

**Universidad Nacional Autónoma de México**

**Facultad de Psicología**

**Propuesta técnica para la Prestación de los Servicios de Diseño, Desarrollo, Validación, Aplicación, Calificación y Análisis de las valoraciones relativas a los Procesos de Selección para la Admisión y la Promoción Vertical en Educación Básica y Educación Media Superior y la Promoción Horizontal en Educación Básica en el marco del Sistema para la Carrera de las Maestras y los Maestros.**

**Elaborada por:**

Dr. Germán Palafox Palafox, Dr. Óscar Zamora Arévalo, Dr. Iwin Leenen, Dra. Georgina Rodríguez García, Ing. David Huerta Saavedra, Dr. Hugo Iván Aburto Pineda, Mtra. María de la Luz Ortíz González, Mtra. Eleonora Rubio Ruiz, Lic. Adriana F. Chávez De la Peña

Coyoacán, Ciudad de México a XX de enero de 2020

**Contenido**

|  |  |
| --- | --- |
| 1. **Introducción** | **3** |
| 1. **Propuesta metodológica** | **5** |
| 2.1 Características generales de la propuesta metodológica | 5 |
| 2.2 Enfoques metodológicos para el desarrollo y calificación de instrumentos | 6 |
| 2.2.1 Modelos de Diagnóstico Cognitivo | 6 |
| 2.2.2 Modelos derivados de la Teoría de Respuesta al Ítem | 9 |
| 2.2.3 Redes Neuronales y algoritmos de aprendizaje profundo | 11 |
| 2.3 Desarrollo del Sistema informático para la administración de instrumentos | 15 |
| 2.3.1 Desarrollo de un Banco de Ítems | 21 |
| 1. **Plan de acción** | **25** |
| 3.1 Fases para el diseño, validación, aplicación y calificación de pruebas | 28 |
| 3.2 Operación del sistema informático | 39 |
| 3.3 Plan general de trabajo para el desarrollo de la aplicación en campo | 41 |
| 3.3.1 Estructura operativa | 41 |
| 3.3.2 Responsabilidades del equipo encargado del operativo en campo | 44 |
| 3.3.3 Momentos clave | 45 |
| 1. **Planteamiento de valor agregado** | **49** |
| 1. **Anexos** | **50** |
| Anexo A: Algoritmos de aprendizaje profundo | 50 |
| Anexo B: Algoritmos del Sistema de Predicción de Calidad | 54 |
| Anexo C: Requerimientos mínimos de Hardware para la aplicación | 55 |
| Anexo D: Propuesta de capacitación para el personal técnico de la USICAMM | 57 |
| Anexo E: Perfiles y funciones del personal encargado del operativo en campo | 61 |
| 1. **Referencias** | **65** |

**I. Introducción**

La Facultad de Psicología de la Universidad Nacional Autónoma de México cuenta con un robusto cuerpo de investigadores, expertos en materia de psicometría y evaluación, con una amplia experiencia en el desarrollo y validación de instrumentos de medición a gran escala y de alto impacto, tanto en el ámbito educativo como en el estudio de variables psicológicas latentes complejas.

Fundada en 1973, la Facultad de Psicología se ha mantenido en constante proceso de evolución para estar al día con las líneas de investigación actualmente en desarrollo, implementando estrategias y enfoques metodológicos y estadísticos de vanguardia que permitan atender las necesidades sociales y de formación señaladas por el contexto. Su basta trayectoria en la formación de psicólogos con bases sólidas en estadística y metodología, así como su participación en proyectos colaborativos, orientados a favorecer la toma de decisiones informada, con base en resultados obtenidos en diversos tipos de instrumentos, dan prueba de la capacidad que tiene la Facultad para coordinar múltiples equipos especializados de trabajo. Lo anterior permite de manera eficaz y cabal la gestión de los recursos necesarios para cumplir con las actividades requeridas para el diseño, desarrollo, validación, aplicación, calificación y análisis de los instrumentos solicitados para la valoración de las distintas figuras educativas en cada uno de los Procesos de selección enmarcados por el Sistema de Carrera de las Maestras y los Maestros (SCMM).

La Facultad de Psicología es considerada como un referente, tanto a nivel nacional como internacional, de amplio rigor metodológico. El estricto apego a los estándares de calidad más altos y vanguardistas, a nivel internacional que presentan las investigaciones y proyectos elaborados en dicha Facultad, han conferido a su cuerpo académico y de investigadores un gran respeto y reconocimiento por parte de la comunidad académica y científica.

Para garantizar el cumplimiento de los más altos estándares de calidad técnica y sustantiva en el diseño, desarrollo, validación, aplicación, calificación y análisis de resultados a obtener a lo largo de los distintos instrumentos que se requiere trabajar como parte del Sistema de Apreciación de las aptitudes y conocimientos docentes (SISAP), la presente propuesta técnica considera la colaboración periférica de expertos nacionales e internacionales en materia de evaluación, desarrollo de pruebas y psicometría.

Es a raíz de todo lo previamente expuesto que los investigadores de la Facultad de Psicología de la UNAM se saben y sienten enteramente preparados para atender, en tiempo y forma, siguiendo los más altos estándares de calidad técnica, estadística y metodológica, a la solicitud presentada en el documento [**TITULO DEL DOCUMENTO]**, siendo conscientes del enorme privilegio que representa el poder contribuir a la mejora constante y valoración de la práctica docente en el país.

**II. Propuesta metodológica, enfoque y plan de acción**

**2.1 Características generales de la propuesta metodológica**

Se conformará un equipo de trabajo central, cuyo espacio de trabajo estará ubicado en las instalaciones de la Facultad de Psicología. Para efectos del trabajo a realizar durante el primer semestre del año 2020, se identifican tres grandes tareas:

1. La primera, tiene que ver con la planificación, coordinación, dirección y supervisión de los procesos relacionados con la aplicación de los instrumentos proporcionados por la USICAMM para los procesos de selección para la Admisión y la Promoción Vertical en los meses de mayo y junio del 2020.
2. La segunda tarea refiere al diseño, desarrollo y mantenimiento de los instrumentos de apreciación del SISAP que serán utilizados para la toma de decisiones en los procesos de selección para la Promoción Horizontal y el Reconocimiento, en el segundo semestre del 2020, así como los procesos de selección para la Admisión y Promoción Vertical que se llevarán a cabo a partir del 2021.
3. La tercera tarea consiste en el diseño y desarrollo de un sistema informático que permita el almacenamiento, la aplicación y la administración general de los instrumentos elaborados como parte del SISAP, así como el resguardo y calificación automática de las respuestas registradas por los sustentantes para la producción de reportes de resultados individualizados.

A continuación se presentan con detalle las características y enfoques metodológicos a emplear para el desarrollo de las últimas dos tareas mencionadas (el desarrollo de nuevos instrumentos de apreciación y el desarrollo del sistema informático). La información correspondiente a las estrategias y protocolos a seguir, para los diversos procesos de aplicación comprometidos en la solicitud presentada, se expone en el apartado referente al Plan de acción.

**2.2 Enfoques metodológicos para el desarrollo y calificación de instrumentos**

A fin de atender las necesidades señaladas en el documento [*Título de la Convocatoria*], se plantean tres distintas propuestas metodológicas que guiarán los procesos involucrados en el diseño, desarrollo, validación, aplicación, calificación y análisis de las pruebas solicitadas para la valoración de las habilidades, conocimientos y aptitudes identificadas como parte del quehacer de las diversas figuras educativas; mismas que serán ponderadas para la toma de decisiones en los procesos de selección para la admisión a las funciones docente y técnico docente en educación básica y media superior, la promoción vertical en educación básica y media superior y la promoción horizontal en educación básica.

Las tres propuestas metodológicas se plantean desde los más altos estándares de calidad técnica, procurando poner al alcance del SCMM mejoras e innovaciones que incidan en la reducción de errores en el diseño, desarrollo, validación, aplicación, calificación y análisis de los resultados de los instrumentos, y que permitan una visión a futuro de mejora progresiva, al buscar promover la estandarización de procesos de aplicación y captura de los resultados obtenidos en los instrumentos, así como a la automatización e inmediatez de su calificación, análisis y reporte. Se propone utilizar el modelo para la Evaluación del Diseño Universal (EDU), propuesto por Thompson, Johnstone y Thurlow (2002), como un referente para garantizar que los instrumentos a desarrollar cumplan con los estándares necesarios de accesibilidad (Johnstone, 2003) y para minimizar la varianza irrelevante del constructo, típicamente atribuida a problemas en el diseño del instrumento, su formato o a cualquier sesgo cultural que pueda presentarse en los ítems (Haladyna, Downing y Rodríguez, 2002).

*2.2.1 Modelos de Diagnóstico Cognitivo*

La primera propuesta metodológica se desarrolla en el marco de los Modelos de Diagnóstico Cognitivo, que constituyen una familia de modelos estadísticos que permiten la extracción de un diagnóstico detallado sobre el grado de dominio que presenta cada sustentante a lo largo de las llamadas “operaciones cognitivas” (en este caso habilidades, conocimientos y aptitudes), requeridas para resolver el conjunto de ítems que componen un instrumento (Lee y Sawaki, 2009). Este enfoque metodológico promueve la medición de estructuras específicas de conocimiento y habilidades de procesamiento, a fin de poder brindar a los sustentantes retroalimentación detallada sobre sus fortalezas y áreas de mejora específicas, lo que permite así el trazo de perfiles individualizados que orientan la toma de decisiones informada para la formación continua.

Los MDC tienen su origen y pueden ser expresados como una variación del Modelo general de clases latentes (Templin y Henson, 2006; von Davier, 2009), siendo identificados en la literatura como Modelos de variables latentes multidimensionales confirmatorias y probabilísticas con una estructura factorial compleja (Rupp y Templin, 2008). Son modelos probabilísticos en el sentido de que cada factor define el dominio que cada sustentante presenta a lo largo de las distintas habilidades o conocimientos evaluados como una probabilidad (Lee y Sawaki, 2009).

Al permitir la medición específica de distintos elementos dentro del objeto de estudio, los MDC abandonan la perspectiva unidimensional adoptada por la Teoría Clásica de los Test y por la mayoría de los modelos desarrollados bajo la Teoría de Respuesta al Ítem, paradigmas predominantes de la Psicometría que suelen interpretar los resultados obtenidos en un instrumento como reflejo de la posición que ocupan los sustentantes a lo largo de un solo continuo, que representa una única variable o habilidad latente evaluada (de la Torre, 2009). Por su parte, los MDC desglosan la interpretación de los resultados a lo largo de las distintas habilidades, conocimientos y estrategias que se asumen como centrales al objeto de estudio, permitiendo hacer inferencias precisas acerca de cuáles de estos elementos se domina y cuáles hace falta reforzar.

De esta forma, los MDC permiten a todo agente interesado en la aplicación de un instrumento conocer con detalle las áreas de fortaleza y áreas de oportunidad identificadas para orientar el desarrollo informado de estrategias de mejora (Jang, 2008). En el ámbito educativo, la aplicación de estos modelos tiene el potencial de permitir a los agentes educativos involucrados tomar mejores decisiones con respecto al diseño, uso, manejo e implementación de recursos y materiales didácticos y de apoyo para la mejora del aprendizaje de los estudiantes y para la formación continua de las figuras educativas (ejemplos concretos de ello se pueden encontrar en Ketterlin-Geller & Yovanoff, 2009 y Pérez-Morán, Vázquez-Lira & Rojas, 2019). En la actualidad, se pueden encontrar ejemplos de la aplicación de MDC al ámbito de las matemáticas (Brown y Burton, 1978; Chen y Macdonald, 2011; Gierl, Leighton, Changjiang, Jiawen, Rebecca & Tan, 2009; Ma, Çetin y Green, 2009; Pérez-Morán, 2014; Pérez-Morán, Contreras-Roldan, Hernández, Olivares, Chan, y Díaz, 2014; Revuelta y Ponsoda, 1998; Romero, Ponsoda y Ximénez, 2008; Birenbaum & Tatsuoka, 1993 ; Ketterlin-Geller & Yovanoff, 2009), la comprensión lectora (Lee & Sawaki, 2009, Jang, 2009; Li, 2011; Li, Hunter & Lei, 2016; Ravand, 2016) y algunas otras evaluaciones generales (Montero, Monfils, Wang, Yen Julian & Moody, 2003).

Los modelos de diagnóstico cognitivo requieren precisar la estructura sustantiva de toda prueba mediante la construcción de lo que se conoce en la literatura como Matriz Q (de la Torre, 2009), al presentar la estructura factorial del instrumento identificando cuáles son las habilidades o conocimientos requeridos por cada ítem contenido en la prueba (Li, 2011). La Matriz Q contiene tantas filas como ítems haya en la prueba y tantas columnas como habilidades o conocimientos se requiera a los sustentantes (Tatsuoka, 1990; Chiu y Douglas, 2013; Templin y Henson, 2006; von Davier, 2005, 2008). Su construcción requiere del trabajo conjunto de expertos en el dominio evaluado, sustentantes que participen en un piloteo activo y reflexivo, que permita identificar los procedimientos seguidos, y de psicómetras que estén constantemente cotejando la información recopilada a la luz de las respuestas observadas (Tatsuoka, 1990). En este sentido, los MDC comparten con los modelos de Análisis Factorial Confirmatorio (AFC) una naturaleza confirmatoria, ya que las variables latentes u operaciones cognitivas que lo componen se definen *a priori* y su relación con cada uno de los ítems queda trazada en la Matriz Q.

Existe una amplia variedad dentro de los MDC (Cohen, 2019; Ferrara, Lai, Reilly, Nichols, Rupp y Leighton, 2017; van der Linden, 2016), en términos de los supuestos específicos que postula cada modelo en torno a la descripción de cómo interactúan o se pondera el dominio de las distintas habilidades y/o conocimientos requeridos por la prueba para producir un acierto o error en cada ítem. Por ejemplo, una distinción importante tiene que ver si el modelo es conjuntivo o disyuntivo (Rupp, Templin y Henson, 2010): los modelos conjuntivos asumen que se requiere el dominio de todos los atributos asociados al ítem para poder obtener un acierto, mientras que los modelos disyuntivos asumen que la falta de dominio de un atributo puede ser compensada por el dominio de otros atributos.

El modelo DINA constituye uno de los modelos conjuntivos más sencillos dentro de la familia de los MDC (Junker y Sijtsma, 2001; de la Torre, 2009). Cuenta únicamente con dos parámetros libres que describen para cada ítem la probabilidad de que los aciertos o errores registrados no estén relacionados con el grado de dominio que los sustentantes tienen en las habilidades requeridas. Estos parámetros, conocidos como de adivinación y desliz, cuantifican la probabilidad de obtener un acierto aún sin dominar las habilidades necesarias, es decir, de “atinarle” a la respuesta correcta; y la probabilidad de errar el ítem aún dominando las habilidades necesarias, (en otras palabras, de cometer un “desliz” al momento de seleccionar la respuesta), respectivamente.

El modelo DINA se expresa a partir de la siguiente ecuación,

(1)

donde es una variable binaria que indica si la persona obtuvo un acierto (1) o un error (0) en el ítem ; es un vector que contiene para toda persona los valores de la variable binaria que señala si la domina (1) o no (0) el subdominio (habilidad o conocimiento) evaluado en la prueba; es un vector que por cada ítem contiene a la variable binaria que señala la respuesta “esperada” por el modelo (acierto o error), para el sustentante en el ítem , dado lo que la matriz Q (, el modelo cognitivo detrás de la prueba, ha establecido acerca de qué habilidades se requieren para obtener un acierto en cada ítem y lo que el vector nos dice sobre las habilidades dominadas por el participante . El parámetro de desliz , describe la probabilidad de que aún si la respuesta esperada para el participante al ítem sea 1 (), el participante cometa un “desliz” (; ). Por su parte, el parámetro de adivinación , representa la probabilidad de que un examinado que no posee todas las habilidades requeridas por el ítem (), “adivine” la respuesta correcta

(

De acuerdo con el modelo DINA, lo examinados que dominan todos los atributos requeridos por cada ítem tienen una probabilidad de acierto igual a , mientras que el resto tiene una probabilidad de acertar igual a .

*2.2.2 Modelos de la Teoría de Respuesta al Ítem*

La segunda propuesta metodológica implica el uso de los principios derivados de la Teoría de Respuesta al Ítem, para orientar el diseño, desarrollo y validación de las escalas planteadas para valorar percepciones y actitudes, utilizando el Modelo de Respuesta Graduada para la calificación y análisis de las respuestas registradas en dichos instrumentos, así como para nutrir la integración del reporte de resultados que se entregará de manera individualizada y automatizada a cada sustentante.

La Teoría de Respuesta al Ítem (TRI) constituye uno de los enfoques psicométricos más reconocidos y desarrollados para el diseño, construcción y calificación de instrumentos de medición. Se distingue de la Teoría Clásica de los Test en que toma a los ítems como unidad de análisis, lo que permite que las propiedades psicométricas del instrumento puedan ser descritas a partir de parámetros invariantes que no dependen de la muestra en la que se apliquen (Abad, 2011).

Existe una gran variedad de modelos que han sido desarrollados a partir de los supuestos planteados por la TRI. Cada modelo presenta su propia estructura matemática, a partir de la cual se describe una teoría sobre el comportamiento probabilístico con que cada participante responda a cada uno de los ítems. En otras palabras, cada modelo describe matemáticamente la probabilidad de que cualquier participante , responda al ítem en cualquiera de las posibles categorías de respuesta . Se distingue así entre los modelos dicotómicos, que admiten únicamente dos categorías de respuesta para cada ítem (0,1; típicamente interpretados como “incorrecto” o “correcto”, o bien, como la “ausencia” o “presencia” señalada de un elemento) y los modelos politómicos, que admiten más de dos categorías de respuesta y que típicamente se presentan en forma de una escala con niveles ordenados de respuesta (Abad, 2011). En este mismo sentido los modelos politómicos, desarrollados bajo el marco de la TRI, son en realidad una generalización de los modelos dicotómicos, al ampliar el número de respuestas que cada participante puede registrar a lo largo de los ítems.

Para el diseño, desarrollo, validación, aplicación, calificación y análisis de resultados obtenidos en los instrumentos de valoración requeridos para dar forma al SISAP, se recomienda emplear el marco metodológico derivado de la TRI para instrumentos politómicos (las encuestas de percepción y de los cuestionarios de actitudes), ya que permite la extracción de información precisa acerca de una única variable latente evaluada en cada prueba. Para el caso de los instrumentos dicotómicos, se recomienda el uso de Modelos de Diagnóstico Cognitivo, en tanto que se considera que éstos presentan un marco metodológico más cercano a los propósitos que se buscan alcanzar con este proyecto: favorecer la toma de decisiones informada para el diseño y selección de estrategias de formación continua.

Dentro de los modelos politómicos derivados de la TRI se propone el uso particular del Modelo de Respuesta Graduada, un modelo matemático propuesto para dar cuenta de la información que aportan las respuestas registradas a lo largo de una serie de ítems politómicos, cuya resolución requiere del sustentante hacer una valoración y la consecuente elección de uno de varios posibles niveles o categorías de respuesta ordenadas acerca del nivel o gradualidad con que el sustentante presenta la variable latente que representa al objeto de estudio del instrumento (van der Linden, 2017).

De acuerdo con la estructura politómica del instrumento, se asume que para cada ítem existen categorías o niveles de respuesta ordenados (0,1,2,…,). El Modelo de Respuesta Graduada (Samejima, 1969) se describe en términos de la siguiente ecuación:

(2)

Donde la probabilidad de que la respuesta () registrada por el participante en el ítem corresponda con el nivel de respuesta o mayor está definida por un modelo logístico, que depende de los parámetros (el grado de discriminación del ítem) y (la “dificultad”, de acuerdo con la nomenclatura empleada en Teoría de Respuesta al Ítem, asociada a este nivel de respuesta en el ítem).

*2.2.3 Redes neuronales y algoritmos de aprendizaje profundo para la calificación de Rúbricas*

Para la revisión y calificación de los instrumentos de corte cualitativo propuestos, para orientar el proceso de selección para la Promoción Horizontal dentro de la función docente y técnico docente en Educación Básica y para la Promoción Vertical en Educación Media Superior, que comprenden la revisión de un proyecto de seguimiento elaborado por los sustentantes y la realización de una entrevista y una observación en clase, se propone como tercera propuesta metodológica: el desarrollo de rúbricas de evaluación y la incorporación de algoritmos de última generación en materia de inteligencia artificial, redes neuronales y análisis neurolingüístico como una herramienta para automatizar y disminuir el sesgo en la calificación de las mismas, al analizar, clasificar y modelar las respuestas registradas a estos instrumentos.

El uso de rúbricas como instrumento de evaluación, permite la valoración sistemática de un conjunto de criterios de calidad asociados de manera sustantiva con el objeto de estudio, favoreciendo la clasificación de los sustentantes en niveles de desempeño delimitados por medio del uso de descriptores o indicadores (García, 2014). El uso de rúbricas se considera particularmente útil para la valoración de los sustentantes en tareas donde no es posible cuantificar con claridad las respuestas registradas, por lo que es necesario el análisis cualitativo de las fortalezas y áreas de mejora que estas presentan (Andrade y Du, 2005).

La incorporación de métodos derivados de la inteligencia artificial y el análisis neurolingüístico busca garantizar el cumplimiento de dos grandes objetivos: 1) Orientar la toma de decisiones en cuanto a los procesos de selección para la Promoción Horizontal en Educación Básica y la Promoción Vertical en Educación Media Superior y 2) Proporcionar un instrumento de clasificación automática, que permita estimar características muy específicas de cualquier sustentante en ausencia de test especializados.

De acuerdo con la Neurolingüística, la forma particular en que los humanos estructuran y desarrollan su lenguaje tiene una relación directa con sus esquemas conductuales y cognitivos. El primer problema que presenta el análisis de producciones lingüísticas tiene que ver con el método elegido para su extracción, especialmente si se busca que tenga algún valor predictivo. Tradicionalmente el análisis del lenguaje natural, mediante el uso de producciones de texto, suele realizarse por medio de jueces, lo que implica un trabajo de lectura individual para su posterior clasificación bajo criterios personales y protocolos semi-estandarizados. No obstante, la implementación de estas técnicas consume una amplia cantidad de recursos humanos y temporales y compromete en gran medida la precisión de las mediciones obtenidas, en tanto que el método empleado por cada juez funge como una fuente de sesgo importante.

En los últimos diez años se han desarrollado un conjunto de algoritmos especializados en el procesamiento de datos no estructurados, como son texto, audio o imágenes, para atender a la creciente necesidad de analizar grandes cantidades de información derivada del acelerado desarrollo de tecnologías de procesamiento de matrices. Estos algoritmos son la base principal de lo que se conoce como *Aprendizaje Profundo* (*Deep Learning)*. En el Anexo A puede consultarse información detallada acerca de su funcionamiento y operación.

Para resolver el problema que representa el análisis del lenguaje natural surge el desarrollo de una herramienta conocida como LIWC (por sus siglas en inglés, *Linguistic Inquiry and Word Count*; Pennebaker, 1996) dentro de la literatura de la Neurolingüística (Francis & Booth, 2001; Tausczik & Pennebaker, 2010), que parte del supuesto central de que toda palabra puede categorizarse en ocho grandes grupos de funciones neurolingüísticas: 1) Procesos cognitivos; 2) Procesos perceptuales; 3) Procesos biológicos; 4) Intenciones; 5) Orientación temporal; 6) Relatividad; 7) Preocupaciones personales; 8) Lenguaje informal.

De acuerdo con el marco del LIWC, cada una de las ocho funciones neurolingüísticas cuenta con su propio conjunto de subgéneros, siendo que cada palabra puede ser asignada a más de una de estas categorías. Posteriormente, mediante el análisis de la frecuencia con que se presenta cada una de estas categorías neurolingüísticas y su estructura de uso, se pueden trazar marcadores que den cuenta del foco de atención, el estado emocional, la complejidad cognitiva, las relaciones sociales, los estilos de pensamiento y diferencias individuales detectadas a partir de las respuestas de los sustentantes.

Los algoritmos propuestos toman como input el texto registrado por los sustentantes en los instrumentos de respuesta construida. Los textos recopilados son traducidos y resguardados en tres formatos distintos: 1) formato tf/idf, que asigna un peso específico a cada palabra como resultado de la implementación de algoritmos de aprendizaje profundo; 2) formato LIWC que asigna cada palabra a una o varias categorías neurolingüísticas; 3) formato de diccionario, que permite asignar a cada palabra un integral.

El texto transformado y resguardado en formato LIWC es ingresado en un Sistema de Selección de Atributos (SSA) que agrupa los textos en grupos semejantes. El SSA está conformado por tres algoritmos derivados del Análisis de Componentes Principales, Fuzzy C-means y Redes Neuronales Recurrentes que se encargan de generar los grupos mediante votación. Una vez agrupados los textos, se analizan sus características neurolingüísticas y se explora su relación con los conocimientos y aptitudes medidos por cada instrumento, robusteciendo las implicaciones de su interpretación y uso para la toma de decisiones informada.

Por su parte, el texto transformado y resguardado en formato diccionario, es procesado por un algoritmo de análisis de tópicos conocido como LDA (por sus siglas en inglés *Latent Dirichlet Allocation*), que funciona como una generalización del análisis semántico de probabilidad latente, con la diferencia de que ésta introduce una probabilidad a priori. El objetivo principal de este algoritmo es identificar las palabras más significativas de cada texto, referidos como tópicos.

Los resultados obtenidos con los análisis descritos pueden mapearse en conjunto con distintas variables para conocer las características específicas de distintos subconjuntos dentro de la muestra evaluada. Por ejemplo, la zona geográfica a la que pertenecen los sustentantes puede mapearse, junto con los grupos neurolingüísticos identificados mediante el LIWC, para conocer cuáles son los temas e intenciones más recurrentes a lo largo de la República.

Finalmente todas las transformaciones del texto son procesadas por un Sistema de Predicción de Calidad (SPC), conformado por seis algoritmos de clasificación (descritos a detalle en el Anexo B) y un módulo de votación, que en conjunto, permiten modelar el conglomerado de textos partiendo de los atributos de calidad obtenidos. El SPC genera como producto final un modelo predictor, capaz de inferir la calidad de las respuestas registradas por cada sustentante en ausencia de test especializados.

El SPC debe entenderse como un proceso que funciona en dos grandes momentos: primero, se modelan los distintos tipos de inputs con los seis algoritmos que lo conforman (fase de entrenamiento), evaluando la calidad predictiva de cada modelo, deteniéndose en el momento en que la prueba resulta positiva (fase de prueba) y se realizan los ajustes necesarios en los algoritmos para reducir el error general (fase de ajuste), en cuyo caso se vuelve a la fase de entrenamiento. Después, se procede a trabajar con un modelo de votación, construyendo una base de datos que permita el mapeo de cada sustentante con la matriz de resultados generada con los modelos. Esta base funciona como el input de una Red Neuronal Recursiva (RNN por sus siglas en inglés *Recursive Neural Network*) que sondea el resultado obtenido por los modelos con la calidad estimada para cada sustentante y su funcionamiento requiere de las mismas tres fases descritas para el modelamiento: una fase de entrenamiento, una fase de prueba y una fase de ajustes.

Integrando las diversas metodologías expuestas previamente, todos los instrumento de medición, con independencia del enfoque metodológico que se emplee para el diseño, desarrollo, validación, aplicación, calificación y análisis, garantizan la satisfacción de las siguiente exigencias (Abad, 2011):

1. Deben ser instrumentos confiables, cuyos resultados sean replicables al arrojar las mismas calificaciones para sustentantes con niveles de conocimiento y aptitud similares.
2. Deben aportar mediciones válidas del objeto de estudio, a fin de que los resultados obtenidos puedan extrapolarse en inferencias sobre los conocimientos y aptitudes de los sustentantes.
3. Su aplicación debe regirse por un protocolo, de manera sistemática, identificando y atendiendo necesidades de mejora y mantenimiento.

**2.3 Sistema informático para el almacenamiento y administración de instrumentos**

Para lograr el desarrollo de un sistema informático integral, que permita almacenar de forma segura los instrumentos de apreciación que conforman el SISAP, así como eficientar su aplicación y automatizar el cómputo y registro de los resultados obtenidos en cuanto a los niveles de conocimientos, aptitudes, actitudes y experiencias de las maestras y los maestros, se propone utilizar una metodología de software ágil denominada *Kanban* (Junior & Godinho, 2010), que permite gestionar la construcción del sistema, visualizando puntualmente el flujo continuo de trabajo y controlando la entrega de funcionalidad en plazos cortos de tiempo. En general, se propone que el sistema informático sea desarrollado para su funcionamiento en una plataforma web debido a que ofrece las siguientes ventajas:

* Compatibilidad: el sistema podrá ser ejecutado en cualquiera de las computadoras instaladas en las sedes de aplicación, siendo sólo necesario contar con servicio de Internet y tener instalado un navegador de internet Firefox y/o Google Chrome.
* Inmediatez de acceso: al estar el sistema instalado en una sola ubicación (servicio de aplicaciones en la nube), no será necesario instalar o realizar cambios en la configuración de las computadoras en las que se realice la aplicación. El sustentante sólo deberá teclear su clave de usuario y la contraseña que le fue asignada para poder acceder al sistema. De esta forma, el sistema se vuelve tan accesible como una red social, un portal de banca en línea o un portal de e-commerce, lo cual lo convierte en la opción más eficiente y segura. Además de ello, se facilita la comunicación inmediata entre el equipo de coordinación central de la aplicación, los coordinadores estatales, los responsables en sedes y el equipo central del sistema, permitiendo así la atención de las potenciales eventualidades del mismo.
* Actualización: debido a que la administración y el control del sistema se realizará siempre de forma central, todas las adecuaciones y mejoras efectuadas estarán disponibles en la versión en línea que los usuarios utilicen.
* Optimización de recursos: tanto el sistema informático como los datos generados por éste serán resguardados en la nube, donde se ejecutará la mayor parte del procesamiento de información. La Tabla 4 presenta de manera detallada las especificaciones técnicas mínimas requeridas de las computadoras a utilizar.
* Escalabilidad: el sistema contará con la flexibilidad necesaria para detectar y atender de manera automática la necesidad de aumentar la capacidad de soporte de carga de trabajo con modificaciones o ampliaciones a su infraestructura de operación (número de servicios en la nube o incremento de espacio en éstos). De esta forma, se podrán atender los casos en que se presenten nuevos requerimientos de explotación de los datos obtenidos durante la aplicación.
* Usuarios concurrentes: el sistema está diseñado para atender hasta 10 millones de usuarios de manera simultánea en las diversas sedes a nivel nacional en las que se presente la aplicación.
* Seguridad: permite implementar protocolos, que garanticen el alcance de los más altos niveles de seguridad, incluyendo controles de autenticación para el acceso a los servidores de aplicación y base de datos por parte de los administradores autorizados en cada caso, a fin de poder realizar los mantenimientos y/o las actualizaciones que se requieran.

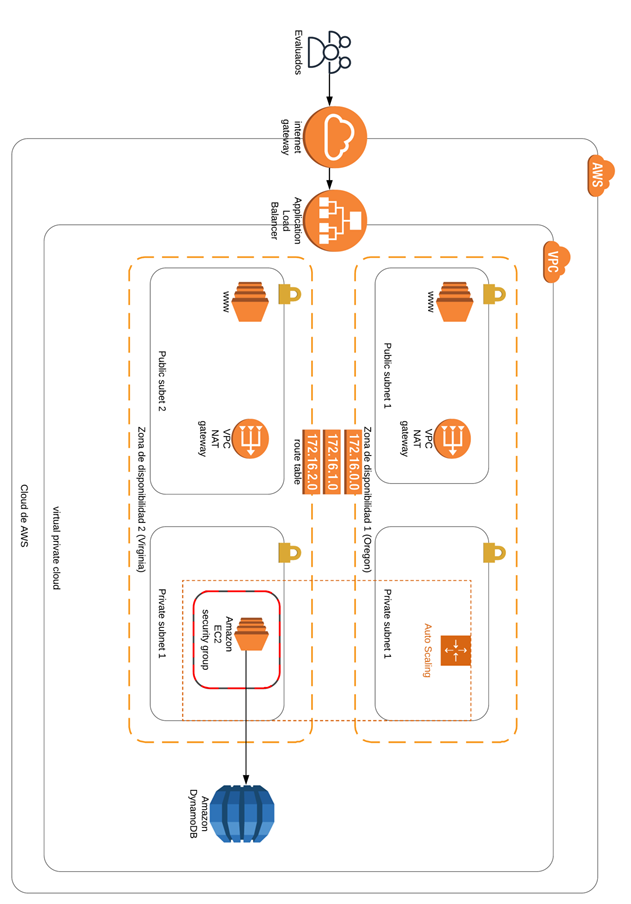
Tabla 4. Requisitos técnicos mínimos (e ideales) con los que deben contar las computadoras en sede para el correcto funcionamiento del Sistema Informático

|  |  |  |
| --- | --- | --- |
| **Requisito** | **Mínimo** | **Escenario Ideal** |
| Ancho de banda. | 100 KB/s por usuario | 500 KB/s por usuario |
| Asumiendo el caso de sede donde se realicen 50 evaluaciones de manera simultánea, el ancho de banda ideal sería de 25 MB y el mínimo de 5MB, siendo el funcionamiento mínimo equivalente a aquel de un café internet regular. | |
| Características Técnicas de las computadoras. | Ver Anexo C | Última versión 2019 |
| En esencia, si las computadoras se encuentran en condiciones de navegar en Google, no habrá problema para presentar la evaluación. | |
| Requerimientos de Software (Tener instalado el navegador Firefox o Chrome). | Versión igual o mayor a 2016. | Última versión 2019 |
| Estos navegadores se actualizan automáticamente, por lo que no es común encontrar versiones anteriores | |
| Requerimientos de la Red de la Sede. | Priorizar el uso de la red alámbrica  Deshabilitar el Wifi para evitar saturar el ancho de banda  Pausar cualquier transmisión de CCTV por la RED. | Dedicar todo el consumo de internet a la aplicación |

El sistema informático funcionará como una aplicación web de evaluación, donde hasta 25,000 usuarios por día podrán acceder desde las sedes de aplicación, como ubicaciones autorizadas por el sistema, en el horario programado en los calendarios publicados por la USICAMM. Para el diseño del sistema, se considera la implementación de arquitecturas de nube, utilizando para ello los servicios de Amazon Web Services (AWS), caracterizados por su alta disponibilidad y escalabilidad, junto con una serie de arquitecturas orientadas a eventos, microservicios y de tipo MVC (ver Figura 1). Se propone el uso de bases de datos de tipo relacional y NoSQL, así como de distintos tipos de frameworks/SDK (AWS SDK, Node 10.x), ServerlesFramework, React y Jest). Todo el trabajo realizado para el desarrollo del sistema se hará utilizando los lenguajes de programación JS, Python y Go.

La idea central es partir de un escenario de alta concurrencia poniendo a disposición un conjunto de servidores para procesar las solicitudes recibidas. De manera inicial, se estima el uso de 20 servidores por cada Centro de Datos, implementando un servicio de auto escalamiento que permita agregar más servidores al sistema en caso de presentarse picos de demanda.

El uso de los servicios de AWS garantiza un acuerdo de disponibilidad de servicios[[1]](#footnote-1) del 99.9999999% (siete nueves), permitiendo desarrollar la aplicación de forma redundante a fin de maximizar dicha disponibilidad, trabajando con dos Centros de Datos ubicados en Estados Unidos (específicamente, en los estados de Oregon y Virginia). Además de ello, el servicio AWS proporciona soporte técnico altamente especializado en atender cualquier necesidad detectada durante el desarrollo del sistema, optimizando su configuración.



Dentro de las configuraciones y propiedades más importantes del sistema informático propuesto para trabajar en la nube, destaca la preparación meticulosa contra ataques de internet. Particularmente se piensa implementar uno de los métodos más comúnmente empleados por su alta efectividad, conocido como Denegación Distribuida de Servicio (DDoS), cuyo funcionamiento se ilustra en la Figura 2.

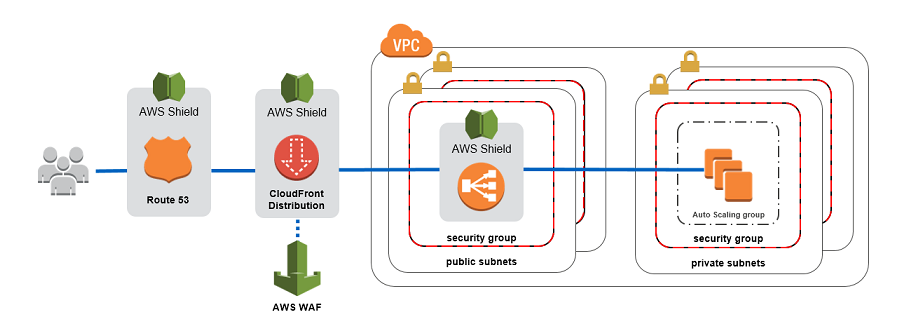
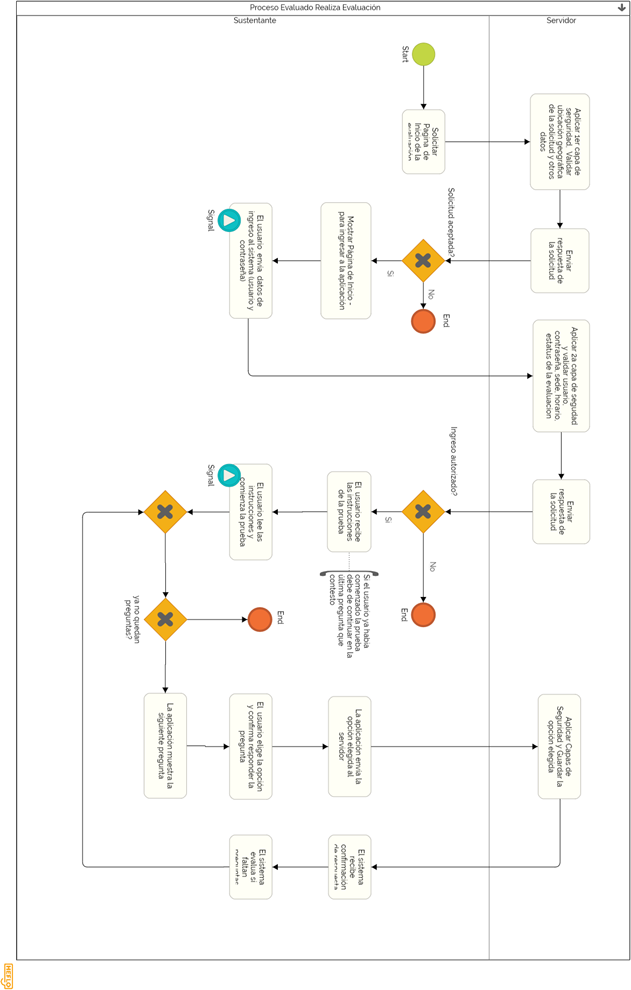


Figura 2. Protección contra ataques DDoS.

El desarrollo de este sistema informático, pensado para permitir la aplicación y resolución de los instrumentos de apreciación por medio de una aplicación web, hará énfasis en la implementación de estrategias que garanticen la seguridad de la aplicación, incorporando mecanismos que permitan el monitoreo en tiempo real del comportamiento de los usuarios para facilitar la detección de operaciones no reconocidas, así como la validación de la contraseña ingresada por cada sustentante, misma que será generada y proporcionada al momento por la aplicación.

Como medidas de seguridad preventivas, se propone incorporar restricciones de acceso al sistema, tales como una restricción por calendarización que sólo permita acceder a los instrumentos en los días específicos en que fue programada su aplicación y una restricción por computadora, que sólo autorice el acceso desde las direcciones de IP y usuario registrados en cada sede de aplicación, de acuerdo con la zona geográfica. El diagrama presentado en la Figura 3 ilustra el proceso de aplicación, considerando las medidas de seguridad ya descritas.



En cuanto a los procesos que se contemplan como parte esencial de las tareas a desarrollar a partir del sistema informático, se distingue de forma inicial siete grandes módulos: 1) La administración de los instrumentos; 2) La administración de los sustentantes; 3) La administración de las Sedes y Coordinadores; 4) La administración de la Aplicación; 5) La administración para la calificación y análisis de resultados; 6) La administración y elaboración de los reportes de resultados; 7) La integración de los reportes y otros entregables del proyecto. En este mismo sentido, los permisos de acceso otorgados a los distintos tipos de usuario admitidos al sistema pueden entenderse a partir de los siguientes módulos de operación:

1. Módulo de Administración (para usuarios “administradores”): este módulo está orientado a la creación y actualización de las claves y contraseñas de los usuarios del sistema. Permite efectuar altas, bajas y modificaciones a los permisos de acceso de los usuarios del sistema.
2. Módulo de Aplicación (para usuarios “sustentantes”): refiere a la presentación de las preguntas correspondientes a los instrumentos de valoración para su resolución, por parte de los sustentantes que han ingresado su usuario y contraseña. La arquitectura básica del Módulo de aplicación puede apreciarse en la Figura 3.
3. Módulo de Análisis (para usuarios “analistas”): permite el procesamiento de las respuestas registradas por los sustentantes para la calibración y el cómputo de las propiedades estadísticas y psicométricas de los instrumentos, así como para realizar el cómputo de las calificaciones de cada sustentante.
4. Módulo de Resultados: módulo desarrollado para facilitar la consulta y generación de reportes individuales o estadísticos de la aplicación de la evaluación.

*2.3.1 Desarrollo de un Banco de Ítems*

Para la administración de los instrumentos, se considera la construcción de un banco de ítems o reactivos que resguarde los avances y productos derivados del desarrollo de las tareas evaluativas y los reactivos que conformarán a los instrumentos del SISAP, permitiendo en todo momento el acceso al personal técnico de la USICAMM, sin necesidad de un tercero especializado. El banco de reactivos estará asegurado de modo que sea el equipo central de sistemas quien, además de llevar el control del histórico de los reactivos, valide y registre los permisos de acceso solicitados por la USICAMM.

En cuanto a la conformación y estructura del banco de ítems, a continuación se presenta un resumen de las variables, típicamente reportadas (Bergstrom & Gershon, 1995; Gierl & Haladyna, 2012; Weiss, 2013) como filtros para guiar de manera óptima la categorización y posterior búsqueda de reactivos y tareas almacenadas:

* ID del ítem, constituido por una serie de números, letras o una combinación de ambas, teniendo en cuenta que:
  + Al hacer ajustes o mejoras a un ítem, cada versión del mismo debe contar con su propio ID, a fin de poder rastrear y evaluar el impacto de los cambios realizados. Típicamente, se recomienda que los últimos dos caracteres del ID señalen el número de versión (01, 02, ….99).
  + El ID del ítem debe permitir la identificación del instrumento al que pertenece.
  + En el caso de los ítems con una estructura factorial claramente identificada (por ejemplo, en el caso de la Escala de Aprendizaje Socioemocional), el ID debe permitir la identificación del factor o dominio al que pertenece cada ítem.
* Estado actual del ítem o tarea (Activo o Inactivo)
* Autor (por Nombre)
* Revisor (por Nombre)
* Fecha única de elaboración
* Fechas de aplicación (periodos de aplicación en los que ha sido utilizado)
* Tipo de respuesta (Opción múltiple, escala ordinal, respuesta construida)
* Categorías de respuesta admitidas por el ítem (Dicotómico y Politómico), recordando que:
  + Se entiende por Ítem Dicotómico aquel que permite catalogar la respuesta registrada por el sustentante en una de dos posibles categorías (“Acierto” o “Error”; “Sí”, “No”)
  + Se entiende por Ítem Politómico aquel en que la respuesta registrada por los participantes implica una valoración a lo largo de una escala ordenada, por lo que cualquier respuesta registrada aporta información sobre la variable a medir.
* Fuentes bibliográficas, procurando:
  + Distinguir entre referentes normativos (Perfiles profesionales publicados por la SEP, la Ley General del Sistema de Carrera para las Maestras y los Maestros y la Ley de Mejora Continua en la Educación).
  + Distinguir entre referentes de la práctica docente (Planes de estudio, entre otros.)
* Contenido: identificar la habilidad, área de conocimiento o actitud que se intenta medir/evaluar con cada ítem.
* Indicadores psicométricos estimados a partir de su aplicación: puede darse la opción al usuario de especificar un rango de valores de interés para los distintos indicadores (Ver Ejemplo 1), o bien, permitirles especificar un valor fijo de referencia (Ver Ejemplo 2).

Algunos de los indicadores psicométricos a incluir son:

* + Índices derivados de la TCT
    - Índice de dificultad de los ítems dicotómicos (con una sola respuesta “correcta”)
  + Índices derivados de la TRI
    - Parámetro de adivinación
    - Parámetro de dificultad
    - Parámetro de discriminación
  + Correlaciones
    - Correlación ítem-total
  + Índices del comportamiento diferencial del ítem
    - Índice de Mantel-Haenszel

En el caso particular de los ítems que conformarán los instrumentos del SISAP, desarrollados como parte del presente proyecto, se sugiere incluir:

* Figuras educativas para las que fue diseñado cada ítem
* Proceso de selección en el que se aplica el ítem
* Nombre del instrumento(s)
* Enfoque metodológico

Además de las variables, previamente presentadas como posibles filtros que guíen y agilicen la búsqueda de reactivos específicos dentro del banco de ítems, es recomendable que la interfaz incluya información detallada sobre su estructura y características. Es decir, una vez que las variables filtro permitan el acceso a un conjunto específico de ítems, es recomendable que se incluya en pantalla la siguiente información sobre cada uno de estos:

* Una casilla de comentarios que permita al usuario rastrear los detalles acerca del desarrollo histórico o cronológico del reactivo, así como sus posibles áreas de mejora.
* En el caso de los ítems dicotómicos (de “acierto” o “error”), es importante señalar cuál de las opciones de respuesta es la correcta.
  + Se debe poder consultar en el banco de ítems la justificación del por qué esta respuesta es correcta, así como una justificación de por qué cada uno de los distractores no puede ser considerado como respuesta correcta.
* En el caso de ítems o tareas que se califican con rúbrica debe poder verse en pantalla la descripción del tipo de respuesta que se asocia con cada nivel de la rúbrica, de acuerdo con el algoritmo de redes neuronales construido para este fin.
* Una lista de ítems equivalentes o bien, cercanos en contenido (a veces llamados, ítems “amigos” o “amigos cercanos”).
* Identificar o enlistar los ID correspondientes a las versiones previas y posteriores, de acuerdo con las mejoras realizadas.
* Número de sustentantes a partir de los cuales se realiza la estimación de los indicadores psicométricos usados como variables filtro.
* Proporción de sustentantes que omitió la resolución del ítem.
* Una casilla que permita al usuario, que revisa el banco de ítems, agregar sus propias etiquetas de clasificación, para facilitar la consulta de ítems que resulten de particular interés.

**III. Plan de acción**

De acuerdo con lo presentado en el Anexo Técnico, se comenzará a trabajar desde el primer semestre del 2020 en la gestión de la aplicación correspondiente a los procesos de selección para la Admisión y la Promoción Vertical, en el diseño de los instrumentos de apreciación que se aplicarán durante el segundo semestre y, a lo largo del año 2021, en el desarrollo del sistema informático que será utilizado para administrar el almacenamiento y aplicación de los instrumentos de apreciación, así como para el resguardo y calificación de las respuestas de los sustentantes.

En la Tabla 1 se muestran los instrumentos que serán proporcionados por la USICAMM para su aplicación durante los meses de Mayo y Junio del 2020, como parte de los procesos de selección para la Admisión a las funciones Docente y técnico docente y la Promoción Vertical a funciones Directivas y de Supervisión. Como se puede apreciar, para la calificación de todos estos instrumentos se utilizarán modelos dicotómicos y politómicos derivados de la Teoría de Respuesta al Ítem.

Por su parte, en la Tabla 2 se presenta con detalle cuáles de los instrumentos que se requiere diseñar, desarrollar y validar desde el primer semestre del 2020 para su incorporación al SISAP a partir de las aplicaciones programadas para el segundo semestre del 2020 e inicios del 2021, estarán alineados con cada uno de los tres enfoques metodológicos ya descritos.

Tabla 1. Instrumentos proporcionados por la USICAMM para su aplicación, calificación y análisis e integración de reportes de resultados, durante el primer semestre del 2020.

|  |  |  |  |
| --- | --- | --- | --- |
| **Educación Básica** | | | |
| **Enfoque metodológico (calificación)** | **Proceso de selección** | **Instrumento** | **Periodo de aplicación** |
| **Modelos de la Teoría de Respuesta la Ítem** | **Admisión** a funciones Docente y técnico docente | Instrumento de acreditación. Curso Habilidades Docentes para la NEM | Mayo – Junio del 2020 |
| Instrumento de valoración de conocimientos y aptitudes |
| **Promoción Vertical** a funciones Directivas y de Supervisión | Instrumento de valoración de conocimientos y aptitudes | Mayo – Junio del 2020 |
| Cuestionario de habilidades directivas |
| Encuesta de percepción sobre el trabajo directivo y aportaciones al colectivo escolar |
| **Educación Media Superior** | | | |
| **Enfoque metodológico (calificación)** | **Proceso de selección** | **Instrumento** | **Periodo de aplicación** |
| **Modelos de la Teoría de Respuesta la Ítem** | **Admisión** a funciones Docente y técnico docente | Instrumento de conocimientos del modelo educativo | Mayo – Junio del 2020 |
| Instrumento de valoración de aptitudes y habilidades |
| **Promoción Vertical** a funciones Directivas y de Supervisión | Encuesta a la comunidad escolar | Mayo – Junio del 2020 |
| Encuesta por un Comité examinador |
| Instrumento de valoración de conocimientos y aptitudes |

Tabla 2. Instrumentos a desarrollar para su aplicación a partir del segundo semestre del 2020.

|  |  |  |  |
| --- | --- | --- | --- |
| **Educación Básica** | | | |
| **Enfoque metodológico** | **Proceso de selección** | **Instrumento** | **Periodo de aplicación** |
| Modelos de Diagnóstico Cognitivo | **Admisión** a funciones Docente y técnico docente | Examen de acreditación Curso | Primer semestre del **2021** |
| Instrumento de valoración de conocimientos y aptitudes |
| **Promoción Vertical** a funciones Directivas y de Supervisión | Instrumento de valoración de conocimientos y aptitudes | Primer semestre del **2021** |
| **Promoción Horizontal** dentro de las funciones Docente, Técnico Docente, Asesoría Técnico Pedagógica, funciones Directivas y de Supervisión | Instrumentos de valoración de conocimientos y aptitudes | Segundo semestre del **2020** |
| Modelos de la Teoría de Respuesta al Ítem | **Promoción Vertical** a funciones Directivas y de Supervisión | Encuesta de percepción sobre el trabajo directivo y aportaciones al colectivo escolar | Primer semestre del **2021** |
| **Promoción Horizontal** dentro de las funciones Docente, Técnico Docente, Asesoría Técnico Pedagógica, funciones Directivas y de Supervisión | Cuestionario de habilidades socioemocionales | Segundo semestre del **2020** |
| Instrumentos de respuesta construida | **Promoción Horizontal** dentro de las funciones Docente, Técnico Docente, Asesoría Técnico Pedagógica, funciones Directivas y de Supervisión | Proyecto de seguimiento | Segundo semestre del **2020** |
| Entrevista sobre el proyecto de seguimiento |
|
| **Promoción Vertical** a funciones Directivas y de Supervisión | Cuestionario de habilidades directivas | Primer semestre del **2021** |
| **Educación Media Superior** | | | |
| **Enfoque metodológico** | **Proceso de selección** | **Instrumento** | **Periodo de aplicación** |
| Modelos de Diagnóstico Cognitivo | **Admisión** a funciones Docente y técnico docente | Instrumento de conocimientos del modelo educativo | Primer semestre del **2021** |
| Instrumento de valoración de aptitudes y habilidades |
| **Promoción Vertical** a funciones Directivas y de Supervisión | *Instrumento de valoración de conocimientos y aptitudes* | Primer semestre del **2021** |
| Modelos de la Teoría de Respuesta al Ítem | **Promoción Vertical** a funciones Directivas y de Supervisión | *Encuesta a la comunidad escolar* | Primer semestre del **2021** |
| Instrumentos de respuesta construida | **Promoción Vertical** a funciones Directivas y de Supervisión | *Entrevista por un Comité Examinador* | Primer semestre del **2021** |

**3.1 Fases para el diseño, validación, aplicación y calificación de pruebas**

El diseño, desarrollo y validación de los instrumentos se realizará bajo el marco de la evaluación formativa con fines de mejora continua, incorporando técnicas y estrategias propuestas en el marco de los Modelos de Diagnóstico Cognitivo, a fin de garantizar la integración de los distintos conocimientos y aptitudes, referidos como vitales para el ejercicio de la práctica educativa por los perfiles profesionales. A fin de cumplir con el propósito que persigue la estructuración del SISAP y el impacto que este tendrá para los distintos procesos de selección que enmarcan al SCMM, el trabajo se ha organizado en seis grandes fases: I) El diseño; II) El desarrollo; III) La validación; IV) La aplicación; V) La calificación; VI) El análisis de resultados y la integración de reportes de resultados. Las actividades centrales a desarrollar a lo largo de cada una de estas fases se describen en la Tabla 3.

Tabla 3. Fases a considerar para el desarrollo de los instrumentos de apreciación del SISAP a aplicar a partir del segundo semestre del 2020 y sus actividades centrales.

|  |  |
| --- | --- |
| **Fase** | **Actividades a desarrollar** |
| **Fase I: Diseño** de los instrumentos de apreciación del SISAP | Diseñar los marcos de referencia y especificaciones generales de las pruebas de logro que forman parte del SISAP, empleando métodos y técnicas derivadas de los Modelos de Diagnóstico Cognitivo (MDC) con fines formativos. |
| Diseñar los marcos de referencia de las encuestas de percepción y cuestionarios de actitudes sobre el trabajo educativo realizado por los sustentantes y aportaciones al colectivo escolar. |
| **Fase II: Desarrollo** de los instrumentos de apreciación del SISAP | Desarrollar las especificaciones de los reactivos. |
| Elaborar un manual para la construcción de ítems y brindar capacitaciones detalladas para promover la homogeneidad y sistematicidad en el trabajo de los desarrolladores de ítems. |
| Elaborar las primeras versiones de los distintos instrumentos de valoración del SISAP. |
| Supervisar la revisión de los ítems con ayuda de expertos de contenido y expertos en medición y evaluación. |
| Supervisar la edición y el formato de los ítems. |
| **Fase III: Validación** de los instrumentos de apreciación del SISAP | Conformación y capacitación de un comité de validación, donde participen expertos de contenido y en medición y evaluación. |
| Piloteo y validación de las primeras versiones de los distintos instrumentos de valoración del SISAP. |
| **Fase IV: Aplicación** de los instrumentos de apreciación del SISAP | Diseño y desarrollo de un sistema informático que permita la aplicación de los instrumentos del SISAP. |
| Diseño y desarrollo de un sistema informático que salvaguarde los instrumentos del SISAP. |
| Capacitación para los coordinadores y supervisores en campo de la aplicación de los distintos instrumentos de valoración del SISAP. |
| Aplicación y coordinación logística del operativo en campo, y aseguramiento de la información de las valoraciones del SISAP. |
| **Fase V: Calificación** de los instrumentos de apreciación del SISAP | Capacitación para el análisis y calificación de los instrumentos de valoración construidos. |
| Desarrollo de tecnologías de última generación, Inteligencia Artificial (IA) para la calificación objetiva de instrumentos de medición cualitativos. |
| Realizar y presentar los análisis psicométricos de las pruebas, presentando indicadores propios de la Teoría Clásica de los Test, de la Teoría de Respuesta al Ítem y de los Modelos de Diagnóstico Cognitivo, según aplique. |
| Presentar un informe con las necesidades de mejora detectadas a partir de la aplicación y una propuesta para atenderlas. |
| **Fase VI: Análisis de resultados** e integración de reportes individualizados | Integración de reportes individualizados de los resultados obtenidos en cada uno de los instrumentos del SISAP por cada sustentante, identificando de manera puntual sus áreas de fortaleza y mejora. |
| Integración de reportes regionales, estatales y nacionales, según sea el caso, para contribuir al desarrollo de materiales, cursos y otros recursos que atiendan las necesidades de formación continua identificadas en la población. |

A continuación, se presenta una descripción detallada del plan de acción a seguir para cada una de las fases señaladas.  *Fase I. Diseño de los instrumentos de apreciación del SISAP.*

La construcción de los instrumentos de apreciación del SISAP, que serán empleados como referente para la toma de decisiones, en cuanto a los procesos de selección para la Admisión, la Promoción Horizontal y la Promoción Vertical en Educación Básica y Educación Media Superior, debe partir del diseño de marcos de referencia, que orienten la construcción de los instrumentos en torno a la definición clara de objetos de estudio, cuyos componentes puedan ser identificados con la precisión suficiente para conseguir una valoración completa y con alta validez de constructo del mismo.

Por cada instrumento, de acuerdo con el marco metodológico que ha sido identificado como más apropiado para su construcción, uso y calificación (ver Tabla 2), deberá trazarse un modelo sustantivo que permita describir la manera en que los distintos elementos, que dan forma al objeto de estudio, se relacionan entre sí, integrando para ello la información derivada de los perfiles profesionales con los modelos teóricos desarrollados desde la Psicología Cognitiva y Educativa, o bien, desde las Ciencias de la Educación. Lo anterior con el fin de garantizar que las mediciones e interpretación de resultados, que se entreguen a partir de la aplicación y calificación de estos instrumentos, sea congruente con el propósito del mismo y permita, bajo los más altos estándares de validez y robustez sustantiva, orientar la toma de decisiones informada y basada en evidencia.

*Fase II. Desarrollo de los instrumentos de apreciación del SISAP.*

Una vez diseñados los modelos sustantivos y cognitivos que dan cuenta de la estructura del objeto de estudio de cada instrumento, se desarrollará una matriz de especificaciones que contribuya a sistematizar la generación de más y nuevos ítems, que puedan ser incorporados al instrumento ya sea en sustitución de los ítems que demuestren tener problemas de validez, o bien, para mantener actualizada la base de ítems.

El diseño de las especificaciones estará íntimamente alineado a la necesidad de sistematización identificada como requisito fundamental para los procesos de evaluación a gran escala, buscando disminuir los errores de medida entre los sistemas educativos participantes. Para ello, se propone trabajar con un conjunto de expertos tanto de contenido como en materia de medición y evaluación, quienes deberán supervisar en todo momento que se conserve intacta la congruencia entre la estructura subyacente plasmada en la matriz de especificaciones y los contenidos de los perfiles profesionales, planteados como el referente principal con el que el personal educativo busca mejorar la propia práctica y el referente normativo a partir del cual se desarrolla el SISAP y cualquier otra política en materia de educación.

Se sugiere aplicar un enfoque metodológico complementario, basado en el análisis de contenido, que permita determinar la correspondencia entre los contenidos plasmados en las tablas de especificaciones y los perfiles profesionales correspondientes. La alineación entre ambos aspectos contribuye a garantizar la equidad en la valoración de las habilidades y conocimientos que poseen los sustentantes, así como la inclusión de la amplia gama de aspectos que abarca su práctica educativa.

Para la construcción de los ítems que conformarán cada uno de los instrumentos del SISAP, se propone trabajar en torno a los siguientes momentos:

1. Construcción individual de los ítems

El desarrollo del total de los ítems (mismos que serán distribuidos entre los miembros del equipo central, en función de la tabla de especificaciones, la estructura de los instrumentos por desarrollar y el número de participantes en los equipos nacionales) se realizará a distancia y con estricto apego al Manual digital elaborado, para homogeneizar la construcción de los mismos. El envío de los ítems se realizará a través del uso de una plataforma informática, que fungirá como elemento de gestión, repositorio y retroalimentación.

b. Revisión interna de los ítems

Se asignará un coordinador dentro del equipo central por cada instrumento, quien será responsable de realizar una primera revisión de los elementos conceptuales contenidos en cada ítem elaborado, así como de su correspondiente convergencia con la tabla de especificaciones. La revisión se realizará con el apoyo de una lista de verificación, con la posibilidad de registrar comentarios específicos que retroalimenten la labor del desarrollo de ítems. Este proceso resulta iterativo entre quienes elaborarán los ítems y cada coordinador.

c. Revisión mediante experto disciplinario y experto en medición

Una vez realizada la revisión de la alineación de los ítems desarrollados con respecto a la matriz de especificaciones propuesta, se realizará una segunda revisión especializada, donde se acreditará la validez técnica y conceptual de cada uno de los ítems ajustados, a partir de su revisión por una diada conformada por un experto de contenido y un experto en materia de medición y evaluación.

d. Edición y formato de los ítems

La versión final de los ítems, posterior a las revisiones realizadas por el coordinador de instrumento y por la diada de expertos, se someterá a un proceso de edición, donde se procurará asegurar que las modificaciones realizadas no han alterado la relación que guarda el ítem propuesto con los contenidos referidos en la tabla de especificaciones. Se revisa y homogeniza el formato y estilo de redacción de la base del ítem y sus distractores, de manera que el instrumento quede compuesto por una serie congruente de elementos.

*Fase III. Validación de los Instrumentos de apreciación del SISAP.*

La tercera fase de nuestro Plan de Acción estará enfocada a la revisión de la calidad técnica de los ítems desarrollados en la fase anterior, como parte de cada uno de los instrumentos de apreciación del SISAP. Para ello, se propone la realización de aplicaciones piloto con una muestra similar a la población objetivo y con un grupo de expertos en contenido, a fin de garantizar la validez del contenido de cada uno de los ítems elaborados, en relación con su objeto de estudio.

Las aplicaciones piloto se realizarán en condiciones similares a aquellas en las que se espera se produzca la aplicación real de los instrumentos, brindando así la oportunidad de evaluar tanto la calidad técnica de los ítems elaborados como la eficacia del sistema informático desarrollado para hospedar los instrumentos para su aplicación en línea y para registrar y guardar las respuestas registradas por cada sustentante.

De acuerdo con lo señalado en el Anexo Técnico, se espera contar con una muestra amplia de sustentantes que presenten las mismas características que las poblaciones objetivo grandes, a fin de poder revisar la estructura interna de los instrumentos y reportar las estimaciones solicitadas correspondientes a los índices psicométricos de la Teoría Clásica de los Test, de la Teoría de Respuesta al Ítem y, en el caso de las pruebas objetivo, de los Modelos de Diagnóstico Cognitivo, con el objetivo de identificar cualquier posible necesidad de mejora que se requiera atender antes de su aplicación real. Por otro lado, en cuanto a los instrumentos desarrollados para su aplicación con una población objetivo reducida, se realizará un piloteo de los instrumentos con un comité de expertos que responderá la prueba, a fin de validar su relación con el objeto de estudio e identificar sus áreas de mejora.

Se contempla también la participación de un comité de validación, compuesto por al menos seis integrantes de la USICAMM, quienes recibirán una capacitación en torno a: 1) los marcos de referencia de los distintos enfoques metodológicos a emplear durante el diseño y desarrollo de los instrumentos; 2) las técnicas psicométricas derivadas de dichos enfoques y que serán utilizadas para la calificación de las respuestas registradas en cada instrumento y 3) el uso y funcionamiento de técnicas cognitivas a aplicar durante el piloteo de los instrumentos, a fin de obtener evidencias de validez de los mismos, a partir de la información extraída durante la resolución de cada ítem. En concreto, se considera una capacitación que promueva el dominio y uso de técnicas psicométricas y cognitivas relacionadas con:

* Calibración en Teoría Clásica de los Test (TCT).
* Análisis Factorial Exploratorio y Confirmatorio
* Calibración en Teoría de Respuesta al Ítem (TRI).
* Reportes y protocolos verbales con técnicas de pensamiento en voz alta (concurrentes y retrospectivos).
* Técnicas de seguimiento del sendero de la vista.
* Análisis cronométrico de respuesta.
* Modelos DINA, DINO y G-DINA para el diagnóstico cognitivo.

En atención a la solicitud hecha por la USICAMM en donde se planea dotar de nuevas capacidades y habilidades a diversos integrantes de su área técnica, se plantea el siguiente temario que aborda los múltiples tópicos con los cuales el personal que tome la capacitación y la acredite satisfactoriamente, podrá aportar en mayor medida desarrollos conceptuales y empíricos al diseño, desarrollo, validación, calificación y análisis de instrumentos de medición estandarizados y de respuesta construida. La capacitación del comité de validación se realizará de manera presencial, con una duración total de 60 horas en aula y bajo la guía de un instructor altamente calificado. Todos los materiales a trabajar durante la capacitación serán elaborados y distribuidos de manera electrónica, e incluirán una serie de manuales, tutoriales y ejercicios sobre simulación computacional y manejo de software especializado de libre acceso. Para una revisión más detallada de los contenidos a abordar durante la capacitación teórico-práctica propuesta, se sugiere revisar el Anexo D.

El objetivo principal de la capacitación del personal técnico de la USICAMM, es familiarizarlos con las teorías, modelos y bases metodológicas empleadas para el diseño y desarrollo de los instrumentos, permitiendo alinear su participación en la validación de los instrumentos piloteados con los marcos de referencia a partir de los cuales fueron elaborados los mismos.

En general, se busca integrar un modelo de validación impulsado por la participación de comités de expertos, coordinados con el personal de la SEP, y considerando la participación de expertos nacionales e internacionales en materia de psicometría y evaluación (ver Contreras, 2004; y Nitko, 1994) que permita asegurar la calidad técnica de los instrumentos (ver OECD, 2012; Mislevy, 2006; y Messick, 1989).

Para optimizar la información a obtener a partir de la ejecución de los participantes en la aplicación piloto, se capacitará al comité de validación en la implementación de protocolos de pensamiento en voz alta concurrentes y recursivos para la integración de reportes detallados; ésto permite identificar las estrategias y procesos de respuesta evocados por los participantes, la realización de entrevistas introspectivas y retrospectivas y la aplicación de técnicas de seguimiento del sendero de la vista. De manera general, el desarrollo de esta aplicación piloto implicará:

• Seleccionar una muestra que emule las características generales de la población objetivo, procurando que se integre la participación del personal educativo de distintos niveles educativos y modalidades de servicio, (por ejemplo, en el caso de los instrumentos de Educación básica, se requiere de la participación de personal educativo de nivel Preescolar, Primaria, Secundaria, Educación especial y Educación física).

• Aplicar los instrumentos a los sustentantes, solicitando en apego al protocolo diseñado para estos fines, que verbalicen en voz alta lo que piensan al tratar de resolver cada uno de los ítems o tareas presentados.

• Contar con el consentimiento para grabar el audio y video de la pantalla de los participantes en el piloto.

Previo a la aplicación piloto se realizará una capacitación presencial dirigida a los equipos nacionales, con el objetivo de presentar los lineamientos y protocolos necesarios para el desarrollo de la aplicación. Una vez desarrollada la aplicación piloto, se revisarán y analizarán los reportes de incidencias, así como el proceso de aplicación para descartar posibles errores de aplicación y detectar necesidades de mejora en el diseño de los ítems.

Posteriormente, el equipo implementador, en conjunto con las coordinaciones disciplinares, desarrollarán un taller para la corrección de los ítems en atención a las necesidades de mejora detectadas a partir de la aplicación piloto.

*Fase IV. Aplicación.*

El personal especializado contratado por la Facultad de Psicología se encargará de la planeación, la organización, la dirección, supervisión y el control de la aplicación de los instrumentos de apreciación del SISAP en las 32 entidades federativas del país, en las sedes establecidas para este fin por las autoridades educativas estatales.

Los instrumentos de apreciación del SISAP serán respondidos por cada sustentante en línea, usando para ello una computadora con conexión a internet ubicada en una de las sedes de aplicación. Es en este sentido que se entiende como responsabilidad de las autoridades educativas estatales garantizar que exista la infraestructura y conectividad a la web necesaria para poder acceder a la aplicación web diseñada como parte del sistema informático a desarrollar por el equipo especializado de la Facultad de Psicología para el almacenamiento y administración de los instrumentos y el resguardo y calificación de las respuestas registradas por los participantes, (ver las secciones 2.b y 3.b para información detallada sobre los requisitos para el correcto funcionamiento del sistema y su operación durante la aplicación, respectivamente).

La exposición de las características de la planeación del operativo en campo, así como la estructura operativa del equipo encargado de coordinar, planificar, dirigir y supervisar los procesos de aplicación comprometidos en la presente propuesta técnica a nivel nacional (es decir, la aplicación de los instrumentos proporcionados por la USICAMM para su aplicación durante los meses de mayo y junio del 2020 y la aplicación de los instrumentos diseñados y desarrollados por el equipo de especialistas de la Facultad de Psicología para su aplicación durante el segundo semestre del año), se encuentra en el apartado 3.c. del presente documento.

*Fase V. Calificación.*

Una vez almacenadas las respuestas de los sustentantes en el sistema informático desarrollado para este fin, se procederá a estimar los índices psicométricos derivados de la Teoría Clásica de los Test y la Teoría de Respuesta al Ítem, que den cuenta de la calidad técnica de los instrumentos elaborados como parte del SISAP, a fin de respaldar la validez de los resultados obtenidos por cada sustentante como un insumo valioso para orientar la toma de decisiones en los diversos procesos de selección, así como para identificar las necesidades de mantenimiento y mejora de los instrumentos.

El reporte de la calidad técnica de los instrumentos aplicados comenzará con un análisis descriptivo de las respuestas registradas por los participantes a lo largo de los ítems que les conforman, con el objetivo de detectar posibles sesgos e imprecisiones en la forma en que se están presentando las instrucciones o las opciones de respuesta que no hayan podido ser detectadas por el comité de validación. Por ejemplo, una revisión de la frecuencia con la que los sustentantes eligen cada una de las opciones de respuesta presentadas en los ítems de las pruebas objetivas proporciona información valiosa para la detección de distractores que pudieran ser demasiado fáciles de descartar, o bien, demasiado cercanos a la respuesta correcta.

En el caso particular de las pruebas objetivas, se propone complementar el análisis descriptivo con la estimación de los indicadores de dificultad propuestos desde la Teoría Clásica de los Test, que refiere a la proporción de la población que acertó cada ítem. Nótese que la interpretación del valor absoluto de estos índices es contraintuitiva: un mayor índice de dificultad señala una mayor proporción de aciertos en la muestra evaluada, permitiendo identificar al ítem como un ítem “más fácil” que aquellos con un índice de dificultad menor.

En aras de evaluar la consistencia interna de los instrumentos, entendida como la correspondencia o concordancia detectada entre las respuestas registradas a lo largo de los ítems y la puntuación obtenida por los sustentantes, el reporte de calidad técnica contendrá información acerca de la estimación del coeficiente de Cronbach (ver ecuación 3; Abad, 2011),

(3) ]

donde J es el número total de ítems contenidos en cada prueba; es la varianza observada entre los puntajes obtenidos por los sustentantes; es la covarianza entre la respuesta o resultado registrado en el ítem , () y el resto de los ítems , ().

Para evaluar la estructura interna de los instrumentos proporcionados por la USICAMM para su aplicación en el primer semestre del 2020, se incluirán Análisis Factoriales Exploratorios (AFE) en el reporte de la calidad técnica de los instrumentos. Además de ello, en el caso de los instrumentos desarrollados para su aplicación a partir del segundo semestre del 2020, se valorará el grado en que la estructura interna revela a partir de las respuestas registradas por los sustentantes y el modelo subyacente diseñado a partir de los distintos enfoques metodológicos empleados, se realizarán además Análisis Factoriales Confirmatorios (AFC), (Matsunaga, 2010).

En cuanto a los índices psicométricos derivados de la Teoría de Respuesta al Ítem, se presentarán los índices de dificultad y discriminación obtenidos para cada uno de los ítems que componen los distintos instrumentos del SISAP a partir de la aplicación. En el caso particular de los instrumentos dicotómicos, se propone acompañar estas estimaciones con las curvas características de cada ítem, a fin de agilizar la lectura de las propiedades psicométricas de cada ítem.

Para los instrumentos desarrollados para su aplicación a partir del segundo semestre del 2020 bajo el enfoque propuesto por los Modelos de Diagnóstico Cognitivo, se propone incluir en el reporte de índices psicométricos la estimación de índices propios de este enfoque, tales como la estimación de los parámetros de adivinación y desliz para cada ítem, así como el índice de discriminación (, que constituye un control para garantizar el correcto funcionamiento de la prueba al verificar que su valor no sea menor a 0, para ninguno de los ítems que constituyen la prueba.

Una vez computados los índices psicométricos previamente descritos, se integrará un reporte de la calificación de cada instrumento donde se precisen sus propiedades estadísticas y psicométricas, identificando de manera puntual cualquier necesidad de mejora detectada, así como las acciones y estrategias que se seguirán para su atención.

*Fase VI. Análisis e integración de reportes de resultados*

El objetivo central del plan de acción presentado en cuanto al diseño, desarrollo y validación de nuevos instrumentos de apreciación para el SISAP, es garantizar que los resultados obtenidos por cada sustentante contribuya de manera clara y fundamentada a la toma de decisiones informada, al ser ponderados junto con el resto de los elementos multifactoriales a considerar para la elaboración de las listas de prelación a partir de las cuales serán asignadas las plazas o incentivos comprometidos en cada proceso de selección.

A fin de promover una lectura clara, intuitiva y significativa para la toma de decisiones tanto en términos de los procesos de selección, como para el desarrollo y selección de estrategias y recursos que atiendan las necesidades de formación continua, la presente propuesta considera la devolución de resultados mediante reportes individualizados que puedan generarse de manera automática por el sistema informático desarrollado para la administración de los instrumentos del SISAP.

Estos reportes individualizados contendrán los datos de identificación del sustentante, de acuerdo con la documentación presentada para validar el cumplimiento de los requisitos marcados en la LGSCMM para participar en cada uno de los procesos de selección, y un desglose detallado de los puntajes obtenidos en cada uno de los instrumentos realizados, donde no sólo se presentarán los puntajes totales, sino una desagregación de los puntajes parciales obtenidos en cada uno de los dominios o factores que conforman la estructura interna de los instrumentos, acompañados de descriptores que faciliten la interpretación general de los resultados obtenidos. Adicionalmente, cada reporte individualizado proporcionará a cada sustentante una serie de recomendaciones para la atención de sus áreas de oportunidad detectadas, presentando para ello un listado de las mismas y de los recursos y materiales preparados por las Autoridades Educativas Estatales para su fortalecimiento.

El grado de desagregación con que puedan devolverse los resultados obtenidos a lo largo de los instrumentos del SISAP, depende en gran medida del enfoque metodológico empleado para su desarrollo y calificación. Los instrumentos desarrollados bajo el enfoque de los MDC permitirán retroalimentar a los sustentantes con el mayor grado de fineza, proporcionándoles información específica sobre el dominio detectado a lo largo de las distintas habilidades y conocimientos contenidos en la prueba. En cuanto a los instrumentos desarrollados bajo el marco de la TRI, se espera que el reporte de resultados permita retroalimentar a los sustentantes en términos de su ejecución a lo largo de los distintos dominios o factores que constituyen la estructura interna de la prueba; por ejemplo, en el caso del Cuestionario de Aprendizaje Socioemocional, se espera proporcionar información sobre el puntaje total alcanzado y sobre los puntajes obtenidos a lo largo de las cinco grandes habilidades socioemocionales que constituyen los cinco dominios del instrumento. Finalmente, en cuanto a los instrumentos de respuesta construida que serán valorados mediante la implementación de algoritmos de aprendizaje profundo para la minería de texto, se espera proporcionar información detallada acerca de las categorías neurolingüísticas en las que son clasificadas las producciones de los sustentantes, permitiendo la identificación y descripción de las intenciones, tópicos principales, complejidad de lenguaje y sofisticación de las respuestas construidas por los sustentantes.

***3.2 Operación del Sistema Informático para el almacenamiento y administración de los instrumentos del SISAP***

De acuerdo con la propuesta metodológica para el desarrollo del sistema informático, se propone que este funcione como una aplicación web construida bajo un diseño computacional de nube, especialmente pensado para soportar ataques en internet y atender una alta concurrencia de usuarios (hasta 3000 solicitudes simultáneamente).

Para el desarrollo, mantenimiento y administración del sistema informático a desarrollar se planea la contratación de un equipo central conformado por seis integrantes: tres programadores seniors, un programador junior, un consultor del servicio AWS y un consultor para el diseño gráfico de todas las interfaces del sistema.

De acuerdo con el diagrama presentado en la Figura 6, se espera que la operación del sistema infográfico durante la aplicación de los instrumentos se lleve a cabo a partir de la secuencia de pasos siguiente:

1. Primero, el usuario deberá ingresar desde la sede de aplicación con su usuario y contraseña.
2. Una vez validados estos datos, el sustentante comenzará a visualizar en pantalla, uno por uno, cada uno de los ítems o tareas que componen el instrumento que le toca responder de acuerdo con los calendarios oficiales. En caso de que el sustentante cerrase la pestaña de aplicación por error, podrá volver a ingresar al sistema y continuar su aplicación desde donde se quedó.
3. Finalmente, en cuanto el sustentante haya terminado de contestar la totalidad de los reactivos o tareas, recibirá un mensaje en pantalla para notificarle que su valoración ha terminado.
4. Una vez que se haya dada por terminada la aplicación de los instrumentos a nivel nacional, se dará inicio a las etapas de análisis y calificación de resultados, durante las cuales se realizarán análisis estadísticos y estimaciones paramétricas derivadas de los marcos metodológicos empleados para el desarrollo de los instrumentos. Como principal producto de esta fase de análisi y calificación, se tiene identificada la generación de un reporte de la calidad técnica de cada instrumento y un reporte de resultados individualizado para cada sustentante.

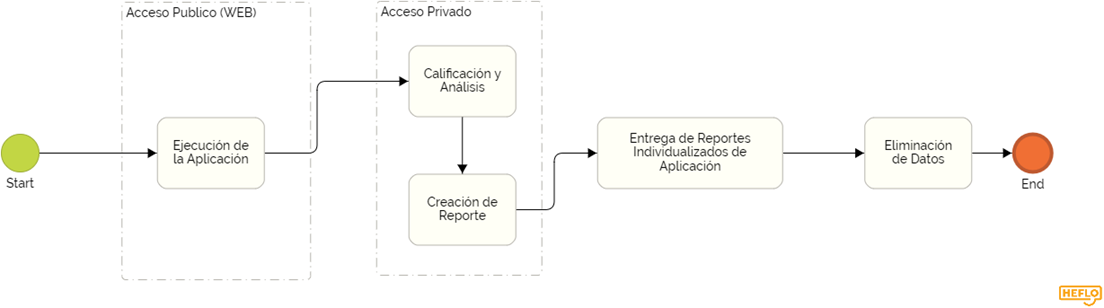


Figura 6. Esquematización del proceso de aplicación.

En cuanto a las restricciones de acceso por zona, dirección IP y calendario, descritas en la propuesta metodológica, se enfatiza que el sistema contará con la flexibilidad suficiente para que, en caso de que por causa de fuerza mayor se requiriera efectuar un cambio de sede, el Responsable Técnico de la sede pueda validar el acceso desde la nueva ubicación en menos de 3 minutos.

Para garantizar la confianza en el correcto funcionamiento del sistema informático, del cual dependen el almacenamiento y aplicación de los instrumentos, así como el resguardo y calificación de las respuestas de los participantes, se desarrollará un ambiente de pruebas que permita en todo momento el acceso de los usuarios registrados por parte del equipo central de trabajo y de la USICAMM, a fin de facilitar la consulta de la visualización y funcionamiento de la última versión validada del sistema (siendo que la validación e incorporación de cada nueva versión del sistema al ambiente de pruebas esté programada para llevarse a cabo en un periodo no mayor a cinco días).

Bajo esta misma lógica, para el diseño y desarrollo del banco de reactivos se contempla el intercambio constante de avances y retroalimentación entre el personal del equipo central del sistema y el personal técnico de la USICAMM. En específico, se plantea comenzar a trabajar en el desarrollo del banco de reactivos proponiendo una primera versión de la interfaz de usuario a fin de garantizar que esta sea intuitiva y contenga los elementos necesarios para optimizar su uso.

**3.3 Plan general de trabajo para el desarrollo de la aplicación en campo**

Como se mencionó en la Propuesta metodológica (Sección II del presente documento), una de las tres grandes tareas identificadas como parte del servicio solicitado por la SEP, a través de la USICAMM, tiene que ver con el trabajo a realizar en pos de la planificación, organización, dirección y supervisión de los procesos de aplicación de los instrumentos que conforman el SISAP, de acuerdo con las fechas programadas para la celebración de los distintos procesos de selección referidos como parte del SCMM.

Como parte de los procesos de selección para la Admisión y la Promoción Vertical, se tienen programadas jornadas de aplicación durante los meses de mayo y junio del 2020 y para ello, la USICAMM deberá entregar los instrumentos ensamblados con al menos dos semanas de anticipación respecto de las fechas de aplicación programadas en el calendario oficial. Bajo esta misma lógica, para todas las aplicaciones subsecuentes que tengan lugar a partir del segundo semestre del 2020, el personal encargado del desarrollo de pruebas deberá entregar los instrumentos desarrollados en la Facultad de Psicología bajo los marcos metodológicos señalados en la sección 2.2 de la presente Propuesta Técnica, con al menos dos semanas de anticipación, para que puedan estar disponibles en el sistema informático.

A continuación, se describe la estructura operativa del equipo de trabajo que estará a cargo de la coordinación de los diversos procesos de aplicación comprometidos como parte del presente proyecto, así como las características generales del plan de trabajo para el desarrollo de los mismos.

**3.3.1 Estructura operativa**

La planeación, organización, dirección, supervisión y control de los procesos de aplicación a nivel nacional correrá a cargo de un equipo de trabajo central y por equipos estatales conformado por el personal señalado en el organigrama que se presenta en la Figura 4. De tal forma, que como líderes del equipo central se considera la participación de dos personas con amplia experiencia en el desarrollo de aplicaciones de instrumentos a gran escala, quienes contarán con un equipo técnico especializado para atender responsabilidades señaladas en el segundo nivel del esquema presentado.

Se considera también la contratación de cuatro coordinadores regionales que estarán encargados de apoyar al equipo central durante todo el proceso, así como de coordinar, asesorar y supervisar a los coordinadores estatales de la región que les sea asignada. Se trabajará con 32 coordinadores estatales, quienes según las dimensiones del levantamiento en la entidad, podrán contar con una persona de apoyo operativo y otra más como enlace administrativo. Finalmente, se estima que será necesaria la contratación de un aproximado de 1548 Aplicadores-supervisores encargados de la atención directa a los grupos de aplicación, siendo que en los grupos de aplicación donde se registren sustentantes con alguna discapacidad que les requiera de algún apoyo humano, se les asignará de manera individual un personal de apoyo para garantizar la igualdad de oportunidades en los procesos de selección.

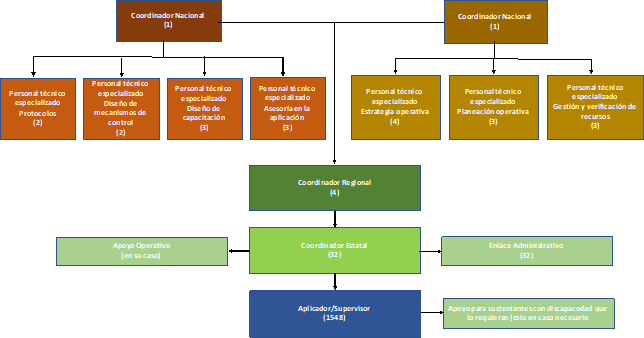


Figura 4. Estructura operativa nacional.

En la Figura 5 se presenta la estructura operativa para la entidad y la sede de aplicación. Muestra la función del coordinador estatal como responsable de dirigir el trabajo operativo en la entidad antes, durante y después de la aplicación; específicamente en los días establecidos para la resolución de instrumentos, su responsabilidad será asesorar y supervisar la actividad en las sedes de aplicación en contacto directo con el personal que se asignará a las mismas.

Para atender el proceso en cada sede, la autoridad educativa estatal asignará un responsable técnico de sede. Por su parte, la Facultad de Psicología asignará un aplicador-supervisor por grupo, asimismo habrá en cada aula, un apoyo para la aplicación designado por la autoridad educativa. Estos elementos serán los encargados de atender la aplicación y dar fe de quienes asistieron a la aplicación de los instrumentos y que ésta, se desarrolló con base en las normas establecidas para el proceso.

Por otra parte y de juzgarse necesario, por cada cierta cantidad de grupos de aplicación y con base en el criterio establecido por la USICAMM, se designará un coordinador de sede de aplicación.

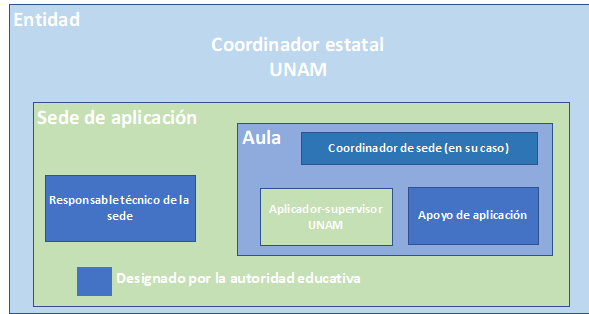


Figura 5. Estructura operativa en la entidad.

En el Anexo E se presentan los perfiles correspondientes a las distintas figuras que conformarán la estructura operativa del equipo encargado de planear, organizar, dirigir y supervisar los procesos de aplicación a nivel nacional y estatal.

**3.3.2 Acciones y responsabilidades del equipo encargado del operativo en campo**

**3.3.2.1 Planeación del operativo en campo**

La planeación del operativo en campo tiene dos propósitos principales: 1) Programar las tareas requeridas antes, durante y después de la aplicación de instrumentos señalando sus responsables y 2) Establecer dónde, cuándo y quiénes participarán en la aplicación como responsables del proceso y como sustentantes en cada sede de aplicación.

**3.3.2.2. Organización**

La organización de la aplicación asegurará que todas las figuras con responsabilidades directas en campo cuenten con la formación y la información necesaria para la realización de sus tareas.

Esto implica la definición de procedimientos a nivel regional, estatal y de sede para efectuar la aplicación de instrumentos, así como la determinación de la información y formación de las figuras con tareas relacionadas directamente con las labores de campo.

Con tal objeto se elaborarán documentos informativos (manuales y protocolos) para los coordinadores regionales, los coordinadores estatales y los aplicadores-supervisores destacados por la Facultad de Psicología; además, siempre previo acuerdo con la USICAMM y si se considerara pertinente, un volante informativo para las maestras y maestros sustentantes que recibirán justo antes de iniciar la resolución de su examen.

**3.3.2.3. Dirección**

La dirección del proceso tiene como propósito la atención personalizada a las labores directas en campo antes, durante y después de la aplicación, para garantizar su apego a los procedimientos establecidos en manuales y protocolos, así como el control y resolución de incidencias o eventualidades con la orientación y apoyo permanente de la coordinación nacional de la aplicación de instrumentos.

### 

### 

**3.3.2.4. Supervisión**

La supervisión del proceso de aplicación de instrumentos tiene como objetivo corroborar que las acciones de organización y gestión antes, durante y después de la resolución de los instrumentos funcionen de acuerdo a lo planeado. La supervisión de la calidad en el cumplimiento de las tareas establecidas en el cronograma de trabajo será en todos los casos una responsabilidad del superior jerárquico, dentro de la estructura operativa y quedará bajo su exclusiva responsabilidad, exclusivamente con dos excepciones: lo que corresponde a la capacitación de aplicadores-supervisores y a la propia aplicación de instrumentos.

**3.3.2.5. Control**

Se incorporan procedimientos que garanticen la calidad del proceso de aplicación en términos de transparencia y orden con lo que se propiciará que todos los sustentantes cuenten con las condiciones óptimas para la resolución del instrumento que les corresponde, y a la vez se documenta a detalle las condiciones bajo las cuales se efectuó la resolución de los instrumentos. En consecución de tal objetivo, se establecerán estrategias específicas para el control de la aplicación en los distintos niveles de aplicación (por región, entidad federativa, sede y/o grupo-turno).

**3.3.3 Momentos clave en el desarrollo de los procesos de aplicación**

A continuación se presenta una breve descripción de las principales actividades a desarrollar como parte del trabajo para la planificación y organización de los procesos de aplicación en tres momentos clave: antes, durante y después de la aplicación.

**3.3.3.1 Antes de la aplicación**

Previo a cada proceso de aplicación, será necesario elaborar un presupuesto detallado que considere las necesidades a cubrir para el desarrollo de la aplicación, así como un cronograma específico que considere todas las etapas de aplicación y permita identificar las fechas clave para el cumplimiento de las actividades comprometidas como parte del proceso de aplicación.

A partir de estos documentos, en conjunto con los referentes normativos que la Secretaría ponga a disposición de la Facultad de Psicología para identificar cualquier acción que se considere de interés particular para garantizar la calidad técnica de los procesos de aplicación de los instrumentos, se integrará un documento medular denominado “Condiciones generales de la aplicación”, donde se explicitará la organización general del proceso y el cronograma de actividades a realizar al interior de las sedes y cada aula de aplicación y que, a su vez, será la base para el desarrollo de las tareas subsecuentes, señaladas a continuación:

1. El diseño y elaboración de la “Estrategia de capacitación para la estructura operativa”, donde se especificarán las características de las acciones de capacitación y actualización a realizar con los coordinadores regionales y estatales, así como los talleres de capacitación desarrollados para los aplicadores-supervisores y los aplicadores de apoyo. Se elaborarán y difundirán materiales dirigidos a los distintos integrantes de la estructura operativa, señalando sus funciones y responsabilidades, así como la calendarizarción de las actividades a realizar en cada una de las etapas del proceso de aplicación. La disposición de estos documentos garantiza que las personas con responsabilidades en las labores de campo cuenten siempre con un material de consulta para orientar su labor.

2) El establecimiento de un “Procedimiento para el control y documentación de la aplicación en campo” que permita documentar lo ocurrido durante la aplicación a nivel nacional y que facilite la disposición de información para la atención oportuna de incidencias. Para ello, se considera el uso de reportes de incidencia organizados en los siguientes formatos de control:

1. Formato por Grupo-turno de aplicación: Lista de sustentantes por grupo-turno de aplicación que podrá ser en formato impreso o electrónico, en el que se presentarán datos de identificación de los sustentantes y claves de acceso, donde se registrará su asistencia. Además este formato deberá contener información respecto a las capacidades para contestar el cuestionario de manera independiente. Si el sustentante tiene alguna discapacidad que requiera de apoyo físico o humano para dar respuesta a la prueba, el formato deberá especificar de qué discapacidad se trata y cuál es el apoyo requerido. Este formato permitirá garantizar que sean exclusivamente los docentes inscritos los que ingresen al aula de aplicación.
2. Formato por Sede de aplicación: Acta de aplicación por sede-turno de aplicación en la que se registrarán los sustentantes atendidos por grupo y las incidencias presentadas durante la aplicación.
3. Formato por Entidad: Reporte y avance de la aplicación por turno y sede en la entidad (formato exclusivamente electrónico).

3) Desarrollo y definición de la “Planeación operativa”, que implica una determinación formal de las sedes y grupos de aplicación a atender en cada uno de los días-turno de aplicación, para establecer la fecha y turno de aplicación que será asignado a cada sustentante.

4) Elaboración de la “Estrategia operativa”, donde los coordinadores estatales proporcionarán información operativa de interés para el control central de la aplicación e identificarán los requerimientos humanos y financieros requeridos de acuerdo con lo establecido en la “Planeación operativa”.

5) Definición de los perfiles específicos de cada uno de los integrantes de la estructura operativa estatal, para orientar el reclutamiento de los coordinadores estatales y su personal de apoyo (en caso de que así se requiera) y coordinadores regionales.

6) Diseñar y desarrollar los contenidos y materiales a utilizar durante las capacitaciones que se realizarán con los coordinadores regionales y estatales, y con los aplicadores-supervisores y de apoyo, así como preparar estrategias para brindar asesoría y seguimiento a las actividades a desarrollar en las entidades antes de la aplicación.

7) Desarrollo e implementación de estrategias para la verificación de las condiciones de infraestructura y conectividad con las que cuentan las sedes de aplicación, que permitan orientar el seguimiento de las acciones administrativas para la disposición oportuna de recursos en campo y de las gestiones administrativas para el desarrollo de estrategias de resolución de problemas de conectividad en campo. Los coordinadores estatales estarán a cargo de garantizar, mediante georreferenciación, la accesibilidad de cada una de las sedes de aplicación, identificando para ello las rutas de acceso posibles.

**3.3.3.2 Durante de la aplicación**

Cada sustentante registrado deberá presentarse en el día y turno que le fue asignado, en la sede que le corresponde. El aplicador-supervisor y el aplicador de apoyo estarán a cargo de verificar la validez de la documentación que presenten los sustentantes para identificarse y posteriormente, de proporcionarles su nombre de usuario y contraseña. En caso de que se presente alguna situación particular, deberán llenarse los reportes del avance e incidencias presentadas en cada sede de aplicación.

En caso de presentarse alguna eventualidad o situación en que se requiriera asesorar, los coordinadores estatales y regionales deberán estar en contacto constante con los responsables técnicos de cada sede de aplicación y del personal que en estas se encuentren laborando, a fin de poder responder de manera oportuna a cualquier nueva necesidad detectada.

**3.3.3.3 Después de la aplicación (al concluir el periodo de aplicaciones del semestre)**

Al terminar los procesos de aplicación, deberá integrarse un expediente que dé cuenta de las eventualidades que pudieran haberse presentado durante el mismo y que pudieran estar poniendo en riesgo la validez de las mediciones obtenidas.

**III. Planteamiento de valor agregado**

Las mejoras e innovaciones metodológicas propuestas en el presente documento para el diseño, desarrollo, validación, aplicación, calificación y análisis de resultados de los instrumentos de valoración a desarrollar para el SISAP, cumplen con los estándares más altos de calidad reconocidos internacionalmente y se encuentran respaldados por la experiencia del cuerpo de investigadores que estarán a cargo de la realización del proyecto. Dichas mejoras e innovaciones metodológicas pueden identificarse a lo largo de distintas líneas de acción:

* Implementación de un programa de validez basado especialmente en evidencias del diseño de los ítems y la implementación de técnicas de pensamiento en voz alta en diferentes subpoblaciones para verificar el diseño y adaptaciones de los ítems, la equidad de la prueba, procesos, así como estrategias de respuesta y tiempos de respuesta.
* Certificación y conformación de grupos de supervisores locales, estatales, regionales y nacionales que vigilen y den constancia de la correcta aplicación de los instrumentos.
* Elaboración de protocolos de actuación y manuales de aplicación (digitales).
* Definición de ruta crítica y visión a futuro con la definición de informes de resultados, basados en modelos de usos y esquemas de tomas de decisiones de actores educativos de primer orden de relevancia para la mejora de la práctica educativa.
* Integración de reportes de resultados automatizados e individualizados que permitan conocer, de manera inmediata y específica, las necesidades de formación y las áreas de fortaleza de los sustentantes, acompañados de modelos para el uso de estos resultados, para orientar la toma de decisiones de los actores involucrados en favor de la formación continua para la mejora efectiva de la práctica educativa.
* Capacitación del personal técnico de la USICAMM con un curso extensivo de corte teórico-práctico para el desarrollo de habilidades técnicas y metodológicas de última generación, permitiéndoles posicionarse como líderes dentro de un equipo especializado dentro del Servicio Público Federal capacitado para atender las demandas de evaluación y medición que se les presenten en un futuro, cumpliendo con los estándares más altos de calidad técnica a nivel internacional.

**Anexo A**

**Algoritmos de Aprendizaje Profundo**

El objetivo central de los algoritmos de aprendizaje profundo (*Deep Learning Algorithms*) es estimar los valores de los parámetros comprendidos en una función: . Digamos que: es un texto del conjunto de textos: con índice . Digamos que para todos los: existe un atributo: que toma uno de los siguientes valores: 0,1. Tenemos: . Digamos que para todo: existe un conjunto de parámetros: , que representan las unidades fundamentales del texto (letras o palabras), tenemos: .

Digamos que *p* tiene siempre 10 unidades, que es decir los textos tienen 10 palabras: . Si modelamos: U para predecir el valor de: para cualquier: , lo primero que asumimos es que para todo con atributo , la semejanza con el subconjunto que cumple será mayor que la semejanza con el subconjunto que cumple , por lo tanto tenemos: .

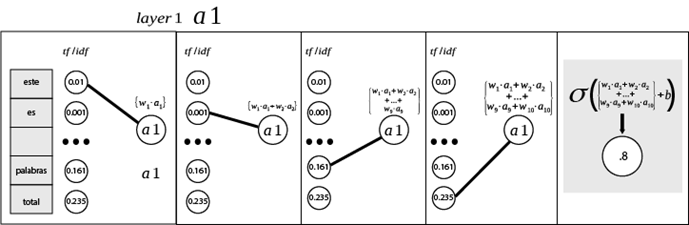
Durante el entrenamiento los algoritmos tienen dos fases generales: propagación frontal y propagación inversa (*Forward Propagation* & *Backward Propagation*). La propagación frontal estima los parámetros necesarios para predecir *y* tenemos: donde es la predicción de , mientras que y son los parámetros a estimar. Teniendo y , el algoritmo calcula el valor de por medio de la función de activación:. Esta función decide el tipo de output o formato de la predicción de . El algoritmo más común para es RELU (*Rectified Linear Unit),* que define como: . En los ejemplos se refiere a la función *tanh*. Funciones como éstas introducen no-linearidad al modelo, que es lo que permite encadenar capas de entrenamiento y es además lo que les da el adjetivo de *profundo.* De otra forma, una función de activación lineal, sin importar el número de capas que tenga, es indistinguible de una simple regresión logística.

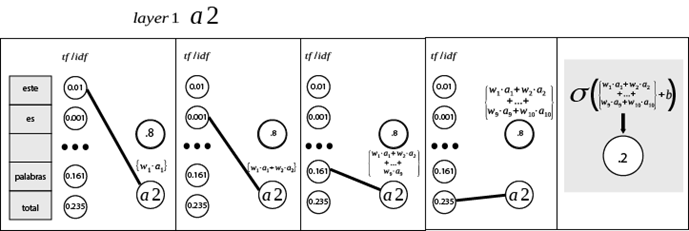
Teniendo , el algoritmo calcula el error total del modelo, referido comúnmente como: *loss function,* que se reduce a la diferencia entre la predicción y el atributo real, tenemos:

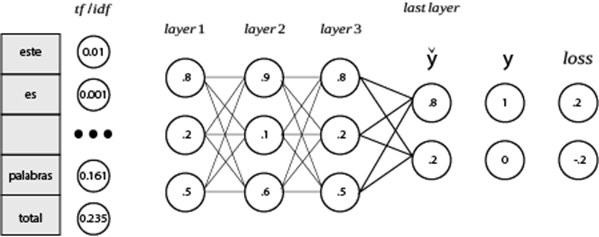
Teniendo: finaliza la propagación frontal.

Digamos que = “este es el texto de x1 tiene 10 palabras total”.

Digamos que: , representa cada nodo y su peso. La propagación frontal puede dibujarse como sigue:

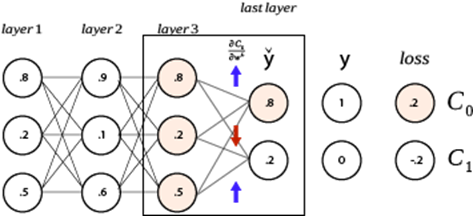




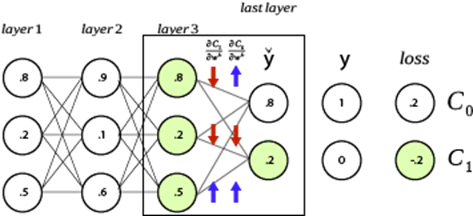


El siguiente proceso: La propagación inversa (*Backwards Propagation)*, es probablemente el proceso esencial de éstas máquinas o algoritmos, como hemos aclarado. Tiene la función de optimizar el error de la última capa. Podemos describir una capa como: , donde L es el índice de la capa, los pesos, es el resultado de la capa previa. Llamaremos cost: , al resultado de *loss function* de la neurona con índice 0 de la última capa, en este caso:

Si queremos encontrar la proporción con que los cambios en se reflejan en , formalmente realizamos la siguiente derivada: , que por la regla de cadena de poder puede definirse como:



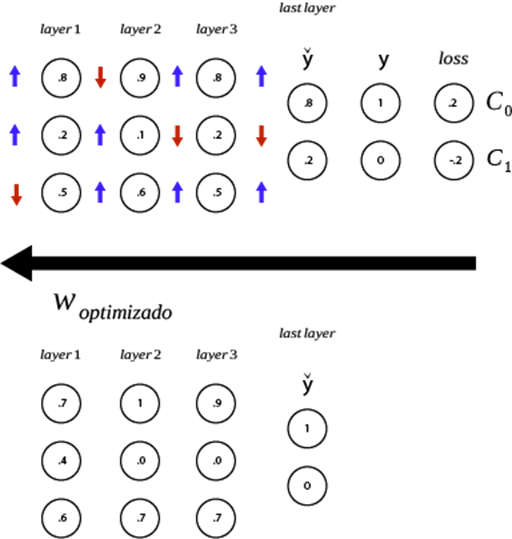
Lo siguiente es encontrar :



Teniendo , y se estima , restando a el producto de y un hiperparámetro: , llamado *tasa de aprendizaje*, tenemos: .

Con podemos encontrar e iterando optimizar todos los de

Tenemos:



Ahora podemos decir que el modelo ha entrenado con la instancia . Iterando sobre todos los elementos de U se completa una época de entrenamiento. Iterando hasta el número de épocas que necesite el modelo, el resultado final es una matriz de pesos.

Esta matriz tiene ahora la función de predecir el valor de para cualquier .

**Anexo B**

**Algoritmos que conforman el Sistema Predictor de Calidad (SPC)**

* **SimpleRNN (*Recursive Neural Network)***. Es un algoritmo con una estructura prácticamente indistinguible de los algoritmos de aprendizaje profundo. Este algoritmo ayuda a marcar un mínimo necesario de eficiencia en el modelo.
* **BidirectionalLSTM** es una clase de RNN, fue creada para solucionar un problema natural de las RNN, en que es tan pequeño que impide que la célula se actualice, esto podía dar lugar a grandes secciones de neuronas perpetuamente inactivas. LSTM reduce ese efecto al introducir reguladores -comúnmente tres, llamados: input gate, output gate y forget gate. Encargados de retener un valor significativo y decidir si debe ser actualizado a o debe permanecer con el valor anterior. Es llamado de memoria temporal y a largo plazo por su capacidad de retener valores en distintas unidades de tiempo. BidirectionalLSTM es una variación que analiza los textos de izquierda a derecha y de derecha a izquierda, lo que permite al algoritmo retener valores importantes sin importar el orden en que aparecen en el texto.
* **SeparableCNN2d** es una clase de CNN, red neuronal de convoluciones**,** creada para el análisis de información visual. Su estructura está basada en el funcionamiento del procesamiento visual humano. Su función general consiste en extraer patrones espaciales por medio de convoluciones, que son transformaciones locales de la matriz de pesos . A diferencia del algoritmo general descrito arriba, las máquinas de convoluciones realizan la propagación frontal construyendo en cada capa una nueva matriz de dimensiones distintas. Se trata de una variación que separa la matriz original a través de convoluciones 1x1, simplificando la matriz y obteniendo w con número menor de operaciones.
* **SVM**, **Gradiant Boosting** y**Regresión Logística** son algoritmos no relacionados con redes neuronales que serán usados como balance en el SPC.

**Anexo C**

**Características Mínimas de Hardware**

**Computadora con sistema operativo Windows**

- De 32-bit o 64-bit

- Sistema operativo Windows 7, Windows 8, Windows 10

- Procesador Pentium 4 o superior que soporte SSE2

- Memoria RAM de 512 MB / 2 GB para la versión de 64-bit

- 200 MB de espacio en disco duro

**Computadora con sistema operativo MACOS**

- Sistema operativo macOS 10.9, macOS 10.10, macOS 10.11, macOS 10.12, macOS 10.13, macOS 10.14, macOS 10.15

- Procesador Intel x86

- Memoria RAM de 512 MB

- 200 MB de espacio en disco duro

- Sistema operativo GNU/Linux[[2]](#footnote-2)

**Requerimientos de Software**

* Firefox no se ejecutará totalmente sin las siguientes bibliotecas o paquetes:

1. GTK+ 3.4 or higher

2. GLib 2.22 or higher

3. Pango 1.22 or higher

4. X.Org 1.0 o superior (se recomienda 1.7 o superior)

5. libstdc++ 4.6.1 o superior

* Para una funcionalidad óptima, se recomienda contar con las siguientes bibliotecas o paquetes:

1. NetworkManager 0.7 o superior

2. DBus 1.0 o superior

3. GNOME 2.16 o superior

4. PulseAudio

Adicional a estos requisitos técnicos, se ejecutarán pruebas simultáneas de la conexión de los equipos a Internet, donde se registre en un listado (tabulado), la velocidad de carga y descarga al igual que la latencia de las respuestas.

**Anexo D**

**Propuesta de Capacitación para el personal técnico de la USICAMM que conformará el Comité de Validación**

|  |
| --- |
| *Objetivo general:*  Promover la comprensión y el desarrollo de conocimientos y habilidades psicométricas, a partir del análisis de la Teoría Clásica de los Tests, la Teoría de Respuesta al Ítem y los Modelos de Diagnóstico Cognitivos, a fin de que los asistentes puedan aplicar estos conocimientos y habilidades en su quehacer diario en el servicio público federal. |
| *Objetivos específicos:*  1. Los asistentes describirán los fundamentos de la Teoría de la Medición en la ciencia y explicará la lógica en diferentes enfoques de la medida en la Evaluación Educativa.  2. Los asistentes comprenderán el modelo y los supuestos de la Teoría Clásica de los Tests (TCT), los supuestos, conceptos e indicadores y sus limitaciones en la medición de variables latentes complejas.  3. Los asistentes comprenderán el modelo de Análisis Factorial Exploratorio y Confirmatorio (AFE, AFC), los supuestos, parámetros y sus aplicaciones en la medición de variables latentes complejas.  4. Los asistentes comprenderán los principales modelos dicotómicos, politómicos y explicativos de la Teoría de Respuesta al Ítem (TRI), como el modelo de Rasch y los modelos de dos y tres parámetros, el modelo LLTM, entre otros, así como los principales procedimientos para estimar parámetros, funciones de información y la bondad de ajuste de los principales modelos de la TRI.  5. Los asistentes aplicarán los principales modelos dicotómicos y politómicos de la Teoría de Respuesta al Ítem (TRI), en escenarios educativos.  6. Los asistentes comprenderán los principales modelos de Diagnóstico Cognitivo (MDC), como el modelo DINA, DINO, G-DINA, entre otros así como los principales procedimientos para estimar parámetros y la bondad de ajuste de los principales modelos de Diagnóstico Cognitivo.  7. Los asistentes razonaran sobre el desarrollo y evolución de la validez através de sus diversas etapas, e incorporarán los nuevos estándares de medición de variables latentes complejas, para la formulación de instrumentos de medición estandarizados. |

**Índice Temático**

|  |  |  |  |
| --- | --- | --- | --- |
| Unidad | Temas | **Horas** | |
| *Teóricas* | *Prácticas* |
| 1 | Marco general de la Teoría Clásica de los Tests (TCT) | 3 | 0 |
| 2 | Modelo de la Puntuación Verdadera | 6 | 3 |
| 3 | Confiabilidad del Test en la TCT | 6 | 6 |
| 4 | Análisis factorial | 6 | 6 |
| 5 | Marco general de Teoría de Respuesta al Ítem (TRI) | 3 | 0 |
| 6 | Modelo de Rasch | 6 | 6 |
| 7 | Modelos logísticos de 2 y 3 parámetros | 6 | 6 |
| 8 | Modelos politómicos de la TRI | 6 | 6 |
| 9 | Modelos explicativos de la TRI | 6 | 6 |
| 10 | Modelos de Diagnóstico Cognitivo (MDC) | 6 | 6 |
| 11 | Teoría de la validez | 6 | 0 |
| *Horas:* | | 60 | 45 |
| *Total de horas* | | 105 | |

|  |  |
| --- | --- |
| **Unidad** | **Tema y Subtemas** |
| 1 | 1 Marco general de Teoría Clásica de los Tests (TCT).  1.1 Antecedentes de la Teoría Clásica de los Tests.  1.2 Desde Spearman hasta Lord y Novick.  1.3 Lee Cronbach. |
| 2 | 2 Modelo de la Puntuación Verdadera.  2.1 Modelo de la puntuación verdadera.  2.2 Supuestos del modelo de puntuación verdadera.  2.3 Paralelismo y equivalencia de las medidas en la TCT. |
| 3 | 3 Confiabilidad del Test en la TCT.  3.1 La confiabilidad en la TCT.  3.2 Concepto de error y tipos de errores en la medición.  3.3 Estimación del error estándar de medida.  3.4 Corrección por atenuación de los errores de medición.  3.5 Limitaciones de la TCT. |
| 4 | 4 Análisis factorial.  4.1 Marco general del análisis factorial.  4.2 Análisis Factorial Exploratorio (AFE).  4.3 Análisis Factorial Confirmatorio (AFC).  4.4 Aplicaciones y ejemplos en el contexto educativo. |
| 5 | 5 Marco general de Teoría de Respuesta al Ítem (TRI).  5.1 Introducción a la Teoría de Respuesta al Ítem.  5.2 Curva característica del ítem CCI.  5.3 Supuestos de los principales modelos de la TRI. |
| 6 | 6 Modelo de Rasch.  6.1 Características del modelo de Rasch.  6.2 Estimación de parámetros.  6.3 Evaluación de bondad de ajuste del modelo de Rasch.  6.4 Programas informáticos para aplicar el modelo de Rasch. |
| 7 | 7 Modelos logísticos de 2 y 3 parámetros.  7.1 Explicación de los modelos de la TRI de dos y tres parámetros.  7.2 Evaluación de bondad de ajuste de los modelos IRT.  7.3 Programas informáticos para aplicar los modelos de 2 y 3 parámetros. |
| 8 | 8 Modelos politómicos de la TRI.  8.1 El modelo de crédito parcial.  8.2 El modelo de respuesta graduada.  8.3 El modelo de respuesta nominal.  8.4 El modelo para ítems de opción múltiple. |
| 9 | 9 Modelos explicativos de la TRI.  9.1 La relación entre los modelos TRI y los modelos lineales generalizados.  9.2 Un modelo explicativo para el parámetro de personas: El modelo de regresión latente de Rasch.  9.3 Un modelo explicativo para el parámetro de los ítems: El modelo logístico lineal de Fischer (1973).  9.4 Modelos explicativos para personas e ítems: Una ilustración. |
| 10 | 10 Modelos de Diagnóstico Cognitivo (MDC).  10.1 El modelo DINA.  10.2 El modelo DINO.  10.3 El modelo G-DINA.  10.4 Aplicaciones de los MDC. |
| 11 | 11 Teoría de la validez.  11.1 Revisión histórica del concepto de validez.  11.2 Etapa de cristalización (1921-1951).  11.3 Etapa de fragmentación (1952-1974).  11.4 Etapa de reunificación (1975-1999).  11.5 Etapa de deconstrucción (2000-Actualidad). |

|  |
| --- |
| *BIBLIOGRAFÍA BÁSICA:*  Abad, F.J., Olea, J., Ponsoda, V., García, C. (2011). *Medición en ciencias sociales y de la salud.* Madrid: Síntesis.  Baker, F.B., & Kim, S.H. (2017). *The basics of item response theory using R*. Madison, WI, USA: Springer Nature.  Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*.  Mahwah, N.J.: Lawrence Erlbaum Associates.  Linn, R. (Ed.) (1989). *Educational measurement*. New York: MacMillan.  Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Beverly Hills, CA: Sage.  Martin, P. & Bateson, P. (1991). *La medición del comportamiento.* Madrid: Alianza  Martínez-Arias, R. (1996). *Psicometría*: Teoría de los tests psicológicos y  educativos. Madrid: Síntesis.  Michell, J. (1990). *An introduction to the logic of psychological measurement*.  Hillsdale New Jersey: Lawrence Erlbaum.  Pendiente handbook classification models    *BIBLIOGRAFÍA COMPLEMENTARIA:*  Baker, F. B. (2004). *Item Response Theory: Parameter Estimation Techniques*. New York: Marcel Dekker.  De Mars, C. (2010). *Item Response Theory.* *Series Understanding Statistics*. New York Oxford University Press.  Lord F.M. y Novick, M.R. (1968). *Statistical* *Theories of Mental Test Scores.* Reading Mass, Addison–Wesley.  Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*.  Madrid: Pirámide.  Muñiz, J. (1996). *Psicometría*. Madrid: Universitas.  Muñiz, J. (2003). *Teoría Clásica de los Tests*. Madrid: Pirámide  Rao, C. R. & Sinharay, S. (2007). *Handbook of Statistics*: *Psycometrics,* Vol. 26,  Amsterdam: Elsevier.  Santisteban, C. (2009). *Principios de Psicometría*. Madrid: Editorial Síntesis.  van der Linden, W. J. (Ed.) (2016). *Handbook of item response theory: Models, statistical tools, and applications* (Vols. 1–3). Boca Raton, FL: Chapman & Hall/CRC.  Wright, B. D & Mock, M. (2004). An overview of the family of Rasch Measurement Models. In B. Wright, et al., *Introduction to Rasch Measurement: Theory, Models and Application.* JAM Press. |

**Anexo E**

**Perfiles y funciones asociadas con las distintas figuras comprendidas en la estructura operativa del equipo encargado de los procesos de aplicación en campo**

1. **Coordinadores regionales**

|  |  |
| --- | --- |
| **Perfil** | **Funciones** |
| * Licenciatura concluida. * Nivel intermedio en el uso de Microsoft Office. * Uso de correo electrónico e Internet. * Experiencia de al menos 3 años en planeación, diseño y seguimiento de operativos en campo, supervisión y aplicación de instrumentos diseñados para el levantamiento de información. * Experiencia en administración de proyectos y liderazgo de equipos de trabajo. * Orientación a resultados, comportamiento ético, trabajo en equipo y amplia capacidad de negociación. * Disponibilidad para viajar. | * Dirigir, orientar y asesorar el correcto desarrollo de las actividades operativas de acuerdo con lo establecido por la coordinación nacional. * Relacionarse con las actividades administrativas de todos los niveles de operación. * Garantizar que los procesos estipulados por la UNAM se lleven a cabo en tiempo y forma a nivel regional de acuerdo con el calendario establecido. * Asesorar y apoyar, en su caso, a las autoridades educativas estatales para el trabajo conjunto relacionado con la aplicación de instrumentos. * Verificar que los coordinadores estatales a su cargo cuenten con los recursos e insumos necesarios para desarrollar las operaciones en campo. * Definir y proponer soluciones a las potenciales contingencias antes, durante y después de la aplicación en las entidades a su cargo. * Ser responsable y garantizar que las comprobaciones de recursos financieros del personal de honorarios que participe en la región a su cargo se realicen con transparencia y en los tiempos establecidos. * Ser responsable y garantizar que el personal asignado a la región a su cargo, se encuentre oportunamente incluido en la póliza del aseguramiento colectivo que será adquirida, y realizará las altas y bajas de personal oportunamente; para que se encuentren asegurados antes, durante y después de desempeñar sus funciones operativas en campo. |

1. **Coordinadores Estatales**

|  |  |
| --- | --- |
| **Perfil** | **Funciones** |
| * Licenciatura trunca. * Nivel intermedio en el uso de Microsoft Office. * Uso de correo electrónico e Internet. * Experiencia de un año en planeación, seguimiento y supervisión de aplicaciones de instrumentos en el ámbito educativo. * Experiencia en la impartición de capacitación para el desarrollo de levantamientos de datos. * Experiencia en la coordinación de equipos de trabajo. * Experiencia en administración de proyectos y liderazgo de equipos de trabajo. * Orientación a resultados, comportamiento ético, trabajo en equipo y amplia capacidad de negociación. * Disponibilidad para viajar. | * Dirigir y supervisar el desarrollo de las actividades operativas relacionadas con la aplicación de instrumentos de acuerdo con lo establecido por la coordinación nacional. * Reclutar, siempre que cuente con la autorización de la coordinación nacional, al personal que le apoyará directamente durante el proceso (apoyo operativo y enlace administrativo). * Reclutar a los aplicadores-supervisores necesarios para el desarrollo de la aplicación. * Reclutar al personal de apoyo a personas con discapacidad. * Capacitar al personal de apoyo a personas con discapacidad. * Capacitar a los aplicadores-supervisores. * Relacionarse con las actividades administrativas de todos los niveles de operación. * Garantizar que los procesos establecidos por la Facultad de Psicología se lleven a cabo en tiempo y forma con base en el calendario establecido. * Coordinarse con la autoridad educativa local para el desarrollo del trabajo conjunto en la aplicación de instrumentos. * Asegurar que se disponga de los recursos e insumos necesarios para desarrollar el trabajo de campo. * Definir y proponer soluciones a las potenciales contingencias antes, durante y después de la aplicación en la entidad. * Responsabilizarse de que el personal asignado a la entidad a su cargo se encuentre oportunamente incluido en la póliza de aseguramiento colectivo correspondiente al proceso de aplicación y de realizar las altas y bajas de personal asegurado de acuerdo a lo indicado para garantizar que todos los participantes que corresponda cuenten con seguro mientras se encuentren efectuando el trabajo de campo o en los traslados inherentes al mismo. * Supervisar y garantizar que las comprobaciones de recursos financieros relacionadas con el desarrollo de la aplicación como las que correspondan al personal de honorarios que participe en la entidad a su cargo se realicen con transparencia y en los periodos establecidos. |

1. **Apoyos operativos**

|  |  |
| --- | --- |
| **Perfil** | **Funciones** |
| * Licenciatura trunca o bachillerato concluido * Uso intermedio de Microsoft Office * Uso de Internet y correo electrónico * Experiencia de un año en el apoyo a la planeación y seguimiento de operativos de levantamiento de datos en campo. * Comportamiento ético, capacidad para el trabajo en equipo y disponibilidad para viajar. | * Participar en la organización de las actividades operativas de la aplicación * Apoyar en el proceso de reclutamiento y selección de aplicadores-supervisores. * Coadyuvar en el control del avance y seguimiento del operativo en campo. * Apoyar en las actividades administrativas, de organización y supervisión en la sede a la que sea asignado. |

1. **Enlaces administrativos**

|  |  |
| --- | --- |
| **Perfil** | **Funciones** |
| * Licenciatura trunca o bachillerato concluido * Uso intermedio de Microsoft Office * Uso de Internet y correo electrónico * Experiencia de un año en la administración de recursos económicos * Experiencia de un año en el apoyo a la realización de levantamientos de datos en el ámbito educativo. * Comportamiento ético, capacidad para el trabajo en equipo y disponibilidad para viajar. | * Apoyar en la revisión de comprobaciones de recursos financieros del personal de honorarios que participe en la entidad asignada. * Coadyuvar en la validación de la inclusión del personal en la póliza del aseguramiento colectiva y en la revisión de que los datos requeridos estén completos y sean correctos. * Apoyar en la realización de altas y bajas de personal con la finalidad de que se garantice el aseguramiento del personal participante en la aplicación. * Apoyar en las actividades propias del proceso de reclutamiento y selección de personal operativo. * Apoyar en las demás actividades administrativas que le sean encomendadas por su superior en el ámbito de su competencia |

1. **Aplicador-supervisor**

|  |  |
| --- | --- |
| **Perfil** | **Funciones** |
| * Licenciatura trunca o bachillerato concluido * Manejo intermedio de office * Uso de correo electrónico e Internet * Experiencia como aplicador de instrumentos en el ámbito educativo * Comportamiento ético, disciplina y disponibilidad para viajar. | * Dirigir el correcto desarrollo de la aplicación de instrumentos en el grupo-turno que le sea asignado. * Coadyuvar en lo que se le requiera para que se cuente con los recursos e insumos necesarios para desarrollar las operaciones en la sede asignada. * Si cuenta con la información de potenciales contingencias en la sede de aplicación que le fue asignada, informarla a su coordinador estatal. * Plasmar en el acta correspondiente a su grupo-turno la información requerida. * Ser responsable y garantizar que sus comprobaciones de recursos financieros se realicen con transparencia en los tiempos establecidos. * Asistir al taller de capacitación correspondiente. |

1. **Aplicador-apoyo**

|  |  |
| --- | --- |
| **Perfil** | **Funciones** |
| * Licenciatura trunca o bachillerato concluido. * Experiencia en aplicación de instrumentos en el ámbito educativo. * Comportamiento ético, disciplina y disponibilidad para viajar. | * Apoyar a aquellos sustentantes que se registraron con discapacidad a ingresar a la sede de aplicación, con base en las instrucciones del aplicador-supervisor. * Asistir al taller de capacitación correspondiente. |

**Referencias:**

* Abad, F. J. (2011). Medición en ciencias sociales y de la salud.
* Andrade, H., & Du, Y. (2005) Student perspectives on rubric-referenced assessment. *Practical Assessment, Research and Evaluation,* 10 (3), 1-11.
* Bergstrom, B. A., & Gershon, R. C. (1995). 8. Item Banking.
* Birenbaum, M., & Tatsuoka, K. K. (1993). Applying an IRT-based cognitive diagnostic model to diagnose students' knowledge states in multiplication and division with exponents. Applied measurement in education, 6(4), 255-268.
* Brown, J. & Burton, R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 2, 155-192.
* Chen, Y. & Macdonald, G. (2011). Validating Cognitive Sources of Mathematics Item Difficulty: Application of the LLTM to Fraction Conceptual Items. Psychological Assessment, 7, 74–93.
* Cohen, Y. (2019). The Handbook of Cognition and Assessment; Frameworks, Methodologies, and Applications.
* Ericsson, K. & Simon, H. (1984). Protocol analysis: verbal reports as data. Cambridge: MIT Press.
* Ericsson, K. A., & Simon, H. A. (1993). Protocol Analysis: Verbal Reports as Data. Cambridge, MA: MIT.
* Ferrara, S., Lai, E., Reilly, A., Nichols, P. D., Rupp, A. A., & Leighton, J. P. (2017). Principled approaches to assessment design, development, and implementation. The Handbook of Cognition and Assessment, 41-74.
* Fredericksen, J. (1980). Component skills in Reading: measurements of individual differences thought chronometric analysis. In R. E. Snow, P-A. Federico & W. E. Montage (Eds.), Aptitude, learning, and instructions: Cognitive process analyses of aptitude, Vol. 1, (pp. 105-138). Hillsdale: Lawrence Erlbaum.
* García Sanz, M. P. (2014). La evaluación de competencias en Educación Superior mediante rúbricas: un caso práctico. Revista Electrónica Interuniversitaria de Formación del Profesorado, 17 (1), 87-106.
* Gierl, M. J., & Haladyna, T. M. (Eds.). (2012). *Automatic item generation: Theory and practice*. Routledge.
* Gierl, M., Leighton, J., Changjiang, W., Jiawen, Z., Rebecca, G. & Tan, A. (2009). Validating Cognitive Models of Task Performance in Algebra on the SAT. Research Report 2009-3. College Board, Research Report, 2009(3), New York.
* Haladyna, T. Downing, S. M. & Rodríguez, M. C. (2002). A review of multiple-choice item writing guidelines for classroom assessment. Applied Measurement in Education, 15(3), 309–334.
* Instituto de Evaluación e Ingeniería Avanzada, (s.f.) Banco de reactivos. Recuperado de:<http://www.ieia.com.mx/web/index.php/software-menu-p/banco-de-reactivos-p>
* Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for Fusion Model application to LanguEdge assessment. Language Testing, 26(1), 031-73.
* Johnstone, C. (2003). Improving validity of large-scale tests: Universal design and student performance (Technical Report 37). Minneapolis: National Center on Educational Outcomes.
* Junior, M. L., & Godinho Filho, M. (2010). Variations of the kanban system: Literature review and classification. International Journal of Production Economics, 125(1), 13-21.
* Ketterlin-Geller, L. R., & Yovanoff, P. (2009). Diagnostic assessments in mathematics to support instructional decision making. Practical Assessment, Research & Evaluation, 14(16), 1-11.
* Lee, Y. W., & Sawaki, Y. (2009). Application of three cognitive diagnosis models to ESL reading and listening assessments. Language Assessment Quarterly, 6(3), 239-263.
* Leighton, J. (2009). Two Types of Think Aloud Interviews for Educational Measurement: Protocol and Verbal Analysis Paper presented for symposium How to Build a Cognitive Model for Educational Assessments at the 2009 annual meeting of the National Council on Measurement in Education (NCME), April, 14-16.
* Leighton, J. & Gierl, M. (2007). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees’ thinking processes. Educational Measurement: Issues and Practice, 26(2), 3-16.
* Li, H. (2011). A cognitive diagnostic analysis of the MELAB reading test. Spaan Fellow, 9, 17-46.
* Li, H., Hunter, C. V., & Lei, P. W. (2016). The selection of cognitive diagnostic models for a reading comprehension test. Language Testing, 33(3), 391-409.
* Ma, L. Çetin, E. y Green, K. (2009, April). Cognitive assessment in Mathematics with the Least Squares Distance Method. Artículo presentado en el Congreso anual de la AERA 2009. San Diego.
* Martínez-Rojas, J. G. (2008). Las rúbricas en la evaluación escolar: su construcción y su uso. Avances en medición, 6(129), 38.
* Matsunaga, M., (2010) How to factor-analyze your data right: Do’s, don’ts, and How-to’s. International Journal of Psychological Research, Vol. 2. No. 1.
* Montero, D. H., Monfils, L., Wang, J., Yen, W. M., Julian, M. W., & Moody, M. (2003, April). Investigation of the application of cognitive diagnostic testing to an end-of-course high school examination. In annual meeting of the National Council on Measurement in Education, Chicago, IL.
* Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates, 71(2001), 2001.
* Pérez-Morán, J. C. (2014). Análisis del aspecto sustantivo de la validez de constructo de una prueba de habilidades cuantitativas (tesis doctoral). Universidad Autónoma de Baja California, Baja California, México.
* Pérez-Morán, J. C., Contreras, S., Hernández, E. M., Olivares, C., Chan, P., y Díaz, K. M. (2014). Análisis de las evidencias de validez basadas en el proceso de respuesta de las pruebas de ENLACE MS de Habilidad lectora y Matemáticas. Reporte técnico. México: INEE.
* Pérez-Morán, J. C.; Larrazolo, N.; Backhoff, E.; y Guaner, R. (2015). Análisis de la estructura cognitiva del área de habilidades cuantitativas del EXHCOBA mediante el modelo LLTM de Fisher. Revista Internacional de Educación y Aprendizaje, 3(1), 25-38. <http://coleccionderevistasdeeducacionyaprendizaje.cgpublisher.com/product/pub.329/prod.5> ISSN 2255-453X.
* Pérez-Morán, J. C.; Vázquez-Lira, R.; & Rojas, G. (2019). Diagnóstico Nacional de las habilidades básicas en Matemáticas de Sexto de Primaria: Resultados de 2015. México: RIMEDIE.
* Posner, M. I. (1978). Chronometric exploration of mind. New York: John Wiley.
* Posner,M. I., & Rogers, M. G. K. (1978). Chronometric analysis of abstraction and recognition. In W. K. Estes (Ed.) (1978). Handbook of learning and cognitive processes (vol. 6). Hillsdale, N. J.: Lawrence Erlbaum Associates.
* Ravand, H. (2016). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. Journal of Psychoeducational Assessment, 34(8), 782-799.
* Revuelta, J. y Ponsoda, V. (1998). Un test adaptativo informatizado de análisis lógico basado en la generación automática de ítems. Psicothema, 10, 753-760.
* Romero, S., Ponsoda, V., y Ximenez, C. (2008). Análisis de un test de aritmética mediante el modelo logístico lineal de rasgo latente 1. Revista Latinoamericana de Psicología, 40, 85–95.
* Rupp, A. A., Templin, J., & Henson, R. A. (2010). Diagnostic assessment: Theory, methods, and applications. New York: Guilford.
* Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometrika monograph supplement.
* Samejima, F. (2016). Graded response models. In Handbook of item response theory, volume one (pp. 123-136). Chapman and Hall/CRC.
* Snow, R. & Lohman, D. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), Educational measurement (3a. ed.), pp. 263-331. New York: Macmillan Publishing Co.
* Sternberg, R. (1977). Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Oxford: Lawrence Erlbaum.
* Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of language and social psychology, 29(1), 24-54.
* Thompson, S., Johnstone, C. & Thurlow, M. (2002). Universal design applied to large scale assessments (Synthesis Report 44). Minneapolis, MN: National Center on Educational Outcomes.
* Van der Linden, W. J. (Ed.). (2017). Handbook of Item Response Theory, Volume Three: Applications. CRC Press.
* Weiss, D. J. (2013). Item banking, test development, and test delivery.
* Yang, X. & Embretson, S. (2007). Construct Validity and Cognitive Diagnostic Assessment. In Leighton, J. y Griel, M. (Edit.). Cognitive diagnostic assessment for education: Theory and applications, pp. 85-118. Cambridge: Cambridge University Press.

1. <https://aws.amazon.com/compute/sla/> [↑](#footnote-ref-1)
2. Los distribuidores GNU / Linux pueden proporcionar paquetes para su distribución que tienen diferentes requisitos. [↑](#footnote-ref-2)