/	3	19
<b>Causa 16</b>	3	3	2	0	0	3	3	2	3	3	1	2	0	2	2	/	29
<b>Suma de dependencias</b>	30	32	30	25	22	15	29	22	28	30	22	33	9	24	32	37	

<b>Variable</b>	<b>ID</b>	<b>II</b>
<b>Causa 1</b>	7.14285714	8.0952381
<b>Causa 2</b>	7.61904762	5
<b>Causa 3</b>	7.14285714	6.9047619
<b>Causa 4</b>	5.95238095	8.0952381
<b>Causa 5</b>	5.23809524	8.80952381
<b>Causa 6</b>	3.57142857	4.52380952
<b>Causa 7</b>	6.9047619	5.47619048
<b>Causa 8</b>	5.23809524	7.85714286
<b>Causa 9</b>	6.66666667	6.9047619
<b>Causa 10</b>	7.14285714	5.95238095
<b>Causa 11</b>	5.23809524	5.95238095
<b>Causa 12</b>	7.85714286	4.04761905
<b>Causa 13</b>	2.14285714	4.76190476
<b>Causa 14</b>	5.71428571	6.19047619
<b>Causa 15</b>	7.61904762	4.52380952
<b>Causa 16</b>	8.80952381	6.9047619

- c) Con base en los índices de influencia y dependencia, sitúe las variables en las distintas zonas, posteriormente señale las que tienen mayor incidencia sobre las otras variables (y sobre los efectos) y que considera pueden atenderse en el menor tiempo posibles y con los recursos que tiene disponible. Sostenga su argumentación con la gráfica de los cuadrantes.

Después de realizar el análisis se identificaron las variables que tienen mayor influencia en otras variables, por ejemplo, una de ellas es la de objetivos de las pruebas. Esta se atiende mediante un taller informativo para el docente, señalando de manera amplia y clara cuales son los propósitos que se tienen para las pruebas estandarizadas que se aplicarán. Así mismo puede atenderse la variable de resistencia a la estandarización de criterios, ya que puede explicarse al docente cuales son las ventajas de aplicar una prueba estandarizada, y que teniendo como precedente también el objetivo de la prueba el docente no corre ningún riesgo de ser señalado públicamente o condicionado por el desempeño en la prueba, sino que los resultados de la prueba pueden aportar información para sugerir cursos de capacitación en áreas específicas para que mejore su práctica docente.

A continuación, se muestra la gráfica de cuadrantes o zonas en las que se clasifican las diferentes variables señaladas con anterioridad, y como se acaba de mencionar, en el cuadrante de zona de conflictos (esta es la zona en la que se ubican las causas sobre las que se tiene que enfocar la intervención puesto que son influyentes ya que trabajar en ellas genera un alto impacto en las demás).



## **Anexo 4 Acciones estratégicas interventivas**

Dirección General de Bibliotecas de la UAQ



**Acciones propuestas**

Tabla 1. Acciones estratégicas propuestas

Nombre de la acción estratégica	Descripción general	Actores involucrados	Participación esperada de los actores	Plazos de realización	Objetivo que atiende	Meta con la cual se relaciona	Productos y resultados esperados
<p><b>1. Valorar la calidad de la prueba</b></p>	<p>Efectuar los análisis correspondientes de la Teoría de Respuesta al Ítem (TRI) para cada uno de los ítems que compondrán la prueba.</p>	<p>Responsable de Evaluación Colegiada del Aprendizaje (ECA)  Comité de evaluación de ítems</p>	<p>El responsable de ECA deberá generar el reporte técnico de las valoraciones psicométricas de cada uno de los ítems que componen la prueba.  El comité de evaluación de ítems emitirá juicio sobre la calidad de cada</p>	<p>Cinco días posteriores a la terminación de la etapa cuatro del proceso de construcción de la prueba (producción y validación de ítems).</p>	<p>1.- Aumentar el índice de confiabilidad y validez de las pruebas que se aplican.</p>	<p>1.- Los índices de confiabilidad (Índice de dificultad de los ítems, Índice de dificultad de la prueba, Índice de relevancia curricular) y validez (índice de discriminación, correlación ítem-total) en las pruebas cumplen con los valores de calidad definidos por la Teoría de</p>	<p>Reporte técnico con la medición de los atributos de calidad de cada ítem.  Reporte con los juicios y valoraciones realizados por el comité de evaluación de ítems.</p>

			uno de los ítems en términos de confiabilidad y validez (índice y coeficiente de discriminación, distractores, índice de dificultad)			Respuesta al Ítem (TRI).	Se espera que todos los ítems cumplan con los criterios de calidad mínimos establecidos en la TRI.
<b>1. Estandarizar métodos de construcción y/o diseño de pruebas</b>	Implementar una metodología que permita estandarizar el proceso de construir una prueba para evaluar el aprendizaje de una asignatura.	Responsable de Evaluación Colegiada del Aprendizaje.	Impartir un taller informativo sobre la metodología a aplicar. Coordinar la elaboración de la prueba, capacitar a los participantes según las	El taller informativo se da el día en que inicia la ejecución de la intervención. La capacitación se realiza en un lapso de dos días al iniciar	3.- Estandarizar el diseño de pruebas para tener un control más efectivo de la calidad de los instrumentos de evaluación que se utilizan. 4.- Eliminar las inconsistencias	3.- Se cuenta con pruebas estandarizadas para las asignaturas de mayor índice de relevancia curricular. 5.- Las inconsistencias de evaluación	Metodología para la elaboración de pruebas estandarizadas referidos a un criterio.  Se espera que la metodología sea la forma institucional para

			<p>necesidades que la metodología requiera para cada una de sus etapas.</p>	<p>cada etapa del ciclo de desarrollo, establecido por la metodología.</p>	<p>en la evaluación de docentes que imparten una misma asignatura.</p> <p>6.- Establecer mecanismos metodológicos que permitan evaluar de manera confiable y valida el dominio de conocimiento por parte de los estudiantes.</p>	<p>entre docentes que imparten una misma materia son erradicadas, al contar con instrumentos que pueden medir el mismo dominio de conocimiento, aunque en su estructura empleen diferentes ítems.</p> <p>6.- Se aplica una metodología para el diseño de pruebas estandarizadas que contemplen los criterios de calidad definidos por la TRI.</p>	<p>el desarrollo de cualquier prueba.</p>
--	--	--	---	--	--	---	---

Dirección General de Bibliotecas de la UAG

<p><b>2. Diseñar reportes técnicos de la ejecución de la prueba</b></p>	<p>Se pretende diseñar un reporte técnico para uno de los involucrados en la ejecución de la prueba, con información pertinente y de interés particular.</p>	<p>Responsable de Evaluación Colegiada del Aprendizaje  Comité diseñador de la prueba</p>	<p>El Responsable de ECA obtiene los datos a presentar en el reporte, asesora y coordina el diseño de los reportes.  El comité diseña el reporte técnico con base en la información proporcionada por el Responsable de ECA.</p>	<p>Un lapso máximo de cinco días después de ejecutar la etapa seis de la metodología de desarrollo de la prueba (Análisis secundario de resultados)</p>	<p>7.- Diseñar reportes técnicos que permitan analizar los atributos de calidad de las pruebas (índice de discriminación, índice de dificultad, índice de correlación de distractores) y con ellos aportar datos válidos para tomas de decisiones oportunas y eficaces.</p>	<p>8.- Se cuenta con tres reportes técnicos, diseñados con información clara, precisa y relevante de cada uno de los aspectos medidos en la prueba, para el alumno, el docente, personal administrativo y responsables de evaluación (comité colegiado del aprendizaje).</p>	<p>Reportes técnicos para el alumno, docente y personal directivo (coordinación académica).  Se espera que estos reportes presenten información detallada con la cual cada involucrado tenga información más significativa con la cual ayudar a la</p>
---	--	---	--	---	---	--	--

							toma de decisiones.
<b>3. Crear un grupo colegiado de evaluación del aprendizaje (GCEA)</b>	Consiste en formar un grupo de docentes que posean las habilidades y conocimientos necesarios para desarrollar pruebas estandarizadas de calidad con base en la TRI.	Director de la unidad académica.  Responsable de Evaluación Colegiada del Aprendizaje.	El director tiene que estimular la participación de los docentes para que se integren en el GCEA, explicar los propósitos y objetivos que se buscan con la creación de dicho grupo.  El Responsable de ECA explica cuáles son las diferentes funciones y/o	El GCEA debe ser establecido en un periodo máximo de treinta días después de concluir el desarrollo de la prueba.	8.- Establecer un programa de capacitación en evaluación estandarizada del aprendizaje a los docentes que conforman la institución educativa.  Contar con un GCEA permitirá la impartición de cursos de capacitación para el aprendizaje de la metodología	7.- En la institución se cuenta con un grupo colegiado del aprendizaje encargado de gestionar los aspectos evaluativos en la institución, expertos en los análisis psicométricos, diseño de pruebas, capacitación entre otros.	Acta de constitución del Grupo Colegiado de Evaluación del Aprendizaje.  Se espera que este grupo sea el encargado de diseñar y evaluar la calidad de las pruebas que se apliquen de forma estandarizada en la institución.

perfiles que se requieren para integrar el GCEA.

de diseño de pruebas, así como capacitación en procesos específicos como diseño y evaluación de ítems, análisis psicométricos entre otros.

Que sean los encargados de fomentar la cultura evaluativa, así como talleres y capacitaciones relacionadas con la evaluación del aprendizaje.

Dirección General de Bibliotecas de la UAQ

<p><b>4. Diseñar el proceso de seguimiento a los resultados de las pruebas.</b></p>	<p>Se pretende establecer de manera sistematizada cual será el procedimiento para dar seguimiento a los resultados de las pruebas, tanto en valoraciones para los alumnos como para el desempeño docente.</p>	<p>Áreas de coordinación (Académica y Formación Básica) y dirección.  Responsable de Evaluación Colegiada del Aprendizaje.</p>	<p>Las áreas de coordinación tienen que establecer cuáles van a ser los mecanismos para dar seguimiento y retroalimentación a los estudiantes después de realizar la prueba, así como en las observaciones realizadas al desempeño de los docentes. Definir las fechas en las que deben entregarse los reportes técnicos.</p>	<p>Este proceso debe ser definido en un periodo máximo de quince días después de haber concluido el desarrollo de la prueba.</p>	<p>2.- Establecer un proceso de seguimiento oportuno de resultados en las pruebas para establecer métodos de retroalimentación efectivos para los estudiantes.</p>	<p>2.- Existe un programa institucional de seguimiento para los resultados obtenidos en las pruebas, cumpliendo con un mínimo de una retroalimentación para cada estudiante en relación al desempeño mostrado en las pruebas.</p>	<p>Programa de seguimiento y retroalimentación de resultados.  Se espera que con este programa se tenga información actualizada de la situación de cada alumno, así como un seguimiento más oportuno de su desempeño y con ello poder tomar decisiones más acertadas en pro de mejorar el aprendizaje de los estudiantes.</p>
---	---	--	---	--	--	---	---

			<p>El Responsable de ECA entrega la información resultante de cada ejecución de la prueba, y/o realiza adecuaciones a la información presentada de acuerdo a las necesidades informativas de las áreas de coordinación y dirección.</p>				<p>Para el caso de los docentes también se espera poder contribuir a la mejora de su función docente al tener información actualizada de los resultados obtenidos en las diferentes asignaturas que imparte.</p>
<p><b>5. Formar docentes con habilidades y competencias en evaluación del aprendizaje.</b></p>	<p>Se pretende contar un programa institucional regular de</p>	<p>Responsable de Evaluación Colegiada del Aprendizaje.</p>	<p>El responsable de ECA coordina y evalúa las necesidades docentes en</p>	<p>Este programa institucional debe ser establecido en un lapso no</p>	<p>8.- Establecer un programa de capacitación en evaluación estandarizada del</p>	<p>4.- Se implementa un programa institucional de capacitación para</p>	<p>Programa institucional de capacitación docente.</p>



<p>capacitación en estrategias de evaluación del aprendizaje, diseño y evaluación de pruebas estandarizadas de calidad.</p>	<p>Grupo colegiado de evaluación del aprendizaje</p>	<p>relación a evaluación del aprendizaje, participa en diseño de talleres, cursos, plenarias, capacitaciones de docentes.</p> <p>El GCEA se encarga de impartir los talleres, capacitaciones, cursos, plenarias a los docentes participantes de la unidad académica.</p>	<p>mayor a sesenta días después de haber concluido el desarrollo de la prueba.</p>	<p>aprendizaje a los docentes que conforman la institución educativa.</p>	<p>el diseño de pruebas estandarizadas y evaluación de calidad de las mismas.</p>	<p>Se espera que, con este programa, semestre tras semestre se realicen un mínimo de capacitaciones a toda la planta docente de la unidad académica para mejorar su formación en el área de evaluación del aprendizaje.</p> <p>Con esto se pretende cambiar la cultura evaluativa que</p>
---	--	--	--	---	---	---

							actualmente predomina en la institución.
<b>6. Integrar tecnología en la aplicación de la prueba.</b>	Se desea que la aplicación de la prueba estandarizada sea realizada de manera sistematizada en un software con el cual se permita un	Responsable de Evaluación Colegiada del Aprendizaje  Responsable de desarrollo	El Responsable de ECA proporciona al RDS la información que desea presentar digitalmente (items, respuestas etc).	El software debe ser terminado en el mismo tiempo que se termina la etapa cuatro del desarrollo de la prueba, para poder hacer el análisis	2.- Establecer un proceso de seguimiento oportuno de resultados en las pruebas para establecer métodos de retroalimentación	9.- El examen se realiza utilizando software computarizado que automatiza la ejecución, y entrega de resultados.	Sistema funcional para la aplicación de pruebas estandarizadas.  Con el sistema se espera reducir los tiempos para

<p>procesamiento de resultados más rápido, ahorro de papel, interacción y representación de reactivos de maneras más atractivas y compuestas, más parecidas a la realidad.</p>	<p>de software (RDS) Personal docente, alumnos, coordinadores y directivos.</p>	<p>Así mismo contribuye dando información sobre requerimientos funcionales y de usabilidad que el software debe tener.</p>	<p>primario y secundario de resultados mediante el software, lo que incluye desde luego un pilotaje del examen.</p>	<p>n efectivos para los estudiantes.</p> <p>5.- Reducir los factores externos que condicionan el desempeño de las pruebas (áreas, condiciones físicas, horarios, días, eventos, entre otros).</p> <p>9.- Integrar tecnología en la aplicación y ejecución de las pruebas estandarizadas.</p>		<p>entrega de resultados, que los reportes técnicos los genere automáticamente el sistema, el software permite generar de manera aleatoria diferentes versiones del examen al ser alimentado con ítems.</p> <p>El sistema genera automáticamente un reporte técnico con el análisis</p>
--	---	--	---	--	--	---

Dirección General de Bibliotecas de la UAQ

							psicométrico de la prueba.
<b>7. Estandarizar las condiciones de ejecución de la prueba.</b>	Se desea que aquellos factores externos que se ven involucrados en la ejecución de la prueba como son el espacio físico, día y hora, eventos, población entre otros elementos sean estandarizados para lograr una igualdad de	Responsable de evaluación colegiada del aprendizaje.  Áreas de coordinación académica y formación básica.  Responsable de soporte técnico (SP).	El Responsable de ECA debe coordinar con el personal involucrado que las condiciones de la prueba sean las adecuadas y similares para los diferentes grupos que realizan la prueba. Las coordinaciones en colaboración con el SP se encargan de	Este plan de control y de establecimiento de condiciones tiene que estar elaborado para antes de que se aplique la prueba piloto de la intervención, es decir al terminar la etapa cuatro del ciclo de	5.- Reducir los factores externos que condicionan el desempeño de las pruebas (áreas, condiciones físicas, horarios, días, eventos, entre otros).	5.- Las inconsistencias de evaluación entre docentes que imparten una misma materia son erradicadas, al contar con instrumentos que pueden medir el mismo dominio de conocimiento, aunque en su estructura empleen diferentes ítems.	Normas establecidas que permitan estandarizar la igualdad de condiciones al momento de ejecutar la prueba.  Estandarizar las condiciones de ejecución de la prueba reducen los factores externos que

	condiciones para desempeñar la prueba.		establecer la fecha y la hora de la prueba para que todos los sustentantes desempeñen la prueba en igualdad de condiciones, el mejor y mismo horario, así como el día el laboratorio donde se realiza la prueba, que el equipo se encuentre en óptimas condiciones.	desarrollo de la prueba (Producción y validación de ítems).			pueden comprometer, influir o afectar los resultados de los alumnos en la prueba, y con ello obtener datos más confiables.
--	--	--	---	---	--	--	--

Dirección General de Bibliotecas de la UAQ

## 2. Análisis de las acciones

Una vez definidas las acciones estratégicas es necesario detallar información asociada a la aplicación de dichas acciones, lo anterior es mostrado en la siguiente tabla:

Acción estratégica	Obstáculos a los que se enfrentaría	Resistencias que pueden presentarse entre los actores involucrados	Recursos requeridos (infraestructura, equipos, materiales, etc.)	Recursos humanos (personal de apoyo que se requiere para el desarrollo de la acción)	Políticas y normas institucionales existentes vinculadas a la idea o que se requieren para desarrollarla.	Valores institucionales que son importantes de recuperar, promover o fortalecer para el correcto desarrollo de la acción estratégica	Consideraciones adicionales
1. <b>Valorar la calidad de la prueba.</b>	Los índices de discriminación, coeficiente de discriminación y correlación ítem-total no cumplen con los valores mínimos de calidad.	El comité no tenga tiempo suficiente para cumplir con la etapa correspondiente por atender actividades propias a sus	Equipo de cómputo con paquetería de oficina básica, procesador de palabras.	Experto en evaluación estandarizada del aprendizaje.  Un docente experto en la asignatura para la cual se	Plan de desarrollo institucional (PDI)  1.2.1. Propiciar las condiciones institucionales para la adecuada	10. Responsabilidad  8. Perseverancia  5. Justicia  1. Confianza	El lograr que la prueba cumpla en todos sus ítems con los criterios mínimos de calidad en cuanto confiabilidad y validez requiere de un esfuerzo

	<p>El comité de producción y evaluación de ítems no desarrolle suficientes ítems para ser probados o incluidos en la prueba.</p>	<p>funciones docentes.</p> <p>El responsable de ECA haga observaciones continuas al comité, esto en razón de que aún no se tiene la experiencia suficiente.</p>	<p>Cubículo o espacio físico reservado para que el comité trabaje sin distracciones.</p> <p>Conectividad a internet.</p> <p>Impresora</p>	<p>construye el examen.</p> <p>Un docente experto en el currículo. Un docente experto en la enseñanza.</p> <p>Un docente experto y con amplia experiencia en el plan de estudios.</p>	<p>operación de los programas educativos y el mejoramiento de su calidad.</p> <p>2.2.8. Establecer mecanismos que permitan conocer el nivel de dominio de las competencias comprometidas en los planes y programas de estudio durante las etapas de formación y en el egreso de los estudiantes.</p>		<p>colaborativo por parte de los actores involucrados.</p> <p>Los valores son importantes puesto que hacer una prueba de calidad requiere mucha responsabilidad y perseverancia, para lograr que los alumnos tengan una prueba justa y de confianza.</p>
--	--	---	---	---	--	--	--

<p><b>2. Estandarizar métodos de construcción y/o diseño de pruebas.</b></p>	<p>Encontrar una metodología que se adecue a las características de los propósitos que se buscan con la prueba.</p> <p>Resistencia al cambio por parte de los docentes.</p> <p>Participación de la comunidad en el proyecto.</p>	<p>Los docentes principalmente no están de acuerdo a unificar criterios de evaluación o atributos de calidad, se tiene la idea de que estandarizar restringe su libertad de cátedra. En un estudio realizado 57.14% de los docentes considera que es complicado estandarizar criterios de evaluación.</p>	<p>Equipo de cómputo con paquetería de oficina básica, procesador de palabras.</p> <p>Software de diseño Visio.</p> <p>Software especializado para evaluación psicométrica.</p> <p>Cubículo o espacio físico reservado para desarrollo de la prueba.</p> <p>Conectividad a internet.</p> <p>Impresora</p>	<p>Experto en evaluación estandarizada del aprendizaje.</p>	<p>Plan de desarrollo institucional (PDI)</p> <p>1.2.1. Propiciar las condiciones institucionales para la adecuada operación de los programas educativos y el mejoramiento de su calidad.</p> <p>Se requiere un mayor fomento a la cultura evaluativa.</p>	<p>10. Responsabilidad</p> <p>2. Democracia</p> <p>3. Honestidad</p>	<p>Para establecer un estándar en la elaboración de pruebas se tiene que lograr una respuesta positiva por parte de los docentes principalmente, es decir que puedan observar de manera clara las ventajas y mejoras para su practica el emplear este tipo de pruebas. Por ello es importante ser honestos con los propósitos de la prueba así como valorar sus opiniones.</p>
--	--	---	---	---	--	--	--



<p><b>3.</b></p> <p><b>Diseñar reportes técnicos de la ejecución de la prueba.</b></p>	<p>Lograr satisfacer en el mayor grado posible las necesidades de información de los actores involucrados, considerando que en algunos casos el tipo de información y la importancia varia de una persona a otra.</p> <p>Que los reportes técnicos sean lo suficientemente claros pero sin perder la profundidad y especialidad.</p>	<p>Diferencias en cuanto al tipo de información a mostrar en los reportes.</p> <p>La presentación del reporte (apariencia, estructura, colores, elementos etc).</p> <p>Utilidad del reporte.</p>	<p>Equipo de cómputo con paquetería de oficina básica, procesador de palabras.</p> <p>Cubículo o espacio físico reservado para que el comité trabaje sin distracciones.</p> <p>Conectividad a internet.</p> <p>Impresora</p>	<p>Experto en evaluación estandarizada del aprendizaje.</p> <p>Un docente experto en la asignatura para la cual se construye el examen.</p> <p>Un docente experto en el currículo.</p> <p>Un docente experto en la enseñanza.</p>	<p>Plan de desarrollo institucional (PDI)</p> <p>1.2.1. Propiciar las condiciones institucionales para la adecuada operación de los programas educativos y el mejoramiento de su calidad.</p> <p>Se requiere estandarizar los formatos de resultados de las ejecuciones</p>	<p>10. Responsabilidad</p> <p>2. Democracia</p> <p>3. Confianza</p>	<p>Los reportes son un elemento crucial en la ejecución de la prueba, ya que son los que plasman la información con la cual el docente dará retroalimentación al estudiante, y a su vez las áreas de coordinación tomaran decisiones sobre la continuidad o cursos de capacitación para docentes.</p> <p>Por ello la importancia de que</p>
--	--	--	--	---	---	---	---

				Un docente experto y con amplia experiencia en el plan de estudios.	de las pruebas, definir la información mínima para considerarse un reporte de calidad y que ayude a la toma de decisiones.		los reportes sean confiables, aquí entran en juego los valores indicados.
<b>4. Crear un grupo colegiado de evaluación del aprendizaje (GCEA).</b>	La disponibilidad de tiempo por parte de la planta docente para realizar las actividades requeridas.  Participación docente insuficiente para	Diferencias en la integración de los miembros del grupo, designación de roles.  El área directiva quiera incluir docentes de	Un cubículo o espacio físico reservado.  Equipo de cómputo para el personal participante.	Experto en evaluación estandarizada del aprendizaje.  Un docente experto en la asignatura para la cual se	Plan de desarrollo institucional (PDI)  1.2.1. Propiciar las condiciones institucionales para la adecuada operación de los programas	10. Responsabilidad 8. Perseverancia 5. Justicia 3. Confianza 2. Democracia 1. Confianza	El grupo colegiado debe existir en cada unidad académica (UA) de la Facultad, para poder emplear docentes de diferentes UA y con ellos enriquecer y diversificar las aportaciones en el

<p>integrar el grupo.</p> <p>La institución no ofrezca las condiciones o posibilidades de que se forme el grupo colegiado, horas de descarga etc.</p>	<p>manera arbitraria.</p> <p>El personal directivo no estimule la participación para la integración del grupo.</p>	<p>Mesa de trabajo y/o para reuniones.</p> <p>Conectividad a internet</p> <p>Impresora</p>	<p>construye el examen.</p> <p>Un docente experto en el currículo.</p> <p>Un docente experto en la enseñanza.</p> <p>Un docente experto y con amplia experiencia en el plan de estudios.</p>	<p>educativos y el mejoramiento de su calidad.</p> <p>Se requiere un mayor fomento a la cultura evaluativa.</p>	<p>desarrollo de las pruebas.</p>
---	--	--	--	---	-----------------------------------

Dirección General de Bibliotecas de la UAQ

<p><b>5. Diseñar el proceso de seguimiento a los resultados de las pruebas.</b></p>	<p>Apreciaciones o juicios personales de los responsables de las áreas de coordinación.</p> <p>Resistencia a establecer un programa institucional de seguimiento, al considerar los métodos actuales como suficientes o funcionales.</p> <p>Poca participación o involucrados en</p>	<p>Resistencia al cambio, no considerar necesario efectuar cambios al proceso que actualmente se realiza.</p> <p>Considerar una carga adicional a sus funciones.</p> <p>No acordar un mismo proceso para todas las áreas involucradas.</p>	<p>Un espacio para efectuar la reunión entre los actores involucrados.</p> <p>Equipo de cómputo con paquetería de oficina básica, procesador de palabras.</p> <p>Impresora</p>	<p>Experto en evaluación estandarizada del aprendizaje.</p> <p>Coordinador Académico</p> <p>Coordinador de Formación Básica.</p>	<p>Plan de desarrollo institucional (PDI)</p> <p>Estrategia 2.2. Fortalecer las trayectorias escolares de los alumnos para asegurar la conclusión exitosa de sus estudios.</p> <p>2.2.7. Implementar esquemas de seguimiento y atención a la trayectoria</p>	<p>3. Honestidad</p> <p>5. Justicia</p> <p>10. Responsabilidad</p>	<p>Un buen examen debe aportar información más allá de solo una escala numérica, pero lo importante de esto es lo que se hace con la información que la prueba genera. Un uso primordial es para el seguimiento del desempeño que los estudiantes tienen, poder hacer valoración de carácter sumativa y formativo. Para lograr lo anterior es imprescindible que la institución cuente con un</p>
---	--	--	--	--	--	--	---

	el diseño del proceso, con lo cual se pueden estar ignorando ciertos elementos.	Considerar necesidades especiales.			escolar de los estudiantes.		programa adecuado de seguimiento al desempeño y trayectoria de los estudiantes, así como de los docentes.  Los reportes técnicos aportan al logro de dicho seguimiento.
<b>6. Formar docentes con habilidades y competencias en evaluación del aprendizaje.</b>	No todos los docentes querrán participar.  Las áreas directivas tienen que facilitar el tiempo y espacio para realizar	Los cursos de capacitación son ofrecidos por el GCEA y el Responsable de ECA, la resistencia se da con el personal docente que según un	Aula o espacio físico reservado para impartir el curso de capacitación.  Proyector, ordenador con procesador de	Experto en evaluación estandarizada del aprendizaje.  Personal con experiencia que haya	Plan de desarrollo institucional (PDI)  Estrategia 6.2. Promover esquemas de formación y	9. Respeto  10. Responsabilidad  11. Solidaridad	Es muy importante que se sigan preparando a docentes en evaluación del aprendizaje de calidad para lograr que todas aquellas asignaturas impartidas por la

<p>estas actividades.</p> <p>Se requieren horas de descarga o designadas especialmente para el desarrollo de este tipo de actividades.</p>	<p>estudio realizado el 100% de los docentes comenta que tiene la capacidad de desarrollar un examen para evaluar cualquier conocimiento de la asignatura que imparte. Por lo que no hay mucho interés por participar en un curso de formación en evaluación.</p> <p>En este sentido un 80.95% de</p>	<p>palabras y presentador de diapositivas.</p> <p>Conectividad a internet.</p> <p>Mesas de trabajo, ordenadores personales para los participantes.</p>	<p>participado en alguna de las etapas del desarrollo de pruebas estandarizadas.</p>	<p>actualización del personal académico, con base en rutas diferenciadas en función de su experiencia, antigüedad y tipo de contratación.</p> <p>6.2.1. Fortalecer los esquemas de formación y actualización docente para el mejorar las capacidades disciplinarias y didácticas del personal</p>		<p>planta docente cuenten por lo menos con un instrumento de evaluación de calidad del aprendizaje.</p> <p>Es crucial que la institución ponga las condiciones adecuadas para que este tipo de programas de capacitación sean continuos.</p> <p>Fortalecer la cultura evaluativa en la institución por medio de</p>
--	---	--	--	---	--	---

		los docentes no está de acuerdo en compartir estrategias para evaluar.			académico de tiempo completo y de asignatura.		programas de esta naturaleza.
<b>7. Integrar tecnología en la aplicación de la prueba.</b>	<p>Contar con el personal adecuado para el desarrollo del software.</p> <p>Contar con los recursos tecnológicos y de infraestructura necesarios para el desarrollo del software.</p>	<p>Resistencia a compartir información para el levantamiento de requerimientos funcionales y de usabilidad del software o sistema que ejecutara las pruebas.</p> <p>Dificultad para comunicar ideas o requisitos de</p>	<p>Cubículo o espacio reservado para el desarrollo del sistema.</p> <p>Servidor de aplicaciones dedicado para albergar el sistema de exámenes.</p> <p>Equipo de cómputo y software especializado para la</p>	<p>Diseñador y programador de la herramienta tecnológica.</p> <p>Responsable de infraestructura tecnológica y de red en la institución educativa.</p> <p>Personal docente, alumnos, responsables de</p>	<p>Plan de desarrollo institucional (PDI)</p> <p>Estrategia 7.1. Favorecer el uso de tecnologías digitales en el desarrollo de las funciones sustantivas y de gestión de la universidad.</p>	<p>10. Responsabilidad</p> <p>3. Honestidad</p> <p>1. Confianza</p>	<p>Esta parte requiere la integración de equipos de trabajo interdisciplinarios, ya que desarrollar un software requiere por si solo de un plan de proyecto, planeación, metodología entre otros elementos.</p> <p>Aunado a lo anterior, actualmente la inclusión de</p>

<p>Disponibilidad de los laboratorios para la aplicación de las pruebas.</p> <p>Equipo de cómputo actualizado y en óptimas condiciones para ejecutar el software.</p> <p>Normas de seguridad de acceso a la información.</p>	<p>funcionalidad al sistema.</p> <p>No todos los requerimientos se podrán incluir en el software en razón de los tiempos de desarrollo y las pruebas piloto que se quieren desarrollar.</p> <p>Retrasos en el desarrollo del software por estar condicionado a los avances en el desarrollo de la prueba.</p>	<p>construcción del sistema.</p> <p>Conectividad a la red interna de la institución.</p> <p>Conectividad a internet.</p>	<p>coordinación académica y formación básica, directivos.</p>	<p>Estrategia 7.2. Propiciar la formación y actualización de la comunidad universitaria en el uso de las tecnologías digitales.</p> <p>7.2.3. Diseñar modelos, materiales y experiencias de aprendizaje que incorporen el uso de tecnologías digitales.</p>		<p>tecnología en casi todo proceso de la vida cotidiana es inevitable, en el caso de la intervención el incluir tecnología también responde a iniciativas del PDI institucional como a la cultura digital, cuidado del medio ambiente (ahorro de papel) entre otros.</p>
--	---	--	---	---	--	--



<p><b>8.</b></p> <p><b>Estandarizar las condiciones de ejecución de la prueba.</b></p>	<p>El número de alumnos es mayor a la capacidad que se tienen en los laboratorios de computo.</p> <p>Actividades académicas que afecten los horarios y/o días definidos para la aplicación de las pruebas.</p> <p>Actividades extracurriculares que impliquen el uso de espacios comunes.</p>	<p>Resistencia de algunos docentes por ajustarse a ciertos criterios que deben cumplirse para poder realizar la prueba.</p> <p>Resistencia a reagendar actividades para no interferir con la ejecución de las pruebas.</p> <p>Personal que considera innecesarias las normas de</p>	<p>Espacio designado para realizar la prueba.</p> <p>Laboratorio de computo con conectividad intranet.</p> <p>Equipo de cómputo en óptimas condiciones.</p> <p>Espacios con ventilación o clima controlado.</p> <p>Espacios con iluminación</p>	<p>Experto en evaluación estandarizada del aprendizaje.</p> <p>Personal de Soporte Técnico.</p> <p>Responsable de Coordinación Académica</p>	<p>Plan de desarrollo institucional (PDI)</p> <p>Estrategia 9.1. Propiciar que la institución cuente con la infraestructura y equipamiento requeridos para el cumplimiento de sus funciones sustantivas y de gestión.</p> <p>9.1.2. Vigilar el cumplimiento</p>	<p>5. Justicia</p> <p>11. Solidaridad</p>	<p>En ocasiones será difícil establecer la misma condición para el desempeño de una prueba entre un grupo de estudiantes y otro, sin embargo la intención de esta estrategia es reducir en la mayor cantidad posible aquellos factores externos que tienen influencia en el desempeño de la prueba por parte de los estudiantes y con ello lograr obtener información más apegada a la</p>
--	---	---	---	--	---	---	--

	<p>Elementos meteorológicos que no pueden ser controlados.</p>	<p>igualdad de condiciones, o minimicen los efectos de factores externos en el desempeño de la prueba.</p>	<p>adecuada y aislamiento de sonido suficiente para minimizar las distracciones externas.</p>	<p>de las normas y estándares de calidad vigentes para la ampliación, conservación, mejoramiento y modernización de la infraestructura y el equipamiento.</p>		<p>realidad y no obtener resultados con los cuales no se puedan hacer inferencias válidas.</p>
--	--	--	---	---	--	--

Dirección General de Bibliotecas de la UAQ

## **ANEXO 5 Plan de evaluación de la intervención**

Dirección General de Bibliotecas de la UAQ

## Fase. Componente 1: Evaluación general

### Criterio: Pertinencia

Se refiere a la estrecha relación que guarda el proyecto de intervención con las necesidades y expectativas de los usuarios, actores o beneficiados. Los instrumentos de valoración para este criterio son los siguientes:

Fase. Componente 1		Evaluación general			
Criterio		Pertinencia			
Indicador: 1		Necesidades de la población y expectativas de la intervención.			
Necesidad o expectativa identificada	Objetivo de intervención que atiende la necesidad	Evaluación del logro			Descripción de las acciones y resultados obtenidos
		Insuficiente	Suficiente	Sobresaliente	
La Facultad de Idiomas no cuenta con una herramienta tecnológica para la evaluación de los aprendizajes.	Integrar tecnología en la aplicación y ejecución de las pruebas estandarizadas.			X	<p>Se ha desarrollado una aplicación tecnológica a la medida que permite construir exámenes de forma digital, y a su vez efectúa de manera automática un análisis psicométrico de la calidad técnica de la prueba.</p> <p>Contar una sistema dedicado a la aplicación de pruebas estímulo cambios en la filosofía evaluativa y políticas de la institución, así como en la formación del personal docente al implementar un programa de</p>

					capacitación en evaluación del estandarizada del aprendizaje.
La facultad no cuenta con una metodología estandarizada o establecida para el diseño de pruebas.	Estandarizar el diseño de pruebas para tener un control más efectivo de la calidad de los instrumentos de evaluación que se utilizan.			X	Se ha implementado una metodología para el diseño de pruebas estandarizadas con base en Teoría de Respuesta al Ítem (TRI), que además, genera una serie de productos que permiten hacer análisis de otro tipo de información, identificar contenidos temáticos relevantes etc.
	Eliminar las inconsistencias en la evaluación de docentes que imparten una misma asignatura.		X		Se estandarizaron los procedimientos para diseñar la prueba, pero solo se utiliza en pruebas finales o semestrales, faltan las pruebas parciales.
	Establecer mecanismos metodológicos que permitan evaluar de manera confiable y valida el dominio de conocimiento por parte de los estudiantes.			X	La metodología implementada permite evaluar el grado de dominio del estudiante con base en un universo de medida definido a partir del currículo de la asignatura. Además, los criterios de validez y confiabilidad son cuantificables y por ende comprobables.
La validez y confiabilidad de las pruebas es baja.	Aumentar el índice de confiabilidad y validez de las pruebas que se aplican.			X	Los criterios de confiabilidad y validez de la prueba con base en la TRI permiten evaluar los atributos de calidad con los siguientes indicadores:

					<p>Índice de dificultad de la prueba</p> <p>Índice de dificultad del ítem</p> <p>Índice de discriminación</p> <p>Coefficiente de discriminación</p> <p>Estos indicadores de calidad fueron medidos con tres softwares especializados en análisis psicométrico por lo que existen el fundamento sólido y comprobado de la calidad de la prueba.</p>
No hay una normativa que regule las condiciones de aplicación de las pruebas.	Reducir los factores externos que condicionan el desempeño de las pruebas (áreas, condiciones físicas, horarios, días, eventos, entre otros).		X		Se diseñó un instrumento normativo que incluye las buenas prácticas que se deben considerar antes de programar una evaluación. El instrumento incluye los principios básicos para asegurar una igual y equitativa condición para el desempeño de la prueba.
No se tienen reportes técnicos de las	Diseñar reportes técnicos que permitan analizar los atributos de calidad de las pruebas y con			X	Se diseñaron tres reportes, cada uno diseñado especial para el tipo de usuario en cuestión (directivo, coordinador, docente), cada uno sometido a

<p>ejecuciones de las pruebas.</p>	<p>ellos aportar datos válidos para tomas de decisiones oportunas y eficaces.</p>				<p>evaluaciones de usabilidad y diseño para garantizar la claridad, pertinencia y calidad de la información presentada.</p>
<p>NO existe un programa institucional de capacitación docente en evaluación del aprendizaje</p>	<p>Establecer un programa de capacitación en evaluación estandarizada del aprendizaje a los docentes que conforman la institución educativa.</p>		<p>X</p>		<p>Se ha realizado el diseño instruccional del curso de capacitación de la implementación de la metodología para la construcción de las pruebas, así como del uso y funcionamiento de la tecnología con la que se ejecutaran las pruebas.</p> <p>Se ha realizado una primera capacitación, el personal directivo ha aprobado el diseño instruccional del programa de capacitación.</p> <p>Sin embargo, aún se requiere de efectuar un mayor número de veces el curso para poder hacer valoraciones más solidas sobre el logro del programa y/o hacer adecuaciones para la mejora del mismo.</p>

Dirección General de Bibliotecas de la UAG

<p>No existe un equipo de trabajo que se encargue de la evaluación colegiada del aprendizaje dentro de la Facultad</p>	<p>Crear un grupo colegiado de evaluación del aprendizaje (GCEA)</p>	<p>X</p>		<p>Esta estrategia para este proyecto de intervención se encuentra fuera del alcance, ya que requiere de ciertos procesos y gestiones administrativas propias de la institución educativa. Situación que es ajena a los objetivos primordiales de la intervención, que es implementar pruebas estandarizadas referidas a un criterio para la evaluación del aprendizaje integrando tecnología. Si bien contar con un grupo facilitaría el desarrollo de más pruebas en menor tiempo, encargarse de actividades de capacitación, talleres entre otros, por el momento no es viable. Además, el trabajo de investigación no trata sobre creación de comités o grupos de trabajo, pero pudiera ser contemplado para una futura intervención.</p>
--	--	----------	--	---

Dirección General de Bibliotecas de la UAG



## Fase. Componente 1

## Evaluación general

Criterio		Pertinencia			
Indicador: 2		Recursos con los que se cuenta.			
Recursos necesarios	Recursos presentes	Evaluación del logro			Descripción de las acciones y resultados obtenidos
		Insuficiente	Suficiente	Sobresaliente	
Se requiere un espacio dedicado o que pueda ser utilizado para las evaluaciones.	Se cuenta con dos laboratorios para uso docente, un laboratorio para consulta estudiantil		X		La institución cuenta con 3 espacios con capacidad tecnológica para ejecutar las pruebas. Si bien no son de uso exclusivo con fines de evaluación, pueden ser acondicionados para utilizarse como centros de evaluación.
Se requiere un espacio designado para el diseño de la prueba, con equipo de cómputo y software especializado.	Se cuenta con un cubículo con capacidad para 5 personas, cuenta con 5 estaciones de trabajo, impresora e internet.		X		La institución cuenta con un espacio que ha sido designado para el equipo de trabajo encargado del diseño de la prueba. Para los casos en los que se requiere una mayor cantidad de participantes en una reunión, la institución proporciona otro espacio con mesas de trabajo para hasta 15 personas.  El software especializado no requirió de su compra, ya que con las versiones gratuitas de los mismos pudieron desarrollarse los análisis.

Se requiere un espacio designado para el desarrollo del sistema de evaluación, con equipo de cómputo y software especializado.	Se cuenta con un cubículo con capacidad para 5 personas, cuenta con 5 estaciones de trabajo, impresora e internet.			X	Además de proporcionar el aula exclusiva para desarrollar estas actividades, proporcionó equipo nuevo, la adquisición de las licencias para el software especializado y las condiciones para que el equipo de desarrollo pudiera trabajar incluso de manera remota.
--	--	--	--	---	---

Dirección General de Bibliotecas de la UAG

## Fase. Componente 1

## Evaluación general

Criterio		Pertinencia			
Indicador: 3		Viabilidad de la intervención			
Elementos identificados	Realidad institucional	Evaluación del logro			Descripción de las acciones y resultados obtenidos
		Insuficiente	Suficiente	Sobresaliente	
La institución debe proveer las condiciones, espacios, personal y tiempo para efectuar la intervención.	La institución posee los espacios que pueden ser adecuados a las necesidades del proyecto, cuenta con suficiente personal para integrar los equipos de trabajo necesarios, cuenta con la disponibilidad de horario para desarrollar las actividades			X	La institución ha otorgado y contribuido con los recursos necesarios para poder implementar e proyecto de intervención, incluso ha facilitado procesos con la participación de docentes de otras unidades académicas, asignando horas específicas para desempeñar actividades del proyecto
Se debe contar con el personal suficiente para el desarrollo de la prueba, la construcción del software y los análisis de calidad de la prueba.	La institución cuenta con una planta docente de 23 Maestros de asignatura y 7 tiempos completos.  La institución tiene presencia en cuatro municipios del			X	La institución proveyó del 70% de la planta docente que forma parte de la unidad académica donde se desarrolla la intervención, y un 30% fue personal que provenía de otras unidades académicas. Esta integración de los participantes enriqueció la

	estado, por lo que hay más personal con el cual se pudiera contar.				construcción de la prueba al recibir aportaciones de una mayor diversidad maestros.
La institución debe garantizar la participación del personal docente, administrativo, directivo y de alumnos en la intervención.	En cuatro de los cinco municipios del estado existe presencia de la Facultad de Idiomas, cada unidad posee personal directivo, de coordinación, administrativo, docente y una población promedio de 2800 alumnos.			X	<p>El director de la Facultad estableció oficialmente la participación incondicional de la planta docente, administrativa y de estudiantes en el proyecto de intervención. Asigno horas específicas para atender las actividades que le proyecto requería, estimulo la participación y compromiso de la planta docente.</p> <p>También efectuó reuniones periódicas para monitorear avances y participación de docentes, administrativos y coordinadores. Así como también en el desarrollo de ciertas actividades.</p>

## Fase. Componente 2: Evaluación del proceso de implementación

### Criterio: Coherencia

El criterio de coherencia, operativamente, se trata del grado de consistencia de la propuesta de intervención y la aplicación en el contexto real. Los instrumentos utilizados para evaluar este grado de consistencia son los siguientes:

Fase. Componente 2		Evaluación del proceso de implementación		
Criterio		Coherencia		
Indicador: 1		Metodología implementada		
Propuesta	Etapas contempladas	Etapas realizadas		Descripción del grado de aplicación o ausencia de la propuesta
		Si	No	
Metodología desarrollada en el Instituto de Investigación y Desarrollo Educativo (IIDE) (Contreras, 2000; Contreras y Backhoff, 2004; Contreras, Encinas y De Las Fuentes, 2005)	Capacitación teórica sobre la mitología a implementar	X		Se aplicó el 100% de las actividades que esta etapa contemplaba.
	Comité de diseño de la prueba	X		Se aplicó el 100% de las actividades que esta etapa contemplaba. Sin embargo de los 6 docente que se había destinado a participar en esta etapa, solo participaron 5.
	Comité de especificaciones de ítems	X		Se aplicó el 100% de las actividades que esta etapa contemplaba.

con base en el modelo psicométrico que propuso originalmente Nitko (1994) para elaborar exámenes de gran escala de referencia criterial orientados por el currículum.	Comité de producción de ítems	X		Se aplicó el 100% de las actividades que esta etapa contemplaba. Sin embargo no se logró producir como mínimo 4 ítems para todos los contenidos temáticos previstos en la prueba, en el 35% solo se produjeron 3 ítems.
	Pilotaje de la prueba	X		Se aplicó el 100% de las actividades que esta etapa contemplaba.
	Comité de evaluación de ítems	X		Se aplicó el 100% de las actividades que esta etapa contemplaba.
	Diseño de reportes técnicos	X		Se aplicó el 100% de las actividades que esta etapa contemplaba.
	Reporte de evaluación de la prueba	X		Se aplicó el 100% de las actividades que esta etapa contemplaba.
<b>Fase. Componente 2</b>		<b>Evaluación del proceso de implementación</b>		
<b>Criterio</b>		<b>Coherencia</b>		
<b>Indicador: 1</b>		<b>Metodología implementada</b>		
<b>Propuesta</b>	<b>Etapas contempladas</b>	<b>Etapas realizadas</b>		<b>Descripción del grado de aplicación o ausencia de la propuesta</b>
		<b>Si</b>	<b>No</b>	
Metodología Rational Unified Process (RUP) para el desarrollo de la	Fase de inicio	X		Se aplicó el 100% de las actividades que esta etapa contemplaba.

tecnología innovadora para la aplicación y análisis psicométrico de las pruebas estandarizadas	Fase de elaboración	X		Se aplicó el 100% de las actividades que esta etapa contemplaba.
	Fase de desarrollo	X		Se aplicó el 100% de las actividades que esta etapa contemplaba. Sin embargo no todas las acciones planeadas para cada actividad se realizaron, por ejemplo, algunos requerimientos funcionales no se implementaron en el software, ya que algunas actividades demandaron más tiempo del previsto y no formaban parte de las funciones críticas para la operación y puesta en práctica de la herramienta tecnológica.
	Fase de transición	X		Se aplicó el 100% de las actividades que esta etapa contemplaba. Se realizó una prueba más de funcionalidad a la red (de las definidas en el plan de pruebas) en el laboratorio donde se realizó el pilotaje de la prueba

Dirección General de Bibliotecas de la UAQ

## Fase. Componente 2

## Evaluación del proceso de implementación

Criterio		Coherencia		
Indicador: 2		Operatividad		
Propuesta	Elementos contemplados	Se cumplió		Descripción del grado de aplicación o ausencia de la propuesta
		Si	No	
Infraestructura tecnológica	Laboratorio de computo con capacidad para 28 usuarios	X		Sin observaciones
	Servidor de base de datos para la herramienta tecnológica	X		Se implementó un segundo servidor de base de datos para que funcionará como respaldo en caso de presentar alguna falla durante la ejecución de la prueba.
	Conectividad a la red interna de la Facultad de Idiomas	X		Sin observaciones
	Cubículo o espacio designado para el desarrollo del software con conectividad a la red interna e internet	X		Sin observaciones
	Software especializado para análisis psicométrico	X		Sin observaciones



	Software especializado para el desarrollo del software	X		Sin observaciones
	Software para modelado y diseño	X		Sin observaciones
	Ordenador central para servidor de aplicaciones	X		Se implementó un segundo servidor de aplicaciones para que funcionará como respaldo en caso de presentar alguna falla durante la ejecución de la prueba.

Dirección General de Bibliotecas de la UAQ

## Fase. Componente 2

## Evaluación del proceso de implementación

Criterio		Coherencia		
Indicador: 2		Operatividad		
Propuesta	Elementos contemplados	Se cumplió		Descripción del grado de aplicación o ausencia de la propuesta
		Si	No	
Recurso humano	Comité diseñador de la prueba (6 integrantes)	X		Sin observaciones
	Comité de especificaciones de ítems (6 integrantes)	X		Solo se integraron cinco participantes de los seis propuestos, uno de ellos no quiso participar
	Comité de producción de ítems (8 integrantes)	X		Solo participaron siete de los 8 propuestos, uno de ellos recibió otra asignación.
	Comité de evaluación (4 integrantes)	X		Sin observaciones
	Comité de diseño de reportes (4 integrantes)	X		Sin observaciones

Dirección General de Bibliotecas de la UAQ

Docentes impartiendo la asignatura para la que se construye el examen  (6)	X		Sin observaciones
Alumnos muestra  (200)	X		La muestra final para el pilotaje de la prueba fue de 238 alumnos.
Personal de áreas de coordinación (8)	X		Solo participaron seis de los ocho propuestos, dos de ellos no participaron por tener otras asignaciones.
Personal para el desarrollo del software (7)	X		Se propusieron siete participantes, pero al final se integraron nueve, se agregó un programador y un analista.

Dirección General de Bibliotecas de la UAQ

## Fase. Componente 2

## Evaluación del proceso de implementación

Criterio		Coherencia		
Indicador: 3		Capacitación		
Propuesta	Capacitación contemplada	Se cumplió		Descripción del grado de aplicación o ausencia de la propuesta
		Si	No	
Capacitar al personal docente y de coordinación que participara en los diferentes comités del diseño de la prueba	Capacitación grupal por cada comité participante en el desarrollo de la prueba	X		Se capacito al 100% de los integrantes de los diferentes comités. En el caso particular del comité de evaluación fue necesario impartir una segunda vez el curso.
Capacitar al personal docente que utilizará el software para la aplicación de las pruebas estandarizadas	Capacitación a los docentes que imparten la asignatura para la cual se desarrolla el examen estandarizado (8 docentes)	X		Se capacito a los ocho docentes que para el semestre en que sería aplicada la prueba impartían la asignatura, sin embargo hay más docentes que pudieran impartir dicha asignatura, y por ende será necesario un nuevo curso de capacitación.
Capacitar al personal docente, de coordinación y directivos en la interpretación de los resultados en los reportes técnicos	Capacitación a los 8 docentes que imparten la asignatura, 4 subdirectores y 8 coordinadores	X		Se capacito al 100% del personal planeado en la propuesta, originalmente se tenía planeado un curso en cada unidad académica, pero se impartió un solo curso para todo el personal propuesto.

### Fase. Componente 3: Evaluación de los resultados

Criterio: Eficacia

Este criterio contempla la valoración de los resultados obtenidos con la implementación de la propuesta de intervención. Formalmente este criterio se define como la distancia entre los objetivos planeados inicialmente en el diseño y los logros obtenidos, la relación entre lo asignado y lo alcanzado. Los formatos utilizados para la evaluación de este criterio son los siguientes:

Fase. Componente 3		Evaluación de los resultados			
Criterio	Eficacia				
Indicador: 1	Logros				
Objetivo de intervención	Meta	Descripción del logro alcanzado	Cumplimiento		
			Insuficiente	Suficiente	Mayor
1.- Aumentar el índice de confiabilidad y validez de las pruebas que se aplican.	Los índices de confiabilidad (Índice de dificultad de los ítems, Índice de dificultad de la prueba promedio entre .4 y .65, Índice de relevancia curricular) y validez (índice de discriminación, correlación ítem-total mayores a .2) en las pruebas cumplen con los valores de calidad definidos por la Teoría de Respuesta al Ítem (TRI).	<p>Índice de dificultad promedio en la prueba: .621</p> <p>Índice de discriminación promedio: .27</p> <p>Coefficiente de discriminación promedio: .23</p> <p>Todos los ítems cumplieron con los mínimos de calidad</p>			X

		<p>establecidos por la TRI, con excepción de 4 de los 63 que componen la prueba, específicamente el índice y coeficiente de discriminación fueron menores a .2, pero esto fue porque los ítems eran muy fáciles (dificultad de .8 a .9), motivo por el cual los ítems fueron contestados en igual proporción.</p>			
--	--	---	--	--	--

<p>2.- Establecer un proceso de seguimiento oportuno de resultados en las pruebas para establecer métodos de retroalimentación efectivos para los estudiantes.</p>	<p>Existe un programa institucional de seguimiento para los resultados obtenidos en las pruebas, cumpliendo con un mínimo de una retroalimentación para cada estudiante en relación al desempeño mostrado en las pruebas.</p>	<p>Aunque si bien este es un complemento clave para poder obtener el mayor beneficio de contar con pruebas de calidad, el trabajo de investigación no contempla información teórica o referentes metodológicos para el diseño de procesos para el seguimiento de trayectoria estudiantil. Sin embargo, este también sería un apartado contemplado para una futura intervención, la cual se enfoque especialmente en el diseño y evaluación de un modelo específico para el seguimiento del rendimiento escolar, un esquema de toma de decisiones entre otros aspectos.</p> <p>La intervención alcanza solamente a unas primeras aproximaciones de tomas de decisiones con los resultados obtenidos en las</p>	<p>X</p>		
--	---	---	----------	--	--

pruebas y representados en los reportes técnicos, pero no tan complejo como un proceso de trayectoria.

Dirección General de Bibliotecas de la UAQ



<p>3.- Estandarizar el diseño de pruebas para tener un control más efectivo de la calidad de los instrumentos de evaluación que se utilizan.</p>	<p>Se cuenta con pruebas estandarizadas para las asignaturas de mayor índice de relevancia curricular.</p>	<p>Se desarrollaron pruebas para las dos asignaturas con mayor índice de relevancia curricular en el tronco común de Idiomas de la Facultad de Idiomas.</p> <p>Se implementó una metodología que se adecuó a las necesidades y características del contexto de la intervención. La metodología además, generó una serie de herramientas que son útiles para otro tipo de situaciones, como la retícula y el análisis de relevancia curricular.</p> <p>Estas pruebas estandarizadas contempladas en el proyecto de intervención son aplicadas específicamente para los exámenes ordinarios o</p>			<p>X</p>
--	--	---	--	--	----------

Dirección General de Bibliotecas de la UAQ

departamentales de las asignaturas. Puesto que estos exámenes evalúan todo el contenido de la asignatura, es decir todo el currículo, situación que empata directamente con las características de la metodología seleccionada para desarrollar pruebas estandarizadas (exámenes referidos a un criterio, por criterio se refiere al currículo de la asignatura). Pero se tienen las condiciones y la capacidad de migrar a exámenes parciales o de cortes de unidades temáticas por mencionar algunos.

<p>4.- Eliminar las inconsistencias en la evaluación de docentes que imparten una misma asignatura.</p>	<p>Las inconsistencias de evaluación entre docentes que imparten una misma materia son erradicadas, al contar con instrumentos que pueden medir el mismo dominio de conocimiento, aunque en su estructura empleen diferentes ítems.</p>	<p>La metodología permite producir ítems que, aunque sean diferentes, evalúan con la misma calidad el mismo dominio de conocimiento. Esto permite que cada docente pueda crear su propia versión de un examen y aunque existan “n” versiones de un examen, todos miden de la misma manera el dominio de conocimiento de un estudiante, lo que contribuye a que el estudiante tenga las mismas probabilidades de acreditar un curso independientemente del docente con el que curse la materia.</p> <p>Las valoraciones arrojadas por la prueba permiten hacer comparaciones entre las evaluaciones parciales que hace el docente con sus propios</p>	<p>X</p>		
---	---	--	----------	--	--

instrumentos, y las evaluaciones realizadas con la prueba estandarizada.

Sin embargo el logro no es sobresaliente, puesto que solo se han eliminado las principales inconsistencias en la evaluación, pero aún restan elementos por corregir, como las puntuaciones asignadas a las actividades, dosificación de horas entre otros.

Dirección General de Bibliotecas de la UAQ

<p>5.- Reducir los factores externos que condicionan el desempeño de las pruebas (áreas, condiciones físicas, horarios, días, eventos, entre otros).</p>	<p>Se elabora y aplica un programa con las normas que regulen las condiciones de ejecución de la prueba.</p>	<p>Se elabora el plan normativo para la ejecución de las pruebas, adaptado a las condiciones contextuales y capacidades de infraestructura y tecnología con las que se cuentan en la institución.</p> <p>Existen espacios suficientes para atender la población estudiantil, considerando que las pruebas en esta primera intervención son las departamentales u ordinarias, se cuenta con una mayor disponibilidad de horario para el uso de los laboratorios, actualmente ningún grupo supera los 28 alumnos, motivo por el cual todos los grupos realizarían la prueba en las mismas condiciones. Los laboratorios cuentan con clima controlado, iluminación recientemente mejorada, y se encuentran ubicados de forma que</p>		<p>X</p>	
--	--	---	--	----------	--

Dirección General de Bibliotecas de la UAQ

		<p>no hay distracciones visuales inmediatas.</p> <p>Referente a interferencias sonoras del exterior, también se encuentran en espacios alejados de plazas o áreas deportivas, por lo que en dado caso de presentarse actividades en alguna de estas áreas no afectarían durante la ejecución de la prueba.</p>			
<p>6.- Establecer mecanismos metodológicos que permitan evaluar de manera confiable y valida el dominio de conocimiento por parte de los estudiantes.</p>	<p>Se aplica una metodología para el diseño de pruebas estandarizadas que contemplen los criterios de calidad definidos por la TRI.</p>	<p>Se implementó una metodología que contempla de manera clara cuales son los criterios con los que se puede medir la calidad de la prueba.</p> <p>Define criterios específicos para evaluar la validez y la confiabilidad de la prueba.</p> <p>Mediante un proceso estadístico – matemático se genera información con la cual se emiten juicios</p>			X

		<p>pertinentes sobre la calidad de la prueba, de esta forma los resultados obtenidos por un estudiante en el examen tienen un fundamento comprobable sobre el grado de dominio que este posee en determinada área del conocimiento.</p> <p>Aunque los ítems de la prueba sean diferentes miden de la misma manera el dominio de conocimiento en cuestión.</p>			
<p>7.- Diseñar reportes técnicos que permitan analizar los atributos de calidad de las pruebas (índice de discriminación, índice de dificultad, índice de correlación de distractores) y con ellos aportar datos válidos para tomas</p>	<p>Se cuenta con tres reportes técnicos, diseñados con información clara, precisa y relevante de cada uno de los aspectos medidos en la prueba, para el docente, personal administrativo y responsables de evaluación (comité colegiado del aprendizaje).</p>	<p>Se capacitó a un grupo de docentes para elaborar reportes técnicos de ejecución de una prueba, con lo cual se tiene la capacidad de construir reportes para cualquier prueba que se quiera desarrollar en un futuro.</p>			

<p>de decisiones oportunas y eficaces.</p>		<p>Los reportes respondieron a las necesidades directas de los usuarios finales, se efectuaron evaluaciones de usabilidad, diseño, estructura, redacción e interpretación de datos, esto con el fin de contemplar la mayor cantidad de criterios de calidad en los reportes y así lograr producir instrumentos pertinentes y que contribuyan a la mejora en la toma de decisiones.</p>			
<p>8.- Establecer un programa de capacitación en evaluación estandarizada del aprendizaje a los docentes que conforman la institución educativa.</p>	<p>Se implementa un programa institucional de capacitación para el diseño de pruebas estandarizadas y evaluación de calidad de las mismas.</p>	<p>Este objetivo viene implícito en la aplicación de la metodología para el desarrollo de pruebas estandarizadas, ya que para cada comité que se integra en cada una de las etapas del desarrollo de la prueba hay un proceso de capacitación, con lo cual los docentes adquieren habilidades para desempeñar ciertas tareas, como hacer un análisis curricular,</p>		<p>X</p>	



		<p>establecer índices de relevancia curricular, justificar contenidos, elaborar especificaciones de ítems, producir ítems entre otros.</p> <p>Se elaboró el diseño instruccional del curso de capacitación para los diferentes participantes en los comités de elaboración de la prueba, así como a los usuarios de los reportes técnicos.</p> <p>El programa ya fue aceptado por el director, a partir del próximo semestre se ofrecerá de manera regular.</p>			
9.- Integrar tecnología en la aplicación y ejecución de las pruebas estandarizadas.	El examen se realiza utilizando software computarizado que automatiza la ejecución, evaluación y entrega de resultados.	Se construyó un software que permite construir de forma digital la prueba, además de que permite generar automáticamente los reportes técnicos de la ejecución			X

de la prueba, entre ellos el reporte con el análisis psicométrico.

Este es un desarrollo innovador, ya que para hacer estos análisis psicométricos se requería software especializados y experticia en el manejo de los mismos para dar formato a los archivos fuente, manipulación de variables y operación general del software.

Con este nuevo software llamado SIEXAES cualquier docente puede efectuar un análisis psicométrico de la prueba que construya en la plataforma, solo requiere conocimientos básicos de psicometría para interpretar los resultados que el sistema arroja.

		Esto reduce los tiempos de calificación de la prueba, de generar reportes para cada sustentante y del proceso de retroalimentación por parte del docente al alumno.			
--	--	---	--	--	--

Dirección General de Bibliotecas de la UAQ

## Fase. Componente 4: Institucionalización de la intervención

### Criterio: Relevancia

El criterio de relevancia se entiende como la importancia que se le otorga a un programa para ser implementado y cubrir las necesidades individuales y/o sociales. Los instrumentos utilizados para evaluar los indicadores que permiten medir este criterio son los siguientes:

Fase. Componente 4		Institucionalización de la intervención		
Criterio		Relevancia		
Indicador: 1		Políticas		
Propuesta	Realidad en las políticas institucionales	Institucionalizada		Políticas y normas institucionales existentes vinculadas a la propuesta (Plan de desarrollo institucional de la universidad)
		Si	No	
Definición de políticas institucionales en relación a la aplicación de las pruebas estandarizadas.	Normativa sobre estandarización de condiciones en la ejecución de la prueba	X		<p>Estrategia 9.1. Propiciar que la institución cuente con la infraestructura y equipamiento requeridos para el cumplimiento de sus funciones sustantivas y de gestión.</p> <p>9.1.2. Vigilar el cumplimiento de las normas y estándares de calidad vigentes para la ampliación, conservación, mejoramiento y modernización de la infraestructura física y equipamiento.</p>

<p>Institucionalización de la metodología y los criterios de calidad que las pruebas deben cumplir.</p>	<p>Metodología definida y aplicada para el diseño y elaboración de pruebas estandarizadas en la evaluación del aprendizaje</p>	<p>X</p>		<p>Estrategia 1.2.1. Propiciar las condiciones institucionales para la adecuada operación de los programas educativos y el mejoramiento de su calidad.</p>
<p>Institucionalización de una capacitación continua en evaluación del aprendizaje (técnicas, estrategias etc.) para la planta docente.</p>	<p>Curso de capacitación elaborado (cuenta con una primera puesta en práctica ), diseño instruccional</p>	<p>X</p>		<p>Estrategia 6.2. Promover esquemas de formación y actualización del personal académico, con base en rutas diferenciadas en función de su experiencia, antigüedad y tipo de contratación.</p> <p>Estrategia 6.2.1. Fortalecer los esquemas de formación y actualización docente para el mejorar las capacidades disciplinarias y didácticas del personal académico de tiempo completo y de asignatura.</p>

Dirección General de Bibliotecas de la UAQ

## Fase. Componente 4

## Institucionalización de la intervención

Criterio	Relevancia				
Indicador: 2	Filosofía				
Objetivo de intervención	Meta	Descripción del logro alcanzado	Cumplimiento		
			Insuficiente	Suficiente	Mayor
<p><b>Crear una cultura evaluativa donde toda acción evaluativa sea precedida de un fundamento teórico y metodológico que garantice la calidad del instrumento evaluativo.</b></p>	<p>El 50% de la planta docente desarrolla sus instrumentos de evaluación con fundamento en la TRI o TCT( Teoría clásica de los test).</p>	<p>Esta etapa aun no puede ser medida, estos cambios se esperan posterior a la conclusión de la intervención. Sin embargo, se puede observar que del 70% de la planta docente que participo en el diseño de la prueba empleando la metodología con base en la TRI, el 100% manifestó el deseo de crear instrumentos de evaluación de la misma manera.</p>			X
	<p>El 50% de la planta docente analiza más que solo la calificación numérica de la prueba, efectúa análisis más detallados y concretos del desempeño del estudiante.</p>	<p>El 80.95% de los docentes que participaron en la capacitación de los diferentes comités de diseño de la prueba y en el uso de los reportes técnicos manifestaron un interés por</p>			X

		efectuar análisis más profundos de los resultados obtenidos en las pruebas.			
	El 50% de la planta docente analiza la calidad técnica de sus instrumentos de evaluación.	El 76.19% de los docentes consideran importante evaluar la calidad de sus instrumentos, pero no la efectuaban por desconocer métodos para hacerlo.			X
	El 60% de la planta docente posee capacitación en materia de evaluación estandarizada del aprendizaje.	Esta meta aún está en proceso, al momento han participado solo 38% de la planta docente en la unidad académica, pero el 100% de personal de coordinación y directivo.	X		

Dirección General de Bibliotecas de la UAQ

**Fase. Componente 4****Institucionalización de la intervención**

<b>Criterio</b>	<b>Relevancia</b>		
<b>Indicador: 1</b>	<b>Adopción general del proceso</b>		
<b>Objetivo de intervención</b>	<b>Atributos</b>	<b>Estado antes de la intervención</b>	<b>Descripción del logro alcanzado</b>
<b>La Facultad integra como parte de su operación académica y ejecución educativa la estandarización de pruebas para la evaluación del aprendizaje.</b>	Resistencia a estandarizar criterios	El 57.14% de los docentes considera que es complicado estandarizar criterios de evaluación.	El 80.95% considera viable estandarizar los criterios de evaluación.
	Auto apreciación de expertos	El 100% de los docentes comenta que tiene la capacidad de desarrollar un examen para evaluar cualquier conocimiento asociado a la asignatura que imparte.	El 66.66% considera que requieren capacitarse para elaborar pruebas válidas y confiables
	Libertad de catedra	El 52.38% considera que estandarizar una prueba es una forma de limitar su libertad de catedra.	El 76.19% acepta las pruebas estandarizadas como método valido para evaluar en su práctica.
	Cultura evaluativa	El 61.90% considera que su clase ha funcionado con los instrumentos y estrategias que han aplicado, por lo que no	El 85.71% considera necesario aplicar otros instrumentos de evaluación y/o poder comparar resultados con los que



	consideran relevante querer implementar algunos otros.	actualmente aplica en el desarrollo de sus clases.
Objetivos de las pruebas	El 85.71% de los docentes consideran como objetivo final del examen el observar la calificación numérica que estos obtienen.	El 19.04% de los docentes solo se limita a considerar como objetivo final de la prueba la calificación numérica obtenida.
Resistencia a ser observado	El 85.71% de los docentes considera que se deben estipular bien los usos de la prueba y garantizar la privacidad de la información para poder participar en una prueba estandarizada.	El 90.47% se siente seguro y confiado del manejo de la información en la aplicación de la prueba.
Estímulo a evaluar la calidad de las pruebas	El 76.19% de los docentes consideran que sus instrumentos son de calidad, aunque no los sometan a ningún tipo de juicio o revisión por pares, ni ningún tipo de análisis.	Solo el 23.80% considera que sus instrumentos son de calidad, aunque aceptan el hecho de que pueden ser mejorados o analizados para validar sus resultados.
Resistencia a compartir estrategias	El 80.95% de los docentes consideran que el material que desarrollan no debe ser compartido libremente ya que requirió de un esfuerzo propio.	El 14.28% manifiesta una resistencia a querer compartir estrategias.

	<p>Impacto de la prueba</p>	<p>El 81.94% de los alumnos considera que es importante saber la calidad de la prueba, y qué tanto valor tiene un examen en relación a la calificación final del semestre. El 91.66% manifestó sentirse más seguro de hacer una prueba en la que tengan la seguridad del contenido específico que se evaluara. En el caso de los docentes el 95.23% expresó que es importante saber cuál será el uso del resultado de las pruebas, si tienen consecuencias para ellos y de qué tipo, o si son formativas. El 100% considero elemental saber para que servirá la prueba en relación a su deseo de participar o no en un proceso de estandarización.</p>	<p>El 100% de los alumnos expresaron sentirse seguros con la calidad de la prueba al recibir información sobre el procedimiento de construcción de la prueba.</p> <p>El 100% de los docentes estuvieron de acuerdo con el propósito de la prueba, el uso y manejo de la información que esta produce.</p>
--	-----------------------------	--	---

## **ANEXO 6 Proyección de la estrategia**

Dirección General de Bibliotecas de la UAQ

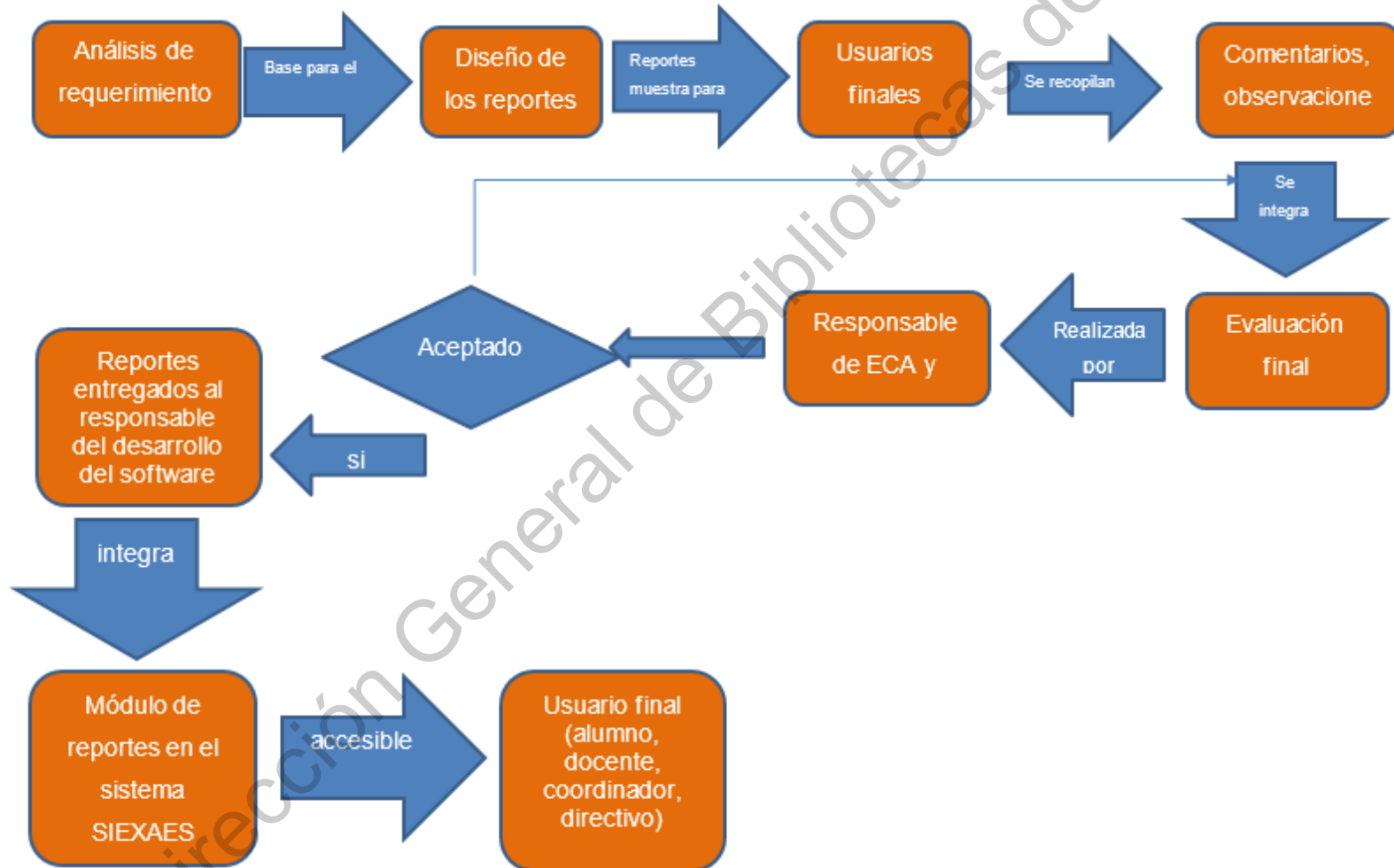
### Estrategia #3: Diseñar reportes técnicos de la ejecución de la prueba

#### Definición de las etapas

ETAPA	OBJETIVO(S)	INDICADORES DE LOGRO
<b>1. Análisis de requerimientos y necesidades de los usuarios finales</b>	<p>Identificar la información más relevante, pertinente y de interés para los usuarios finales de la prueba, alumnos, docentes, coordinadores y directivos.</p> <p>Esta etapa es realizada por el responsable de ECA</p>	Requerimientos identificados
<b>2. Diseño de los reportes técnicos</b>	<p>Elaboración de los reportes técnicos en los cuales se presentará la información resultante de la ejecución de la prueba estandarizada.</p> <p>Esta etapa es realizada por el responsable de ECA y del comité diseñador de la prueba.</p>	Reporte del alumno diseñado Reporte del docente diseñado Reporte del coordinador diseñado Reporte del director diseñado
<b>3. Jueceo de los reportes por una muestra de los usuarios finales</b>	<p>Obtener una valoración por parte de los usuarios finales de la prueba para identificar la satisfacción, la usabilidad, comprensión e impacto del reporte en cada uno de ellos.</p> <p>Etapa realizada por el comité diseñador de la prueba</p>	Reporte con las valoraciones, comentarios y observaciones realizadas a cada uno de los reportes.
<b>4.- Evaluación de los reportes técnicos</b>	<p>Se realiza una evaluación técnica sobre la estructura del reporte, la claridad de la información, efectividad y funcionalidad, además de realizar el análisis del jueceo para considerar dichas contribuciones en la evaluación.</p> <p>Etapa realizada por el responsable de ECA y el comité diseñador de la prueba.</p>	Reporte técnico con la evaluación de cada reporte.
<b>5.- Integración de los reportes en el sistema SIEXAES</b>	Los reportes ya validados se integran al sistema computacional SIEXAES para que puedan ser generados automáticamente desde el sistema, facilitando la consulta de información por parte de los usuarios interesados.	Modulo o función del sistema SIEXAS habilitada para generar los reportes.

En esta etapa participa el responsable de ECA, Responsable de desarrollo de software, personal de programación y responsable de administración de la base de datos.

## Vinculación de las etapas en una cadena lógica de resultados



Dirección General de Bibliotecas de la UAQ

<b>ETAPA 1: Análisis de requerimientos y necesidades de los usuarios finales</b>				
<b>OBJETIVO(S):</b>	Identificar la información más relevante, pertinente y de interés para los usuarios finales de la prueba, alumnos, docentes, coordinadores y directivos.			
<b>INDICADORE(S):</b>	Requerimientos identificados			
Actividades	<b>Tiempo</b>	<b>Participantes</b>	<b>Estrategias de seguimiento</b>	<b>Recurso</b>
<b>1.- Diseño y validación de la encuesta para identificación de requerimientos funcionales y de usabilidad</b>	1 día	Responsable de ECA: se encarga de diseñar y validar la encuesta para levantamiento de requerimientos (debe tener experiencia en diseño y buenas prácticas para la construcción de encuestas)  Comité de diseño de la prueba: participan en el diseño de la encuesta (debe tener experiencia en diseño y buenas prácticas para la construcción de encuestas)	Reporte del diseño de la encuesta  Evaluación de la encuesta  Encuesta elaborada	Computadora  Internet  Software de procesador de palabras  Impresora  Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>2.- Determinación de la muestra y aplicación de encuestas con los usuarios finales para levantamiento de requerimientos</b>	1 día	Responsable de ECA: calcula la muestra representativa para aplicar la encuesta (conocimientos y experiencia en cálculo de muestras poblacionales)  Comité de diseño de la prueba: aplican la encuesta para levantamiento de requerimientos (se requiere experiencia en análisis de requerimientos para poder identificar necesidades primarias, secundarias y de usabilidad)	Reporte con el desglose del cálculo de la muestra  Reporte con los participantes entrevistados.  Reporte con los requerimientos funcionales identificados  Reporte con los requerimientos de usabilidad identificados	Personal a encuestar  Disponibilidad de un lugar para realizar las encuestas o los lugares a donde se les ira a entrevistar.  Tablet para capturar las respuestas.  Internet
<b>3.- Redacción de los requerimientos funcionales y de usabilidad</b>	1 día	Comité de diseño de la prueba: redacta con la estructura y el formato adecuado los requerimientos obtenidos con las encuestas (habilidades para redacción correcta de requerimientos funcionales y de usabilidad)	Lista de cotejo con los requerimientos funcionales y de usabilidad identificados (modelo canvas requerimientos por hacer, requerimientos en proceso, requerimientos terminados)	Computadora  Internet  Software de procesador de palabras

<b>4.- Integración de un reporte con los requerimientos identificados y priorizados para cada uno de los diferentes usuarios del examen (docente, alumno, coordinador, director)</b>	4 hrs	Responsable de ECA: realiza la documentación necesaria para integrar en un reporte todos los requerimientos obtenidos, así como organizarlos de acuerdo a su prioridad (experiencia amplia en análisis de requerimientos, redacción de requerimientos, manejo del procesador de palabras o software a fin)	Lista de cotejo con los requerimientos funcionales y de usabilidad integrados en el reporte así como su priorización correspondiente (modelo canvas requerimientos por hacer, requerimientos en proceso, requerimientos terminados)	Computadora Internet Software de procesador de palabras Impresora

Dirección General de Bibliotecas de la UMO



ETAPA 2:		Diseño de los reportes técnicos		
<b>OBJETIVO(S):</b>	Elaboración de los reportes técnicos en los cuales se presentará la información resultante de la ejecución de la prueba estandarizada.			
<b>INDICADORE(S):</b>	Reporte del alumno diseñado Reporte del docente diseñado Reporte del coordinador diseñado Reporte del director diseñado			
Actividades	Tiempo	Participantes	Estrategias de seguimiento	Recurso
<b>1.- Revisión de los requerimientos funcionales y de usabilidad identificados</b>	2 hrs	Responsable de ECA: realiza una revisión del documento – reporte de los requerimientos identificados (debe tener experiencia en análisis de requerimientos para poder identificar necesidades primarias, secundarias y de usabilidad)  Comité de diseño de la prueba: participa revisando requerimientos (debe tener experiencia en análisis de requerimientos para poder identificar necesidades primarias, secundarias y de usabilidad)	Reporte de observaciones y/o comentarios de los requerimientos identificados.	Computadora Internet Software de procesador de palabras Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>2.- Diseño de los reportes técnicos</b>	1 día	Responsable de ECA: diseña, evalúa, revisa, monitorea y asesora el diseño de los reportes técnicos (debe tener experiencia en el manejo del procesador de palabras o software para diseño)  Comité de diseño de la prueba: participan diseñando los reportes (debe tener experiencia en	Lista de cotejo con el estatus de desarrollo de cada reporte.  Lista de cotejo con los reportes a desarrollar (modelo Canvas reportes por hacer, reportes en desarrollo, reportes terminados)	Computadora Internet Software de procesador de palabras Lugar (cubículo, laboratorio etc) para desempeñar la actividad.

		el manejo del procesador de palabras o software para diseño)		
<b>3.- Producción de reporte muestra</b>	1 hr	Comité de diseño de la prueba: producen un reporte con información de muestra para cada uno de los usuarios (debe tener experiencia en el manejo del procesador de palabras o software para diseño)	Lista de cotejo con los reportes muestra a desarrollar (modelo Canvas reportes por hacer, reportes en desarrollo, reportes terminados)	Computadora Internet Software de procesador de palabras Lugar (cubículo, laboratorio etc) para desempeñar la actividad.

Dirección General de Bibliotecas de la UAQ

<b>ETAPA 3: Jueceo de los reportes por una muestra de los usuarios finales</b>				
<b>OBJETIVO(S):</b>	Obtener una valoración por parte de los usuarios finales de la prueba para identificar la satisfacción, la usabilidad, comprensión e impacto del reporte en cada uno de ellos.			
<b>INDICADORE(S):</b>	Reporte con las valoraciones, comentarios y observaciones realizadas a cada uno de los reportes.			
Actividades	<b>Tiempo</b>	<b>Participantes</b>	<b>Estrategias de seguimiento</b>	<b>Recurso</b>
<b>1.- Exposición de los reportes muestra a cada uno de los diferentes usuarios finales de la prueba.</b>	1 hr	Comité de diseño de la prueba: exponen de manera específica a cada uno de los usuarios finales del examen los reportes muestra generados en la etapa anterior (habilidades para el manejo de software de presentaciones)	Presentación digital o impresa de la exposición Lista de participantes de la exposición	Computadora Internet Software de presentación Proyector Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>2.- Recopilar los comentarios, apreciaciones, sugerencias de los usuarios sobre los reportes.</b>	0 – 1 hr	Comité de diseño de la prueba: documentan los comentarios o sugerencias de los reportes por parte de los usuarios (debe tener experiencia en el manejo del procesador de palabras, habilidades para redactar, habilidades para agrupar elementos)	Lista de participantes	Computadora Internet Software de procesador de palabras Proyector Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>3.- Integrar un reporte con la recopilación de los datos del jueceo por cada tipo de reporte</b>	1 hr	Comité de diseño de la prueba: producen un reporte que integra todos los comentarios, observaciones, sugerencias de todos los reportes muestra diseñados (debe tener experiencia en el manejo del procesador de palabras o software para diseño)	Lista de cotejo con la identificación de reportes que obtuvieron comentarios desfavorables. Reporte con las observaciones, comentarios y sugerencias de los participantes, separado por cada tipo de reporte.	Computadora Internet, Software de procesador de palabras Lugar (cubículo, laboratorio etc) para desempeñar la actividad.

<b>ETAPA 4:</b>		<b>Evaluación de los reportes técnicos</b>		
<b>OBJETIVO(S):</b>	Obtener una evaluación técnica sobre la estructura del reporte, la claridad de la información, efectividad y funcionalidad.			
<b>INDICADORE(S):</b>	Reporte técnico con la evaluación de cada reporte (alumno, docente, coordinador, directivo).			
Actividades	Tiempo	Participantes	Estrategias de seguimiento	Recurso
<b>1.- Análisis del reporte de jueceo</b>	2 hr	<p>Responsable de ECA: efectúa una valoración de la información presentada en el reporte de jueceo obtenido en la etapa anterior (debe tener experiencia en jueceo de requerimientos de funcionalidad y usabilidad)</p> <p>Comité de diseño de la prueba: analizan la información del reporte de jueceo, aportan ideas u opiniones (debe tener experiencia en jueceo de requerimientos de funcionalidad y usabilidad)</p>	Reporte de observaciones y/o comentarios de los requerimientos identificados.	<p>Computadora</p> <p>Internet</p> <p>Software de procesador de palabras</p> <p>Lugar (cubículo, laboratorio etc) para desempeñar la actividad.</p>
<b>2.- Evaluación técnica de los reportes</b>	0 – 1 hr	<p>Responsable de ECA: evalúa la calidad técnica del reporte en cuanto su estructura, formato, información, funcionalidad, usabilidad, pertinencia entre otros criterios (debe tener experiencia en diseño y aplicación de reportes técnicos con base en ejecuciones de pruebas estandarizadas)</p> <p>Comité de diseño de la prueba: aportan ideas, opiniones, críticas (experiencia en aplicación de pruebas y recopilación de información)</p>	<p>Reporte con los dictámenes de cada uno de los reportes.</p> <p>Lista de cotejo con los reportes evaluados (modelo canvas reportes por evaluar, reportes en evaluación, reportes evaluados)</p>	<p>Computadora</p> <p>Internet</p> <p>Software de procesador de palabras</p> <p>Lugar (cubículo, laboratorio etc) para desempeñar la actividad.</p>
<b>3.- Correcciones y/o adecuaciones a los reportes</b>	Dependiendo de la naturaleza y el impacto de los cambios puede ir desde 1hr hasta 1 día.	Comité de diseño de la prueba: se encarga de hacer las correcciones necesarias de acuerdo a la evaluación producto de la acción anterior (debe tener experiencia en el manejo del procesador de palabras o software para diseño)	<p>Lista de cotejo con la identificación de reportes que obtuvieron comentarios desfavorables.</p> <p>Lista de cotejo con los reportes corregidos (modelo canvas reportes por corregir, reportes en corrección, reportes corregidos)</p>	<p>Computadora</p> <p>Internet</p> <p>Software de procesador de palabras o diseño</p> <p>Lugar (cubículo, laboratorio etc) para desempeñar la actividad.</p>

<p><b>4.- Producir un documento donde se integren los reportes finales para cada tipo de usuario</b></p>	<p>2 hrs</p>	<p>Comité de diseño de la prueba: realizan la documentación correspondiente de los reportes finales (debe tener experiencia en el manejo del procesador de palabras o software para diseño)</p>	<p>Lista de cotejo con los reportes (modelo canvas reportes integrados por integrar, reportes en integración, reportes integrados)</p>	<p>Computadora Internet Software de procesador de palabras</p>
--	--------------	---	--	--

Dirección General de Bibliotecas de la UANL

**ETAPA 5:****Integración de los reportes en el sistema SIEXAES**

<b>OBJETIVO(S):</b>	Integrar los reportes ya validados al sistema computacional SIEXAES para que puedan ser generados automáticamente, facilitando la consulta de información por parte de los usuarios interesados.			
<b>INDICADORE(S):</b>	Modulo o función del sistema SIEXAS habilitada para generar los reportes.			
<b>Actividades</b>	<b>Tiempo</b>	<b>Participantes</b>	<b>Estrategias de seguimiento</b>	<b>Recurso</b>
<b>1.- Programar los reportes técnicos diseñados en las interfaces del módulo de reportes del sistema Siexaes</b>	3-5 días	Responsable del desarrollo del software: coordina, revisa y monitorea la programación (amplia experiencia en programación, desarrollo de sistemas, uso de módulos de reportes, dirección de proyectos, arquitectura y arquitecto de software)  Responsable de programación: codifica en la plataforma de programación las funcionalidades requeridas (amplia experiencia y experticia en desarrollo de software)	Lista de cotejo con los reportes programados en el sistema (modelo canvas reportes por programar, reportes en programación, reportes ya programados)	Computadora  Internet  Plataforma para el desarrollo de software.  Servidor de Base de datos  Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>2.- Programación de la base de datos para conectividad al módulo de reportes del sistema SIEXAES</b>	6 hr	Responsable de administración de la base de datos: programa la conectividad de los reportes (amplia experiencia y experticia en desarrollo de bases de datos y conectividad con sistemas desarrollados en diferentes lenguajes de programación)	Lista de cotejo con las conectividades de cada módulo programadas modelo canvas reportes por programar, reportes en programación, reportes ya programados)	Computadora  Internet  Plataforma para el desarrollo de software.  Servidor de Base de datos  Lugar para desempeñar la actividad.
<b>3.- Pruebas de funcionalidad</b>	12 hrs	Responsable del desarrollo del software: coordina y monitorea las pruebas (amplia experiencia en programación, desarrollo de sistemas, pruebas de software, arquitectura y arquitecto de software)  Responsable de pruebas: ejecuta el plan de pruebas del software (amplia experiencia y experticia en desarrollo y pruebas de software)	Reporte con las situaciones presentadas durante las pruebas.	Computadora  Internet  Sistema SIEXAES

				Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>4.- Reporte con los resultados de las pruebas</b>	2 hrs	Responsable del desarrollo del software: documenta en un reporte los resultados de la ejecución del plan de pruebas (amplia experiencia en programación, desarrollo de sistemas, pruebas de software, arquitectura y arquitecto de software, documentación)	<p>Lista con los sucesos presentados en las pruebas de cada uno de los reportes.</p> <p>Lista con los dictámenes de cada reporte.</p>	<p>Computadora</p> <p>Internet</p> <p>Software de procesador de palabras</p> <p>Lugar) para desempeñar la actividad.</p>
<b>5.- Correcciones y/o adecuaciones a los reportes en caso de ser necesario</b>	Según sea la corrección, desde 1hr hasta 2 días	<p>Responsable del desarrollo del software: coordinar y monitorear las correcciones.</p> <p>Responsable de programación: programar las correcciones.</p> <p>Responsable de administración de la base de datos: programar las correcciones.</p>	Lista de cotejo con los reportes corregidos (modelo canvas reportes por corregir, reportes en corrección, reportes corregidos)	<p>Computadora</p> <p>Plataforma para el desarrollo de software.</p> <p>Servidor de Base de datos</p> <p>Lugar (cubículo, laboratorio etc) para desempeñar la actividad.</p>

Estrategia #4: *Formar docentes con habilidades y competencias en evaluación del aprendizaje (institucionalizar curso de capacitación docente).*

#### Definición de las etapas

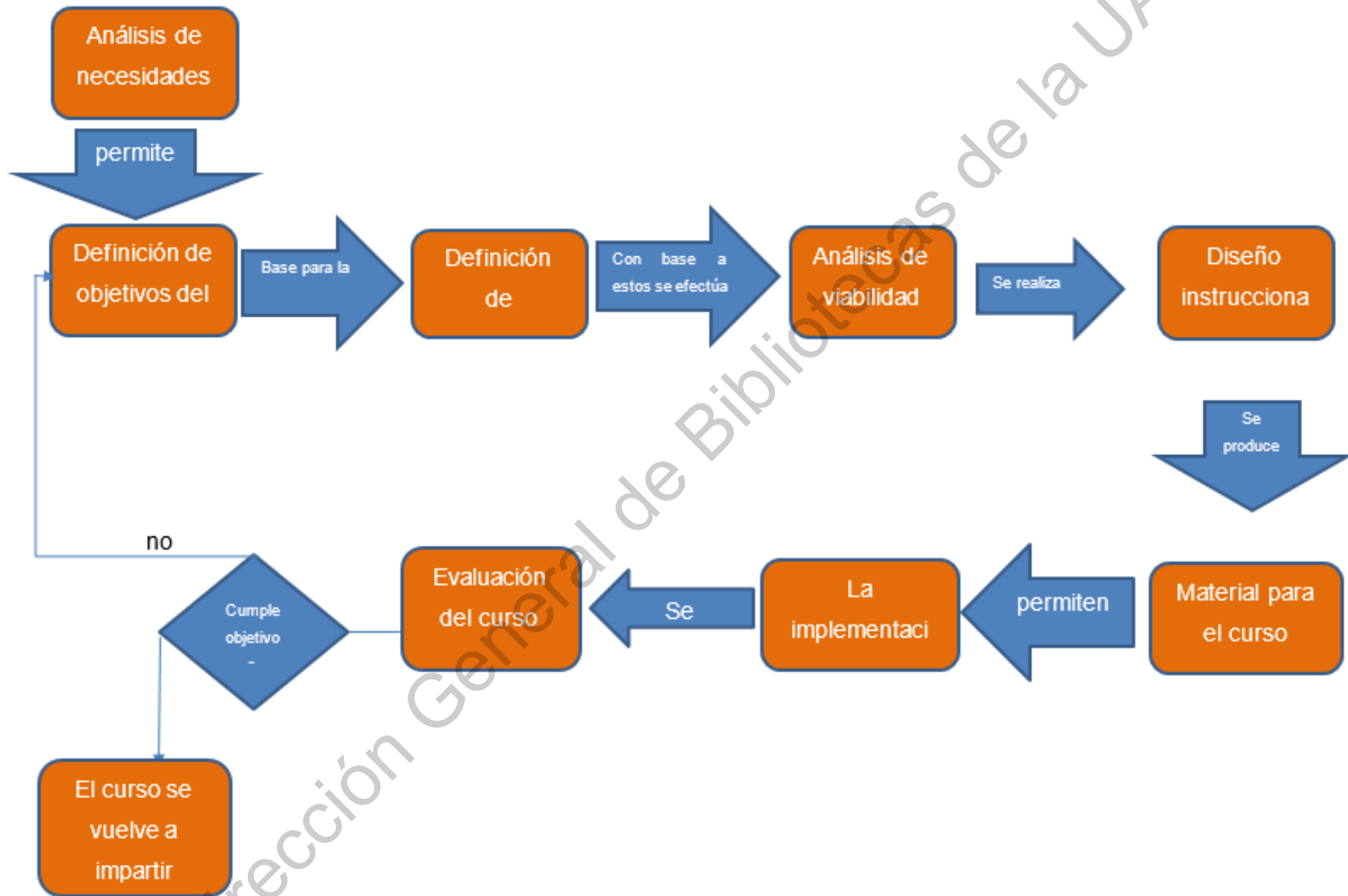
ETAPA	OBJETIVO(S)	INDICADORES DE LOGRO
<b>1. Definición de objetivos</b>	<p>Se trata de identificar específicamente cuales son los objetivos que se pretenden alcanzar con el curso de capacitación.</p> <p>Esta etapa se encarga de desarrollarla el responsable de ECA</p>	Objetivos del curso definidos
<b>2. Determinar contenidos</b>	<p>Determinar claramente cuáles son los contenidos temáticos que se abordaran en el curso de capacitación y que estos estén relacionados con los objetivos del curso.</p> <p>Esta etapa se encarga de desarrollarla el responsable de ECA</p>	Mapa del contenido temático a incluirse en el curso.
<b>3. Análisis de viabilidad (medios y recursos)</b>	<p>Determinar si se cuentan con los recursos tecnológicos, de infraestructura, económicos, materiales, disponibilidad entre otros para la realización del curso de capacitación de la manera más adecuada.</p> <p>Esta etapa se encarga de desarrollarla el responsable de ECA</p>	Lista de cotejo con la valoración de los requisitos mínimos para la puesta en práctica del curso.
<b>4. Diseño instruccional</b>	<p>Crear el ambiente de aprendizaje necesario, para que los docentes desarrollen las capacidades necesarias para lograr las tareas específicas del curso.</p> <p>Esta etapa se encarga de desarrollarla el responsable de ECA</p>	<p>Modelo de diseño instruccional</p> <p>Plan del diseño instruccional</p>
<b>5. Desarrollo de materiales</b>	<p>Elaborar los materiales necesarios para facilitar y apoyar el proceso de aprendizaje e instrucción en el curso de capacitación.</p> <p>Esta etapa se encarga de desarrollarla el responsable de ECA</p>	<p>Materiales producidos</p> <p>Evaluación de calidad de los materiales</p>



<b>6.- Implementación</b>	<p>Puesta en acción del curso de capacitación docente</p> <p>Etapa realizada por el responsable de ECA</p>	<p>Reporte técnico de las incidencias, situaciones presentadas y actividades realizadas durante la ejecución del curso.</p>
<b>7.- Evaluación del curso</b>	<p>Establecer la relación entre los objetivos del curso y el logro de estos. Determinar la calidad general del curso.</p> <p>Etapa realizada por el responsable de ECA</p>	<p>Numero de objetivos completados.</p> <p>Reacciones</p> <p>Comportamiento</p> <p>Aprendizaje (examen)</p>

Dirección General de Bibliotecas de la UFMQ

## Vinculación de las etapas en una cadena lógica de resultados



Dirección General de Bibliotecas de la UAQ

ETAPA 1:		Definición de objetivos		
<b>OBJETIVO(S):</b>	Identificar específicamente cuales son los objetivos que se pretenden alcanzar con el curso de capacitación.			
<b>INDICADORE(S):</b>	Objetivos del curso definidos			
Actividades	<b>Tiempo</b>	<b>Participantes</b>	<b>Estrategias de seguimiento</b>	<b>Recurso</b>
<b>1.- Diseño y validación de la encuesta para identificación de necesidades de la planta docente</b>	1 día	Responsable de ECA: se encarga de diseñar y validar la encuesta para recopilación de información (debe tener experiencia en diseño y buenas prácticas para la construcción de encuestas)	Reporte del diseño de la encuesta Evaluación de la encuesta Encuesta elaborada	Computadora Internet Software de procesador de palabras Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>2.- Aplicación de encuesta y análisis de necesidades de la planta docente</b>	1 día	Responsable de ECA: se encarga de aplicar las encuestas para recopilar la información necesaria con la cual efectuar un análisis de necesidades de la planta docente en cuanto su formación en el área de evaluación estandarizada y de calidad del aprendizaje (debe tener experiencia en análisis de intervenciones educativas, evaluación)	Reporte del análisis de necesidades Lista con las 3 necesidades más relevantes o mayor interés por parte de la planta docente	Personal a encuestar Disponibilidad de un lugar para realizar las encuestas o los lugares a donde se les ira a entrevistar. Tablet para capturar las respuestas. Internet
<b>3.- Determinar los objetivos del curso</b>	4 hrs	Responsable de ECA: analiza y codifica las necesidades detectadas en objetivos del curso, apoyado en el producto de la acción anterior ( se requiere experiencia en detección y valoración de áreas de oportunidad en contextos	Reporte de codificación de necesidades en objetivos	Computadora Internet Software de procesador de palabras Lugar (cubículo, laboratorio etc)

		educativos específicamente del área de evaluación)		para desempeñar la actividad
<b>4.- Documentar los objetivos planeados</b>	2 hrs	Responsable de ECA: elabora reporte con la lista de los objetivos principales del curso ( se requieren habilidades de redacción y manejo de la aplicación de procesador de palabras)	Lista de cotejo con los objetivos del curso identificados y correctamente redactados (modelo canvas, objetivos por documentar, objetivos en documentación, objetivos documentados).	Computadora Internet Software de procesador de palabras

Dirección General de Bibliotecas de la UAQ

**ETAPA 2:****Determinar contenidos**

<b>OBJETIVO(S):</b>	Identificar cuáles son los contenidos temáticos que se abordaran en el curso de capacitación y que estos estén relacionados con los objetivos del curso.			
<b>INDICADORE(S):</b>	Mapa del contenido temático a incluirse y la relación con los objetivos del curso.			
Actividades	<b>Tiempo</b>	<b>Participantes</b>	<b>Estrategias de seguimiento</b>	<b>Recurso</b>
<b>1.- Análisis de contenido temático asociado a los objetivos del curso</b>	6 hrs	Responsable de ECA: se encarga de seleccionar los contenidos de acuerdo a los criterios de nivel, claridad, grado de profundidad, funcionalidad y actualidad (debe tener experiencia valoración y evaluación de pertinencia de contenidos temáticos para la enseñanza - aprendizaje)	Reporte de evaluación de los contenidos temáticos con base en cada uno de los criterios	Computadora Internet Software de procesador de palabras Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>2.- Definición de los contenidos temáticos a incluir</b>	2 hrs	Responsable de ECA: con base en el análisis de la acción anterior se decide cuáles son los contenidos más apropiados para incluir en el curso en pro del logro de los objetivos planeados (debe tener experiencia en análisis de intervenciones educativas, evaluación)	Lista de cotejo con los contenidos temáticos incluidos por objetivo del curso  Mapa de contenido temático a incluir	Computadora Internet Software de procesador de palabras o de diseño Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>3.- Representación gráfica de los contenidos temáticos en relación a los objetivos</b>	2 hrs	Responsable de ECA: produce un diagrama o representación gráfica que permita identificar de manera clara como cada contenido temático está asociado al cumplimiento de un(os) objetivo(s) del curso ( se requiere experiencia en el manejo de software de diseño)	Reporte o documentación que incluya el diagrama o representación gráfica de la relación entre contenidos y objetivos.  Lista de cotejo de la relación entre contenidos	Computadora Internet Software de procesador de palabras Lugar (cubículo, laboratorio etc)

			temáticos y objetivo del curso	para desempeñar la actividad
<b>4.- Reporte con la definición de contenidos y justificación de los mismos</b>	4 hrs	Responsable de ECA: elabora reporte con la lista de los contenidos temáticos del curso y detalla la justificación de la inclusión de estos (se requiere experiencia en la elaboración de justificaciones de contenidos para cursos, asignaturas, pruebas etc)	Lista de cotejo con los contenidos del curso definidos y correctamente justificados (modelo canvas, contenidos por justificar, contenido en justificación, contenido justificado).	Computadora Internet Software de procesador de palabras de

Dirección General de Bibliotecas de la UAQ

**ETAPA 3:****Análisis de viabilidad (medios y recursos)**

<b>OBJETIVO(S):</b>	Determinar si se cuentan con los recursos tecnológicos, de infraestructura, económicos, materiales, disponibilidad entre otros para la realización del curso de capacitación de la manera más adecuada.			
<b>INDICADORE(S):</b>	Lista de cotejo con la valoración de los requisitos mínimos para la puesta en práctica del curso.			
<b>Actividades</b>	<b>Tiempo</b>	<b>Participantes</b>	<b>Estrategias de seguimiento</b>	<b>de Recurso</b>
<b>1.- Análisis de recursos tecnológicos, infraestructura, materiales y disponibilidad para cada contenido temático</b>	2 hrs	Responsable de ECA: se encarga analizar si para los contenidos temáticos que se desean impartir en el curso se cuenta con los recursos necesarios (experiencia en análisis de requerimientos)	Listado con el estatus actual de recursos disponibles para el curso	Computadora Internet Software de procesador de palabras Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>2.- Evaluación de los recursos disponibles</b>	1 hrs	Responsable de ECA: con base en el análisis de la acción anterior se evalúa de manera cuantitativa si se poseen los recursos necesarios mínimos para poder impartir el curso (debe tener experiencia en análisis de intervenciones educativas, evaluación de recursos)	Lista de cotejo con los recursos mínimos necesarios para la implementación del curso	Computadora Internet Software de procesador de palabras o hoja de calculo Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>3.- Reporte con los resultados obtenidos de la evaluación de viabilidad</b>	1 hrs	Responsable de ECA: produce un documento que integra los resultados obtenidos de la evaluación de los recursos y junto con ello el dictamen de la viabilidad de puesta en práctica del curso ( se requiere experiencia en el manejo de software de diseño)	Reporte o documentación que incluye el dictamen del análisis de recursos.  Lista de cotejo de la relación de contenidos temáticos que cuentan con recursos y aquellos que no los disponen.	Computadora Internet Software de procesador de palabras Lugar (cubículo, laboratorio etc) para desempeñar la actividad



ETAPA 4:		Diseño instruccional		
<b>OBJETIVO(S):</b>	Crear el ambiente de aprendizaje necesario, para que los docentes desarrollen las capacidades necesarias para lograr las tareas específicas del curso.			
<b>INDICADORE(S):</b>	Modelo de diseño instruccional Plan del diseño instruccional			
Actividades	Tiempo	Participantes	Estrategias de seguimiento	Recurso
<b>1.- Análisis de diseños instruccionales existentes</b>	1 día	Responsable de ECA: se encarga analizar diferentes modelos de diseño instruccional con mira a elegir el que más se adecue a las necesidades de los participantes del curso (experiencia en análisis de requerimientos, amplio conocimiento y aplicación de diseños instruccionales)	Reporte de metodologías identificadas y analizadas.	Computadora Internet Software de procesador de palabras Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>2.- Elección y justificación del modelo instruccional</b>	4 hrs	Responsable de ECA: con base en el análisis de la acción anterior se evalúan los diferentes modelos instruccionales y se elige el que más se adecue al curso, además de justificar la elección (amplia experiencia en modelos de diseño instruccional y aplicación de los mismos)	Lista de cotejo con la valoración de cada uno de los modelos de diseño instruccional analizados  Justificación del modelo instruccional elegido	Computadora Internet Software de procesador de palabras u hoja de calculo Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>3. Elaboración del diseño instruccional según el modelo elegido</b>	Dependiendo de la naturaleza de las actividades el tiempo puede ir de horas hasta días (máximo 5 días)	Responsable de ECA: elabora el diseño instruccional de acuerdo al modelo elegido, por ejemplo, si se elige el modelo ADDIE las etapas del diseño instruccional serian análisis, diseño, desarrollo, implementación y evaluación. ( se requiere amplia experiencia en	Reporte de estado de cada una de las fases del modelo de diseño instruccional a implementar	Computadora Internet Software de procesador de palabras u hoja de calculo

		modelos de diseño instruccional y aplicación de los mismos)		Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>4.- Ejecución de las etapas del modelo de diseño instruccional</b>	Dependiendo de la naturaleza de las etapas el tiempo puede ir de horas hasta días (máximo 5 días)	<p>Responsable de ECA: tomando como referencia el marco general del modelo ADDIE se encargará de ejecutar las siguientes fases:</p> <ol style="list-style-type: none"> <li>Análisis: analizar el alumnado, el contenido y el entorno, obteniendo como resultado una descripción de un problema y una solución propuesta, el perfil de un alumno y la descripción de las restricciones de los recursos.</li> <li>Diseño: Determinar el enfoque didáctico general, planificando la formación (contenido temático y orden), diseñar las actividades del alumno.</li> <li>Desarrollo: Producción real de los contenidos y materiales de aprendizaje basados en la fase anterior (diseño). Estructurar los planes de las lecciones y los materiales de las mismas. Se desarrollará la instrucción y los medios que serán usados, así como documentos o equipo de apoyo.</li> <li>Implementación: ejecuta y pone en práctica la acción formativa con la participación de los docentes en el curso.</li> <li>Evaluación: formativa (durante las fases para mejorar la instrucción antes de implementar la</li> </ol>	Documentación de las acciones realizadas y resultados obtenidos de cada una de las etapas del diseño instruccional	<p>Computadora</p> <p>Internet</p> <p>Software de procesador de palabras</p> <p>Lugar (cubículo, laboratorio etc) para desempeñar la actividad.</p> <p>Recursos definidos en el diseño instruccional</p>

		versión final y sumativa para valorar la eficacia total de la instrucción)		
--	--	--	--	--

Dirección General de Bibliotecas de la UAQ

<b>ETAPA 5:</b>		<b>Desarrollo de materiales</b>		
<b>OBJETIVO(S):</b>	Elaborar los materiales necesarios para facilitar y apoyar el proceso de aprendizaje e instrucción en el curso de capacitación.			
<b>INDICADORE(S):</b>	Materiales producidos Evaluación de calidad de los materiales			
<b>Actividades</b>	<b>Tiempo</b>	<b>Participantes</b>	<b>Estrategias de seguimiento</b>	<b>de Recurso</b>
<b>1.- Análisis de materiales necesarios para cada contenido temático del curso</b>	2- 4 hrs	Responsable de ECA: se encarga analizar el tipo de recursos materiales que son necesarios para la correcta enseñanza - aprendizaje de los contenidos temáticos que el curso incluye (experiencia en análisis de requerimientos materiales para determinadas estrategias de enseñanza-aprendizaje)	Lista de cotejo con la identificación de materiales para cada contenido temático	Computadora Internet Software de procesador de palabras u hoja de cálculo. Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>2.- Elaboración de los materiales necesarios</b>	Dependiendo de la naturaleza de las actividades el tiempo puede ir de horas hasta días (máximo 5 días)	Responsable de ECA: con base en el análisis de la acción anterior se encarga de producir o elaborar los materiales necesarios para las estrategias de enseñanza - aprendizaje definidas y representadas en el diseño instruccional (se requiere conocimiento y experiencia en buenas prácticas para la elaboración de materiales).	Lista de cotejo con los recursos materiales elaborados (modelo canvas recursos por desarrollar, recursos en desarrollo y recursos materiales ya elaborados).	Computadora Internet Software de procesador de palabras o hoja de calculo Lugar (cubículo, laboratorio etc) para desempeñar la actividad. Material necesario para los diferentes tipos de recursos a elaborar (cartulinas, software especializado,

				mesas de trabajo etc)
<b>3.- Evaluación de calidad de los materiales elaborados</b>	2-4 hrs	Responsable de ECA: se encarga de efectuar la evaluación de la calidad del material elaborado, considerando los criterios de aspectos técnicos, interactividad, impacto, pertinencia, accesibilidad, aprendizaje producido ( se requiere amplia experiencia en evaluación y producción de materiales didácticos)	Reporte o documentación que incluye el dictamen de la evaluación de los materiales.  Lista de cotejo de los materiales evaluados (modelo canvas material por evaluar, material en evaluación, material ya evaluado)	Computadora  Internet  Software de procesador de palabras  Lugar (cubículo, laboratorio etc) para desempeñar la actividad

Dirección General de Bibliotecas de la UNQ

ETAPA 6:		Implementación		
<b>OBJETIVO(S):</b>	Puesta en acción del curso de capacitación docente			
<b>INDICADORE(S):</b>	Instrumento de evaluación diseñado con criterios definidos. Reporte de evaluación del curso.			
Actividades	Tiempo	Participantes	Estrategias de seguimiento	Recurso
<b>1.- Ejecutar y poner en práctica la acción formativa con la participación de los docentes en el curso</b>	2- 4 hrs	Responsable de ECA: imparte el curso de capacitación de acuerdo a el análisis de necesidades efectuado en la etapa 1, trabajando los contenidos temáticos definidos en la etapa 2, aplicando el diseño instruccional planeado en la etapa 4 y utilizando los recursos elaborados en la etapa 5 (experiencia en cursos de capacitación de formación docente en áreas de evaluación del aprendizaje)	Lista de participantes en la capacitación.	Computadora Internet Proyector Aula o laboratorio Material específico definido en el plan de diseño instruccional.
<b>2.- Elaboración de bitácora de ejecución del curso</b>	1 hr	Responsable de ECA: elabora un documento que registra a manera de bitácora todos aquellos sucesos presentes en la capacitación para ser considerados en futuras repeticiones del curso (se requiere conocimiento y experiencia en buenas prácticas para la elaboración de bitácoras, redacción y manejo del procesador de palabras).	Reporte con las incidencias presentadas durante la ejecución del curso	Computadora Internet Software de procesador de palabras o hoja de calculo Lugar (cubículo, laboratorio etc) para desempeñar la actividad.
<b>3.- Elaboración de reporte de evaluación implementación</b>	2 hrs	Responsable de ECA: se encarga elaborar un documento que incluye cuales, de las actividades incluidas en	Lista de cotejo de las actividades, materiales, recursos y herramientas diseñadas o planeadas	Computadora Internet

		<p>el diseño instruccional, así como los materiales o recursos planeados realmente fueron aplicados. Básicamente se evalúa cuanto de lo diseñado realmente fue puesto en practica.</p>	<p>fueron aplicadas y cuáles no.</p> <p>Reporte de evaluación.</p>	<p>Software de procesador de palabras</p> <p>Lugar (cubículo, laboratorio etc) para desempeñar la actividad</p>
--	--	--	--	---

Dirección General de Bibliotecas de la UAQ

**ETAPA 7:****Evaluación**

<b>OBJETIVO(S):</b>	Establecer la relación entre los objetivos del curso y el logro de estos. Determinar la calidad general del curso.			
<b>INDICADORE(S):</b>	Numero de objetivos completados. Reacciones Comportamiento Aprendizaje (examen)			
Actividades	<b>Tiempo</b>	<b>Participantes</b>	<b>Estrategias de seguimiento</b>	<b>Recurso</b>
<b>1.- Analizar los objetivos completados en la ejecución del curso</b>	1 – 2 hrs	Responsable de ECA: analiza los resultados obtenidos con el curso y determina cuantos de los objetivos planeados fueron alcanzados por los participantes (experiencia evaluación de la relación resultado - objetivo)	Lista de objetivos logrados con el curso	Computadora Internet Software de hoja de cálculo o procesador de palabras
<b>2.- Evaluar las reacciones de los participantes</b>	.5 hr	Responsable de ECA: elabora una breve encuesta de satisfacción con la cual se evalúa el gusto del curso en los participantes (se requiere conocimiento y experiencia en diseño y aplicación de encuestas de satisfacción).	Reporte con el grado de reacción o satisfacción de cada uno de los participantes en el curso.	Computadora Software de procesador de palabras o hoja de calculo Lugar (cubículo, laboratorio etc) para desempeñar la actividad. Tablet
<b>3.- Evaluar el comportamiento de los participantes</b>	.5 hrs	Responsable de ECA: elabora una breve encuesta de comportamiento con la cual se pueda analizar el grado influencia que tuvo el curso en el cambio de conducta o actitudes ante determinadas situaciones asociadas a la evaluación del aprendizaje (se requiere conocimiento y experiencia en diseño y aplicación de encuestas de comportamiento).	Reporte con la evaluación del comportamiento de cada uno de los participantes en el curso.	Computadora Internet Software de procesador de palabras Lugar (cubículo, laboratorio etc) para desempeñar la actividad



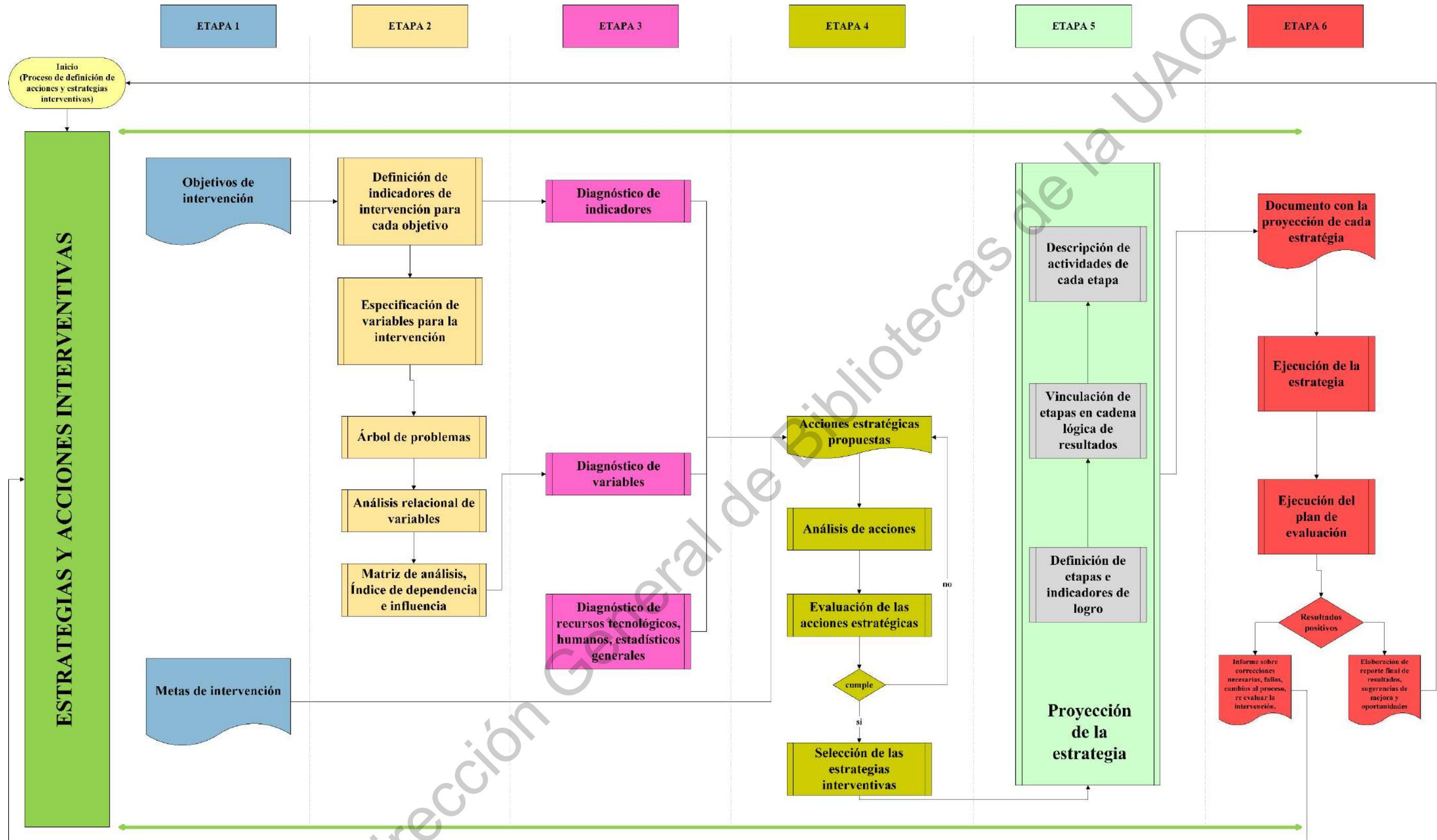
<p><b>4.- Evaluar el aprendizaje de los participantes con una prueba</b></p>	<p>.5 hrs</p>	<p>Responsable de ECA: elabora un breve examen con la finalidad de comprobar si en realidad los participantes aprendieron algo en términos de conocimientos y habilidades específicamente sobre las temáticas abordadas en el curso (se requiere conocimiento y experiencia en diseño y aplicación de evaluaciones del aprendizaje).</p>	<p>Reporte con los resultados de la evaluación del aprendizaje.</p>	<p>Computadora para cada participante</p> <p>Internet</p> <p>Aplicación en la cual realizan el examen</p> <p>Laboratorio de cómputo para desempeñar la actividad</p>
--	---------------	--	---	--

Dirección General de Bibliotecas de la UNAM

**ANEXO 7 Diagramas de subprocesos de sistematización de la  
intervención**

Dirección General de Bibliotecas de la UAQ

## DIAGRAMA DE SISTEMATIZACIÓN DEL SUB PROCESO ESTRATEGIAS Y ACCIONES INTERVENTIVAS



En el diagrama general del proyecto de intervención se observó de manera global la ejecución de esta fase, como se ilustra en la siguiente figura:

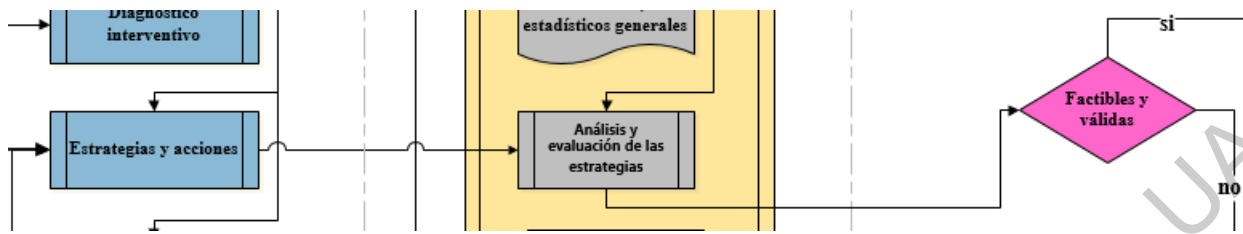


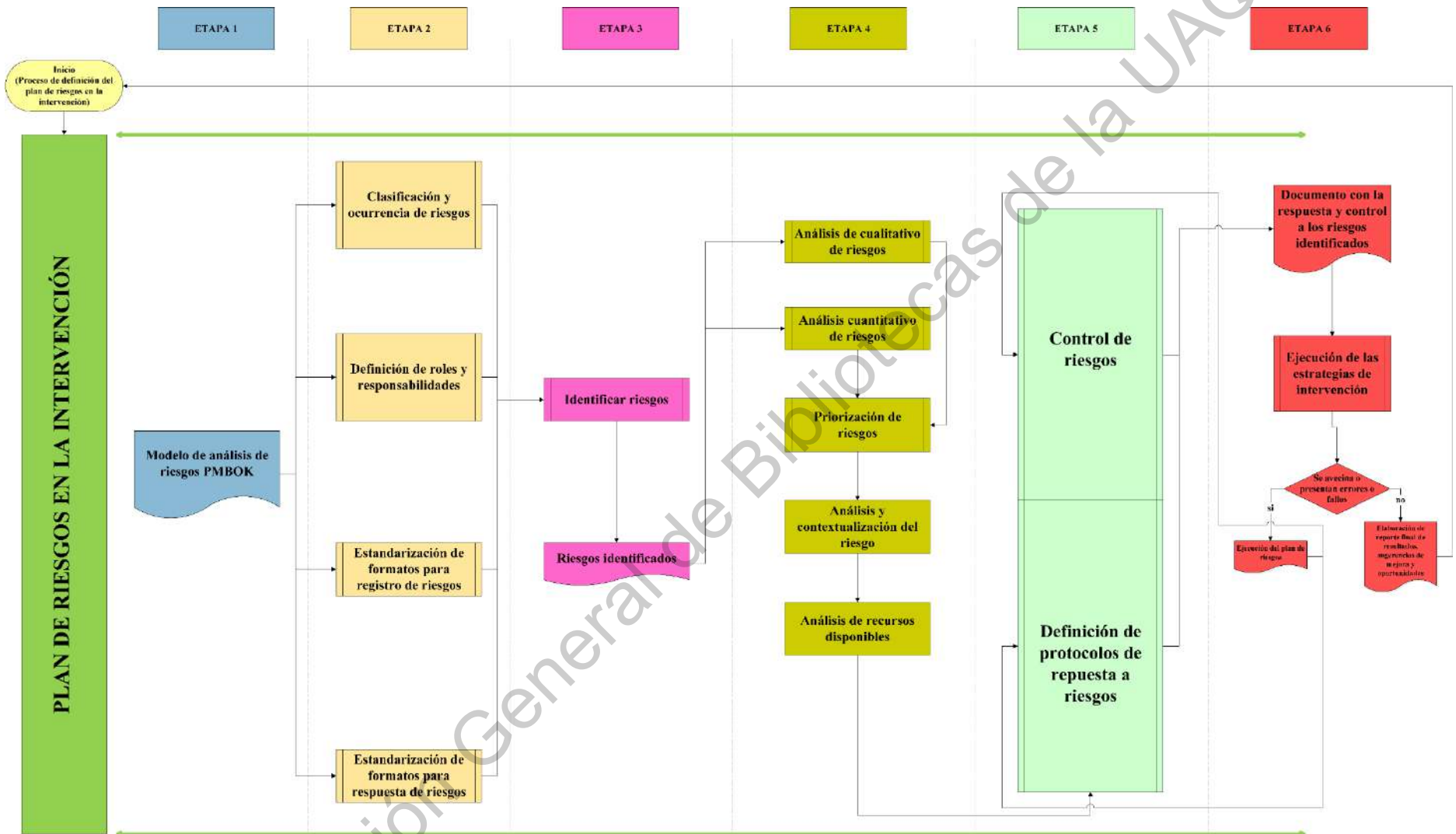
Figura 1. Ilustración del proceso general de Estrategias y Acciones.

Fuente: elaboración propia.

Sin embargo, en el diagrama detallado del proceso se nota como este sub proceso se divide a su vez en varias etapas. Por ejemplo, el proceso de Análisis y evaluación de las estrategias requiere que previamente se efectúen siete actividades, y que a su vez estas actividades requieren como insumo los documentos generados como producto de la ejecución de otras fases del plan de gestión de intervención.

Además, el proceso de la proyección de las estrategias requiere a su vez de tres sub procesos, los cuales tienen como fin producir el documento que integra la proyección de todas las acciones estratégicas que cumplieron con los criterios de evaluación, para posteriormente proceder a su ejecución. Y continuando con el diagrama general de la intervención se pasa a la etapa de evaluación de los resultados de las mismas.

# DIAGRAMA DE SISTEMATIZACIÓN DEL SUB PROCESO PLAN DE RIESGOS DE LA INTERVENCIÓN



Dirección General de Bibliotecas de la UAQ

En el diagrama general del proyecto de intervención se observó de manera global la ejecución de esta fase, como se ilustra en la siguiente figura:

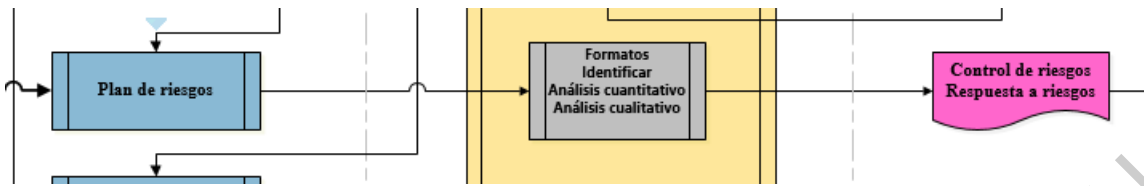


Figura 2. Ilustración del proceso general de Plan de Riesgos de la Intervención.

Fuente: elaboración propia.

Al igual que en la fase de *Estrategias y Acciones*, la fase de Plan de Riesgos requiere a su vez una mayor serie de actividades y sub procesos. Se inicia con un documento como insumo, este es la guía del PMBOK para establecer el plan de riesgos, posteriormente la siguiente etapa se compone de cuatro subprocesos entre los cuales destacan la estandarización de formatos, definición de roles y la clasificación de riesgos. La siguiente etapa en la sistematización implica la ejecución de un sub proceso que consiste en la identificación de riesgos para la intervención y esto genera como producto un documento con los riesgos identificados.

Subsecuentemente en la siguiente etapa se efectúan los procesos más complejos de esta fase, ya que hay que realizar los análisis cuantitativo y cualitativo de los riesgos, su priorización y el análisis de disponibilidad de recursos con los que se cuenta para responder ante los riesgos. Como producto de este proceso se cuenta el documento con las respuestas y el control a los riesgos identificados.

Este documento es insumo para la ejecución del proceso de *Ejecución de las Acciones Estratégicas*, en el caso de presentarse algún error o fallo se ejecuta el proceso de protocolos de respuesta a riesgos, o bien si se prevé alguna situación se ejecuta el proceso de control de riesgos.

**ANEXO 8 Reportes técnicos generados por el sistema  
SIEXAES**

Dirección General de Bibliotecas de la UAQ



## REPORTE DE RESULTADOS DEL ALUMNO

Alumno: n#n#n#

El examen de morfología aplicado contenía 63 ítems. Tu puntaje obtenido fue de:

47 ítems acertados

Obteniendo un promedio (en escala de 0 a 100) general de:

74.60

El promedio general del grupo fue de:

83.78

### Informe detallado por unidad de aprendizaje

Distribución de aciertos	UNIDAD DE APRENDIZAJE 1	Temas de mayor relevancia curricular (IRC)	Temas dominados de mayor IRC	Temas no dominados de mayor IRC
	Ítems acertados: 10/13  Calificación en la unidad: 76.92 /100  Nivel de dominio: Satisfactorio	Ítems acertados: 2/3  Porcentaje de dominio: 66.66%	1.3.1 Morphemes and allomorphs  1.3.3 Roots, affixes	1.2.2 Signs and morphemes



Distribución de aciertos	UNIDAD DE APRENDIZAJE 2	Temas de mayor relevancia curricular (IRC)	Temas dominados de mayor IRC	Temas no dominados de mayor IRC
	<p>Ítems acertados: 17/22</p> <p>Calificación en la unidad: 77.27 /100</p> <p>Nivel de dominio: Satisfactorio</p>	<p>Ítems acertados: 3/5</p> <p>Porcentaje de dominio: 60%</p>	<p>2.5 Nominal inflection</p> <p>2.3 Inflection</p> <p>2.1 Derivation</p>	<p>2.6.1 Tense</p> <p>2.2 Compounding</p>

Dirección General de Bibliotecas de la UAG

## REPORTE DE RESULTADOS DEL DOCENTE



Grupo: n

El examen de morfología aplicado contenía 63 ítems. El promedio de aciertos obtenido por el grupo fue de (promedio de aciertos / aciertos posibles):

42.37 / 63

El promedio (en una escala de 0 a 100) ponderado del grupo fue de:

67.25

A continuación, se presenta el informe detallado:

Alumno	Aciertos Totales	Promedio Ponderado
Ejemplo 1	40	63.49
Ejemplo 2	54	85.71
Ejemplo n	32	50.79

**Unidad(es) con mejor rendimiento (promedio de aciertos / aciertos posibles):**

Unidad 1: 9.68 / 13      Unidad 3: 9.42 / 12

**Unidad(es) con rendimiento insuficiente:**

Unidad 5: 5.53 / 10

**Contenido Temático con mayor insuficiencia (aciertos / aciertos posibles):**

Unidad 1	1.3.2 History	15/36
Unidad 2	2.6.2 Aspect	9/18
Unidad 5	5.1 Compounds versus phrases	1/18
Unidad 5	5.2 Compound Verbs	20/36
Unidad 5	5.5 Headed and headless compounds	9/18

**Comportamiento grupal en los temas de mayor Índice de relevancia curricular**

Contenido de mayor IRC	Promedio aciertos
2.5 Nominal inflection	n/2
2.3 Inflection	n/3
1.3.1 Morphemes and allomorphs	n/2
3.1 Nouns	n/3

3.2 Adjectives	n/3
2.1 Derivation	n/2
2.6.1 Tense	n/3
3.4 Verbs	n/3
2.2 Compounding	n/2

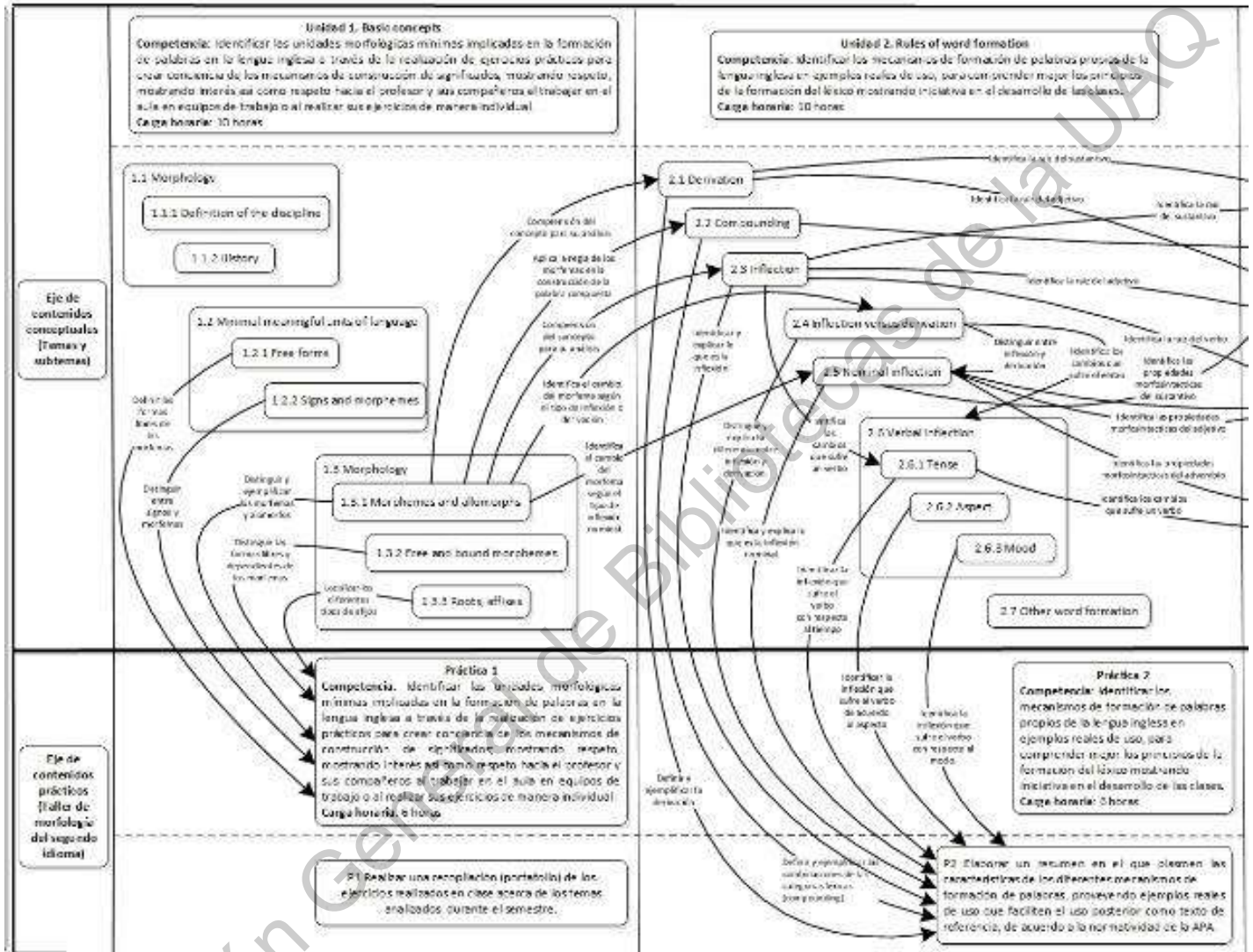
**Comportamiento grupal en todo el universo de medida (promedios):**

Contenido temático	Promedio
1.1.1 Definition of the discipline	0.00
1.1.2 History	0.00
1.2.1 Free forms	0.00
1.2.2 Signs and morphemes	0.00
1.3.1 Morphemes and allomorphs	0.00
1.3.2 Free and bound morphemes	0.00
1.3.3 Roots, affixes	0.00
2.1 Derivation	0.00
2.2 Compounding	0.00
2.3 Inflection	0.00
2.4 Inflection versus derivation	0.00
2.5 Nominal inflection	0.00
2.6.1 Tense	0.00
2.6.2 Aspect	0.00
2.6.3 Mood	0.00
2.7 Other word formation	0.00
3.1 Nouns	0.00
3.2 Adjectives	0.00
3.3 Adverbs	0.00
3.4 Verbs	0.00
4.1 Pronouns	0.00
4.2 Prepositions	0.00
4.3 Conjunctions	0.00
4.4 Interjections	0.00
5.1 Compounds versus phrases	0.00
5.2 Compound verbs	0.00
5.3 Compound adjectives	0.00
5.4 Compound nouns	0.00
5.5 Headed and headless compounds	0.00
5.6 Blends and acronyms	0.00
5.7 Phrasal words	0.00

**ANEXO 9 Retícula de la asignatura morfología de la segunda  
lengua**

Dirección General de Bibliotecas de la UAQ

Debido a que el instrumento es muy extenso, solo se muestran dos unidades temáticas, sin embargo, en este mismo formato se observan las cinco unidades interactuando entre sí, además de que hay una descripción de la asignatura y su pertinencia en el plan de estudios.



Dirección General de Bibliotecas de la UNQ