

Desarrollo y Validación de un Instrumento para Evaluar la Práctica Docente en Educación Preescolar

Development and Validation of an Instrument for Evaluating the Teaching Practice in Early Childhood Education

Luis Horacio Pedroza Zúñiga *
Edna Luna

Universidad Autónoma de Baja California

Actualmente se reconoce que la calidad de la enseñanza contribuye a elevar los aprendizajes de los niños, por ello se tiene un interés creciente en evaluar de forma eficiente y precisa lo que realizan los docentes en las aulas. Este artículo da cuenta de las evidencias de validez de contenido y constructo de un instrumento para evaluar la práctica docente en educación preescolar, así como la consistencia de la medición y del proceso de observación. El instrumento en su versión más acabada se compone de 30 rúbricas divididas en tres dimensiones: planeación, intervención y evaluación. Para ello se analiza la planeación didáctica y la práctica docente en el aula. La información brindada por el instrumento logró una alta consistencia interna y buen ajuste al modelo de Rasch. El análisis factorial exploratorio permitió identificar tres factores subyacentes, con diferencias respecto a la estructura teórica. Los análisis para identificar la confiabilidad entre los observadores, Kappa cuadrática ponderada y análisis de la varianza a partir de la Teoría de la Generalizabilidad ofrecieron resultados coincidentes: adecuada consistencia en las distintas observaciones.

Palabras clave: Evaluación docente, Educación preescolar, Observación sistemática, Rúbricas, Confiabilidad, Validez.

This article reports the content and construct-related validity evidence of an instrument for evaluating teaching practice in early childhood education. In addition, it shows evidence of the consistency of measurement and observation process. The instrument consists of 30 rubrics divided into three dimensions: planning, intervention and evaluation. To do this, the planning and the teaching practice is analyzed. The information provided by the instrument achieved a high internal consistency and a good fit to the Rasch model. Exploratory factor analysis identified three underlying factors, with differences in the theoretical structure. Analyses to identify reliability among observers, quadratic weighted Kappa and analysis of variance based on the generalizability theory, showed consistent results: proper consistency among the different observations. Further analyses are needed with a larger sample of examinees, in order to confirm the factorial structure and adjust the instrument to increase its accuracy.

Keywords: Teacher evaluation, Early childhood education, Measurement of instruction, Rubrics, Reliability, Validity.

*Contacto: horaciopedroza@hotmail.com

1. Introducción

Evaluar la enseñanza que realizan los docentes en las aulas es una tarea demandante. Esto es así porque la enseñanza es una actividad multidimensional y compleja. La profesión docente requiere del dominio de competencias especializadas, es necesario un conocimiento de lo que se enseña, de cómo enseñar y de las características de los aprendices. En particular en la enseñanza en preescolar se caracteriza por ser el primer ciclo de la educación obligatoria, en la cual se plantea la utilización de prácticas adecuadas al desarrollo y aprendizaje de los niños.

En general, los programas de evaluación de la docencia de los profesores de educación básica incorporan información de tres fuentes principales: 1) del aprendizaje de los estudiantes al final del curso; 2) evaluaciones hechas por pares académicos y directivos; y 3) autoevaluaciones, autorreportes y portafolios. La elección de una estrategia particular de evaluación depende en gran medida de los propósitos de la evaluación y de las condiciones institucionales de aplicación. La evaluación por pares académicos, como estrategia, ha sido ampliamente utilizada con fines sumativos y formativos (Millis, 2006). Aunque en la actualidad se ha privilegiado su uso en contextos orientados a la mejora de la actividad docente.

La presente investigación tiene como propósito diseñar y aportar evidencias de validez de un instrumento de observación en el aula para evaluar la práctica de las docentes de educación preescolar en México.

En el plano internacional existen dos instrumentos de observación ampliamente utilizados en educación preescolar. El primero es el *Early Childhood Environment Rating Scale* (ECERS) cuyo propósito es medir la calidad de los servicios educativos en edad preescolar, es un instrumento estandarizado de observación, los referentes de evaluación que le subyacen están vinculados con el desarrollo de prácticas adecuadas para este nivel educativo (Harms, Clifford y Debby, 1988).

El segundo es el *Classroom Assessment Scoring System* (CLASS); es un protocolo estandarizado de observación de la práctica docente, su propósito es medir la calidad de las interacciones del docente relacionadas con un mayor aprendizaje de los alumnos, el cual ha sido utilizado principalmente en EUA (Pianta y Hamre, 2009). Ambos instrumentos se caracterizan por tomar como marco de referencia la investigación sobre el desarrollo y aprendizaje infantil, no están referidos a un currículo en particular.

En México, Myers, Martínez y Linares (2003) desarrollaron un instrumento de observación similar al ECERS que evalúa el ambiente educativo y tiene como referente el concepto de calidad de la educación preescolar, cuyo fin último es el aprendizaje para una cultura democrática. El instrumento se ha utilizado principalmente en proyectos de investigación. No obstante, aunque se haya utilizado en distintos momentos no se han presentado las propiedades métricas de la escala.

La Secretaría de Educación Pública (SEP) ha realizado investigaciones de la enseñanza en el marco de la Reforma a la Educación Preescolar. El equipo responsable en la Dirección de Desarrollo Curricular para Preescolar emprendió un conjunto de acciones para promover el conocimiento y análisis de –entre otros aspectos– la práctica pedagógica en las aulas del país. Derivado de estas acciones se realizó un diagnóstico de la situación del preescolar en México (SEP, 2006) el cual sirvió principalmente para

identificar avancen en la implementación del Programa de Educación Preescolar 2004 (PEP 2004) y establecer estrategias de mejora.

Pedroza, Álvarez y Jiménez (2013), en el marco de una evaluación de la implementación del currículo de preescolar [PEP 2004], desarrollaron una bitácora para evaluar de manera estandarizada las prácticas docentes. Para ello se utilizó un auto registro con preguntas abiertas para identificar lo que hacían las educadoras en tres jornadas escolares, el cual fue codificado con un conjunto de rúbricas. Se identificaron los propósitos educativos, las características del enfoque pedagógico y la demanda cognitiva de las actividades.

Una evaluación alineada al currículo y al marco normativo de la evaluación docente es fundamental, puesto que la evaluación orienta la acción educativa. Darling-Hamond (2012) señaló que para que la evaluación docente tenga un efecto positivo en el aprendizaje de los niños tendría que partir de una estrategia sistémica de evaluación, compuesta por: estándares comunes de evaluación; evaluaciones de desempeño basados en estos estándares; sistemas locales de evaluación inspirados en los mismos estándares para evaluar la calidad de la enseñanza *in situ*; estructuras de apoyo, evaluadores y mentores para docentes que requieran ayuda adicional; y oportunidades de desarrollo profesional.

A partir de la revisión de los antecedentes, es claro que en México no existe un instrumento en preescolar que evalúe la práctica docente mediante observación y análisis de evidencias tomando como referente el marco normativo. El instrumento que se presenta en este artículo se diseñó con la particularidad de estar alineado al currículo y al marco normativo para evaluar a los docentes de educación preescolar en México. Por una parte el instrumento retoma el Perfil, Parámetros e Indicadores [PPI] (SEP, 2015), que son los estándares para evaluar a los docentes, y por otra el Programa de Educación Preescolar 2011 [PEP 2011] (SEP, 2011), puesto que en este se establecen los principios pedagógicos que guían la práctica de los docentes de este nivel educativo.

2. Método

En este apartado se describe el proceso seguido para en el desarrollo de un instrumento de evaluación de la práctica docente en la educación preescolar.

2.1. Participantes

Se contó con un total de 28 participantes, incluyendo a los participantes en el diseño del instrumento como a los participantes en la muestra, a saber: ocho integrantes del comité de jueces de rúbricas, una educadora experta quien participó en la elaboración de las rúbricas y 19 educadoras a quienes se aplicó el instrumento. Las características de los participantes son las siguientes:

El comité de jueceo: se conformó por profesionales con experiencia en algunos de los siguientes aspectos: diseño del currículo nacional, evaluación en educación preescolar y asesoría técnico pedagógica al profesorado. Su media de edad es 46 y los años de experiencia 24 años.

La educadora experta: participó en la elaboración de las rúbricas, es jubilada con 28 años de servicio; 24 frente a grupo y con 15 de docente en educación normal y 4 de asesor técnico pedagógico del nivel preescolar.

Educadoras observadas: el grupo se conformó por 19 docentes con formación específica para la docencia en el nivel preescolar; todas del sexo femenino. Diez fueron educadoras en formación en su primer año frente a grupo y las nueve restante con más de seis años de experiencia. Centra (1993) identifica a los docentes en la etapa de noveles (0-3 años) hasta de mayor experiencia (más de 8 años). Las docentes laboraban en escuelas públicas ubicadas en zonas urbanas de alta marginación.

2.2. Materiales

- ✓ Dos videocámaras, equipadas con tripí y micrófonos inalámbricos de solapa.
- ✓ Equipo de cómputo (hardware y software) para procesamiento, almacenamiento y acceso en línea al material audiovisual (Google drive) proporcionado por la Universidad Autónoma de Baja California.
- ✓ Cuadernillo del instrumento y hojas de registro.

2.3. Procedimiento

Este estudio comprendió tres etapas: desarrollo del instrumento, recolección de la información y análisis de los resultados. Cada una de ellas se describe a continuación.

2.3.1. Etapa 1: Desarrollo del instrumento

El desarrollo del instrumento tuvo las siguientes fases: operacionalización, elaboración de rúbricas, jueceo de las rúbricas por el comité de validación, prueba en campo y elaboración de manuales.

- ✓ *Fase 1. Operacionalización.* La tarea consistió en definir constitutivamente el constructo que mide el instrumento: la enseñanza en educación preescolar. Se retomaron los postulados de Shulman (2005), y se dividió el constructo en tres dimensiones: planeación, intervención educativa y evaluación. A continuación se presentan las definiciones de las mismas.

Planeación. Incluye dos procesos tal como los señaló Shulman (2005), el de comprensión y el de transformación. La comprensión es donde el docente conoce y comprende críticamente un conjunto de ideas que van a enseñarse. Además de la comprensión del contenido a enseñar, es fundamental que el docente comprenda los objetivos educativos, que si bien pueden partir de un texto o currículo, estos tiene que adaptarse o modificarse en función del grupo de educandos.

El segundo proceso es la transformación, en este las ideas comprendidas deben ser modificadas para poder enseñarlas. En general implican varios pasos: preparación (materiales, textos, etc.); representación de las ideas a nuevas formas como analogías o metáforas; selección de un método didáctico; y adecuación de estas adaptaciones a las características generales del grupo y a características específicas de cada niño de la clase.

Para el PEP 2011, la planeación de la intervención educativa es entendida como “un conjunto de supuestos fundamentados que la educadora considera pertinentes y viables para que niñas y niños avancen en su proceso de aprendizaje” (SEP, 2011: 25). Asimismo, señala que la planeación es indispensable para que el docente realice una práctica eficaz, ya que le permite definir la intención educativa, prever los recursos didácticos, la organización del

grupo e identificar los criterios para evaluar el proceso de aprendizaje en los alumnos que conforman su escolar. Es claro que el programa visualiza este proceso, al igual como lo señala la literatura, como una tarea reflexiva que implica distintos elementos a considerar para poder realizarse adecuadamente.

Intervención educativa. Es la actividad que comprende el desempeño observable del docente en el aula. Incluye distintas acciones de la didáctica como: la organización y manejo del grupo; las explicaciones; la asignación de trabajo o tareas y su revisión; el monitoreo a los alumnos; y la interacción por medio de preguntas, reacciones, elogios o críticas Shulman (2005). A esta fase de la enseñanza otros autores denominan como conducción del proceso de enseñanza - aprendizaje (García, Loredó, Luna, Rueda, 2008). Los distintos elementos de la intervención educativa se organizaron en dos grandes aspectos: a) principios pedagógicos del programa y b) ambiente de aprendizaje.

La evaluación. Es un proceso que refiere a la verificación de la comprensión en los alumnos, incluye las distintas actividades mediante las cuales los docentes identifican el progreso de sus alumnos y reconocen el logro de las metas de aprendizaje, como verificar o monitorear la comprensión de los alumnos durante el proceso de enseñanza, retroalimentar la comprensión de los alumnos al finalizar las situaciones didácticas o unidades de aprendizaje. En otros niveles educativos, este proceso también incluye una valoración para emitir una calificación. Para la comprobación de los aprendizajes en los niños se requiere que el docente comprenda el contenido a enseñar y los procesos de aprendizaje del niño, es decir, las formas de comprensión y transformación descritas anteriormente. Otra orientación de la evaluación, es hacia el propio docente y su desempeño logrado, lo que conduce a otro proceso de la enseñanza, que es el de reflexión sobre la práctica.

Posteriormente se definió operacionalmente el constructo, para ello se generó una tabla donde se ordenaron los elementos del más abstracto al más concreto: dimensiones, subdimensiones e indicadores. Esta tabla de contenidos se acompañó de un documento de fundamentación, donde se describieron los contenidos a evaluar y se identificó su relación con el PEP 2011.

- ✓ *Fase 2. Elaboración de rúbricas.* En el diseño de las rúbricas se siguió una metodología de elaboración colectiva, a partir de la deliberación colegiada de los participantes (Jornet, González, Suárez y Perales, 2011; Pedroza, Vilchis, Álvarez, López y García, 2013). El proceso de construcción de las rúbricas se realizó de acuerdo con el siguiente orden: descripción del nivel 3 que corresponde a lo esperado por el programa; luego se describe el nivel 1, características de una práctica inadecuada; después del nivel 2, en donde se mezclan elementos del nivel 1 y 3; y al final el nivel más avanzado, el nivel 4, donde se incluye un *plus* respecto a lo que señala el programa. En la elaboración, se consideraron los criterios de calidad de las rúbricas descritos por Arter (2010). La elaboración tomó aproximadamente 90 horas de trabajo. El producto de esta fase es una primera versión de las rúbricas.
- ✓ *Fase 3. Jueceo del instrumento.* Para esta fase se conformó el Comité de Jueceo, el cual, durante dos sesiones colectivas, estuvo encargado de hacer la valoración de la primera versión de las rúbricas en función de su relevancia y suficiencia con el

PEP 2011. Para esto se siguió un procedimiento donde se privilegió llegar a consensos sobre las modificaciones a realizar en el instrumento.

- ✓ *Fase 4. Piloteo de las rúbricas.* En esta fase se realizaron dos acciones, en primer lugar se realizó una prueba del instrumento a partir de evidencias como las planeaciones de docentes, expedientes de los niños y videos de la práctica docente. Se analizaron 12 planes de trabajo, tres videos y ocho expedientes. Se siguió un proceso interactivo, donde se evaluaba la evidencia y se hacían modificaciones al instrumento. En la segunda acción se llevó a cabo una prueba en campo, la cual consistió en la observación de cuatro jornadas escolares completas a distintos docentes. Esto permitió examinar el funcionamiento del instrumento en condiciones de aplicación y realizar los ajustes pertinentes del mismo. La etapa culminó con la versión final del instrumento y su manual de uso.

2.3.2. Etapa 2: Recolección de información y retroalimentación a educadoras

- ✓ *Fase 1. La conformación de la muestra.* La conformación de la muestra de educadoras se dio a partir de aquellas que dieron acceso a que fueran grabadas en video. En el diseño del proyecto inicialmente se contempló que fueran 20 educadoras, sin embargo, de una no fue posible completar todas las observaciones. La aplicación del instrumento se realizó a 19 educadoras (tabla 1).

Tabla 1. Características de la muestra de educadoras observadas

	N
Jardines de niños	3
Educadoras	19
Sesiones observadas por educadora	3
Codificaciones por sesión	2
Total de registros de observación	114

Fuente: Elaboración propia.

Se estableció un acuerdo en el que los investigadores se comprometieron a devolver una retroalimentación para cada una de las observaciones y poner a disposición de las docentes una copia electrónica de los videos. Además, se contó con un convenio en el que las educadoras dan su consentimiento para la grabación de su clase y en el que se establecen los usos de la información recabada, los cuales se rigen bajo la Ley Federal de Protección de Datos Personales en Posesión de Particulares (DOF, 2010). La información proporcionada es utilizada bajo los principios de confidencialidad, consentimiento, información, calidad, finalidad, lealtad, proporcionalidad y responsabilidad.

- ✓ *Fase 2. Calibración de la observación.* Dado que los dos observadores participaron en la elaboración de las rúbricas, no se requirió un proceso de capacitación para el uso del instrumento. En vez de ello se realizó una codificación colegiada a partir de videos, lo que permitió unificar los criterios de valoración.
- ✓ *Fase 3. Trabajo de campo.* La recolección de información se llevó a cabo en un periodo de tres meses. Los observadores estuvieron encargado de recabar las planeaciones didácticas y grabar la intervención de cada docente durante toda la jornada escolar, exceptuando el recreo y clases que la docente titular no dirigía

(música, educación física, etc.). Cada educadora fue grabada en video durante tres sesiones de trabajo, y cada ocasión se calificó dos veces, una *in situ* y otra a partir del video. De esta forma se cuenta con seis medidas de desempeño de cada educadora (ver Tabla 1).

- ✓ *Fase 4. Retroalimentación a los evaluados.* Los observadores estuvieron encargados de entregar un reporte de resultados y de dar retroalimentación individual a las educadoras al finalizar los ciclos de grabación. Además, los videos se pusieron a disposición de las docentes a través de un acceso personalizado en línea.

2.3.3. Etapa 3. Análisis de la información

En esta etapa se sistematizó la información para conocer la estructura y confiabilidad del instrumento, así como la precisión de las estrategias de recolección de información. La información recabada se analizó en tres fases:

- ✓ *Fase 1. Identificación de la dimensionalidad del instrumento.* Los ítems del instrumento se sometieron a un análisis con el Modelo de Crédito Parcial de Rasch Master (Masters, 1982), con la finalidad de identificar ítems anómalos en su funcionamiento, este modelo es utilizado para variables ordinales. Si bien se reconoce que el modelo Rasch para ítems dicotómicos es sensible al tamaño de la muestra, Smith, Fallowfield, Velikova y Sharpe (2008) encontraron que los modelos politómicos de Rasch no presentan la misma sensibilidad, aunque se identificó mayor estabilidad en muestras mayores a 200 sujetos.

Posteriormente se realizó un análisis factorial exploratorio (AFE) para probar la estructura subyacente a partir de la evidencia empírica. Dodou y Wieringa (2009 citados en Frías-Navarro, Pascual-Soler, 2012) señalaron que un $N=50$ es un valor mínimo razonable para este tipo de análisis, en este estudio tuvieron 114 observaciones. Para el AFE se utilizó un módulo del paquete R denominado *R Factor* (Basto y Pereira, 2012), el cual pone a disposición distintas técnicas de extracción, rotación y selección del número de factores. Debido a que la métrica de las rúbricas es ordinal, se decidió utilizar el análisis factorial a partir de matrices policóricas. El método de extracción fue Factorización de Ejes Principales. Se hizo una rotación oblicua (Quartimin) ya que es más pertinente para variables sociales donde se espera que los factores tengan algún grado de asociación (Basto y Pereira, 2012). Para la selección del número de factores se valoraron cinco métodos: Regla de Kaiser, Análisis de Paralelo, Coordinación Óptima, Factor de aceleración y Prueba de Velicer. Asimismo, se estimaron varios coeficientes de confiabilidad como: alfa de Cronbach, coeficiente alfa ordinal y theta de Armor's.

- ✓ *Fase 2. Confiabilidad de las dimensiones.* Se calcularon cuatro coeficientes de confiabilidad: alfa de Cronbach, el coeficiente alfa ordinal, Theta de Armor y Coeficiente Theta ordinal. Los tres últimos coeficientes son más apropiados para una métrica ordinal.
- ✓ *Fase 2. Confiabilidad de la observación.* Se realizaron tres análisis para dar cuenta del grado de acuerdo que presentan los evaluadores: porcentaje de acuerdo entre los jueces, Kappa de Cohen y coeficientes de determinación y generalizabilidad. En primer lugar se calculó el porcentaje de acuerdo exacto entre los dos observadores. En un segundo momento se estimaron los coeficientes Kappa de

Cohen en su versión cuadrático ponderado. Por último, a partir de la teoría de la generalizabilidad se estimó la confiabilidad entre los observadores y entre las distintas sesiones de observación, así como las fuentes de varianza principales. El análisis de generalizabilidad ofrece una mejor medición de los errores que otras formas de análisis porque permite identificar las variaciones atribuidas a distintas facetas de medida como juez, ocasión, ítems, y las interacciones entre estas de forma simultánea (Shavelson y Webb, 1991).

3. Resultados

3.1. Elaboración de rúbricas

El propósito del instrumento es evaluar la práctica pedagógica para ofrecer una retroalimentación al desempeño del docente. Enseguida se presentan algunas de sus características:

- ✓ *Población objetivo.* Docentes de educación preescolar.
- ✓ *Uso.* Evaluación formativa de docentes e investigación.
- ✓ *Tipo de instrumento.* Protocolo de observación, compuesto por un conjunto de rúbricas que especifican los criterios de calidad de la enseñanza para distintos niveles de desempeño.
- ✓ *Forma de administración:* mediante la observación, ya sea in situ, o por medio de grabaciones en video de unidades didácticas completas.
- ✓ *Usuarios:* agentes del sistema educativo encargados de hacer una supervisión o acompañamiento de la práctica pedagógica como supervisores, asesores técnico pedagógicos ATP's, directores y colegas docentes, así como observadores externos con conocimiento del PEP 2011.
- ✓ *Perfil de los evaluadores:* conocimiento del currículo de preescolar y experiencia en la observación de la práctica docente. Asimismo, acreditar un proceso de capacitación, en el cual deberán mostrar un grado de acuerdo superior a 80% con los criterios del diseñador.
- ✓ *Calificación:* el puntaje del instrumento está determinado por las rúbricas, donde cada nivel de desempeño, corresponde a un puntaje distinto.
- ✓ *Devolución de resultados:* posterior a la calificación del instrumento se genera un reporte individualizado, donde además del puntaje de cada rúbrica se incluye la descripción del nivel de desempeño. La devolución de la información a las educadoras se da en entrevista presencial, privilegiando el diálogo entre el evaluador y el evaluado.

En la primera versión del instrumento se construyeron 38 matrices de valoración. Las matrices tienen cuatro niveles de calidad para cada rubro, salvo algunos casos donde es más pertinente tener tres niveles (tabla 2). Estos niveles reflejan un continuo de experiencia en la enseñanza, una progresión de niveles menos desarrollados a niveles más desarrollados. El nivel "insatisfactorio", el más bajo, representa una práctica inadecuada o errónea desde el PEP 2011; el nivel "en proceso", es una práctica que tiene rasgos de niveles más avanzados pero con mezcla de elementos del nivel insatisfactorio;

el nivel “competente” es una práctica que está de acuerdo a lo que señala el PEP 2011; y el nivel “experto” representa una práctica de un docente con amplia experiencia y que manifiesta una práctica ejemplar por su dominio y ejecución. El instrumento consta de una hoja de registro para una jornada de observación y un cuadernillo con las matrices de valoración. Enseguida se presentan sus características principales.

Tabla 2. Ejemplo de las matrices de valoración del instrumento

DIMENSIÓN	INSATISFACTORIO	EN PROCESO	COMPETENTE	EXPERTO
Intención educativa	Las intenciones educativas no están vinculadas con las competencias o aprendizajes esperados del programa.	Las intenciones educativas están vinculadas a las competencias del programa o con alguno de los aprendizajes esperados.	Las intenciones educativas están vinculadas a las competencias del programa y con alguno los aprendizajes esperados.	Las intenciones educativas están vinculadas a las competencias del programa y con algunos de los aprendizajes esperados. Además se identifican competencias o aprendizajes esperados que se favorecen indirectamente con la situación didáctica.
Conocimientos previos	La docente no recupera los conocimientos previos de los alumnos. No hace preguntas o actividades de indagación.	La docente recupera los conocimientos previos de los alumnos en forma parcial. Utiliza preguntas que no permiten indagar la intención educativa a trabajar.	La docente recupera los conocimientos previos de los alumnos y corresponden a la intención educativa a trabajar.	La docente recupera los conocimientos previos de los alumnos, corresponden a la intención educativa a trabajar y los vincula con la situación didáctica.
Promueve la interacción entre niños	La educadora no promueve la interacción entre los niños o les limita que se comuniquen entre ellos.	La educadora permite la interacción entre los niños. Existe un acomodo de los alumnos en pequeños grupos. Pero se designa sólo trabajos individuales, o de manera grupal todos tienen el mismo objetivo y producto.	La educadora promueve la interacción entre los niños. Existe un acomodo de los alumnos en pequeños grupos y trabajan colaborativamente (realizan un producto en equipo).	La educadora promueve la interacción colaborativa de los niños en distintas formas, por ejemplo: en parejas, en pequeños grupos, en trabajo grupal, sesiones plenarias.

Fuente: Elaboración propia.

3.2. Juceo del instrumento

A partir de la retroalimentación realizada por el comité de jueceo se realizaron los siguientes cambios: diez rúbricas fueron aprobadas sin cambios, la mayoría tuvieron modificaciones y se eliminaron siete de ellas. La eliminación de las rúbricas se debió a que estas indagaban aspectos poco relevantes de la práctica docente o el contenido no

podía evaluarse adecuadamente mediante las evidencias: observación de tres sesiones, plan de clase y expedientes.

Las modificaciones de las rúbricas obedecieron principalmente a las razones: mejorar la descripción de los indicadores a observar, ya sea por una redacción ambigua o porque se incluían ejemplos que no eran apropiados; las descripciones, principalmente de los niveles más altos, creaban una imagen poco real de la práctica docente, se incluían elementos que se realizan con poca frecuencia en los jardines de niños; se modificaron dos rúbricas que no representaban de forma fiel lo que el PEP 2011 señala; otro grupo numeroso de cambios, se debió a observaciones de forma, que incluyeron cambios en la redacción, de una palabra por otra y cambios sintácticos-gramaticales. Cabe mencionar que no se sugirieron cambios a la estructura de la operacionalización.

Tabla 3. Cambios realizados en las rúbricas del instrumento por el comité de jueceo

DIMENSIÓN	PRESENTADAS A REVISIÓN	APROBADAS	MODIFICADAS	ELIMINADOS
Planeación	16	1	8	7
Intervención	11	5	6	0
Principios pedagógicos	7	1	6	0
Ambiente de aprendizaje	4	0	4	0
Evaluación	38	7	24	7
Totales				

Fuente: Elaboración propia.

3.3. Pilotaje del instrumento

El pilotaje del instrumento se realizó en cuatro aulas, esto permitió identificar elementos que no son factibles de observar durante una sesión de clase y precisar las descripciones de las rúbricas. A partir de la observación en campo se eliminó una rúbrica más, puesto que en una sesión de clase no se lograba identificar los aspectos a observar. Asimismo, se hicieron ajustes a once rúbricas. La estructura final del instrumento se presenta en la tabla 4.

El estudio piloto también sirvió para probar el proceso de recolección de la evidencia para la aplicación definitiva, como el plan de trabajo y la grabación de la práctica de la docente en video. El proceso de aplicación del instrumento es el siguiente:

1. *Obtención de la planeación de clase.* El observador solicita la planeación a la docente antes de iniciar la sesión. Se califican las primeras rúbricas correspondientes a la planeación, a excepción de las relacionadas con el conocimiento de los niños. En este paso también se fotografía la planeación.
2. *Observación de toda la jornada escolar,* durante este tiempo se graba en video la sesión y se califica las rúbricas del instrumento.
3. *Procesamiento y almacenamiento del material audiovisual.* Se organiza y almacena en medios electrónicos las evidencias obtenidas.
4. *Segunda codificación.* Se analiza la planeación didáctica y se observa la práctica docente por medio del video.
5. *Retroalimentación.* Se genera un reporte individualizados por sesión de observación y en entrevista con la docente se le ofrece una retroalimentación.

Tabla 4. Estructura del instrumento para su aplicación extensiva

PROCESO	DIMENSIÓN	SUB-DIMENSIONES
1. Planeación	1.1. Intención educativa	1.1a. Intención congruente con el PEP 2011 1.1b. Intención acorde a las necesidades de niños
	1.2. Conocimiento de los alumnos	1.1c. Intención clara
	1.3. Diseño de la situación de aprendizaje	1.3a. Actividad congruente con intenciones 1.3b. Actividad congruente con enfoque del campo 1.3c Plan para evaluación
	1.4. Demanda cognitiva	
2. Intervención educativa	2.1. Principios pedagógicos del programa	2.1. Recupera conocimientos previos 2.2. Promueve la interacción entre los niños 2.3. Fomenta el deseo por aprender 2.4a. Hace el contenido interesante (estrategias) 2.4b. Interés de los niños en la actividad 2.5. Promueve la participación y responsabilidad en el aprendizaje 2.6. Reglas 2.7. Demanda cognitiva de la actividad 2.8. Dominio del contenido 2.9. Congruencia con lo planeado
	2.2. Ambiente del aula para aprendizaje	2.10a. Comunicación cálida (verbal) 2.10b. Comunicación cálida (no verbal) 2.11. Consigna clara 2.12. Manejo de errores 2.13. Orden del grupo 2.14a. Uso del tiempo en las actividades 2.14b. Uso del tiempo en la jornada 2.15. Atención a la diversidad
3. Evaluación de la interacción en aula		3.1. Monitoreo 3.2. Retroalimentación 3.3. Reflexión sobre el proceso de aprendizaje 3.4. Reconocimiento

Fuente: Elaboración propia.

3.4. Análisis descriptivos

La distribución de los datos muestra que, en la gran mayoría de los casos, las docentes observadas se distribuyeron en todos los niveles definidos en cada una de las rúbricas de Planeación (tabla 5). Sin embargo, se observa que en dos de ellas, en la 1.1a y 1.1c, un alto porcentaje de las docentes se concentra en el nivel más alto de la rúbrica (experto); esto se relaciona con que se encontraron planes de trabajo muy exhaustivos entre las educadoras observadas, debido a que las directoras exigen a ese nivel de detalle las planeaciones. No obstante, esto no sucedió así en la muestra de planeaciones que se revisó en la construcción del instrumento.

Tabla 5. Distribución de los evaluados por nivel de dominio. Planeación

RÚBRICAS	PORCENTAJE DE EVALUADOS POR NIVEL				TOTAL
	Insatisfactorio	En proceso	Competente	Experto	
1.1a. Intención congruente con el PEP 2011	3.5%	2.6%	8.8%	85.1%	100.0%
1.1b. Intención congruente con necesidades de niños	5.3%	14.0%	33.3%	47.4%	100.0%
1.1c. Intención clara	2.6%	2.6%	2.6%	92.1%	100.0%
1.2. Conocimiento de los alumnos	4.4%	35.1%	34.2%	26.3%	100.0%
1.3a. Actividad congruente con intenciones	15.8%	32.5%	7.0%	44.7%	100.0%
1.3b. Actividad congruente con enfoque del campo	14.0%	40.4%	45.6%	N/A	100.0%
1.3c. Plan para evaluación	11.4%	39.5%	36.8%	12.3%	100.0%
1.4. Demanda cognitiva de planeación	10.5%	48.2%	28.1%	13.2%	100.0%

Fuente: Elaboración propia.

Una distribución similar se observa en las rúbricas de Intervención y Evaluación, los informantes se distribuyen, por lo general, en todos los niveles (Tabla 6). Sólo en cuatro de ellas se presenta una alta concentración de evaluados en uno de los niveles (2.2, 2.12, 2.15 y 2.9); destaca el caso de la rúbrica 2.12, en la cual los extremos están prácticamente vacíos, esto sugiere que sería factible colapsar, para este caso, los 4 niveles a sólo 2. Además, se identifican dos rúbricas en las que el nivel más alto está desierto (2.5 y 3.3), lo que invita a revisar la pertinencia de dicho nivel.

Tabla 6. Distribución de los evaluados por nivel de dominio. Intervención y Evaluación

RÚBRICAS	PORCENTAJE DE EVALUADOS POR NIVEL				TOTAL
	Insatisfactorio	En proceso	Competente	Experto	
2.1. Recupera conocimientos previos	13.2%	34.2%	24.6%	28.1%	100%
2.2. Promueve la interacción entre los niños	0.9%	79.8%	10.5%	8.8%	100%
2.3. Fomenta el deseo por aprender	7.9%	64.0%	28.1%	N/A	100%
2.4a. Hace el contenido interesante (estrategias)	1.8%	34.2%	36.0%	28.1%	100%
2.4b. Interés de los niños en la actividad	5.3%	21.1%	48.2%	25.4%	100%
2.5. Promueve la participación y responsabilidad en el aprendizaje	11.4%	60.5%	28.1%	0.0%	100%
2.6. Reglas	1.8%	29.8%	65.8%	2.6%	100%
2.7. Demanda cognitiva de la actividad	14.9%	44.7%	36.0%	4.4%	100%
2.8. Dominio del contenido	3.5%	14.9%	65.8%	15.8%	100%
2.9. Congruencia con lo planeado	8.3%	4.6%	87.2%	N/A	100%
2.10a. Comunicación cálida (verbal)	0.9%	34.2%	50.0%	14.9%	100%
2.10b. Comunicación cálida (no verbal)	0.0%	58.8%	39.5%	1.8%	100%
2.11. Consigna clara	0.9%	27.2%	47.4%	24.6%	100%
2.12. Manejo de errores	0.0%	71.1%	28.1%	0.9%	100%
2.13. Orden del grupo	7.0%	37.7%	51.8%	3.5%	100%
2.14a. Uso del tiempo en las actividades	3.5%	42.1%	44.7%	9.6%	100%
2.14b. Uso del tiempo en la jornada	5.3%	42.1%	52.6%	N/A	100%
2.15. Atención a la diversidad	9.7%	13.3%	74.3%	2.7%	100%
3.1. Monitoreo	5.3%	43.9%	29.8%	21.1%	100%
3.2. Retroalimentación	20.2%	53.5%	25.4%	0.9%	100%
3.3. Reflexión sobre el proceso de aprendizaje	57.9%	24.6%	17.5%	0.0%	100%
3.4. Reconocimiento	30.7%	43.0%	18.4%	7.9%	100%

Fuente: Elaboración propia.

3.5. Análisis según el modelo de Rasch Masters

El primer análisis del funcionamiento de las rúbricas del instrumento se realizó con Rasch. Los resultados indican que la mayoría de las rúbricas se ajustan de forma adecuada al modelo. Es decir que los estadígrafos INFIT y OUTFIT se encuentran entre un rango de .7 y 1.3, o la correlación punto biserial es mayor a .40 (Linacre, 2007). Sólo cuatro de ellas están fuera de estos parámetros. En la tabla 7 se puede identificar que el nivel de dificultad que van desde -2.2 a 3.35 en escala *logit*.

Tabla 7. Estadígrafos de ajuste del modelo Rasch para las rúbricas del instrumento

ÍTEMS	DIFI- CULTAD	IN- FIT	OUT- FIT	CORRELA- CIÓN- PUNTO BISERIAL	DISCRI- MINA
1.1a. Intención congruente con el PEP 2011	-1.95	0.78	1.52	0.48	1.02
1.1b. Intención congruente con necesidades de niños	-1.04	0.94	0.92	0.63	1.06
1.1c. Intención clara	-2.2	0.72	1.38	0.42	1.03
1.2. Conocimiento de los alumnos	-0.52	0.95	0.93	0.64	1.08
1.3a. Actividad congruente con intenciones	-0.21	1.18	1.09	0.64	0.89
1.3b. Actividad congruente con enfoque del campo	-0.56	0.84	0.78	0.65	1.24
1.3c. Plan para evaluación	0.35	1.2	1.19	0.52	0.76
1.4. Demanda cognitiva de planeación	0.36	1.4	1.53	0.4	0.52
2.1. Recupera conocimientos previos	-0.02	1.15	1.17	0.59	0.84
2.2. Promueve la interacción entre los niños	-0.17	0.81	0.62	0.52	1.13
2.3. Fomenta el deseo por aprender	1.14	0.86	0.83	0.61	1.16
2.4a. Hace el contenido interesante (estrategias)	-0.93	0.95	0.94	0.62	1.06
2.4b. Interés de los niños en la actividad	-0.58	0.98	0.99	0.59	1
2.5. Promueve la participación y responsabilidad en el aprendizaje	-0.27	0.83	0.82	0.62	1.22
2.6. Reglas	0.00	0.85	0.82	0.58	1.11
2.7. Demanda cognitiva de la actividad	0.96	1.10	1.10	0.52	0.88
2.8. Dominio del contenido	-0.60	1.35	1.35	0.33	0.73
2.9. Congruencia con lo planeado	-1.77	2	4.84	0.14	0.54
2.10a. Comunicación cálida (verbal)	-0.88	0.97	0.94	0.57	1.05
2.10b. Comunicación cálida (no verbal)	2.76	0.96	0.84	0.51	1.09
2.11. Consigna clara	-1.20	1.21	1.23	0.45	0.69
2.12. Manejo de errores	3.35	0.82	0.63	0.56	1.25
2.13. Orden del grupo	0.56	0.82	0.82	0.65	1.18
2.14a. Uso del tiempo en las actividades	-0.08	0.74	0.74	0.71	1.29
2.14b. Uso del tiempo en la jornada	-1.32	0.97	0.9	0.54	1.06
2.15. Atención a la diversidad	0.57	1.32	1.31	0.42	0.88
3.1. Monitoreo	-0.24	0.83	0.76	0.69	1.2
3.2. Retroalimentación	1.82	0.86	0.86	0.62	1.16
3.3. Reflexión sobre el proceso de aprendizaje	1.42	0.83	0.86	0.61	1.16
3.4. Reconocimiento	1.25	1.02	0.98	0.60	1.01

Fuente: Elaboración propia.

Los ítems que se salen de los parámetros de funcionamiento adecuado se excluyeron para los análisis subsecuentes. Entre ellos podemos destacar el primero del instrumento, el alto valor de outfit indica que las docentes que no tenían la habilidad para puntuar alto en la rúbrica, sí lo hicieron.

3.6. Dimensionalidad del instrumento

La Tabla 8 presenta la estructura factorial del instrumento, en ella se puede apreciar que se conformaron tres factores. El primer factor está relacionado con el Ambiente de aprendizaje y las prácticas de evaluación, el segundo con la Planeación de la enseñanza y el tercero con los Principios pedagógicos del programa. Esto hace una diferencia respecto a la estructura teórica planteada, pues se consideraba a la evaluación como un factor independiente.

Tabla 8. Cargas factoriales de los ítems en los tres factores rotados

ÍTEMS	FACTORES		
	1	2	3
1.1b. Intención congruente con necesidades de niños	.227	.694	.018
1.2. Conocimiento de los alumnos	.153	.727	.077
1.3a. Actividad congruente con intenciones	.233	.869	-.115
1.3b. Actividad congruente con enfoque del campo	.210	.747	.034
1.3c. Plan para evaluación	-.026	.692	.082
2.7. Demanda cognitiva de la actividad	-.190	.630	.377
2.2. Promueve la interacción entre los niños	.247	.154	.566
2.3. Fomenta el deseo por aprender	.130	.168	.636
2.4a. Hace el contenido interesante (estrategias)	-.112	.304	.780
2.4b. Interés de los niños en la actividad	.156	.006	.672
2.5. Promueve la participación y responsabilidad en el aprendizaje	.217	.068	.629
2.10b. Comunicación cálida (no verbal)	.338	.025	.438
2.14a. Uso del tiempo en las actividades	.344	.250	.423
2.1. Recupera conocimientos previos	.639	.115	.029
2.6. Reglas	.712	.151	-.092
2.10a. Comunicación cálida (verbal)	.612	.026	.108
2.11. Consigna clara	.422	-.098	.283
2.12. Manejo de errores	.717	.195	-.029
2.13. Orden del grupo	.754	.013	.110
2.14b. Uso del tiempo en la jornada	.491	-.071	.372
2.15. Atención a la diversidad	.666	.136	-.137
3.1. Monitoreo	.860	-.051	.070
3.2. Retroalimentación	.612	-.020	.260
3.3. Reflexión sobre el proceso de aprendizaje	.599	.223	.045
3.4. Reconocimiento	.660	-.278	.416

Fuente: Elaboración propia.

La Tabla 9 muestra los porcentajes de varianza a partir del Análisis Factorial Exploratorio. En ella se puede apreciar que la solución factorial final explica el 66% de la varianza, el factor de Evaluación y ambientes de aprendizaje obtuvo el porcentaje superior con 50%, seguido de Planeación con 9.8 % y Principios pedagógicos con 6 %.

Tabla 9. Porcentaje de varianza explicada y autovalores del instrumento

FACTOR	AUTOVALORES	% OF VARIANZA	ACUMULADA %
Evaluación y ambientes de aprendizaje	12.536	50.143	50.143
Planeación	2.462	9.849	59.991
Principios pedagógicos	1.555	6.220	66.211

Fuente: Elaboración propia.

La solución de tres factores se determinó a partir de la comparación de varios métodos, como se aprecia en la Figura 1. La regla de Kaiser (*eigenvalues*) retiene cinco factores, el análisis de Paralelo establece el primer punto de corte en tres factores, al igual que el

análisis de Coordinación Óptima, la prueba de Velicer's arrojó un resultado de dos factores y el Factor de aceleración de uno (ver figura 1).

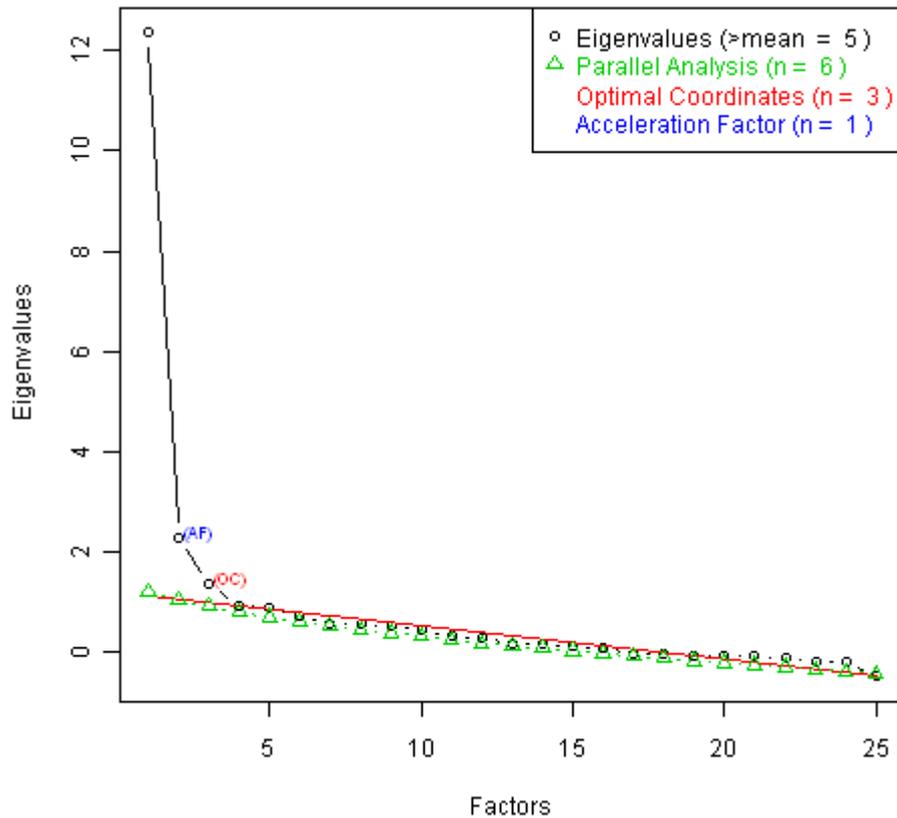


Figura 1. Estimación del número de factores a partir de cuatro métodos
Fuente: Elaboración propia.

Los métodos aproximaron a una selección de dos o tres factores. A partir de los estadísticos de bondad de ajuste de los modelos se identificó un mejor ajuste para la solución de tres factores. El índice de bondad de ajuste Goodness of Fit Index [GFI] tuvo un valor de .739 (valores cercanos al uno son considerados muy buenos), y el RMSR (Root Mean Square Residual) obtuvo un valor de .063 (valores menores a .05 son considerados excelentes).

3.7. Confiabilidad de las dimensiones

Se calculó la confiabilidad de los tres factores, mostrando valores cercanos al .90, lo que significa una consistencia alta para los ítems seleccionados. En la Tabla 10, además del usual alfa de Cronbach, se presenta el coeficiente alfa ordinal. Además se presentan los valores de la confiabilidad utilizando dos estadísticos específicos para variables de tipo ordinal: Theta de Armor y Coeficiente Theta ordinal. En todos los coeficientes se muestran valores adecuados, mayores a 0.85.

Tabla 10. Coeficientes de confiabilidad para los factores del modelo

	FACTOR 1	FACTOR 2	FACTOR 3
Alfa de Cronbach	.893	.882	.852
Coefficiente alfa ordinal	.932	.918	.915
Theta de Armor	.896	.884	.855
Coefficiente theta ordinal	.933	.920	.915

Fuente: Elaboración propia.

3.8. Confiabilidad de la observación

En este apartado se presentan los resultados de la confiabilidad de la aplicación del instrumento. La Tabla 11 contiene el porcentaje de acuerdo exacto entre los jueces y el coeficiente de concordancia Kappa de Cohen en su versión cuadrática ponderada. Asimismo, se presenta el porcentaje de varianza atribuida a distintas facetas de error.

Tabla 11. Estadísticos sobre la confiabilidad de la codificación

ÍTEM	Porcentaje de acuerdo	Kappa cuadrática ponderada	PORCENTAJE DE VARIANZA							Total
			Educadora (E)	Ocasión (O)	Codificación (C)	E*O	E*C	O*C	E*O*C	
1.1a.	91	.84	38	0	0	47	2	1	12	100
1.1b.	68	.50	23	3	0	24	0	0	50	100
1.1c.	89	.54	25	1	0	29	10	0	36	100
1.2.	70	.70	31	12	0	28	1	0	28	100
1.3a.	58	.64	15	6	0	44	0	0	36	100
1.3b.	72	.61	27	3	0	29	0	0	41	100
1.3c.	63	.67	3	0	0	60	0	0	36	100
1.4.	65	.61	5	0	0	56	1	3	35	100
2.1.	56	.58	32	0	1	25	0	2	39	100
2.2.	81	.62	13	5	2	42	0	0	38	100
2.3.	67	.48	19	2	0	18	0	4	58	100
2.4a.	61	.60	25	3	1	32	0	0	40	100
2.4b.	65	.70	28	0	0	42	2	0	27	100
2.5.	58	.37	25	9	2	4	13	0	46	100
2.6.	77	.62	34	0	1	28	5	0	32	100
2.7.	72	.68	12	7	1	46	0	0	35	100
2.8.	70	.55	0	0	0	57	10	3	29	100
2.9.	89	.70	6	2	1	61	0	0	30	100
2.10a.	65	.53	26	4	0	24	2	2	41	100
2.10b.	53	.22	11	0	12	12	8	3	55	100
2.11.	53	.39	18	0	1	24	1	4	53	100
2.12.	86	.46	24	2	0	20	2	3	48	100
2.13.	84	.52	35	0	0	16	0	1	48	100
2.14a.	79	.53	22	0	2	32	5	0	39	100
2.14b.	91	.66	43	2	1	21	4	0	29	100
2.15.	95	.57	26	1	0	31	4	0	39	100
3.1.	81	.72	54	2	0	14	0	0	29	100
3.2.	77	.43	33	3	0	5	0	5	54	100
3.3.	91	.65	15	10	1	42	0	0	32	100
3.4.	81	.59	24	0	0	33	0	3	40	100
Evaluación y ambiente de aprendizaje			57	1	0	22	0	4	16	100
Planeación			28	8	0	36	0	0	28	100
Principios pedagógicos			42	5	0	33	0	1	19	100

Fuente: Elaboración propia.

A partir de los resultados se puede identificar que el porcentaje de acuerdo entre los jueces fue alto, el rango oscila entre 53% y 91%. El estadístico Kappa muestra grados de acuerdo un poco menores, esto debido a que esta medida ajusta el grado de acuerdo considerando el azar al contestar la escala. La escala de valoración para el coeficiente Kappa de Landis y Koch (1977) establece los siguientes valores: pobre (menos de .01), leve (entre .01 - .20), aceptable (.21-.40), moderada (.41-.60), considerable (.61-.80) y casi perfecta (mayor a .81).

Los resultados muestran que sólo tres valores caen en el rango aceptable, 11 como aceptable y el resto como moderados. Adicionalmente, en la Tabla 11 se presenta la descomposición de las variaciones de las puntuaciones para el estudio G. Para este tipo de diseño existen siete tipos de fuentes de varianza: entre las educadoras (E), entre las ocasiones de observación (O), entre las codificaciones (C), entre las educadoras en las distintas ocasiones (E*O), entre las codificaciones con las educadoras (E*C), la codificación en las distintas ocasiones (O*C) y una variación no identificada (E*O*C). En esta tabla 11 se incluyen todos los ítems del instrumento y al final tres índices correspondientes a cada factor resultantes de los análisis previos.

A partir del análisis de descomposición de la varianza, se identifica que algunos ítems tienen un funcionamiento deficiente, esto coincide con los ya identificados por los análisis con Rasch y factorial exploratorio: 1.1c, 1.4, 2.8 y 2.9. Además se identifican que algunos de ellos tienen dificultades para su codificación esto debido a que muestran una alta variación en los puntajes de los jueces, en este caso encontramos al 2.10b. A diferencia del estadístico Kappa, con el análisis de descomposición de la varianza, la variación entre los observadores no parece ser importante, de hecho, para la mayoría de los reactivos, las diferencias se aproximan a cero.

Por otra parte, los resultados permiten identificar dos fuentes principales de variación: la varianza real de las educadoras (E) y las educadoras en las distintas ocasiones (E*O). Para el primer caso significa que las educadoras mostraron un desempeño muy distinto entre ellas. La variación de las educadoras en las distintas ocasiones fue la segunda más alta para la mayoría de ítems, lo que significa que una misma educadora puede variar mucho su desempeño cada vez que se le evalúa. Para algunos ítems, la proporción de varianza fue mayor que incluso la variación entre educadoras.

En la tabla 12 se muestran los coeficientes de generalizabilidad para el diseño propuesto en esta investigación, el cual incluye tres sesiones de observación por dos observadores. La confiabilidad es mayor en el factor de ambientes de aprendizaje, el de principios del PEP es el segundo más alto. Por su parte, el relacionado con la planeación tuvo el valor más bajo, lo que indica que el diseño propuesto no permite obtener una medición precisa de esta dimensión de la enseñanza.

Tabla 12. Coeficientes de generalizabilidad para los índices del instrumento

ÍNDICE	COEFICIENTE G
Planeación	.63
Principios PEP	.75
Evaluación y ambiente de aprendizaje	.85

Fuente: Elaboración propia.

4. Discusión y conclusiones

El propósito de este estudio fue desarrollar y aportar evidencias de validez de un instrumento para la evaluación de la enseñanza de las docentes de educación preescolar. Los resultados de este artículo mostraron que con algunas modificaciones, el instrumento desarrollado ofrece información relevante de la enseñanza que realizan las educadoras en el aula.

Un instrumento como el aquí desarrollado es de utilidad para el sistema educativo por varias razones: (a) Los principios pedagógicos que evalúa el instrumento son los mismos que las docentes deben implementar en el aula, por lo tanto, los resultados obtenidos son pertinentes para conocer en qué medida los docentes muestran el desempeño esperado; (b) Los criterios de evaluación del instrumento están alineados con el currículo y la evaluación del desempeño docente, por esto, podría convertirse en una herramienta útil para apoyar a las educadoras durante su formación en servicio; y (c), El manejo del instrumento es relativamente sencillo, dado que no requiere del conocimiento de aspectos conceptuales adicionales a los que el personal del sistema educativo maneja en el desempeño de sus funciones (sólo se requiere conocer el PEP 2011).

Además de la pertinencia del instrumento con la normatividad vigente, los análisis realizados mostraron que cuenta con propiedades psicométricas adecuadas. La mayoría de ítems presentaron valores apropiados en el análisis Rasch Masters. Esto significa que las rúbricas tuvieron un correcto funcionamiento. Los análisis también permitieron identificar que algunos de los ítems requieren modificaciones para que puedan medir adecuadamente el constructo.

Los resultados obtenidos permitieron identificar una estructura subyacente con tres factores, la estructura teórica propuesta incluía tres dimensiones pero con otra configuración. Permaneció la dimensión de Planeación, pero no se identificó una dimensión de evaluación como en un inicio se tenía contemplado. En su lugar emergieron dos factores uno relacionado con los Principios pedagógicos del programa y otro con la Evaluación y ambientes de aprendizaje. Estos resultados hacen pensar que desde la perspectiva de las docentes estos dos factores forman parte de un evento pedagógico que no es independiente uno del otro. Por otra parte, también se identificó un factor dominante en la estructura subyacente: Evaluación y ambiente de aprendizaje, puesto que retiene una gran proporción de la varianza, dejando al resto de factores con una contribución proporcionalmente reducida. Los estudios recientes de la estructura factorial para el instrumento CLASS se han ensayado otras estructuras factoriales con dos factores y un factor utilizando Análisis Factorial Confirmatorio. Estas dos estructuras muestran ajustes adecuados, aunque la estructura de tres factores muestra un ajuste ligeramente mejor. A la estructura de un solo factor se le ha denominado de "Enseñanza Efectiva" (Hamre et al. 2013). Estos resultados apuntan a seguir explorando la estructura que subyace a la práctica docente de las educadoras, incorporando muestras más grandes y otras técnicas analíticas.

Se reconoce como limitación del estudio la reducida muestra de docentes observados. Si bien se cumple con los requerimientos en cuanto al tamaño de muestra para utilizar el AFE, es deseable tener muestras más grandes para los análisis Rasch de Crédito Parcial. Futuras indagaciones con el instrumento tendrán que aplicarse a una muestra más

amplia de sujetos en aras de ofrecer datos más precisos sobre el funcionamiento del instrumento y los ítems en particular.

Otro de los hallazgos relevantes del estudio es que las docentes muestran desempeños muy distintos de una sesión a otra. Se identificó que la segunda mayor fuente de variaciones en la práctica pedagógica, solo después de la variación entre educadoras, se concentra en las diferentes ocasiones en que se observó a las educadoras. Estos resultados coinciden con lo encontrado por otros investigadores (Martínez, Borko y Stecher, 2012). A partir de ello se confirma que si se quiere observar la práctica docente de forma precisa, es necesario observar en varias ocasiones a las docentes, pues es alta la variación que existe entre las distintas ocasiones en que se observa.

Referencias

- Arter, J. A. (2010). Scoring rubrics. En P. Peterson, E. Baker y B. McGaw (Eds.), *International encyclopedia of education* (pp. 123-139). Nueva York, NY: Elsevier.
- Basto, M. y Pereira, J. M. (2012). An SPSS R-menu for ordinal factor analysis. *Journal of Statistical Software*, 46(4), 1-29. doi:10.18637/jss.v046.i04
- Centra, J. A. (1993). *Reflective faculty evaluation*. San Francisco, CA: Jossey-Bass.
- Darling-Hammond, L. (2012). Desarrollo de un enfoque sistémico para evaluar la docencia y fomentar una enseñanza eficaz. *Pensamiento Educativo: Revista de Investigación Educativa Latinoamericana*, 49(2), 1-20. doi:10.7764/PEL.49.2.2012.1
- Diario Oficial de la Federación [DOF], (2010). Ley Federal de Protección de Datos Personales en Posesión de Particulares. Recuperado de: <http://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>
- Frías-Navarro, D. y Pascual Soler, M. (2012). Prácticas del análisis factorial exploratorio (AFE) en la investigación sobre conducta del consumidor y marketing. *Suma Psicológica*, 19(1), 47-58.
- García, B., Loredó, J., Luna, E. y Rueda, M. (2008). Modelo de evaluación de competencias docentes para la educación media y superior. *Revista Iberoamericana de Evaluación Educativa*, 1(3), 124-136.
- Hamre, B., Pianta, R., Downer, J., DeCoster, J., Mashburn, A., Jones, S. y Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal*, 113(4), 461-487. doi:10.1086/669616
- Harms, T., Clifford, R. y Debby C. (1988). *Early Childhood Environment Rating Scale, Revised Edition (ECERS-R)*. Williston, VT: Teachers College Press.
- Gordon, R. A., Hofer, K. G., Fujimoto, K. A., Risk, N., Kaestner, R. y Korenman, S. (2015). Identifying high-quality preschool programs: New evidence on the validity of the Early Childhood Environment Rating Scale-Revised (ECERS-R) in relation to school readiness goals. *Early Education and Development*, 26(8), 1086-1110. doi:10.1080/10409289.2015.1036348
- Jornet, J. M., González Such, J. y Suárez, J. M. y Perales, M. J. (2011). Diseño de procesos de evaluación de competencias: Consideraciones acerca de los estándares en el dominio de las competencias. *Revista Bordón*, 63(1), 125-145.

- Landis, J. R. y Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. doi:10.2307/2529310
- Linacre, J. M. (2007). *A user's guide to Winsteps Ministeps: Rasch-model computer programs*. Chicago, IL: Electronic Publication.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174. doi:10.1007/bf02296272
- Martínez, J. F., Borko, H. y Stecher, B. M. (2012). Measuring instructional practice in science using classroom artifacts: Lessons learned from two validation studies. *Journal of Research in Science Teaching*, 49(1), 38-67. doi:10.1002/tea.20447
- Millis, B. J. (2006). Peer observations as a catalyst for faculty development. En P. Seldin (Ed.), *Evaluating faculty performance. A practical guide to assessing teaching, research, and service* (pp. 82-95). San Francisco, CA: Anker Publishing Company.
- Myers, R., Martínez J. F. & Linares M.E., (2003). *En Búsqueda de la Calidad Educativa en Centros Preescolares*. Informe presentado a la Secretaría de Educación Pública. Dirección General de Investigación Educativa. Hacia una Cultura Democrática, A.C. México, D.F. Documento no publicado.
- Pedroza, L. H., Álvarez, A. C. y Jiménez, A. B. (2013). *La implementación del PEP 2004 en las aulas*. En INEE (Eds.), *Prácticas pedagógicas y desarrollo profesional docente en Preescolar* (pp. 15-57). Ciudad de México: Instituto Nacional para la Evaluación de la Educación.
- Pedroza, L. H., Vilchis, J. E., Álvarez, A. C., López, A. Y. y García, M. A. (2013). *Prácticas pedagógicas y desarrollo profesional docente en preescolar. Reporte técnico*. Ciudad de México: Instituto Nacional para la Evaluación de la educación. Recuperado de http://publicaciones.inee.edu.mx/buscadorPub//P1/D/240/P1D240_13E13.pdf
- Pianta, R. C. y Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109-119. doi:10.3102/0013189X09332374
- Secretaría de Educación Pública (SEP). (2006). *La implementación de la reforma curricular en la educación preescolar: Orientaciones para fortalecer el proceso en las entidades federativas. Programa de Renovación Curricular y Pedagógica en la Educación Preescolar*. Ciudad de México: Autor.
- Secretaría de Educación Pública (SEP). (2011). *Programa de Educación Preescolar 2011. Guía para la educadora*. Ciudad de México: Autor.
- Secretaría de Educación Pública (SEP). (2015). *Evaluación del desempeño docente. Ciclo escolar 2015-2016. Perfil, parámetros e indicadores para docentes*. Ciudad de México: Autor.
- Shavelson, R. J. y Webb, N. M. (1991). *A primer on generalizability theory*. Thousand Oaks, CA: Sage Publications.
- Shulman, L. (2005). Conocimiento y enseñanza: Fundamentos de la nueva Reforma. *Revista de Currículum y Formación del Profesorado*, 9(2), 1-30.
- Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G. y Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, 8(1), 1-18. doi:10.1186/1471-2288-8-33.

Breve CV de los autores

Luis Horacio Pedroza Zúñiga

Estudiante de doctorado del Instituto de Investigación y Desarrollo Educativo, Universidad Autónoma de Baja California, México. ORCID ID: 0000-0002-5256-2967. Email: horaciopedroza@hotmail.com

Edna Luna Serrano

Investigadora del Instituto de Investigación y Desarrollo Educativo, Universidad Autónoma de Baja California, México. ORCID ID: 0000-003-1496-548X. Email: eluna@uabc.edu.mx