

**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. Arturo Bouzas Riaño, Dra. Lucía Monroy, Dr. Iwen Leenen, Dra. Georgina (¿?), Mtra. Eleonora Rubio Ruiz, Lic. Adriana F. Chávez De la Peña

Coyoacán, Ciudad de México a XX de enero de 2020

**I. Introducción.**

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

Fundada en 1973, la Facultad de Psicología se ha mantenido en constante proceso de evolución para estar al día con las líneas de investigación 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 vasta trayectoria en la formación de psicólogos, con bases sólidas en estadística y metodología, así como en la 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, llevando a cabo de manera eficaz 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 la Facultad, han conferido a su cuerpo académico y de investigadores un gran respeto y reconocimiento por parte de la comunidad académica y científica.

En términos de la experiencia que tiene la Facultad de Psicología participando, coordinando y dirigiendo proyectos de alto impacto donde el foco de atención se ha centrado en el desarrollo, aplicación y calificación e interpretación de instrumentos de evaluación aplicados a gran escala, se destacan los siguientes:

* En 2018, se trabajó el proyecto PENUD…
* En 2015,

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.

Con sustento en todo lo previamente expuesto, 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 valoración de la práctica docente en el país.

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

**Enfoque y características generales de la propuesta metodológica.**

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 y 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).

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, permitiendo así el trazo de perfiles individualizados que orienten 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 simple o compleja (Rupp y Templin, 2008). Son modelos probabilísticos en el sentido de que cada uno 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 granular 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/o estrategias que se asumen como centrales al objeto de estudio, lo que permite hacer inferencias precisas acerca de cuáles de estos elementos se dominan y cuáles hace falta reforzarse.

De esta forma, los MDC permiten a todo agente interesado en la aplicación de un instrumento conocer con detalle las áreas de fortaleza y debilidad 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 (ejemplo de ello se puede 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; Pérez-Morán, Larrazolo, Backhoff y Guaner, 2015; 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).

De acuerdo con Leighton y Gierl (2007a) la psicología cognitiva puede aportar al avance de la medición educativa, en tanto que cuenta con un sólido cuerpo de teorías y modelos cognitivos que pueden contribuir a identificar los procesos de respuesta evocados por pruebas ya existentes, y a organizar la elaboración de pruebas que permitan medir el dominio y la aplicación de las habilidades y conocimientos que forman parte del constructo que se quiere medir. Es en este sentido que Nichols (1994) se refiere a la Evaluación Diagnóstica Cognitiva (EDC), como un punto de inflexión donde la psicología cognitiva guía tanto el diseño de instrumentos de evaluación, como la interpretación de los resultados observados, a fin de poder elaborar diagnósticos específicos sobre los procesos cognitivos y las estrategias de respuesta aplicados por los sustentantes para su resolución (Rupp, 2007).

Para realizar una EDC con alta validez sustantiva es indispensable que todo el trabajo realizado, desde el diseño y desarrollo del instrumento hasta su calificación e integración de resultados, se realice bajo el marco de los MDC, aplicando estrategias que incidan en una mayor calidad técnica derivadas de la Psicometría y, construyendo un modelo sumativo derivado de las teorías y modelos trazados desde la Psicología Cognitiva y toda disciplina que guarde relación con el objeto de estudio (en el caso del ámbito educativo, la Pedagogía y otras Ciencias de la educación). Es decir, debe existir una teoría cognitiva que esté guiando la construcción de la prueba, definiendo qué atributos latentes son requeridos por cada reactivo y describiendo el proceso a través del cual dichas habilidades están ligadas para producir una respuesta correcta (Henson, Templin y Willse, 2009).

Es posible trazar la estructura subyacente de un instrumento ya existente, a fin de identificar cuáles son las habilidades, conocimientos, procesos u operaciones cognitivas requeridas para su resolución y poder trazar un diagnóstico cognitivo a partir de los resultados observados. Estas técnicas, conocidas en la literatura como de “retrofitting” permiten aprovechar la información recopilada en aplicaciones a gran escala para obtener un diagnóstico de alto impacto, que permita identificar las áreas de fortaleza y de mejora de cada sustentante a partir de su desempeño. Como un ejemplo, tómese el estudio realizado por Pérez-Morán, Vázquez-Lira y Rojas (2019), quienes aplicaron técnicas de retrofitting sobre la Prueba de Matemáticas para primaria (06) del PLANEA ELCE 2015 (INEE, 2015), aplicada a gran escala a los estudiantes de sexto año de primaria de México, para elaborar un diagnóstico nacional del dominio que tienen los estudiantes del país a lo largo de las habilidades básicas en matemáticas evaluadas en la prueba. Como parte de este estudio, los autores realizaron estudios cognitivos para identificar las habilidades cognitivas requeridas por los distintos ítems que conforman la prueba, mediante la aplicación de técnicas de retrofitting. Otros ejemplos pueden encontrarse en Jang (2009) ó Li (2011).

De acuerdo con Yang y Embretson, (2007), toda EDC diseñada con fines de mejora, debe estar respaldada en su diseño y validación por la aplicación de modelos cognitivos que permitan identificar y describir de manera precisa y confiable la interacción entre los distintos procesos de respuesta comprometidos con cada ítem. Para ello es imperante revisar la genealogía, la congruencia y la alineación de los ítems con el apoyo de un grupo de expertos en los contenidos sustantivos de la prueba y en materia de construcción de instrumentos de medición (Rupp, Templin y Henson, 2010).

Es imperante realizar un piloteo de los ítems que permita obtener reportes verbales exhaustivos de los procesos de respuesta a empleados para su resolución (Ericsson & Simon, 1993; Castillo y Padilla, 2013; Cui y Roduta, 2013; Ercikan et al., 2010; Taylor y Dionne, 2000). La obtención de reportes verbales se realiza de manera sistemática (Gorin, 2006), aplicando protocolos de pensamiento en voz alta con técnicas concurrentes y retrospectivas (Ericsson y Simon, 1984, 1993; Leighton, 2009; Leighton y Gierl, 2007), en las que idealmente deben participar tanto sustentantes que emulen las características de la población objetivo, como expertos de contenido. Como análisis complementarios, se propone la aplicación de análisis del sendero de la vista (*eye-tracking*; Snow y Lohman, 1989; Sternberg, 1977; Allen y Horsley, 2014), así como la revisión de las latencias de respuesta (Fredericksen, 1980; Posner, 1978; Posner y Rogers, 1978) y la realización de entrevistas con técnicas introspectivas y retrospectivas (Farr, Pritchard y Smitten, 1990; Powers y Wilson, 1995; Rupp, Ferne y Choi, 2006). Estas medidas se consideran de gran ayuda para obtener información en los casos en que se presentan dificultades para evocar el reporte verbal (Sternberg, 1977). Su aplicación contribuye a la verificación de la relación entre el modelo cognitivo elaborado por los expertos y los procesos reportados por los examinados (Messick, 1989), así como para detectar posibles fuentes de varianza irrelevante al constructo que se pretende evaluar (Haladyna y Downing, 2004; Messick, 1995).

Los modelos de diagnóstico cognitivo requieren de una matriz Q que dé cuenta de la estructura sustantiva de la prueba (de la Torre, 2009), al presentar la estructura factorial del instrumento presentando las habilidades requeridas por cada ítem contenido en la prueba (Li, 2011). La matriz contiene tantas filas como ítems haya en la prueba y tantas columnas como habilidades o conocimientos se esté requiriendo 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 con 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 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 o, en otras palabras, de cometer un “desliz” al momento seleccionar la respuesta, respectivamente.

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

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 .

1. *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 aplique (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 responde a cada uno de los ítems, *i.e*. 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).

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 busca 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:

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).

--------------------------------------

1. *Modelos de Aprendizaje Profundo (Deep Learning)*

La tercera propuesta metodológica proviene de la tradición de la inteligencia artificial y el análisis neurolingüístico, tiene la función de analizar, clasificar y modelar los instrumentos de *respuesta construida*, me referiré a estos instrumentos como: *rúbricas*. Las rúbricas comprenden: un Proyecto de Seguimiento elaborado por los sustentantes y una entrevista.

Los resultados de este método cumplen dos funciones generales:

1.- Orientar el proceso de selección para una:

\* Promoción Horizontal de la función docente y técnico docente en Educación Básica.

\* Promoción Vertical en Educación Media Superior.

2.- Producen un instrumento de clasificación automática, que estima la *calidad* de cualquier sustentante en ausencia de test especializados.

***Acerca del lenguaje.***

La forma particular en que los humanos estructuran y desarrollan su lenguaje tiene una relación directa con la estructura de sus conductas y capacidades cognitivas. Grupos humanos distintos tienen estilos distintos de lenguaje y miembros de un mismo grupo tienen lenguajes semejantes. El problema natural de su análisis radica en el método con que se extrae la información de los textos, especialmente si se busca que tenga algún valor predictivo. Tradicionalmente el análisis del lenguaje natural se realiza por medio de jueces, lo que implica leer uno a uno los textos y clasificarlos por medio de metodologías personales o semi-estandarizadas. Ésta técnica consume demasiados recursos temporales y humanos, además el llevar el sesgo necesario del método de cada juez.

Los últimos diez años, de las necesidades de analizar grandes cantidades de información, y del boom del desarrollo de tecnologías de procesamiento de matrices, se ha desarrollado un grupo de algoritmos, especializados en procesar datos no estructurados, como: texto, audio o imágenes. Estas máquinas son el sustento al que se refiere el título de *Aprendizaje Profundo* (*Deep Learning)*. A partir de aquí me referiré a éstas máquinas como: *algoritmos.*

***Acerca de los algoritmos de Aprendizaje Profundo.***

La conducta esencial de estos algoritmos es estimar los parámetros de la función: que mapean más eficientemente: . 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*).

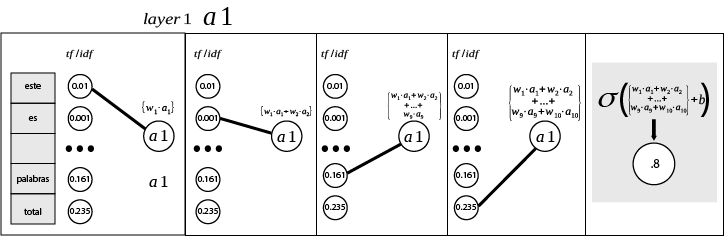
La propagación frontal estima los parámetros necesarios para predecir *y* tenemos: donde es la predicción de , mientras que y son los parámetros a estimar. Teniendo y , el algoritmo calcula el valor de por medio de la función de activación: . Esta función decide el tipo de output o formato de la predicción 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.

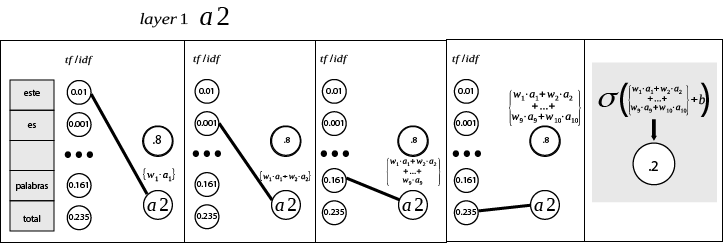
Teniendo , el algoritmo calcula el error total del modelo, referido comúnmente como: *loss function,* que se reduce a la diferencia entre la predicción y el atributo real, tenemos: .

Teniendo: finaliza la propagación frontal.

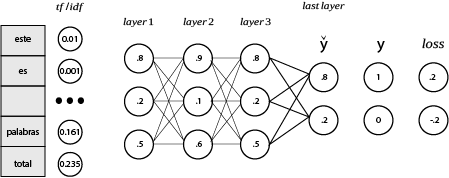
Digamos que = “este es el texto de x1 tiene 10 palabras total”.

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



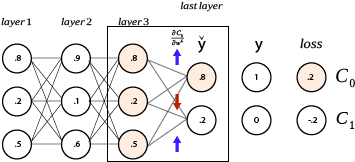


Iterando tenemos:

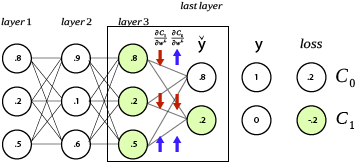


El siguiente proceso: La propagación inversa (*Backwards Propagation)*, es probablemente el proceso esencial de éstas máquinas o algoritmos, como hemos aclarado. Tiene la función de optimizar el error de la última capa. Podemos describir una capa como: , donde L es el índice de la capa, los pesos, es 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: .



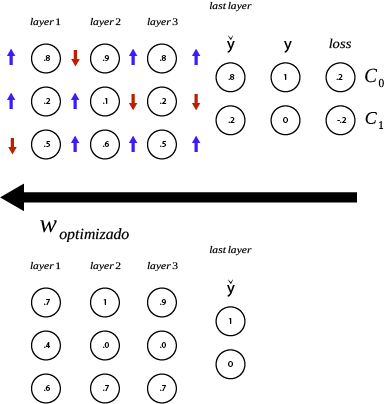
Lo siguiente es encontrar :



Teniendo , y se estima , a través de restar a el producto de y un hyperparámetro: , 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 .

***Acerca de Liwc.***

Desde la neurolingüística tenemos Liwc (Linguistic Iquiry and Word Count). Es una herramienta desarrollada por James W. Pennebaker como una solución al problema del análisis del lenguaje natural.

Sus bases parten de categorizar las palabras en ocho grupos de funciones neurolingüísticas: procesos cognitivos, procesos perceptuales, procesos biológicos, pulsiones, orientación temporal, relatividad, preocupaciones personales y lenguaje informal. Cada uno de ellos tiene subgéneros, como por ejemplo:

1.- Procesos Cognitivos

\*Insight

\*discrepancia

\*causa

…

2.- Pulsiones

\*afiliación

\*riesgo

\*logro

...

Cada palabra es asociada a una o varias categorías. A través del análisis de las frecuencias de estas categorías neurolingüísticas y su estructura de uso podemos obtener marcadores acerca del foco de atención, estado emocional, complejidad cognitiva, relaciones sociales, estilos de pensamiento y diferencias individuales de los sustentantes.

**Detalles de la técnica.**

El input:

El texto de las rúbricas se **transforma** en tres formatos distintos:

\* tf/idf: asigna un peso específico a cada palabra dependiendo de la proporción interna y externa de su frecuencia en el corpus.

\*liwc: asigna una o varias categorías neurolingüísticas a cada palabra.

\*diccionarios: asigna un integral a cada palabra.

El proceso:

La **transformación** de liwc entra en un sistema de selección de atributos (SSA) que clasifica los textos en grupos semejantes. SSA está formado por tres algoritmos que generan los grupos por medio de votación. Los algoritmos son variaciones de Análisis de Componentes Principales, Fuzzy C-means y Redes Neurales Recurrentes.

Teniendo esos grupos se analizan sus características neurolingüísticas y se asocian con *{las cosas que quieren ud como complejidad en el lenguaje, baja agresividad, buenas relaciones sociales, etc.}*. El resultado es un índice de *{más adecuado*} para *{obtener reforzador}.*

La **transformación** por medio de diccionarios entra a una máquina de análisis de tópicos, LDA (Latent Dirichlet allocation), que es una generalización del análisis semántico de probabilidad latente, con la diferencia de que ésta introduce una probabilidad a priori. Su función es estimar por medio de análisis estadísticos las palabras más significativas de los textos, a los que se refiere como tópicos. El resultado se mapea con distintas variables como el lugar geográfico del sustentando o los grupos de la transformación de liwc para conocer desde dónde y de qué hablan los distintos grupos.

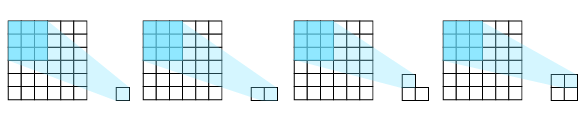
Finalmente **todas las transformaciones** entran a un sistema de predicción de calidad, SPC, formado por 6 algoritmos de clasificación más un modulo de votación, que modelan el conjunto de textos partiendo de los atributos de calidad obtenidos de {el output tuyo}. El resultado de SPC es un modelo que predice la calidad del sustentante en ausencia de test especializados.

Las maquinas de SPC:

**simpleRNN,** *red neuronal recursiva***,** es un algoritmo con una estructura prácticamente indistinguible del ejemplo expuesto en: *acerca de las máquinas 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.

SeparableCNN2d es 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.

**El proceso de SPC:**

1.- Modelaje:

Entrenamiento:

Se modelan los distintos tipos de inputs en los seis algoritmos.

Prueba:

Se prueba el valor predictivo de los modelos. Si la prueba es positiva se cierra el modelo.

Ajuste:

Se ajustan los algoritmos para reducir el error general. Se vuelve a entrenamiento.

2.- Votación:

Modelo de votación:

Se construye una base de datos que mapea cada sustentante con la matriz de resultados de los modelos. Esa base entra como input a una RNN que mapea el resultado de los modelos con la calidad del sustentante. Esta máquina (algoritmo) sigue el mismo proceso que los . modelos anteriores: entrenamiento, prueba y ajuste.

Finalmente la información de todos los modelos y todos los análisis, se estructuran en una aplicación de exploración y/o modelaje para uso del personal autorizado, para quien estos datos resultan útiles.

-------------------------------------

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 son 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, y que 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 bajo cada uno de los cuatro enfoques metodológicos durante el primer semestre del 2020 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** |
| Redes Neuronales (instrumentos cualitativos) | **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 |
| Observación de la práctica profesional |
| **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** |
| Redes Neuronales (instrumentos cualitativos) | **Promoción Vertical** a funciones Directivas y de Supervisión | *Entrevista por un Comité Examinador* | Primer semestre del **2021** |

De acuerdo con Abad (2011), todo instrumento de medición, con independencia del enfoque metodológico que se emplee para su diseño y calificación, debe garantizar la satisfacción de las siguiente exigencias:

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.

La secuencia de acciones específicas a seguir como parte del presente proyecto, para garantizar el cumplimiento de las exigencias ya mencionadas, se presentan a continuación:

**Plan de acción.**

Se conformará un equipo de trabajo central, cuyo espacio de trabajo estará ubicado en las instalaciones de la Facultad de Psicología de la Universidad Nacional Autónoma de México. Para efectos del trabajo a realizar durante el primer semestre del año 2020, se identifican dos grandes tareas: 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. La segunda tarea tiene que ver con el 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.

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 y el diseño de los instrumentos de apreciación que se aplicarán durante el segundo semestre y a lo largo del año 2021.

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 y 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. Esto 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 alienación 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.

Una vez definida la matriz de especificaciones se elaborará un manual digital para homogeneizar la construcción de ítems. El desarrollo de las actividades descritas hasta ahora como parte de la Fase II considera la participación de expertos y especialistas que coordinarán los trabajos realizados. Cada uno de ellos cuenta con una amplia experiencia en el desarrollo de evaluaciones a gran escala relacionadas con la valoración de la práctica educativa de las distintas figuras que componen al SCMM.

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 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, 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, como 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 planteados por los enfoques metodológicos empleados durante el diseño y desarrollo de los instrumentos; 2) las técnicas psicométricas derivadas de dichos enfoques metodológicos y que serán utilizadas para la calificación de las respuestas registradas en cada instrumento, incluyendo la estimación e interpretación de los índices psicométricos derivados de la Teoría Clásica de los Test y la Teoría de Respuesta al Ítem, así como análisis factoriales confirmatorios como medida de revisión de la estructura interna de los instrumentos; 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.

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.

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, lo que alinea 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 (ver Contreras, 2004; y Nitkoxa, 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, que permitan 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 de las respuestas.

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 piloto, resaltando la importancia de tomar en cuenta las sugerencias de operación, el llenado de registros de incidencias, entre otros. Una vez desarrollada la aplicación piloto, se revisarán y analizarán los reportes de incidencias, así como el proceso de aplicación para detectar mejoras en el diseño de los ítems y descartar posibles errores de aplicación.

Posterior a estas actividades, el equipo implementador, en conjunto con las coordinaciones disciplinares, desarrollarán un taller para la corrección de respuestas construidas para cada una de las áreas que comprende el estudio con ejercicios prácticos para la corrección de ítems. El desarrollo de esta capacitación requerirá la elaboración de manuales para el piloteo.

*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. Este servicio será coordinado a lo largo de las 32 entidades federativas del país, en las sedes establecidas para este fin.

La USICAMM proporcionará los instrumentos de apreciación que serán aplicados en mayo y junio del presente año y que serán utilizados para la toma de decisiones en los procesos de selección de Admisión y Promoción Vertical. Cada uno de estos instrumentos presentará una versión maestra a partir de la cual se elaborarán distintas versiones para su aplicación, indicando la correspondencia entre los ítems contenidos en cada prueba en una tabla de equivalencia.

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. Para ello se entiende como responsabilidad de las autoridades educativas estatales el garantizar que exista la infraestructura necesaria para que la aplicación pueda llevarse a cabo de manera eficiente y sin complicaciones. A su vez, el personal técnico aportado por la Facultad de Psicología estará encargado del diseño de un sistema que permita el almacenamiento de los instrumentos de apreciación del SISAP, su aplicación en línea, así como el registro y resguardo de las respuestas registradas por cada uno de los sustentantes.

La planificación, organización, gestión y control de los procesos de aplicación a nivel nacional correrá a cargo del equipo de trabajo central. La Figura 1 presenta la estructura operativa del personal que va a estar encargado de orquestar las aplicaciones nacionales de manera medular. Para liderar esta tarea, se cuenta con la participación de dos expertos con amplia experiencia en proyectos de evaluación a gran escala, quienes se harán cargo de la gestión administrativa y operación práctica de las aplicaciones y quienes contarán con su propio equipo de personal especializado en atender las responsabilidades señaladas en el segundo nivel del organigrama. Se considera la contratación de cuatro coordinadores regionales (zona norte, zona sur, *et. al*.) quienes se encargarán de dirigir, supervisar y verificar el trabajo realizado por los 32 coordinadores estatales, con quienes se busca agilizar la comunicación a partir de la contratación de personal de enlace.

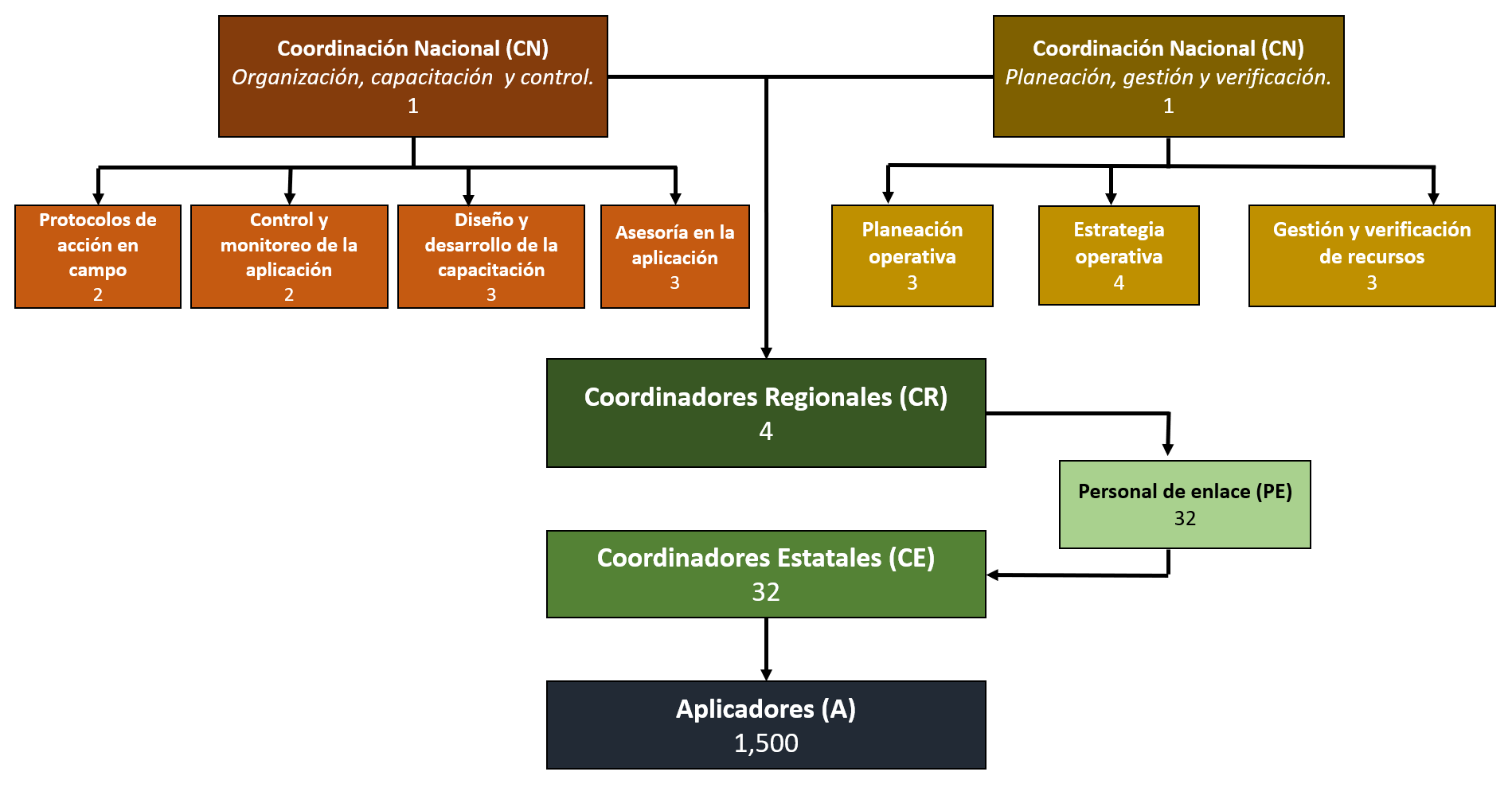
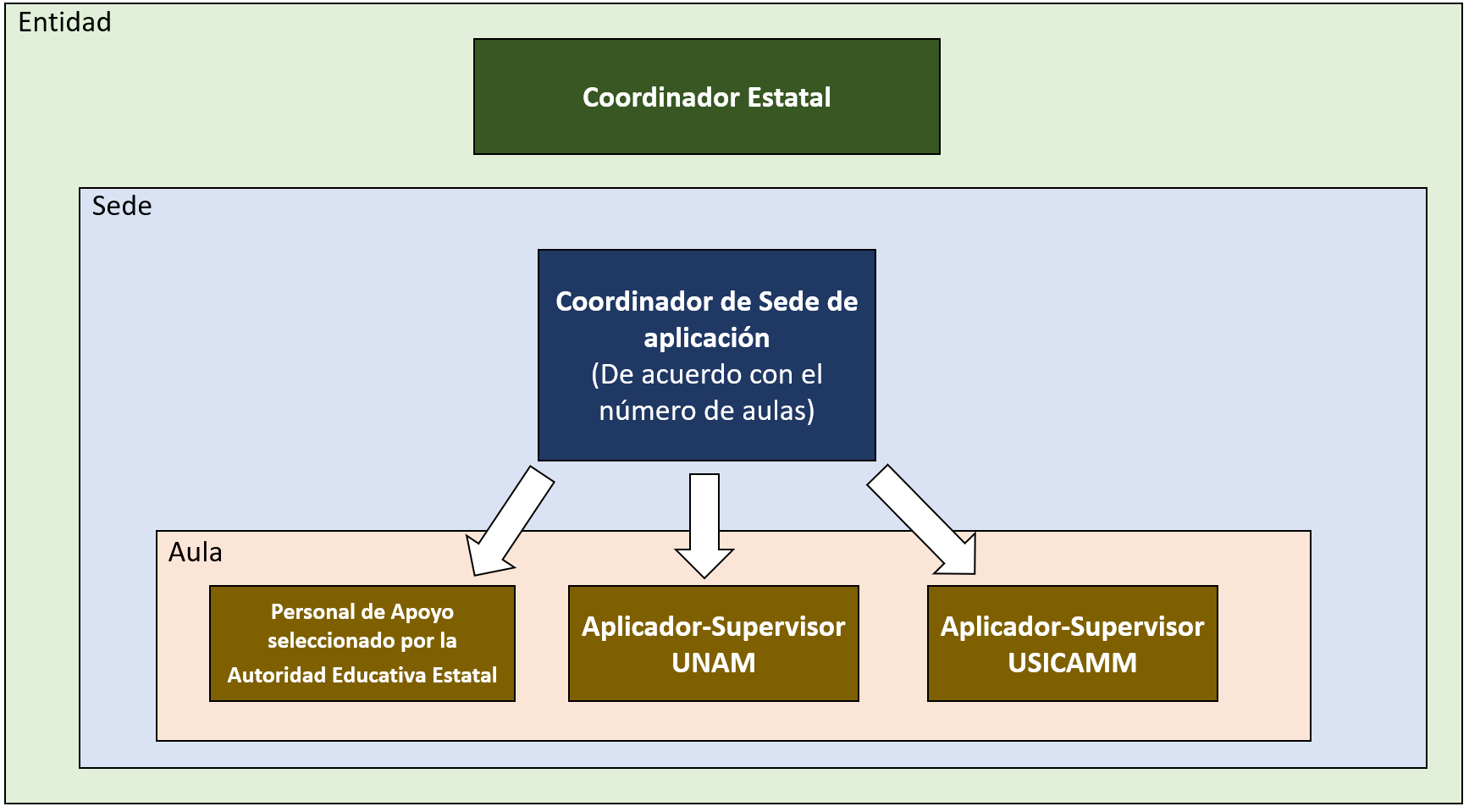


Figura 1. Estructura operativa nacional

De manera más específica, la Figura 2 presenta la estructura operativa a nivel del trabajo a realizar en cada una de las sedes designadas para la aplicación de los instrumentos. De acuerdo con este organigrama, cada coordinador estatal estará encargado de trabajar con los distintos coordinadores de sede (se contratará un coordinador por cada X aulas), que a su vez estarán encargados de orquestar el trabajo que se realice en las distintas aulas, donde de manera presencial se encontrarán trabajando tanto un supervisor contratado por la Facultad de Psicología, un supervisor designado por la USICAMM y un elemento de personal de apoyo seleccionado por la Autoridad Educativa Estatal. Estos últimos tres elementos serán los encargados de dar fe de la legalidad con que se responda cada instrumento, comprobando la identidad de los sustentantes que se presenten en sede y supervisando así el cumplimiento de las normas establecidas por la USICAMM para la aplicación de los instrumentos

Figura 2. Estructura operativa a nivel estatal (con coordinadores por sede y supervisores en aula).

Toda la información que se recopile durante el proceso de aplicación, desde las respuestas de los sustentantes hasta el reporte de incidentes de cada aplicación, serán tratados con estricta confidencialidad, entregando el reporte integrado de incidencias presentadas durante la aplicación a más tardar tres semanas después de haber concluido con la misma.

*Fase V. Calificación.*

Las respuestas de los sustentantes almacenadas en el sistema informático desarrollado.

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

Uno de los elementos más

**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:

• 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.

• 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.

• 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 y estrategias de respuesta y tiempos de respuesta.

• Elaboración de protocolos de actuación y manuales de aplicación (digitales).

• Certificación y conformación de grupos de supervisores locales, estatales, regionales y nacionales que vigilen y den constancia de la correcta aplicación de los instrumentos.

**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.
* 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., 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.
* 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.
* 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 examines’ 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.
* 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.
* 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: Jhon 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.: Lawrecence 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.
* Thompson, S., Johnstone, C. & Thurlow, M. (2002). Universal design applied to large scale assessments (Synthesis Report 44). Minneapolis, MN: National Centre on Educational Outcomes.
* Van der Linden, W. J. (Ed.). (2017). Handbook of Item Response Theory, Volume Three: Applications. CRC Press.
* Yang, X. & Embretson, S. (2007). Construct Validity and Cognitivity Diagnostic Assesment. In Leighton, J. y Griel, M. (Edit.). Cognitive diagnostic assessment for education: Theory and applications, pp. 85-118. Cambridge: Cambridge University Press.