

# MANUAL DE USO DE RESULTADOS DE LA EVALUACIÓN NACIONAL DIAGNÓSTICA DE LA FORMACIÓN INICIAL DOCENTE

(Resumen)

ESTUDIO DE CARACTERIZACIÓN DE CARRERAS DE  
PEDAGOGÍA A NIVEL NACIONAL Y PROPUESTA DE PLAN  
ESTRATÉGICO PARA EL FORTALECIMIENTO DE LA  
FORMACIÓN INICIAL DOCENTE



**Manual de uso de resultados de la  
Evaluación Nacional Diagnóstica  
de la Formación Inicial Docente**

**Material elaborado por:**

Área de Formación Inicial Docente  
Centro de Perfeccionamiento,  
Experimentación e Investigaciones  
Pedagógicas  
Ministerio de Educación

Coordinador:

Christian Libeer Brouckaert

Encargado Evaluación Nacional  
Diagnóstica:

Christian Bustos Cárdenas

Profesional:

Pabla Benavides Garrido

Con la colaboración de:

Instituto de Informática Educativa de  
la Universidad de la Frontera

Diseño y diagramación:

Rita Galleguillos Zamora

Enero 2025

Este documento utiliza el lenguaje en masculino de manera general, esto se hace únicamente con fines de simplicidad y para evitar la sobrecarga lingüística. Este uso debe interpretarse de manera inclusiva, abarcando todas las identidades de género, sin pretender exclusión alguna.

## 1. PRESENTACIÓN

La Evaluación Nacional Diagnóstica (en adelante END) es parte de la política de fortalecimiento de la calidad de la formación inicial docente. Normada por la Ley N°20.903 del año 2016 que crea el Sistema de Desarrollo Profesional Docente y se incorporó a la Ley N°20.129 del año 2006, que establece un Sistema Nacional de Aseguramiento de la Calidad de la Educación Superior.

Su marco evaluativo tiene como referente a los Estándares de la Profesión Docente (en adelante EPD) y transita -de manera progresiva- desde los Estándares Orientadores para Carreras de Pedagogía del año 2012 a los EPD del año 2021; Pedagógicos a contar del año 2024 y Disciplinarios a contar del año 2025. En el caso de Educación Parvularia mantiene el referente en los Estándares Orientadores.

Mediante la aplicación de cuatro instrumentos de evaluación a los estudiantes de pedagogía, aporta información acerca de la calidad de los programas de formación de docentes y educadoras/es, a saber: Prueba de Conocimientos Pedagógicos Generales (en adelante PCPG), Pregunta de respuesta abierta para la Reflexión Pedagógica (en adelante PRP), Prueba de Conocimientos Didácticos y Disciplinarios (en adelante PCDD), y un Cuestionario Complementario (en adelante CC).

La END proporciona información para una reflexión académica permanente basada en los desempeños evidenciados por las diversas cohortes de estudiantes. Este manual aporta criterios de interpretación y uso de dicha información para que las universidades, en función de su contexto, identidad y territorio, examinen, actualicen y desarrollen acciones en respuesta a sus desafíos. Con ello, esperamos atender la necesidad de fortalecer la calidad de las carreras de pedagogía en pos de lo que debe alcanzar cada estudiante al finalizar su formación profesional.

## 2. ASPECTOS METODOLÓGICOS

### 2.1 Criterios de aproximación

La END es un sistema evaluativo estandarizado, diseñado para evaluar los conocimientos pedagógicos y disciplinares de los estudiantes de pedagogía en Chile. Su objetivo principal es proporcionar información de la calidad de la Formación Inicial Docente, identificando fortalezas y áreas de mejora. Esta evaluación se rige por técnicas estadísticas que aseguran su validez, confiabilidad y justicia evaluativa para los participantes. Se basa en un enfoque integral de la validez (Messick, 1989, 1995; Cronbach 1988; Sireci, 2009). Mide conocimientos específicos y considera la capacidad de los futuros docentes para aplicarlos en contextos reales de enseñanza. Estas técnicas aseguran que los resultados de la medición reflejen los conocimientos necesarios para ejercer la docencia en el sistema educativo chileno (Taut & Sun, 2014).

Otro aspecto clave de la END es su confiabilidad, a través de procesos rigurosos de diseño, pilotaje y análisis de ítems aplicados en régimen, que aseguran la consistencia de las mediciones y aplicabilidad a diversos contextos.

En esta evaluación los aspectos éticos son fundamentales para garantizar que el proceso sea justo, transparente y respetuoso con los derechos de los estudiantes evaluados (AERA, APA & NCME, 2014; APA, 2017). La ética en este dispositivo no solo implica el diseño de las pruebas, también las condiciones de aplicación y uso adecuado de sus resultados para promover la mejora institucional de los procesos formativos, la equidad y el bienestar de los evaluados.

Por ello, si bien la rigurosidad estadística en el procesamiento de datos es fundamental para garantizar que la END sea válida y confiable, es tanto o más importante interpretar los resultados desde lo que en la actualidad se conoce como una **perspectiva edumétrica** (Murillo & Román, 2011; García & Yot, 2018; Brookhart & McMillan, 2020). Esto es, focalizar el uso de los resultados para mejorar los procesos educativos y tomar decisiones que impacten positivamente en el aprendizaje y la formación de los futuros docentes. Lo anterior implica no solo analizar los datos desde la técnica evaluativa (p. ej., puntuaciones, percentiles o niveles de logro), sino también considerar el contexto educativo, la coherencia con lo observado localmente en las prácticas pedagógicas y las necesidades específicas de los estudiantes y las instituciones.

### 2.2 Procesamiento

Todo instrumento de la END ha sido diseñado siguiendo una tabla de especificaciones que operacionaliza los constructos a evaluar y ha sido sometido a pilotaje y validación de expertos. La evidencia empírica sobre el funcionamiento de los ítems y el juicio experto

permiten asegurar su calidad técnica. Se observa así tanto el nivel de dificultad y capacidad de discriminación de los ítems (Downing & Haladyna, 2006). Luego se procede al ensamblaje de las pruebas considerando los procesos anteriores y su alineación con los estándares de la FID. Finalmente, definida la estructura se incorporan las preguntas ancla que son ítems comunes entre diferentes versiones de la prueba aplicada en distintos años (Kolen & Brennan, 2014).

Las preguntas ancla se incluyen con el propósito de permitir la equiparación de los resultados; asegurar que entre las cohortes evaluadas los instrumentos usados no son ni más ni menos difíciles. De este modo se cuenta con un punto de referencia para una misma prueba en distintos años. Esto es especialmente relevante en la END donde se busca monitorear el progreso de los programas de formación al comparar resultados entre cohortes.

Como se ha indicado, los instrumentos aplicados son:

- Prueba de Conocimientos Pedagógicos Generales (PCPG): Evalúa conocimientos y habilidades pedagógicas generales.
- Prueba de Conocimientos Disciplinarios y Didácticos (PCDD): Evalúa conocimientos disciplinarios y su aplicación didáctica.
- Pregunta de Respuesta Abierta para la Reflexión Pedagógica (PRP): Examina la capacidad para resolver problemas pedagógicos y la comunicación escrita.
- Cuestionario Complementario (CC): Recoge información sobre los estudiantes y sus expectativas.

### 2.3 Aplicación

Los instrumentos de la END se aplican en laboratorios computacionales, con pruebas informatizadas y las acomodaciones necesarias en caso de necesidades especiales.

### 2.4 Corrección

Respecto de las respuestas obtenidas en las PCPG y PCDD se realiza el siguiente procedimiento de corrección y obtención de puntajes:

#### A) Primera etapa

**Paso uno:** Se aplica **Teoría Clásica de Test (TCT)**. Es un procedimiento estadístico que estima la confiabilidad de la prueba. Se calcula el parámetro "alfa de Cronbach", que mide la consistencia interna de los resultados obtenidos de los instrumentos. Este parámetro va de 0 a 1 y un valor aceptable es 0,7 o superior.

**Paso dos:** Se calcula la **correlación biserial** de cada ítem, esto observa la relación entre las respuestas dadas y la puntuación total de la prueba. Este es un coeficiente cuyos valores van entre -1 y 1, calcula la varianza de cada ítem y su contribución a la varianza total de la prueba. Mientras más se acerque al valor 1 (0,3 o superior) más robusta es la relación entre los elementos antes mencionados.

**Paso tres:** Dificultad del ítem; se procesa los datos de la proporción de evaluados que responden correctamente cada ítem. En la END se acepta un rango que va entre un 20% y un 80% de respuestas correctas. Valores fuera de rango implican la eliminación del ítem, pues no aporta información útil para diferenciar entre estudiantes evaluados.

#### **B) Segunda etapa**

**Paso cuatro:** Se aplica Teoría de Respuesta al Ítem (IRT) de un parámetro, en este caso el Modelo Rash. Este, permite enfocar el análisis en la dificultad del ítem como parámetro para predecir la probabilidad de que un evaluado responda correctamente, esto permite estimar la **habilidad de los evaluados** y la dificultad de los ítems en una misma escala. Luego, se aplica el método WLE, Estimación de Máxima Verosimilitud Ponderada, que calcula la **habilidad de los evaluados**. Resulta ser el método más adecuado, en esta etapa, dado el tamaño de los instrumentos de la END.

**Paso cinco:** los datos se ordenan en una escala común, en este caso escala Logit (que permite identificar a los estudiantes con mayor habilidad que tienen mayor probabilidad de responder ítems con mayor dificultad y estudiantes con menores habilidades tienen mayores probabilidades de respuestas de ítems con menor dificultad) lo que robustece la invarianza de la medición y permite entender mejor el desempeño de los estudiantes y la calidad de los ítems, facilitando el monitoreo del progreso, a través del tiempo, de los programas formativos.

**En la actualidad, dada la transición de estándares y cambio del marco evaluativo, es necesario realizar una distinción en el sexto paso:**

**Paso seis:** en el caso de las PCDD, se transforma a una escala estándar multiplicando los resultados por 20 y sumando una constante de 100 puntos.

En el caso de las PCPG, se aplica un procedimiento de Standard Setting, Scale Anchoring, para la obtención de **puntos de corte** que determinan la ubicación del estudiante dentro de un nivel de desempeño de la rúbrica. Esta técnica describe, de acuerdo con el nivel obtenido, lo que estudiantes saben y son capaces de hacer.

Este procedimiento inicialmente considera tres niveles. En términos generales las descripciones distinguen lo siguiente:

- **Primer nivel:** el estudiante evidencia resultados por debajo de un punto crítico de desempeño, se estima que presentará dificultades importantes para madurar su proceso formativo al momento de finalizar su carrera y que, por lo tanto, requiere apoyo inmediato e intensivo.
- **Segundo nivel:** el estudiante evidencia pertenecer a un rango intermedio esperable; debería ser apoyado para consolidar sus conocimientos.
- **Tercer nivel:** el estudiante evidencia resultados destacados y cuyos apoyos deben orientarse a desplegar sus conocimientos y profundizar sus capacidades.

**Preguntas de respuesta abierta:** son corregidas por un equipo experto especialmente capacitado. Esta sección evalúa tanto la resolución de situaciones pedagógicas como la comunicación escrita. Se corrige mediante una rúbrica holística para la reflexión pedagógica y una específica centrada en los aspectos formales de la redacción.

Finalmente, el **Cuestionario Complementario** es anonimizado y sometido a análisis factorial para identificar aspectos relevantes que puedan ser de utilidad para las instituciones formadoras. Luego las respuestas agregadas son puestas a disposición de las universidades.

### 3. INTERPRETACIÓN DE RESULTADOS

Los resultados obtenidos por los estudiantes en los diversos instrumentos que componen la END son presentados a las instituciones formadoras mediante la plataforma [www.futurodocente.cl](http://www.futurodocente.cl). Es importante recordar que, dadas las características de los estándares y los instrumentos de la END no es posible realizar comparaciones entre instituciones.

#### 3.1 Criterios generales

Para la interpretación de dichos resultados deben considerarse, al menos, tres topos, primero, los individuos (estudiantes), segundo, los contenidos de la evaluación y tercero, el tiempo:

**Estudiantes:** Observar la heterogeneidad de los resultados para una cohorte, es decir, como se agrupan los resultados de diferentes niveles de desarrollo. Por ejemplo, como se distribuyen los niveles de logro de las PCPG, o la distribución de los resultados de las PCDD. Una mayor dispersión indica que los estudiantes no atienden de la misma manera al esfuerzo formativo y, por tanto, surge la necesidad de estrategias pedagógicas diferenciadas que permitan el avance del grupo en su conjunto. Las hipótesis acerca de las causas de las diferencias (heterogeneidad) en el desarrollo formativo pueden ser varias, y es propio del análisis en contexto determinar estrategias que podrían resolverlas. Por ejemplo, las explicaciones pueden estar en el diseño formativo o en las características de los estudiantes. En el primer caso las metodologías de formación podrían requerir mayor diversidad, en el segundo, las características de los estudiantes indicarían la necesidad de un mayor apoyo específico a algunos de ellos. Por otro lado, una menor dispersión, pero altos puntajes, sugiere estabilidad en la formación y un diseño formativo adecuado. En el caso de obtener puntajes menores, nuevamente, las hipótesis explicativas no pueden encontrar respuestas en la evaluación estandarizada, sino en elementos internos del ordenamiento del programa de formación como, por ejemplo, las metodologías de enseñanza o las capacidades de base que han desarrollado los estudiantes durante todo su proceso formativo, incluyendo el escolar y el capital académico en sus familias de origen.

**Contenidos:** La calidad del diagnóstico no solo se relaciona con el tamaño y representación de la población que participa de una evaluación respecto del programa de formación, también depende de las distinciones que se realizan al analizar los resultados. Observar la coherencia o, uno por uno, los resultados de dominios disciplinares, temas curriculares o estándares de la formación inicial para una misma prueba, son distinciones que precisan aquellos aspectos que la gestión institucional debe mantener por los logros obtenidos, pero también especifican las dimensiones de la evaluación que requieren de una intervención institucional para la superación de sus desafíos. Por ejemplo, qué tan similares o disímiles son los resultados a nivel de dominio, tema o estándar, en qué aspectos se obtienen logros superiores y distinguiéndolos de aquellos que evidencien mayores desafíos.

**Evolución en el tiempo:** Observar si un desempeño, a nivel de dominio, tema o estándar viene sistemáticamente al alza o en descenso con independencia de si se trata de un nivel de logro alto o bajo, constatando el crecimiento o decrecimiento de un puntaje. Los resultados de las carreras son comparables consigo mismos y una buena perspectiva de análisis considera responder a: ¿cuál ha sido el mayor nivel de logro entre las cohortes y a través del tiempo? O ¿Cuál es el menor grado de dispersión de los resultados? y, por lo tanto, ¿qué fortalezas deben ser reforzadas y qué desafíos deben ser superados?

Las medidas de tendencia central, como el promedio, son apropiadas para analizar la evolución de los resultados evaluativos en el tiempo y podrán ser una herramienta de gestión que, en la observación del punto por punto, o estándar por estándar, facilita identificar las áreas que se han desarrollado adecuadamente de aquellas que presentan mayores desafíos. Combinado con las medidas de dispersión, facilitan una comprensión más amplia de lo que ocurre con los estudiantes. En el uso de los resultados de la END una lectura estratégica, que combine y comprenda los distintos estadígrafos en conjunto con la elaboración de preguntas e hipótesis que orienten los análisis, hace que este instrumento de evaluación sea más robusto, no solo desde su construcción técnica, también desde el uso que hacen los académicos, estudiantes y autoridades universitarias.

### 3.2 Presentación de Puntajes

Los resultados se presentan mediante gráficos.

**Gráfico de nivel de desempeño de la PCPG:** Indica el porcentaje de estudiantes con resultados descendidos bajo un punto considerado crítico, el de estudiantes con resultados sobre ese punto crítico y de estudiantes con resultados sobresalientes.

**Gráfico de nivel de desempeño de la PCDD:** Muestra la dispersión de los puntajes de los estudiantes, permitiendo identificar heterogeneidad o homogeneidad dentro de los programas de formación evaluados con una prueba.

**Rendimiento institucional en relación con el promedio nacional desde 2016 hasta 2023:** el gráfico de Puntajes Promedio ayuda a detectar tendencias generales, sin embargo, no debe usarse para establecer comparaciones ya que cada institución se desenvuelve en concordancia con su propio proyecto educativo y estructura curricular, en contextos y condiciones distintas que pueden influir significativamente en los resultados y hacen que dichas comparaciones no sean válidas. De esta manera, los resultados de una prueba no indican que una carrera o institución se encuentra por sobre o bajo otra, o si es pertinente considerar sus resultados a la luz del promedio de otras instituciones, pues no es su propósito.

Como se ha indicado, mientras mayor sea la dispersión de puntajes que obtienen los estudiantes sugiere la necesidad de apoyo individualizado, por el contrario, una menor dispersión, pero con altos puntajes, puede indicar consistencia en la enseñanza de determinados estándares.

Finalmente, la combinación de análisis de promedios y distribución permite tomar decisiones informadas sobre programas formativos y priorizar recursos de manera efectiva. Mientras el promedio se puede relacionar con la evolución en el tiempo, las distribuciones explican mejor la composición interna de los grupos evaluados.

### 3.3 Porcentajes de logro de dominios y temas

El porcentaje de logro de dominios pedagógicos y de temas disciplinarios de los instrumentos PCPG y PCDD permite evaluar el desempeño de los estudiantes en áreas específicas. Se calcula como la proporción de respuestas correctas en relación con el total de ítems asociados a un dominio o tema, proporcionando información detallada sobre fortalezas y áreas de mejora.

Los resultados se presentan en gráficos de caja o box plot, que muestran la dispersión de los puntajes a través de la mediana, los percentiles 25 y 75, y los valores mínimo y máximo. Cada segmento de la caja, "cuerpo" y "bigotes", representa al 25% de la población evaluada y el tamaño de cada segmento representa qué tan parecidos o diferentes son sus resultados.

Estos gráficos permiten diferenciar tres situaciones clave:

- Dominios o temas con desempeño homogéneo (caja achatada): Se observan resultados consistentes entre los estudiantes, lo que sugiere una comprensión uniforme del contenido.
- Dominios o temas con mayor variabilidad (caja alargada): La dispersión significativa de los puntajes indica diferencias en la preparación de los estudiantes.
- Dominios o temas con bajo porcentaje de logro (caja cuyo cuerpo se ubica en la parte de abajo del gráfico): Identifican áreas que requieren mayor atención pedagógica.

Aunque los promedios puedan ser similares, las diferencias en dispersión revelan que los temas no se aprenden de la misma manera por todos los estudiantes, lo que plantea preguntas sobre estrategias de enseñanza más eficaces. El análisis permite distinguir entre desafíos en la enseñanza y en el aprendizaje:

- Si la dispersión es baja (caja estrecha), pero el promedio es bajo, se puede señalar como hipótesis que la dificultad radica en la **enseñanza**.

- Si la dispersión es alta (caja alargada), se puede señalar como hipótesis que los desafíos están en el **aprendizaje** o en la diversificación de las **estrategias de enseñanza**.

Estos gráficos permiten observar y comparar tanto cuáles temas se encuentran mejor desarrollados y cuáles presentan mayores desafíos respecto de una misma cohorte, como la caracterización de la evolución de cada dominio o tema en el tiempo.

Cada gráfico de caja permite, respecto a los dominios o temas, comparar los porcentajes de logro entre distintas cohortes, identificando tendencias y variaciones en los temas evaluados. Estas comparaciones deben considerarse con cautela, ya que las cohortes pueden diferir en tamaño, contexto educativo y condiciones de evaluación. Para una interpretación adecuada, es necesario analizar la variabilidad interna de las cohortes, los factores contextuales que influyen en el desempeño y las áreas de mejora en la enseñanza. Cambios curriculares, recursos pedagógicos, condiciones socioeconómicas o eventos externos pueden influir en los resultados, por lo que es fundamental contextualizar las interpretaciones.

Finalmente, esta información permite focalizar esfuerzos en mejorar la equidad educativa, asegurando que la formación docente responda a los estándares de calidad y atienda las necesidades de los estudiantes.

### 3.4 Porcentaje de logro por Estándares

El porcentaje de logro permite observar el desempeño de los estudiantes en el grado de conocimiento de aquellos aspectos esenciales definidos en los Estándares de Formación Inicial Docente. A diferencia de otros análisis, este enfoque centra la atención y responsabilidad en la enseñanza y formación de los futuros docentes.

Los resultados se presentan mediante gráficos de barra, donde el porcentaje promedio de logro por estándar se compara con el promedio nacional. Esta visualización facilita la identificación de fortalezas y áreas de mejora en cada institución, permitiendo el diseño de estrategias para optimizar la enseñanza. Si bien, el nivel de logro se expresa en porcentaje, dado que la END es un dispositivo diagnóstico no se espera que se alcance un 100% de logro, sino que se busca reflejar el estado del conocimiento del estudiante respecto de cada estándar evaluado en particular.

El análisis de los resultados por estándares puede realizarse por cohortes, lo que permite evaluar la evolución del desempeño de los estudiantes a lo largo del tiempo. Este enfoque facilita la identificación de tendencias y brinda a las universidades información para implementar medidas preventivas y remediales y evaluar su impacto. Los resultados deben interpretarse considerando factores contextuales y características propias de cada cohorte que es posible sistematizar a partir del cuestionario complementario.

### **3.5 Pregunta de Reflexión Pedagógica de Respuesta Abierta**

La PRP evalúa las capacidades de los estudiantes en la resolución de problemas pedagógicos y en habilidades de comunicación escrita. Los resultados se presentan mediante gráficos que muestran el desempeño en cuatro niveles: A, indica el logro más alto; B, medio alto; C, medio bajo; y D, indica un nivel de logro descendido..

A diferencia de los instrumentos explicados con anterioridad (PCPG y PCDD), centrados por esencia en el conocimiento, estas preguntas de respuesta abierta ofrecen una mayor aproximación al desarrollo de habilidades y capacidades. Por tanto, complementan la descripción evaluativa del proceso de formación profesional de los estudiantes de pedagogía.

### **3.6 Correlación**

El gráfico de correlación que se despliega en la página [www.futurodocente.cl](http://www.futurodocente.cl) muestra la relación entre los puntajes obtenidos por los estudiantes en las pruebas PCPG y PCDD.

Lo esperable es una correlación de los puntajes en sincronía entre ambos instrumentos (PCPG y PCDD). Los resultados generales observados muestran que existe una tendencia general de incremento conjunto. Los cuadrantes que dividen el área del gráfico identifican patrones específicos de desempeño: el cuadrante inferior izquierdo refleja a los estudiantes con bajo rendimiento y el superior derecho a los de mejor desempeño. Estos hallazgos orientan estrategias pedagógicas para mejorar el rendimiento académico, por ejemplo, al evidenciar una alta correlación entre los instrumentos es posible establecer estrategias de gestión estudiantil para el mejoramiento de su desempeño en materias disciplinares, puesto que ese esfuerzo instalaría una cultura académica que beneficiaría la dimensión complementaria, es decir, en este caso, la pedagógica; por otro lado, resultados con una baja correlación invitan a examinar los criterios de calidad mínimos comunes a todas las etapas y áreas del proceso formativo.

### **3.7 Contextualización del desempeño con características de las cohortes**

El Cuestionario Complementario permite contextualizar los resultados de los instrumentos PCPG y PCDD, ayudando a interpretar el desempeño de cada carrera de manera integral.

Este instrumento evidencia los grados de acuerdo con algunos aspectos fundamentales de la pedagogía y los modos concretos en que se hace práctica, así como las percepciones sobre diversos aspectos de su propio proceso formativo. Por otra parte, estos datos deben ser ubicados en contexto a partir de la información que maneja la propia institución formadora.

Conocer las características de los estudiantes, como su capital cultural, antecedentes de salud mental y estilos de vida, es esencial para diseñar estrategias de apoyo focalizadas y adaptadas

a sus realidades. Además, se recomienda considerar factores motivacionales, estrategias de estudio y acceso a recursos educativos, ya que influyen en el rendimiento académico. Es igualmente importante registrar y evaluar el impacto de innovaciones o cambios curriculares, analizando tanto los resultados de aprendizaje como la percepción estudiantil. Estos análisis contextuales enriquecen la retroalimentación institucional y promueven la colaboración entre académicos y estudiantes, favoreciendo la creación de un entorno educativo más empático con las distintas realidades que los estudiantes enfrentan su proceso formativo.

## 4. USOS SUGERIDOS

El objetivo de la END es proporcionar información relevante para la mejora continua de los procesos de formación inicial docente. Esto implica que los resultados de esta evaluación deben interpretarse adecuadamente. De la misma manera, debe distinguirse entre los datos que son útiles para fortalecer los procesos de enseñanza y aquellos que permiten comprender el contexto en que sucede el aprendizaje y también identificar las áreas donde los datos pueden servir de insumo para el trabajo colaborativo entre académicos.

Dado su carácter diagnóstico y su referencia amplia a todo el cuerpo de estándares, el propósito de la END no es certificar a estudiantes o instituciones, sino entregar antecedentes para la gestión académica en dos sentidos: por una parte, el apoyo con remediales inmediatos para los estudiantes, y por otra, la revisión constante de los procesos internos de los programas de formación de los futuros docentes. De este modo, los focos fundamentales a partir de los cuales debe ser interpretada y utilizada la END se enmarcan en las preguntas:

- ¿Qué observamos en los resultados respecto de fortalezas y debilidades de nuestros estudiantes?
- ¿Qué desafíos nos proponen los resultados respecto de nuestra gestión estudiantil?
- ¿Cómo alimentan estos resultados nuestra reflexión acerca de la gestión académica para fortalecer los procesos formativos de las próximas cohortes?
- ¿Qué estrategias podemos implementar para abordar las preguntas anteriores?

### 4.1 Uso de los resultados para la gestión académica

El diagnóstico de necesidades de formación del cuerpo académico se puede enriquecer con el análisis de los niveles de logro en los temas y estándares evaluados. Los EPD reflejan una visión acerca de lo concordado como necesario en relación con el saber y el saber hacer de los futuros docentes. Si bien existen contenidos específicos, estos aspectos, en

la mayoría de las veces, son transversales e implican un proceso de apropiación por parte de los estudiantes de pedagogía. Por lo tanto, tributan no solo a diversas asignaturas, sino que también a la cultura académica institucional. Los resultados de la END no se explican por lo que un académico pueda haber enseñado puntualmente en una clase, sino por la comprensión del ethos docente, que articula los contenidos, dándoles sentido y uso práctico; por ello, los contenidos evaluados y sus resultados involucran a toda la comunidad.

Para el uso de los resultados en la gestión académica es recomendable sistematizar los logros y desafíos a nivel de carreras e institución. Aquellos dominios pedagógicos o temas disciplinarios con alto desempeño, homogéneo en el grupo evaluado y consistente en el tiempo y entre las cohortes, reflejan un trabajo pedagógico exitoso que la gestión académica debiese consolidar, mientras que los de bajo rendimiento o que evidencian un descenso continuo, requieren revisión y discusión de la comunidad académica en torno a las prácticas pedagógicas con las que se forman a los futuros profesionales.

#### **4.2 Uso de los resultados para la gestión estudiantil**

Se recomienda considerar las características de contexto de los estudiantes, como las sociodemográficas, académicas y psicoeducativas, junto con los resultados de pruebas de admisión o de la PCPG y la PCDD y otros registros institucionales, para obtener una visión más completa de sus necesidades.

Entre las estrategias recomendadas, se destaca la identificación y apoyo a los grupos de estudiantes con bajo rendimiento en comparación con su cohorte, especialmente aquellos situados en la categoría inferior en las pruebas pedagógicas. Por otro lado, la información que proporciona la PCDD permite reconocer estándares que deben ser fortalecidos focalizando los apoyos inmediatos para los estudiantes.

Asimismo, el análisis contextual de las condiciones de los estudiantes, incluidas sus experiencias y acceso a recursos académicos, facilita el diseño de estrategias para abordar inequidades estructurales. También es crucial identificar qué estrategias de estudio son efectivas para los estudiantes con mejor desempeño, promoviendo su adopción a través de tutorías de pares o grupos de estudio.

La identificación temprana de grupos en riesgo, mediante el uso de indicadores institucionales, permite enfocar los recursos y derivar estudiantes a unidades de apoyo si es necesario. El objetivo final es mitigar las inequidades estructurales y mejorar el proceso formativo, asegurando que todos los estudiantes reciban el apoyo necesario para optimizar su desempeño académico.

### 4.3 Uso de los resultados para la gestión curricular

Los resultados de la END son una herramienta clave para evaluar y ajustar los planes curriculares, alineándolos con las competencias y estándares necesarios para la formación docente. Es fundamental reconocer que estos abarcan habilidades y conocimientos amplios, por lo que no se puede vincular asignaturas específicas a cada estándar de manera directa. No obstante, sí debe observarse la transición entre el perfil de ingreso y el perfil de egreso y la articulación entre el propio proyecto educativo institucional y los EPD como parámetro global. Por tanto, se espera que la información proporcionada por la END contribuya a ajustes curriculares de carácter global, considerando al conjunto de asignaturas y su contribución al desarrollo de los conocimientos, capacidades y disposiciones de los futuros docentes.

Los resultados de los instrumentos PCPG y PCDD permiten identificar fortalezas y áreas que requieren reforzarse en las carreras. La correlación entre ambos resultados ayuda a detectar tanto la coherencia en los resultados obtenidos por cada estudiante, como posibles discrepancias en el desempeño pedagógico y disciplinar, lo cual puede reflejar diferencias en las oportunidades de aprendizaje. Esta información orienta la priorización de mejoras curriculares, ya sea en términos pedagógicos o disciplinarios.

El análisis de los porcentajes de logro por tema y estándar puede señalar áreas específicas del currículo que necesitan atención. Para ello, resulta relevante indagar cómo los programas de formación, mediante su estructura curricular y diversas prácticas académicas se aproxima a dichos estándares. La información que proporciona la END muestran que dentro del conjunto de experiencias formativas existen algunas cuyos resultados podrán ser replicables, con las adecuaciones correspondientes, en otras carreras o incluso otras instituciones de educación superior.

Además, es crucial incorporar actividades de nivelación para los estudiantes que ingresen con brechas significativas y revisar los objetivos de formación para asegurarse de que estén alineados con los estándares evaluados.

## REFERENCIAS

American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME]. (2014). Standards for educational and psychological testing. American Educational Research Association.

American Psychological Association. (2017). Ethical principles of psychologists and code of conduct. <https://www.apa.org/ethics/code>

Brookhart, S. M., & McMillan, J. H. (2020). Educational measurement: From foundations to future. Routledge. <https://doi.org/10.4324/9781315709307>

CNA. (s.f). Criterios y Estándares para la Acreditación de Carreras y Programas de Pedagogía. Comisión Nacional de Acreditación.

CPEIP. (2022). Estándares Pedagógicos y Disciplinarios para Carreras de Pedagogía en Lenguaje Educación Media. Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas.

Cronbach, L. J. (1988). Five perspectives on validity argument. En H. Wainer & H. Braun (Eds.), Test Validity (pp. 3-17). Hillsdale, NJ: Erlbaum.

Downing, S. M., & Haladyna, T. M. (2006). Handbook of test development. Mahwah, NJ: Lawrence Erlbaum Associates.

García, C. M., & Yot, C. (2018). Edumetría y evaluación educativa: Una revisión sistemática. Revista Electrónica de Investigación y Evaluación Educativa (RELIEVE), 24(2), 1-20. <https://doi.org/10.7203/relieve.24.2.12923>

Kolen, M. J., & Brennan, R. L. (2014). Test equating, scaling, and linking: Methods and practices. Nueva York: Springer.

Messick, S. (1989). Validity. En R. L. Linn (Ed.), Educational Measurement (3ª ed., pp. 13-103). Nueva York: American Council on Education/Macmillan.

Murillo, J., & Román, M. (2011). De la psicometría a la edumetría: Un cambio de paradigma en la evaluación educativa. Revista Iberoamericana de Evaluación Educativa, 4(2), 1-15. <https://doi.org/10.15366/riee2011.4.2.001>

Prieto G. y Delgado, A. (2010). Fiabilidad y validez. Papeles del psicólogo. Vol 31(1), pp. 67 – 74.

Rasch, G. (1960). Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests. Nielsen & Lydiche.

Sireci, S. G. (2009). Packing and unpacking sources of validity evidence: History repeats itself again. En R. W. Lissitz (Ed.), *The Concept of Validity: Revisions, New Directions, and Applications* (pp. 19-37). Charlotte, NC: Information Age Publishing.

Taut, S., & Sun, Y. (2014). The development and implementation of a national, standards-based, multi-method teacher performance assessment system in Chile. *Education Policy Analysis Archives*, 22(71), 1-30. <https://doi.org/10.14507/epaa.v22.1574>

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427–450. <https://doi.org/10.1007/BF02294627>

**Manual de uso de resultados  
de la Evaluación Nacional Diagnóstica  
de la Formación Inicial Docente**