

**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 docentes relativas a los Procesos de Admisión al Sistema de Carrera de las Maestras y los Maestros, y de Promoción Horizontal para las Funciones Docentes y de Promoción Vertical para las Funciones Directivas y de Supervisión en Educación Básica y Educación Media Superior.**

**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.

Consultar con Óscar sobre los proyectos previos en los que ha participado la UNAM, fuentes y confirmar si entran los del Observatorio. Pensarlo como una cuartilla de “LA UNAM ha coilaborado en estos proyectos” y luego, especificando la Facultad de PSicologia (Observatorio, Penud, adicciones, proyectos de medición a gran escala).

Cabe señalar que la presente propuesta técnica considera también la colaboración periférica del Instituto de Investigación y Desarrollo Educativo (IIDE) adscrito a la Universidad Autónoma de Baja California (UABC), quien a su vez constituye un referente importante en materia de investigación y evaluación educativa a nivel local, regional, nacional e internacional. La participación del IIDE como socio implementador representa la oportunidad de abonar, bajo una perspectiva metodológica robusta, hacia el análisis de la calidad educativa bajo un enfoque de equidad.

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 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 permiten 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 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, 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 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, (ejemplos 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; Pérez-Morán, Larrazolo, Backhoff, y Guaner, 2015; Revuelta y Ponsoda, 1998; Romero, Ponsoda y Ximénez, 2008; Birenbaum & Tatsuoka, 1993 ; Pérez-Morán, Vázquez-Lira, Rojas, 2019; 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, 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 puede 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 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, (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.

Distinción / relación entre modelos dicotómicos y politómicos, presentando los politómicos como una generalización de los primeros

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, permitiendo 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. 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).

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. *Rúbricas de calificación*

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, 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.

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).

Finalmente, la cuarta propuesta metodológica está relacionada con la incorporación de algoritmos de última generación en materia de inteligencia artificial, redes neuronales y aprendizaje de máquinas, como una herramienta para automatizar y disminuir el sesgo en la calificación de las respuestas elaboradas por los participantes, optimizando la información que se puede obtener a partir de estas acerca de sus habilidades y conocimientos y permitiendo así un respaldo sólido para la toma de decisiones por parte de las autoridades educativas. Se propone implementar algoritmos de reconocimiento de patrones en texto que abonen a la integración de una base de referencia para la interpretación de intenciones detectadas en los sustentantes.

Los instrumentos de corte cualitativo que permiten la elaboración por parte de los sustentantes de las respuestas registradas serán piloteados a fin de identificar posibles sesgos en los propios ítems y rescatar ejemplos de respuestas representativos. Para ello se emplearán técnicas de resolución en voz alta (Navas y Urdaneta, 2011) que abonan a la validez de la prueba.

La Tabla 1 permite identificar qué instrumentos, para cada proceso de selección, serán diseñados bajo cada uno de los cuatro enfoques metodológicos propuestos.

Tabla 1. Instrumentos a desarrollar bajo cada uno de los cuatro enfoques metodológicos .

|  |  |  |
| --- | --- | --- |
| **Educación Básica** | | |
| **Enfoque metodológico** | **Proceso de selección** | **Instrumento** |
| Modelos de Diagnóstico Cognitivo | **Admisión** a funciones Docente y técnico docente | Examen de acreditación Curso |
| Instrumento de valoración de conocimientos y aptitudes |
| **Promoción Vertical** a funciones Directivas y de Supervisión | Instrumento de valoración de conocimientos y aptitudes |
| **Promoción Horizontal** dentro de las funciones Docente, Técnico Docente, Asesoría Técnico Pedagógica, funciones Directivas y de Supervisión. | Instrumentos de valoración de conocimientos y aptitudes |
| Modelos politómicos 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 |
| **Promoción Horizontal** dentro de las funciones Docente, Técnico Docente, Asesoría Técnico Pedagógica, funciones Directivas y de Supervisión. | Cuestionario de habilidades socioemocionales |
| Rúbricas  Redes Neuronales (instrumentos cualitativos) | **Promoción Horizontal** dentro de las funciones Docente, Técnico Docente, Asesoría Técnico Pedagógica, funciones Directivas y de Supervisión | Proyecto de seguimiento |
| Entrevista sobre el proyecto de seguimiento |
| Observación de la práctica profesional |
| **Promoción Vertical** a funciones Directivas y de Supervisión | Cuestionario de habilidades directivas |
| **Educación Media Superior** | | |
| **Enfoque metodológico** | **Proceso de selección** | **Instrumento** |
| Modelos de Diagnóstico Cognitivo | **Admisión** a funciones Docente y técnico docente | Instrumento de conocimientos del modelo educativo |
| Instrumento de valoración de aptitudes y habilidades |
| **Promoción Vertical** a funciones Directivas y de Supervisión | *Instrumento de valoración de conocimientos y aptitudes* |
| Modelos politómicos de la Teoría de Respuesta al Ítem | **Promoción Vertical** a funciones Directivas y de Supervisión | *Encuesta a la comunidad escolar* |
| Rúbricas | **Promoción Vertical** a funciones Directivas y de Supervisión | *Entrevista por un Comité Examinador* |

**Plan de acción**

A fin de cumplir con el propósito que persigue la estructuración del SISAP y la contratación de un Prestador de Servicios especializados para tal tarea, 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 2.

Para llevar a cabo el plan de acción se conformarán dos grandes equipos centrales. Un primer equipo dedicará toda su atención al diseño y mantenimiento de los instrumentos del SISAP 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. Este equipo estará integrado por expertos de la Facultad de Psicología, quienes trabajarán de manera conjunta, siendo coordinados por el Dr. Iwen y el Dr Arturo Bouzas. Por su parte, el segundo equipo central operará desde las instalaciones de la Facultad de Psicología y se encargará de coordinar, gestionar y supervisar todos los procesos relacionados con la aplicación de los instrumentos a nivel nacional.

De acuerdo con lo presentado en el Anexo Técnico, se comenzará a trabajar desde el primer semestre del 2020 sobre el diseño de los instrumentos de apreciación que se aplicarán durante el segundo semestre como parte de los procesos de selección para la Promoción Horizontal dentro de las funciones docentes y técnico docente y durante el primer semestre del 2021 para la Admisión y Promoción Vertical en Educación Básica y Educación Media Superior. Por su parte, en cuanto a los instrumentos de apreciación que serán aplicados en los meses de Mayo y Junio del año en curso para los procesos de Admisión y Promoción Vertical, se trabajará con los instrumentos elaborados y proporcionados por la USICAMM.

Tabla 2. Fases de construcción del SISAP 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 1), 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, a 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 1, 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 de los ítems por parte del coordinador

Se asignará un coordinador disciplinar para cada instrumento, quien se encargará 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 dentro de la plataforma de gestión de ítems, con la posibilidad de introducir comentarios 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 compartan las características de la población objetivo, 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, a fin de identificar cualquier posible necesidad de mejora que se requiera atender antes de su aplicación real.

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 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. Se busca que el personal técnico de la USICAMM se familiarice 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 cuanto a las técnicas psicométricas para la calificación y análisis de los resultados obtenidos en los instrumentos, se contempla la capacitación en materia de la estimación e interpretación de los índices psicométricos derivados de la Teoría Clásica de los Test y la Teoría de Respuesta al Ítem, de la importancia que tiene el análisis factorial confirmatorio en la revisión de la estructura interna de los instrumentos y la estructura matemática de los modelos DINA, DINO y G-DINA.

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.

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.

cognitivo previo a la capacitación en cuestión. En este ejercicio el equipo implementador recopilará y organizará los resultados, comunicando a través del manual de corrección las líneas de trabajo para la capacitación presencial. El desarrollo de este ejercicio requerirá que cada equipo nacional recopile información de las respuestas de niñas y niños de tercer y sexto grados a los ítems de respuesta construida de las áreas por evaluar (incluyendo escritura). De manera general, el desarrollo de este ejercicio implicará:

• Seleccionar de una muestra de estudiantes de tercero de primaria y otra de sexto; al menos un grupo de cada grado deberá estudiar en escuela pública y otro en escuela privada.

• Aplicar los ítems de respuesta construida a los estudiantes, solicitando –bajo un protocolo en común para los equipos nacionales– que los estudiantes hablen en voz alta lo que piensan una vez que se enfrentan al ítem (o a la tarea de escritura, en su caso).

• Contar con el consentimiento de las escuelas y los padres o tutores para grabar el audio de las respuestas de los estudiantes.

• Enviar al equipo implementador las respuestas de los estudiantes a los ítems en cuestión.

En general, se busca integrar un modelo de validación impulsado por la participación de comités de expertos (ver Contreras, 2004; y Nitko, 1994), que permita asegurar la calidad técnica de los instrumentos, (ver OECD, 2012; Mislevy, 2006; y Messick, 1989).

*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 y el registro y resguardo de las respuestas registradas por cada uno de los sustentantes.

Para gestionar

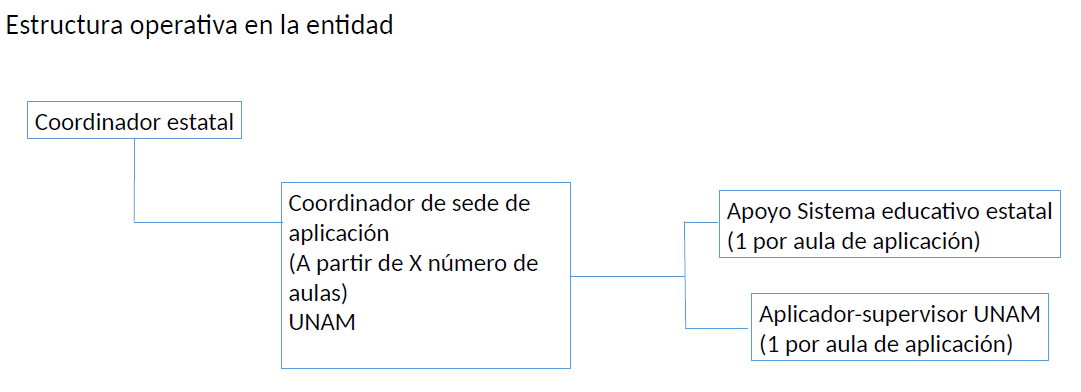
eficiente del proceso de aplicación, se contempla la contratación del personal contenido en el organigrama plasmado en la Figura 1.

Coordinadores regionales 4

Enlace para los estatales (32)



Coordinadores estatales y personal de apoyo que pueda supervisar, de manera presencial, que las aplicaciones realizadas en cada aula de las distintas sedes de aplicación, se lleven a cabo de acuerdo con los tiempos y procedimientos señalados. De tal forma, que por cada aula en la que se lleve a cabo una aplicación, deberá haber al menos un elemento del personal de apoyo designado por la autoridad educativa estatal, un supervisor contratado por el Coordinador Estatal designado por la Facultad de Psicología y un supervisor asignado por el personal de la USICAMM, para dar fe de la legalidad con que se responda cada instrumento, comprobando la identidad de los sustentantes que se presenten en sede y supervisando el cumplimiento de las normas establecidas por la USICAMM para la aplicación de los instrumentos

.

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 final de la aplicación a más tardar dos semanas después de haber concluido con la misma.

Como parte de su propuesta y a fin de documentar sus capacidades para el desarrollo del servicio, el licitante deberá presentar la estructura operativa correspondiente considerando organigrama, cantidad de integrantes por puesto, responsabilidades generales, percepción económica y vigencia de contratación en cada caso.

Además, y a fin de establecer las actividades de organización previas a la aplicación, entregará al contratante el programa general de trabajo para el desarrollo del proceso en campo, al menos 60 días antes del primer fin de semana establecido para la aplicación de instrumentos.

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 el uso y ajuste nacional de los manuales requeridos para ello, sugerencias de operación, 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 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 V. Análisis de datos del estudio piloto y plan de análisis*

La labor del IIDE en el análisis de datos resultantes de la prueba piloto se enfocará en determinar los valores psicométricos de los ítems utilizados, en aras de delimitar las posibles modificaciones por realizar a cada uno de ellos. Como parte de ese proceso, se enviará a los equipos nacionales un reporte de análisis psicométrico con los ítems *dodgy*, con el fin de identificar posibles modificaciones que pretendan la mejora de los instrumentos para la versión definitiva. Adicionalmente, se entregará una propuesta de plan de análisis para la aplicación definitiva, junto con las versiones finales de los instrumentos de ERCE. El tiempo destinado para el desarrollo de las fases del proyecto, así como los productos por entregar se describen en la tabla 2:

(ver OECD, 2012; Mislevy, 2006; y Messick, 1989) y la pertinencia de los resultados mediante la definición de esquemas de toma de decisiones de actores educativos relevantes para la mejora educativa, la inclusión, la equidad y la calidad de la educación en la región.

**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 analisys: 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 diferences thought chronometric analisys. 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.
* Garcia--‐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 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.

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 Center 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 Cognitivy Diagnostic Assesment. In Leighton, J. y Griel, M. (Edit.). Cognitive diagnostic assessment for education: Theory and applications, pp. 85-118. Cambridge: Cambrige University Press.