**Plan de Diseño, Desarrollo y Validación del Instrumento de valoración de conocimientos y aptitudes para aspirantes a funciones de Supervisor como parte del proceso de selección para la promoción horizontal.**

(Versión Preliminar)

Coyoacán, Ciudad de México a 4 de marzo de 2020

**Índice de contenidos**

**Índice de contenidos**

**Índice de figuras y tablas**

**Glosario, siglas y acrónimos**

**I. Introducción**

**II. Marco metodológico**

**2.1. Enfoque de apreciación**

**2.2. Modelo de diseño, desarrollo y validación del instrumento de apreciación**

**2.2.1. Fase de diseño**

**2.2.2. Fase de desarrollo**

**2.2.3 Fase de validación**

**III. Referencias**

**IV. Anexos o apéndice**

**Índice de figuras y tablas**

|  |  |
| --- | --- |
| Tabla 1. Modelo de diseño, desarrollo y validación de Instrumento de valoración de conocimientos y aptitudes docentes alineado a los Perfiles profesionales publicados en el Marco para la excelencia en la enseñanza y la gestión escolar en Educación Básica | ---------------- 9 |

**Glosario, siglas y acrónimos**

**Desarrollo.** Refiere a las acciones que se llevan a cabo para construir el instrumento de valoración, de acuerdo con el marco normativos y sus referentes teóricos; con los enfoques metodológicos, psicométricos y de medición planteados y con el modelo sustantivo trazado para definir y delimitar su objeto de estudio.

**Diseño.** Comprende las acciones relacionadas con la definición del objeto de medición y la delimitación del mismo a través de un modelo sustantivo que retome las teorías y modelos desarrollados en la literatura especializada, identificando dominios y componentes cuya medición a partir de la aplicación del instrumento de apreciación a construir permite interpretaciones que den cuenta del mismo.

**Estructura factorial**. Refiere a la estructura interna de todo instrumento de apreciación construído, en términos de la organización o agrupación de los distintos componentes o dimensiones medidas de acuerdo con el modelo sustantivo, a lo largo de los ítems componen al instrumento.

**Instrumentos de apreciación**. Herramientas de medición diseñadas, desarrolladas y validadas para su aplicación como uno de los multifactores de los procesos de selección, fungiendo como mecanismo para explorar las habilidades, conocimientos y actitudes con las que cuentan los aspirantes.

**Ley General del Sistema de Carrera para las Maestras y los Maestros (LGSCMM)**. Ley secundaria que establece el Sistema de Carrera para las Maestras y los Maestros como una herramienta del estado para tipificar la trayectoria docente a partir de procesos de selección para la Admisión, la Promoción Horizontal o Vertical y el Reconocimiento. La participación en dichos procesos se condiciona al cumplimiento de una serie de requisitos y su resolución depende de la ponderación de un conjunto de multifactores, dentro de los cuales se encuentra un Sistema de Apreciación de conocimientos y habilidades compuesto por un conjunto de instrumentos de apreciación elaborados a partir de los Perfiles, Criterios e Indicadores publicados por la Secretaría de Educación Pública como referente oficial para la práctica de las distintas figuras educativas.

**Modelo sustantivo.** Refiere a la estructura teórica elaborada a partir de la revisión exhaustiva de modelos y teorías desarrolladas en la literatura especializada para dar cuenta del objeto de medición del instrumento de apreciación a construir, donde se identifican sus principales componentes y las relaciones existentes entre estos que orienten el diseño y desarrollo del mismo.

**Marco metodológico.** En el caso particular de los instrumentos de apreciación del SISAP se tienen considerados tres grandes marcos metodológicos: los Modelos de Diagnóstico Cognitivo, la Teoría de Respuesta al Ítem y el Aprendizaje Profundo en Redes Neuronales.

**Perfiles, Criterios e Indicadores.** Documento elaborado por la Secretaría de Educación Pública para describir las características deseables en la figura docente, técnico docente, asesor técnico pedagógico, con funciones directivas o de supervisión.

**Piloteo.** Aplicación del instrumento en desarrollo a una muestra de individuos que tienen las mismas características que la población objetivo

**Procesos de Selección.** Se refiere a los procesos en los que participan los aspirantes a ingresar, ser reconocidos o ser promovidos en el Sistema de Carrera de las Maestras y los Maestros.

**Sistema de Apreciación (SISAP).** Refiere al cuerpo de instrumentos de apreciación que son utilizados en los distintos procesos de selección para aportar información sobre el ordenamiento de las listas de prelación.

**Sistema para la Carrera de las Maestras y los Maestros (SCMM).** Instrumento del Estado para que el personal docente, técnico docente, asesor técnico pedagógico, con funciones directivas y de supervisión acceda a una carrera justa y equitativa.

**Unidad del Sistema de Carrera de las Maestras y los Maestros (USICAMM).** Es un organismo administrativo desconcentrado con autonomía técnica, operativa y de gestión adscrito a la SEP, que tiene a su cargo las atribuciones le confiere la LGSCMM.

**Validez.** Refiere al grado en que la teoría y la evidencia empírica respalda las interpretaciones de los resultados de un test para los usos propuestos. La validez no se ve como un dato o una cualidad intrínseca, sino como un juicio valorativo basado en las evidencias recopiladas.

**I. Introducción**

En el presente documento se expone el Plan de Diseño, Desarrollo y Validación del Instrumento de valoración de conocimientos y aptitudes docentes, instrumento de apreciación aplicado como parte del proceso de selección para la promoción horizontal dentro de la función de Supervisor con el objetivo particular de valorar en los aspirantes los conocimientos y aptitudes que se consideran necesarias para el ejercicio de la función docente, conforme a los Perfiles profesionales que constituyen el referente oficial en cuanto a las características que se consideran esenciales para el desempeño de esta función.

El Plan de Diseño, Desarrollo y Validación detalla las acciones a realizar, de acuerdo con los más altos estándares de calidad técnica a nivel internacional en cuanto al desarrollo de instrumentos de valoración con un diseño universal, libre de sesgos, que procura en todo momento priorizar la recolección en espiral de evidencias de validez que empaten con el enfoque metodológico planteado para el análisis y calificación de los resultados obtenidos, cuya interpretación será uno de los insumos para la elaboración de las listas de prelación con base en las cuales se dará resolución a los procesos de selección descritos por la Ley General del Sistema para la Carrera de las Maestras y los Maestros (LGSCMM), (DOF, 2019).

El Instrumento de valoración de conocimientos y aptitudes docentes está pensado dentro del esquema del SISAP como una prueba cognitiva ,por lo que su calibración, calificación, análisis e interpretación de resultados serán realizados bajo el marco metodológico planteado por modelos de diagnóstico cognitivo tal y como se detalla en el Plan de Análisis correspondiente.

El desglose de las acciones a desarrollar como parte de las fases de diseño, desarrollo y validación, cuenta con un sustento empírico y teórico, basado en modelos de diseño y desarrollo, modelos evaluativos y protocolos de validez reportados en la literatura especializada. De esta forma, el presente documento aborda desde la definición y delimitación del objeto de estudio de acuerdo con los propósitos establecidos en la LGSCMM, alineando los dominios y criterios a medir dentro de los Perfiles profesionales (SEP, 2019), que constituyen el referente oficial para la valoración de la práctica educativa, con el cuerpo de teorías y modelos desarrollados en la literatura especializada para dar cuenta del objeto de medición; hasta la identificación de las relaciones existentes entre las habilidades y conocimientos que se requiere valorar para esquematizar la estructura interna de la prueba con base en la cual se desarrollará el contenido del instrumento, describiendo e incorporando en todo momento las estrategias a seguir para garantizar la validez de la interpretación de las mediciones a obtener tras su aplicación.

**II. Marco Metodológico**

Para el presente plan de diseño, desarrollo y validación se trabajó en la alineación de los dominios y criterios a medir contenidos en el perfil profesional para funciones de Supervisor en el *Marco para la excelencia en la enseñanza y la gestión escolar en la Educación Básica* (SEP, 2019), con un modelo sustantivo que identifica claramente los componentes principales que subyacen al objeto de medición y su interacción, a fin de orientar el desarrollo del Instrumento de valoración de conocimientos y aptitudes docentes como un instrumento de valoración que permita dar cuenta de su complejidad, al obtener mediciones cuya interpretación dé cuenta de las características de los aspirantes.

Durante la fase de diseño, se trabajó con un modelo de diagnóstico cognitivo que integra las propuestas de Nitko (1994) en cuanto al diseño de pruebas alineadas a un marco normativo, y las propuestas para el diseño de pruebas cognitivas de Mislevy (2007) y Embretson y Gorin, (2001) donde se enfatiza la recolección de evidencias observables sobre los procesos de respuesta empleados por los sustentantes al requerirles la aplicación de las habilidades y conocimientos que se consideran parte del constructo a valorar, más allá de la valoración mnémica del conocimiento declarativo, priorizando en todo momento la recolección en espiral de evidencias de validez, basadas en el trabajo con comités de expertos y la obtención de reportes de procesos de respuesta con una muestra similar a la población objetivo. La aplicación de este tipo de modelos suele ser recomendada bajo el enfoque unitario de la validez de constructo (Kane, 2006), al ser referenciado en los estándares 1.1, 1.2, 1.9, 1.10, 1.12, 3.13, 3.14, 4.1 y 4.2 de la APA, AERA y NCME (2014) correspondientes al diseño, desarrollo, validez y equidad de las pruebas.

Durante la fase de desarrollo, el énfasis principal estuvo en el poder garantizar la correspondencia entre el instrumento construído a partir del modelo de diseño planteado en la primera fase y los supuestos técnicos y estadísticos que subyacen a las teorías y modelos estadísticos que comprenden los marcos metodológicos a utilizar para la calibración de los instrumentos y el análisis de los resultados obtenidos por los participantes (la Teoría de Respuesta al Ítem y los Modelos de Diagnóstico Cognitivo), a fin de garantizar la validez de su interpretación como guía para la toma de decisiones.

Finalmente, es importante enfatizar que el esfuerzo concentrado en la fase de validación puede también encontrarse de manera transversal durante el desarrollo de las fases que le anteceden, al promover la validación del trabajo realizado en cuanto al diseño y desarrollo del instrumento por medio de un Comité de Expertos en contenido y en materia de medición y psicometría. La tercera fase del presente plan de diseño, desarrollo y validación, comprende la recopilación de reportes de procesos de respuesta asociados con la resolución de las versiones preliminares del instrumento por un grupo de expertos en contenido, así como la calibración de los resultados obtenidos tras una aplicación piloto con una muestra de la población objetivo.

A continuación, se presentan los detalles correspondientes al modelo de diseño, desarrollo y validación del Instrumento de valoración de conocimientos y aptitudes docentes aplicado como parte del proceso de selección para la promoción horizontal a funciones de Supervisor, las fases y acciones que lo componen, y el sustento teórico y metodológico que le subyace.

**2.1. Enfoque de apreciación**

De manera general, la valoración de los distintos aspectos de la práctica educativa a realizar mediante la aplicación de los instrumentos de apreciación que conforman al SISAP, son entendidos como una actividad que promueve la mejora continua y sistemática de la práctica educativa cotidiana de las funciones docente, técnico docente, directiva, de asesoría técnico pedagógica y de supervisión, en tanto que permiten a los aspirantes obtener información precisa para la identificación de sus propias áreas de fortaleza y mejora, orientando la toma de decisiones en materia del uso de estrategias que contribuyan a su formación, (Abreu, Muñoz, Monter & Cobas Vilches, 2015).

Los dominios y criterios propuestos en los perfiles profesionales (SEP, 2019), identificados como el referente principal para la valoración de la práctica educativa de las maestras y los maestros del país, guardan una relación estrecha con esta concepción de la mejora continua (Fullan, 2007), según la cual debe incorporarse los siguientes principios básicos en la educación: 1) el principio del compromiso ético pedagógico con el cambio; 2) el principio de la participación plena y la responsabilidad y 3) el principio del carácter continuo y prospectivo del proceso.

El proceso de apreciación tiene por objetivo proporcionar la máxima información, procurando en todo momento respetar los más altos estándares de calidad técnica en cuanto a los planes de diseño, desarrollo y validación propuestos, las metodologías empleadas y los recursos administrados.

En función del propósito de medición que subyace a cada instrumento de apreciación comprendido como parte del SISAP, se distinguen tres grandes tipos de instrumentos: 1) Las escalas, que registran las valoraciones de los aspirantes a lo largo de una serie ordenada de opciones de respuesta que reflejan sus percepciones en cuanto a la intensidad, frecuencia o eficiencia con que se suscitan los elementos, eventos o habilidades y aptitudes presentadas; 2) Las pruebas cognitivas, que requieren a los participantes aplicar los conocimientos y habilidades que se asumen como esenciales para el desarrollo de la función educativa a la que pretenden ser admitidos, promovidos o bien, dentro de la cual aspiran a recibir una promoción horizontal, para acertar los distintos ítems que componen la prueba y 3) Los instrumentos de respuesta construida, que brindan la oportunidad a los participantes de elaborar, con el detalle, tono y énfasis que consideren pertinente, sus propias producciones, como evidencia de su quehacer educativa cotidiana. Bajo esta misma lógica, para cada tipo de instrumento se utilizan distintos modelos de diseño, desarrollo y validación: 1) El modelo nomológico para las escalas, 2) El enfoque cognitivo basado en evidencia para los instrumentos de apreciación y 3) Los instrumentos de respuesta construida acuerdo a las propuestas de Nitko (1994).

En los últimos años ha crecido el interés, el uso y desarrollo de pruebas de diagnóstico cognitivo en el campo educativo, en el reclutamiento de personal y en la investigación en psicología, entre otros. El objetivo de estas pruebas (Al-Shamrani & Dimitorv, 2017) es identificar el grado de experticia o dominio que una persona posee a lo largo de una serie de habilidades, conocimientos, destrezas o atributos a valorar, cuya aplicación resulta esencial para poder responder de manera acertada los ítems que las componen. Una de las bondades que presentan las pruebas de diagnóstico cognitivo es que permiten la recolección y devolución de información precisa sobre los resultados obtenidos por los aspirantes, dado que la elaboración de los reactivos que las componen cuenta con un componente cognitivo basado en evidencia. El Instrumento de valoración de conocimientos y aptitudes, está pensado como una prueba de este tipo y se espera que el acotar su diseño, desarrollo, validación y calificación al marco metodológico planteado por los modelos de diagnóstico cognitivo, permitan devolver información granulada a los aspirantes acerca de sus áreas específicas de mejora y fortaleza, a fin de orientar la toma de decisiones en cuanto al diseño y uso de estrategias de mejora continua.

Uno de los componentes esenciales de los modelos de diagnóstico cognitivo (Sorrel et al., 2016) es que captura de manera explícita la estructura interna de la prueba y establece la relación de los atributos necesarios para contestar satisfactoriamente cada reactivo, en lo que se conoce como matriz Q. En contraste, la teoría clásica de los test y la teoría de respuesta al ítem suelen utilizar una estructura simple, donde se asume que cada ítem aporta información para la medición de un sólo factor o atributo, mientras que los modelos de diagnóstico cognitivo permiten tomar en cuenta la interacción entre estos, de manera que se asume que cada ítem requiere del participante la aplicación de más de un atributo, permitiendo trazar una estructura más compleja.

**2.2. Modelo de Diseño, Desarrollo y Validación del Instrumento de apreciación**

En la Tabla 1 se presentan las fases que comprenden el presente modelo de diseño, desarrollo y validación del Instrumento de valoración de conocimientos y aptitudes docentes y las actividades a desarrollar como parte de cada una de ellas.

**Tabla 1. Modelo de diseño, desarrollo y validación de Instrumento de valoración de conocimientos y aptitudes docentes alineado a los Perfiles profesionales publicados en el Marco de excelencia**

|  |  |
| --- | --- |
| **Fase** | **Actividades a desarrollar** |
| I. Diseño | 1.1. Crear el Comité técnico del desarrollo del instrumento de apreciación |
| 1.2. Analizar los Perfiles profesionales y el contenido de la Ley General del Sistema para la Carrera de las Maestras y los Maestros |
| 1.3. Elaborar el Marco de Referencia y el Plan de Diseño, Desarrollo y Validación del instrumento de apreciación. |
| 1.4. Delimitar y alinear el contenido de los Perfiles profesionales con teoría sustantiva que respalde la definición de los rasgos a medir |
| 1.5. Definir la estructura interna y los componentes principales del instrumento de apreciación. |
|
|
| 1.6. Crear el Comité de Expertos para la validación del Diseño (CEVDI) del instrumento de apreciación |
| 1.7. Someter el Marco de Referencia, el Plan de Diseño, Desarrollo y Validación y la propuesta de estructura interna a revisión y validación por parte del CEVDI |
| 1.8. Ajustar los documentos sujetos a revisión con base en la valoración y retroalimentación proporcionada por el CEVDI. |
| 1.9. Elaborar como primer producto el modelo nomológico y el modelo cognitivo que orientará el desarrollo del instrumento de apreciación |
| II. Desarrollo | 2.1. Desarrollo de la matriz de especificaciones |
| 2.2. Elaborar las primeras versiones de los ítems |
| 2.3. Crear el Comité de Expertos para la Validación del Desarrollo de los Ítems (CVEDEI) |
| 2.4.Someter las primeras versiones de los ítems a revisión por el CVEDEI |
| 2.5. Ajustar los ítems con base en la valoración de los ítems y la retroalimentación otorgada por el CVEDEI. |
| 2.6. Pilotear los ítems con Jueces expertos que respondan el instrumento utilizando protocolos de resolución en voz alta |
| 2.7. Ajustar las versiones de los ítems en función a la información recolectada a partir de la resolución de los mismos por el grupo de jueces |
| 2.8. Pilotear los ítems con una muestra pequeña de sustentantes, aplicando un protocolo de Laboratorio Cognitivo donde se recojan evidencias de procesos de respuesta |
| 2.9. Ajustar las versiones de los ítems en función a la información recolectada sobre los procesos de respuesta evocados en el laboratorio cognitivo. |
| 2.10. Ensamblar las primeras versiones del instrumento. |
| III, Validación | 3.1. Crear el Comité de Expertos para la Validación del Contenido de los Ítems (CEVCI) |
| 3.2. Someter las versiones ensambladas a revisión y validación por parte del CEVCI |
| 3.3. Ajustar los ítems con base en la valoración y la retroalimentación realizada por el CEVCI. |
| 3.4. Ensamblar una nueva versión del instrumento para su aplicación piloto |
| 3.5. Definir la muestra de participantes para el piloteo de la prueba. |
| 3.6. Reclutar, seleccionar y capacitar a los aplicadores del piloteo. |
| 3.7. Realizar la aplicación piloto del instrumento |
| 3.8. Recolectar y codificar las respuestas de los participantes del piloteo. |
| 3.9 Realizar análisis preliminares sobre las respuestas recopiladas y estimar indicadores psicométricos derivados de la Teoría Clásica de los Test, |

**2.2.1. Fase de Diseño**

Como todo proyecto de investigación, la elaboración de un instrumento de apreciación que permita hacer inferencias sobre las características de los aspirantes, debe contar con una fase de diseño donde se detalle minuciosamente los objetivos que se pretenden conseguir con su aplicación y las características que debe tener. Es necesario especificar el constructo, atributo o característica psicológica a evaluar con el instrumento a elaborar, alineando el contenido y las características requisitadas en los referentes normativos con la literatura especializada.

De acuerdo con lo propuesto por Embretson y Gorin (2001), el diseño debe de estar basado en modelos teóricos que cuenten con un sustento teórico fuerte y con evidencia empírica del proceso de respuesta, es decir, que permitan representar el constructo con base en los modelos obtenidos de la investigación en Psicología. Al utilizar evidencia empírica para el diseño de las pruebas se pueden evitar errores (Leighton, 2004), tales como: concepciones erróneas, malversaciones y oportunidades no aprovechadas. La validez de la prueba comienza a construirse desde el diseño de la misma, valorando constantemente la adecuación del instrumento a desarrollar con respecto del uso e impacto que tendrán las interpretaciones que se realicen a partir de su aplicación para garantizar la validez de diseño (Thomas,1985).

En términos de las acciones a desarrollar, la fase de diseño comprende todo lo relacionado con el desarrollo y esquematización de la estructura interna del instrumento de apreciación a construir, empatando los contenidos de los Perfiles profesionales y los propósitos establecidos en la LGSCMM para cada proceso de selección, con las teorías sustantivas y evidencias empíricas reportadas en la literatura. De esta forma, es posible identificar los elementos y componentes principales del objeto de estudio y su interrelación, para trazar así el modelo sustantivo que orientará el desarrollo del instrumento de apreciación como una herramienta para la recolección de evidencias observables, cuya interpretación permita hacer inferencias sobre las habilidades, conocimientos, aptitudes y actitudes de los aspirantes. En otras palabras, es durante esta fase cuando se define la estructura interna del instrumento dentro de la matriz Q, donde se especifica la relación existente entre los distintos conocimientos y habilidades a medir y cada uno de los ítems a desarrollar como parte del instrumento.

En su conjunto, la fase de diseño debe ayudar a especificar los siguientes elementos: propósito (Schmeiser y Welch, 2006; DOF, 2017, 28 de abril, p. 5), objeto (DOF, 2017, 28 de abril, p. 5), población objetivo (Mellenbergh, 2011), interpretación y uso de resultados (Mellenbergh, 2011; DOF, 2017, 28 de abril, p. 5), tipo de instrumento, modelo de medición, tipo de reactivos, longitud del instrumento y tiempo de administración (DOF, 2017, 28 de abril, p. 5; Schmeiser y Welch, 2006), contar con un cuerpo de expertos rector del instrumento y especificaciones y el marco teórico (AERA, APA y NCME, 2014; Mellenbergh, 2011).

**2.2.2. Fase de Desarrollo**

La fase de desarrollo comprende el compendio de acciones a realizar, una vez definido el modelo de diseño, para dar forma al instrumento de apreciación de acuerdo con las definiciones y delimitaciones trazadas por el modelo sustantivo elaborado. En el caso particular del Instrumento de valoración de conocimientos y aptitudes docentes, se habla del desarrollo de la matriz Q que funge como el esqueleto de la estructura interna de la prueba, donde se describe de manera detallada cuáles de los ítems a desarrollar como parte de la prueba requieren del participante la aplicación de los conocimientos y habilidades identificados como parte esencial del objeto de medición y el consecuente desarrollo de los ítems.

Los instrumentos basados en modelos (Bejar, 2002) también conocidos como representaciones del constructo (Embretson, 2002) se han vuelto un referente estándar, tanto en términos teóricos como prácticos, de alta calidad técnica. Los ítems que componen al Instrumento de valoración de conocimientos y aptitudes docentes deben de ser congruentes a los dominios,criterios e indicadores y capturar las habilidades y destrezas que se definen en los perfiles profesionales publicados por la SEP (SEP, 2019).

La recolección sistemática de evidencias de validez concurrente durante la fase de desarrollo es posible mediante el trabajo con un comité de expertos que valide los productos generados en la fase de diseño, tales como la matriz Q construída y el conjunto de ítems diseñados para su medición Adicionalmente, se considera la integración de reportes de procesos de respuesta recopilados de sustentantes que comparten las características de la población objetivo, mediante la aplicación de técnicas de resolución en voz alta, entendiendo que las evidencias recolectadas deben arrojar luz tanto sobre el modelo explicativo como la teoría psicológica sustantiva, que guía el proceso de construcción del instrumento, de acuerdo con el modelo de diseño.

**2.2.3 Fase de Validación**

El concepto de validez (Kane, 2006; Lissitz, 2009) comprende aspectos relacionados con la interpretación y uso de los resultados observados y señala la necesidad de recolectar sistemáticamente evidencias de los atributos observables, los atributos definidos operacionalmente y los dominios, en función de su alcance. En otras palabras, garantizar la validez del instrumento implica especificar tanto los propósitos y los usos de la prueba, como el proveer evidencia empírica que respalde la teoría sustantiva que subyace a toda interpretación propuesta (NCME,2014) y el tomar en cuenta las posibles implicaciones a futuro de la aplicación del instrumento, entendida como la validez consecuencial del instrumento, (Abad, 2011).

El proceso de validación de todo instrumento plantea como objetivo recopilar evidencias que respalden las interpretaciones y posteriores usos que se le dará a la información recabada por los distintos métodos de medición que se utilicen en este proceso valorativo. Retomando lo propuesto por los Estándares para Pruebas Educativas y Psicológicas (AERA, APA y NCME, 2018) la validación ahora se ve como un proceso de recopilación de evidencias que sustenten las interpretaciones propuestas bajo los usos previamente delimitados, evidencias que pueden ser de 5 tipos: Evidencias basadas en el contenido del instrumento, evidencias basadas en los procesos de respuesta, evidencias basadas en la estructura interna del instrumento, evidencias basadas en su relación con otras variables y evidencias basadas en las consecuencias del instrumento. Las evidencias necesarias para justificar la validez dependen directamente de las características de las interpretaciones y de las propiedades de la información recabada por el instrumento de medición.

Distintos métodos de validación se han propuesto en la literatura, entre ellos el método argumentativo de Kane (1992; 2015) el cual plantea una que toda validación comienza con la generación de un *Argumento de interpretaciones/usos (AIU)* en donde se especifica toda la red de supuestos e inferencias que llevan del desempeño de los evaluando en la prueba a las conclusiones y decisiones basadas en los resultados. Este argumento de interpretaciones/usos denota un marco de referencia para la validación, especificando qué se está afirmando y bajo qué condiciones teóricas y metodológicas. Posteriormente, se genera el *Argumento de Validez* el cual evalúa críticamente al argumento de interpretaciones/usos al utilizar análisis que generen información relevante y pertinente para sustentar el AIU, dicha información es posteriormente presentada como evidencia empírica que junto con un análisis lógico de la coherencia y plausibilidad, fungen como sustento del AIU. En este modelo de validación la evidencia requerida depende la interpretación y usos propuestos, por lo que sí la interpretación es simple, no incluye un gran número de inferencias y supuestos, no requeriría de mucha evidencia para respaldar.

Para validar el diseño y desarrollo del Instrumento de valoración de conocimientos y aptitudes docentes, se considera una aplicación piloto con una muestra de características similares a la población objetivo, incorporando técnicas de pensamiento en voz alta retroactivas y concurrentes que permitan valorar el grado en que la aplicación del instrumento arroja información relevante y de útil interpretación, respecto del estado del conglomerado de habilidades y conocimientos que se pretende valorar con su aplicación.

**III. Referencias**

1. Abad, F. J. (2011). Medición en ciencias sociales y de la salud.
2. Abreu, O. L. L., Muñoz, J. J. G., Monter, I. B., & Cobas Vilches, M. E. (2015). La mejora continua: objetivo determinante para alcanzar la excelencia en instituciones de educación superior. *Revista Educación Médica del Centro*, *7*(4), 196-215.
3. AERA, APA y NCME. American Educational Research Association, American Psychological Association y National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.
4. AERA, APA y NCME (2018). Estándares para Pruebas Educativas y Psicológicas. Washington, DC: American Educational Research Association. Retrieved March 4, 2020, from www.jstor.org/stable/j.ctvr43hg2
5. Al-Shamrani, A., & Dimitorv, D. M. (2017). Cognitive Diagnostic Analysis of Reading Comprehension Items: The Case of English Proficiency Assessment in Saudi Arabia. *Internasional Journal Of School And Cognitive Psycholigi*, *196*.
6. Bejar, I. I. (2002). Generative testing: From conception to implementation. *Item generation for test development*, 199-217.
7. DOF (2017, 28 de abril). Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación. México. Recuperado de: http://www.dof.gob.mx/nota\_detalle. php?codigo=5481062&fecha=28/04/2017
8. Diario Oficial de la Federación (2019) Ley General del Sistema de Carrera para las Maestras y los Maestros. México, Ciudad de México, 30 de septiembre de 2019.
9. Embretson, S.E. (2002) Generating abstract reasoning items with cognitive theory. *Item generation for test development*, 219-250.
10. Embretson, S., & Gorin, J. (2001) Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement*, 38(4), 345-368.
11. Fullan, M. (2007). Educational reform as continuous improvement. *The keys to effective schools: Educational reform as continuous improvement*, 1-12.
12. Kane, M. T. (1992). An argument-based approach to validity. *Psychological bulletin*, *112*(3), 527.
13. Kane, M. (2006). Validation. *Educational measurement*, 4(2), 17-64.
14. Kane, M. (2015) Explicating Validity. Assessment in education: Principles, Policy & practice. DOI: 10.1080/0969594X.2015.1060192
15. Leighton, J.P. (2004) Avoiding misconception, misuse and missed opportunities: The collection of verbal reports in educational achievement testing. *Educational Measurement: Issues and Practice*, 23(4), 6-15.
16. Lissitz, R. W. (Ed.). (2009). *The concept of validity: Revisions, new directions and applications*. IAP.
17. Mellenbergh, G. J. (2011). *A conceptual introduction to psychometrics: development, analysis, and application of psychological and educational tests*. The Hague, Netherlands: Eleven International Publishing.
18. Mislevy, R. J. (2007). Validity by design. *Educational Researcher*, *36*(8), 463-469.
19. Nitko, A. J. (1994). A Model for Curriculum-Driven Criterion-Referenced and Norm-Referenced National Examinations for Certification and Selection of Students.
20. Schmeiser, C. y Welch, C. (2006). Test Development. En: Brennan, R. L. (ed.), *Educational Measurement* (pp. 307-354). Washington, DC: American Council on Education. Praeger Series on Higher Education.
21. Secretaría de Educación Pública (2019) Marco para la excelencia en la enseñanza y la gestión escolar en la Educación Básica: Perfiles profesionales, criterios e indicadores para docentes, técnicos docentes y personal con funciones de dirección y supervisión. Ciclo escolar 2020-2021.
22. Sorrel, M. A., Olea, J., Abad, F. J., de la Torre, J., Aguado, D., & Lievens, F. (2016).
23. Thomas, E. J. (1985, July). The validity of design and development and related concepts in developmental research. In *Social Work Research and Abstracts* (Vol. 21, No. 2, pp. 50-55). Oxford University Press.

**IV. Anexos o Apéndice**