



# Evaluación diagnóstica inicial en la formación inicial docente en Chile y su relación con contextos institucionales

## Initial Diagnostic Assessment in Pre-Service Teacher Training in Chile and its Relationship With Institutional Contexts

Valentina Giaconi<sup>1,2</sup>, Gabriela Gómez<sup>1</sup>, Daniela Jiménez<sup>2</sup>, Benjamín Gareca<sup>1</sup>, Francisco Durán del Fierro<sup>3</sup>, y María Leonor Varas<sup>2</sup>

<sup>1</sup> Universidad de O'Higgins

<sup>2</sup> Universidad de Chile

<sup>3</sup> University College London

### Resumen

En este artículo se estudia la implementación, en cinco universidades chilenas, de una ley que mandata la aplicación de una evaluación diagnóstica inicial en las carreras de pedagogía. La ley indica que los resultados de esta evaluación serán referenciales y tendrán un uso formativo para los estudiantes; asimismo, da autonomía a las universidades para definir qué y cómo evaluar. Considerando la importancia de esta evaluación y su alcance, el objetivo de este artículo es entender cómo se implementan estas evaluaciones y cómo su aplicación se relaciona con el contexto de cada institución. La metodología utilizada corresponde al estudio de casos y los datos analizados corresponden a entrevistas a representantes institucionales, al análisis de las evaluaciones efectuadas y a sus resultados psicométricos. Se concluye que los aspectos contextuales más importantes para definir las evaluaciones diagnósticas fueron las capacidades técnicas y los recursos a los que pudieron acceder las escuelas de pedagogía y se observó una desarticulación con los perfiles de ingreso o programas formativos.

**Palabras clave:** evaluación diagnóstica, formación inicial docente, estudio de caso

---

#### Correspondencia a:

Valentina Giaconi  
Avenida Libertador Bernardo O'Higgins 611, Rancagua, Chile  
valentina.giaconi@uoh.cl  
ORCID: 0000-0002-5166-5673

---

© 2022 PEL, <http://www.pensamientoeducativo.org> - <http://www.pel.cl>

---

ISSN:0719-0409 DDI:203.262, Santiago, Chile doi: 10.7764/PEL.59.1.2022.4

---

## Abstract

---

This paper examines the way in which five Chilean universities have implemented the mandatory initial diagnostic assessment in pre-service teacher training. The law states that the outcomes of this assessment will be used to enhance students' learning processes. In addition, universities are entitled to autonomy to define what to evaluate and how to carry it out. Considering the purpose and scope of this evaluation, the aim of this paper is to better understand how these assessments have been implemented by universities and how institutional settings have affected the way in which universities have put this assessment into practice. In order to explore this phenomenon, the research employed a multiple case study approach. In this context, semi-structured interviews were carried out with institutional representatives, along with analysis of a set of diagnostic assessments and their psychometric test results. Drawing upon the triangulation of the information generated, the paper concludes that the most important elements affecting the implementation of this initial diagnostic assessment are, on the one hand, professionals' technical skills and, on the other, the resources that teacher training programs or faculties have. There was a lack of relationship between the assessment and the admission profiles and training programs.

*Keywords:* diagnostic assessment, teacher training programs, case study

## Ley chilena vigente y evaluación diagnóstica en programas de pedagogía

---

En este artículo se estudia la implementación de una política en Chile que obliga legalmente a todas las instituciones que forman profesores a realizar evaluaciones diagnósticas al ingreso a sus programas. La ley establece que los resultados de dichas evaluaciones serán referenciales y tendrán un uso formativo para los estudiantes (Ley N° 20.903). La evaluación formativa en la formación inicial docente es clave por múltiples razones. Primero, es una herramienta esencial para promover el aprendizaje entre los estudiantes de pedagogía (Brookhart, 2017). Segundo, modelar prácticas de evaluación de esta naturaleza en la formación inicial docente es fundamental para tener futuros profesores que usen esta práctica apropiadamente (Brookhart, 2017). Tercero, su uso permite identificar los aspectos clave que es necesario monitorear durante la formación para asegurar el logro de aprendizajes, al igual que en la evaluación sumativa (Klassen et al., 2017). Sin embargo, no se tiene conocimiento respecto de la implementación de este tipo de evaluación diagnóstica inicial con foco formativo.

La política estudiada en este artículo comenzó su implementación en 2017. Si la universidad no aplica la evaluación, incumple una de las normas establecidas en la acreditación obligatoria de programas de pedagogía, lo cual podría resultar en su cierre. Esta política es original, en el sentido en el que usualmente la evaluación al inicio de la formación inicial es usada con un propósito de filtro y selección y no con un propósito formativo (Klassen et al., 2017; Sato & Kemper, 2017; Klassen & Kim, 2018). En Chile, los estudiantes que ingresan a la formación inicial docente ya pasaron por un sistema de admisión nacional que es usado por la mayoría de los programas de pedagogía, e incluye la aplicación de exámenes basados en el currículo escolar (Santelices et al., 2019).

Por otro lado, la evaluación en general tiene que cumplir con estándares de medición y contar con evidencias de validez y de confiabilidad (American Educational Research Association et al., 2018). En el caso de la evaluación diagnóstica, la Ley N° 20.903 no considera si existen los equipos, capacidades y recursos necesarios en las instituciones formadoras de profesores para implementar adecuadamente estas evaluaciones. Luego, es posible que las instituciones enfrenten desafíos para implementar adecuadamente estas evaluaciones y que no reflejen el espíritu de la ley. Esto se debiera de evidenciar en los textos y discursos asociados con los sistemas evaluativos de cada institución (Daugherty & Ecclestone, 2006).

Para cumplir con estándares de calidad en la medición y tener evaluaciones pertinentes, el rol de las instituciones es clave, en donde distintos estamentos (estudiantes, académicos y docentes, administrativos y mercado laboral) deben participar y articularse (Banta & Palomba, 2014). Asimismo, las evaluaciones son afectadas por influencias externas (por ejemplo, políticas nacionales e internacionales) junto con el contexto institucional interno (Kohoutek, 2014; Raaper, 2017; Daugherty & Ecclestone, 2006). Caracterizar el efecto de estos contextos es necesario para comprender mejor la implementación de estas evaluaciones que se enmarcan en un sistema complejo (Flórez Petour, 2015; Daugherty & Ecclestone, 2006). Para estas evaluaciones específicamente, la ley que las mandata entrega mucha autonomía a las instituciones para el diseño e implementación de las evaluaciones diagnósticas, lo cual da una oportunidad de analizar cómo se ha establecido la relación entre las evaluaciones y las instituciones. Considerando la importancia de la evaluación formativa en la formación inicial docente (FID) y la originalidad y alcance de la política educativa que obliga a las instituciones chilenas a realizar evaluaciones diagnósticas de pedagogía, se ha realizado un estudio, a partir del cual surge este artículo, en el que se consideran las siguientes preguntas de investigación:

1. ¿Cómo se han desarrollado las evaluaciones diagnósticas de pedagogía?
2. ¿Cómo afecta el contexto institucional al desarrollo de las evaluaciones diagnósticas?

## **Evaluación y formación inicial docente en Chile**

A partir de la década de 1990, se han realizado múltiples esfuerzos para mejorar la calidad de la formación inicial docente en Chile (Ávalos, 2014). Por ejemplo, durante los años 1997 a 2002 se implementó el Programa de Fortalecimiento de la Formación Inicial Docente (FID), que otorgaba financiamiento competitivo a instituciones que quisieran mejorar sus programas. Sin embargo, durante la década del 2000, la FID fue parte de una expansión fuertemente desregulada, creándose múltiples programas, algunos de los cuales ofrecían procesos formativos insatisfactorios (Ávalos, 2014). En los últimos 15 años, a partir de demandas sociales, se instala la necesidad de mejorar radicalmente la educación en Chile, y, en consecuencia, la calidad de la enseñanza (Ávalos, 2014). Esto vuelve a relevar la necesidad de contar con mejores profesores, lo que implica la implementación de políticas de fortalecimiento de la FID basadas en su efectividad según estudios comparados, las cuales pertenecen principalmente a tres ámbitos (Ingvarson & Rowley, 2017). El primer ámbito corresponde a elevar la selección de estudiantes con altos rendimientos académicos, realizado mediante el incremento de requisitos mínimos para el ingreso a las carreras de pedagogía. El segundo ámbito se relaciona con el monitoreo y el establecimiento de estándares para el funcionamiento de programas de FID, materializado mediante el requerimiento de acreditación obligatoria y reforma de nuevas exigencias y estándares para su logro. Finalmente, el tercer ámbito considera la evaluación y certificación de estudiantes cercanos a su egreso, según lo cual se implementó la Evaluación Nacional Diagnóstica de la Formación Inicial Docente (ENDFID), que se aplica cerca del egreso (Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas, 2018), pero no habilitante para el ejercicio de la profesión. La evaluación inicial diagnóstica de la que nos ocupamos responde al primer y segundo punto,

ya que (1) permite hacerse cargo de estudiantes que podrían no estar suficientemente preparados para el ingreso y (2) permite mejorar la calidad de los programas de FID, ya que las evaluaciones diagnósticas entregan información que puede permitir ajustar los planes de estudios y las estrategias de enseñanza.

Por último, la evaluación diagnóstica inicial se vincula indirectamente con otras dos que obligatoriamente se aplicarán a la misma población de estudiantes. Por una parte, en el proceso de admisión a la universidad, inmediatamente anterior al ingreso, deben haber rendido las pruebas de selección (Santelices et al., 2019), cuyos resultados pueden mostrar vacíos en áreas de conocimiento y habilidades, por lo que una evaluación diagnóstica aplicada pocos meses después solo será útil si agrega información a la ya recabada. Por otra parte, los objetivos formativos de las carreras de pedagogía se enmarcan en estándares establecidos a nivel nacional, con base en los cuales se realiza la Evaluación Nacional Diagnóstica ENDFID, que se aplica cerca del egreso (Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas, 2018).

Las pruebas de selección universitaria y la ENDFID son estandarizadas; en contraste, la evaluación inicial diagnóstica es completamente abierta y las universidades tienen autonomía para implementarla. Sin embargo, esto genera riesgos, porque no es seguro que las universidades tengan las capacidades y los equipos para la implementación adecuada de la evaluación. Yorke (1998) establece que para tener una gestión efectiva de la evaluación en educación superior es necesario que se cumplan tres condiciones: (1) una definición clara del propósito de la evaluación, (2) una estrategia para lograr este propósito, y (3) una implementación que funcione. Esto es también considerado por los Estándares para Pruebas Educativas y Psicológicas (American Educational Research Association et al., 2018), que definen validez como “grado en que la evidencia y teoría respaldan las interpretaciones de los puntajes de una prueba para los usos propuestos” (p. 11), lo que implica que si una prueba no tiene un propósito definido es imposible realizar un juicio de validez. En este sentido, es clave identificar propósitos claros y acotados. Es más, la evaluación diagnóstica inicial se define con un objetivo formativo; luego, para que estas evaluaciones sean válidas, tienen que promover aprendizajes y estar asociadas con alguna acción (Heritage et al., 2009). Por lo tanto, el argumento de validez asociado tiene que considerar las consecuencias de estas evaluaciones (Stobart, 2006). Asimismo, la interpretación de la confiabilidad de estas evaluaciones tiene que considerar su foco formativo. Black y Wiliam (2006) indican que, en evaluaciones formativas, la confiabilidad es un aspecto menos relevante en el caso que sean conformadas por interacciones frecuentes entre estudiantes y docentes; sin embargo, en el caso de un test formativo que va a permitir tomar decisiones sobre cuánto tiempo tiene que pasar un estudiante en un curso remedial, la confiabilidad sí es un aspecto crítico. Existen evidencias de evaluaciones formativas exitosas en contextos universitarios, por ejemplo, experiencias con evaluaciones de escritura al inicio de las carreras (Fisher et al., 2011), evaluaciones formativas aplicadas en carreras de pedagogía (Cosi et al., 2020), e intervenciones asociadas con la evaluación formativa, con diversas características en relación con el involucramiento de profesores y estudiantes y al uso o no de tecnología (Stull et al., 2011).

En el caso de la evaluación diagnóstica inicial para pedagogía, la ley indica usos formativos y, más específicamente, remediales relacionados con la nivelación de conocimientos previos. Aparte de estos usos, la Ley N° 20.903 no señala otros propósitos ni precisa los conocimientos y habilidades necesarios de evaluar, por lo que cada universidad debe definir sus evaluaciones iniciales. Sin embargo, por lo expuesto anteriormente, no es claro que las universidades cuenten con los recursos necesarios ni que logren implementar evaluaciones con suficientes evidencias de validez y confiabilidad. Por lo tanto, es necesario valorar si los usos formativos de estas evaluaciones están definidos y si cuentan con evidencias de validez y confiabilidad.

Finalmente, es importante recalcar que esta política de evaluación diagnóstica inicial es original en el contexto internacional, en donde las evaluaciones para pedagogía suelen tener propósitos de selección y certificación, y no de diagnóstico (Klassen et al., 2017; Sato & Kemper, 2017; Klassen & Kim, 2018).

En función de lo anterior, es clave entender cómo se han desarrollado estas evaluaciones diagnósticas iniciales para las carreras de pedagogía.

## **Evaluación y su relación con contextos externos e institucionales**

Es necesario considerar el contexto externo e institucional para entender el funcionamiento de la evaluación diagnóstica y sus diferencias entre distintas instituciones (Kohoutek, 2014; Biggs 1996). Kohoutek (2014) propone un marco conceptual con cuatro elementos para comprender la evaluación de estudiantes en educación superior, el cual incluye las políticas y regulaciones institucionales, la gestión y organización académica, las estrategias de verificación (por ejemplo, la revisión doble de evaluaciones) y las influencias externas en la evaluación, que contemplan regulaciones nacionales y leyes internacionales. Raaper (2017) estudia dos universidades europeas y muestra cómo aspectos propios de las instituciones definen sus políticas evaluativas, como por ejemplo la edad, el tradicionalismo y las políticas internas, así como contextos nacionales y globales relacionados con el neoliberalismo. Por otro lado, Flórez Petour (2015) muestra cómo estructuras más amplias que las mismas instituciones (como partidos políticos, el sector económico, las familias y los medios) determinan también las reformas asociadas a la evaluación formativa en Chile. Estos estudios muestran la relevancia de estudiar la implementación de las evaluaciones, considerando el contexto político e institucional.

En el caso de la evaluación diagnóstica inicial a estudiantes de pedagogía, las instituciones tienen que ajustarse a la ley que mandata la implementación de estas evaluaciones, considerando sus políticas internas y su organización (Ley N° 20.903). Biggs (1996) indica que una institución y sus prácticas educacionales tienen que estar en equilibrio para lograr las metas educacionales. De hecho, para que una evaluación sea usada de manera efectiva para mejorar el aprendizaje del estudiantado, es fundamental contar con la participación de académicos, administrativos, estudiantes y empleadores (Banta & Palomba, 2014; Richmond et al., 2019). En este grupo, Banta y Palomba (2014) definen el rol de cada uno de ellos; estos roles se describen a continuación: consideran que los académicos y estudiantes tienen un rol fundamental y pivote, y debieran de estar fuertemente involucrados en las evaluaciones. Los académicos tienen que ser responsables de llevar a cabo todas las etapas del proceso evaluativo, desde definir los propósitos hasta tomar acciones respecto de sus resultados. El rol principal de los estudiantes es participar en las evaluaciones y en las acciones sobre sus resultados, pero también es importante que participen en comités y den retroalimentación sobre las evaluaciones y la enseñanza. Para promover la participación, es necesario el rol de los administrativos para apoyar a académicos y estudiantes, promover el uso de los resultados en el desarrollo del currículo, traer conocimiento (por ejemplo, expertos, promover la asistencia a conferencias), y proveer financiamiento. Respecto de los empleadores, estos pueden entregar importantes perspectivas respecto de qué evaluar.

En síntesis, el rol de los contextos generales (globales, a nivel país) e institucionales es clave para comprender la implementación de las evaluaciones diagnósticas. Específicamente, es necesario obtener información relacionada con los actores relevantes que participan de las evaluaciones, como administrativos y docentes, y los estudiantes, cuya responsabilidad principal es responder a las evaluaciones.

## **Métodos**

---

La metodología utilizada es cualitativa y consiste en un estudio de casos múltiple. Cada caso corresponde a un proceso de evaluación diagnóstica para pedagogía, implementado en una universidad que provee programas de formación inicial docente.

Este enfoque metodológico es apropiado porque (1) el foco de la investigación es comprender cómo se implementan estas evaluaciones, que son un proceso complejo que depende del contexto institucional, y (2) se tienen múltiples fuentes de información para cada caso (entrevistas, documentos, evaluaciones y resultados psicométricos). El estudio de casos múltiple permite comprender cada proceso evaluativo en su contexto institucional (Yin, 2011), pero también obtener una mayor generalización de los hallazgos, en tanto pretende observar procesos y resultados comunes entre los casos, entendiendo aquello que está condicionado por las realidades locales (Miles et al., 2013).

## Muestra

La muestra consiste en cinco universidades que cubren la mayoría de la matrícula en pedagogía de Chile. Las universidades son diversas, ya que son privadas y públicas (tres son privadas y dos, públicas), sus ubicaciones geográficas son diferentes (dos de la Región Metropolitana y tres de otras regiones), respecto de los años de acreditación institucional (desde no tener acreditación hasta el máximo de años de acreditación), según la variedad de programas y pertenencia al Consejo de Rectores de las Universidades Chilenas (CRUCH) (cuatro pertenecen y una, no). La muestra fue seleccionada de manera tal de contar con instituciones diversas de acuerdo con las características anteriores y, además, se consideraron universidades con matrículas altas y variedad de programas de pedagogía para asegurar el análisis psicométrico de los instrumentos.

## Recolección de datos

Los datos fueron recolectados en 2018 y corresponden a información sobre el proceso de evaluación diagnóstica implementado en 2017.

Las instituciones seleccionadas fueron contactadas e invitadas a participar en el estudio. El (la) rector (a) de cada institución autorizó la participación de la institución en el estudio. Luego de lograr el acuerdo, se solicitó lo siguiente a cada institución:

- Entrevistar a tres perfiles institucionales con el objetivo de comprender el desarrollo e implementación de las evaluaciones diagnósticas. Los perfiles corresponden a (1) un cargo directivo, encargado de la toma de decisiones respecto de las evaluaciones, (2) un cargo técnico-operativo encargado de la implementación de las evaluaciones, y (3) un docente que potencialmente podría usar los resultados de las evaluaciones. Las entrevistas fueron semiestructuradas e incluían preguntas sobre el proceso evaluativo (marco de referencia de la evaluación diagnóstica, características de los instrumentos y aspectos operativos) y la institución (estudiantes e intereses y necesidades).
- Los instrumentos utilizados para la evaluación diagnóstica.
- Toda la documentación disponible asociada con la evaluación (marcos de evaluación, protocolos de aplicación, rúbricas, etc.).
- Las bases de datos con los resultados de los instrumentos aplicados.

Los procesos de recolección de información contaron con estrictos protocolos de seguridad y protección de información. Todas las personas entrevistadas consintieron su participación en la investigación por medio de un formulario de consentimiento informado.

## Análisis

Para desarrollar el estudio de casos, se consideraron todas las fuentes de información (entrevistas, instrumentos, etc.), lo que permite describir el desarrollo de cada sistema evaluativo según su contexto.

Para cada caso se analizaron las entrevistas siguiendo los fundamentos de la técnica denominada análisis temático. Se tomó esta decisión puesto que este enfoque analítico permite la producción e interpretación de temas centrales y el análisis de tópicos en vez de experiencias individuales (Joffe, 2011). En ese sentido, se construyó un sistema de códigos cuyas categorías de análisis fueron definidas de manera inductiva a partir de la lectura de un subconjunto de las entrevistas.

Para validar los códigos, el primer caso fue codificado de manera independiente por dos personas del equipo de investigación. Las codificaciones fueron comparadas y se precisó la definición de los códigos a partir de las diferencias encontradas. Con los códigos redefinidos, el primer caso fue analizado nuevamente, obteniéndose un acuerdo total entre investigadores, por lo que se consideró que el sistema de códigos estaba validado. Luego, se codificaron todas las entrevistas y, posteriormente, se elaboraron informes para cada caso. De esta forma, los textos producidos permitieron la reconstrucción de los contextos y temáticas asociadas a las evaluaciones diagnósticas, las cuales contribuyeron a responder a las preguntas de investigación.

El análisis de los casos incluyó, además, el examen de un conjunto de materiales desarrollados por las instituciones orientados a apoyar los procesos de diseño y aplicación de las pruebas. También se hizo un análisis de contenido a las evaluaciones de lenguaje, matemáticas y aspectos no disciplinares por medio de una revisión por parte de un panel de expertos (Sireci & Faulkner-Bond, 2014). Se analizaron estas áreas por un criterio de factibilidad y porque lenguaje, matemáticas y aspectos no disciplinares fueron áreas evaluadas en todas las instituciones. Para cada área se formó un panel de expertos constituido por profesores de matemática, de lenguaje y psicólogas educacionales, respectivamente. Este panel evaluó la pertinencia conceptual de los instrumentos y sus ítems en relación con la disciplina evaluada y al uso que se le iba a dar al instrumento. Asimismo, el panel evaluó aspectos de medición (claridad de los ítems, número de ítems por dimensión, etc.). Para esto, se utilizó un protocolo de revisión de ítems e instrumentos y se realizaron análisis psicométricos de los instrumentos de matemáticas y lenguaje. Estos análisis se basaron en la teoría clásica de tests (Martínez-Arias, 1996), y se utilizó el coeficiente alfa de Cronbach para estimar confiabilidad. El detalle de estos análisis se puede encontrar en Giaconi et al. (2019).

## Resultados

---

En la tabla 1 se sintetizan las características principales de cada caso de acuerdo con las dimensiones de análisis. En esta sección se utilizan los resultados de la tabla 1 para describir los casos en torno a las dimensiones de análisis que se relacionan claramente con los procesos evaluativos, y luego se menciona un análisis general de las dimensiones que no evidenciaron una relación explícita con las evaluaciones. El primer caso presentado (Caso 2) tiene el sistema evaluativo más desarrollado y el último caso (Caso 5), el menos desarrollado.

### Caso 2

Este caso es el que mostró mayor desarrollo en su sistema de evaluación. El desarrollo de evaluaciones diagnósticas corresponde a una decisión institucional que había sido tomada mucho antes de la implementación de la Ley 20.903. Esta universidad corresponde a una universidad muy compleja, con un alto número de años de acreditación. La mayoría de sus instrumentos se aplicaba a toda la universidad, y los entrevistados destacaron que la articulación entre distintas unidades institucionales con diversas capacidades técnicas (por ejemplo, en evaluación y en desarrollos tecnológicos) fue fundamental para el logro de un buen sistema de evaluación. Asimismo, tenían explícitamente definidos los propósitos y usos de sus evaluaciones. Es posible que estos factores expliquen por qué fue la única institución que tuvo buenos indicadores psicométricos en sus evaluaciones y en donde además el panel de expertos consideró adecuado el contenido de las evaluaciones.

### Caso 3

Este caso mostró un trabajo muy arduo para implementar evaluaciones específicas para sus carreras de pedagogía. Fue el único caso que definió un perfil de ingreso para estas carreras que permitió apoyar el proceso de construcción de instrumentos. Esta institución logró además implementar una aplicación en línea. Este caso corresponde también a una universidad compleja con varios años de acreditación. Los avances logrados son valiosos, ya que van en línea con tener un sistema muy ajustado a las necesidades de las carreras de pedagogía. Sin embargo, el caso en su globalidad presenta un nivel de desarrollo medio y requiere mejoras, ya que las evaluaciones no pudieron ser piloteadas previamente; estas presentaron algunos problemas de construcción, los indicadores psicométricos fueron poco adecuados y el uso de las evaluaciones es relativamente inespecífico. También reportaron que la carga de trabajo tensionó a un equipo reducido asociado a las carreras de pedagogía.

### Caso 1

En este caso construyeron evaluaciones y adaptaron evaluaciones de otros proyectos. Contaban también con soporte tecnológico para la aplicación y uso de las evaluaciones. Sin embargo, los propósitos y usos son descritos con poca especificidad y no se reporta un uso en acciones específicas. Se observaron problemas de construcción y en las propiedades psicométricas de los instrumentos; por esta razón, consideramos que el nivel de desarrollo también es medio. No se observa tampoco una articulación con el modelo formativo ni se constató la existencia de un proceso de revisión de estas evaluaciones.

### Caso 4

Este caso, al igual que el caso 3, también destaca por haber realizado un esfuerzo para tener un sistema específico y ajustado a sus carreras de pedagogía. Asimismo, contaba con procesos sistematizados y protocolos de seguimiento y registro de procesos. Sin embargo, había importantes fallas en el diseño de las evaluaciones y en sus indicadores psicométricos, por lo que consideramos que presenta un desarrollo medio. Se observaron también usos bastante generales, que no presentaban una articulación explícita con las evaluaciones.

### Caso 5

Este caso representa el nivel de desarrollo menos avanzado, ya que la mayoría del proceso estaba externalizado y lo realizaba otra institución. Además, ellos aplicaron las mismas pruebas en toda la universidad. Esto se justificaba porque la institución aún no desarrollaba las capacidades para diseñar e implementar un sistema de evaluación diagnóstica. Se observó también que los usos estaban poco especificados, y de hecho no había una articulación entre las nivelaciones que desarrollaba la universidad y los resultados de la evaluación diagnóstica.

Se observa en los resultados que los sistemas de evaluación necesitan más desarrollo en la mayoría de los casos, especialmente en relación con (1) la definición de propósitos y usos claros que se articulen con el diseño de las evaluaciones, y (2) con el mejoramiento de las características psicométricas y de diseño. Esto permite inferir que faltan capacidades y recursos en las universidades para desarrollar mejores evaluaciones y sistemas de aplicación y uso. Asimismo, necesitan tiempo para evaluar y mejorar los sistemas evaluativos.

Finalmente, es importante destacar que no se observó una relación o articulación entre las dimensiones 1 (política institucional y modelo formativo) y 2 (perfil del estudiantado) con los sistemas de evaluación diagnóstica. Estos temas fueron mencionados por los entrevistados, pero luego no se utilizaron al describir los procesos de diseño ni el uso de las evaluaciones.

Tabla 1  
*Síntesis de resultados de los casos*

Dimensión	Caso 1	Caso 2	Caso 3	Caso 4	Caso 5
<b>1. Política institucional y modelo formativo</b> <b>2. Perfil del estudiantado</b> <b>3. Experiencia en evaluaciones diagnósticas a estudiantes de primer año previa a la Ley 20.903.</b>	<p>1. Perspectiva sociocrítica asociada con una pedagogía social cuyo objetivo es la transformación social.</p> <p>2. Perfil homogéneo, puesto que la mayoría de los estudiantes proviene de establecimientos subvencionados y municipales. Sin embargo, la distribución de puntajes en la prueba de selección universitaria no es homogénea.</p> <p>3. Mencionan experiencias previas en el diseño de evaluaciones diagnósticas.</p>	<p>1. Formación teórico-disciplinaria basada en conocimientos pedagógicos, didácticas de la especialidad y conocimiento disciplinar. Prácticas pedagógicas progresivas.</p> <p>2. Perfil marcado por proveniencia de escuelas municipales y particulares subvencionadas y familias dentro de los primeros tres quintiles de ingreso. Proviene predominantemente de la misma región.</p> <p>3. Aplican evaluaciones diagnósticas desde 2014. Institución pionera en la elaboración de evaluaciones y en la utilización de sus resultados. Mecanismos establecidos de colaboración entre distintos organismos institucionales.</p>	<p>1. Modelo basado en competencias.</p> <p>2. Perfil con alto rendimiento en la prueba de selección universitaria (PSU), proveniente de la misma región y de escuelas particulares subvencionadas.</p> <p>3. Experiencia en aplicación de evaluaciones diagnósticas en todas las carreras para comprender la deserción universitaria. Por esto, no fueron consideradas para la evaluación diagnóstica en pedagogía.</p>	<p>1. Práctica como el eje central del aprendizaje, como el aspecto articulador del proceso de enseñanza y de la producción de conocimiento.</p> <p>2. El perfil del estudiantado cambió a raíz de la incorporación de la universidad al Sistema Único de Admisión, aumentando los puntajes promedio de los estudiantes y la cantidad de estudiantes que se matriculan.</p> <p>3. No mencionan experiencias previas.</p>	<p>1. Modelo formativo común para las carreras de pedagogía, definido por: énfasis en formación disciplinar, vínculo con la práctica, dominio de las tecnologías de la información y de las comunicaciones (TIC), atención a la diversidad y reconocimiento del trabajo colaborativo.</p> <p>2. Perfil marcado por estudiantes de la misma región, primera generación en la universidad y provenientes de escuelas municipales y particulares subvencionadas.</p> <p>3. No mencionan experiencias previas.</p>

Dimensión	Caso 1	Caso 2	Caso 3	Caso 4	Caso 5
<b>4. Áreas evaluadas por los instrumentos</b>	La mayoría de los instrumentos eran aplicados en toda la universidad.	La mayoría de los instrumentos eran aplicados para toda la universidad.	La mayoría de los instrumentos eran específicos para pedagogía.	La mayoría de los instrumentos eran específicos para pedagogía.	Los instrumentos eran aplicados en toda la universidad.
<b>4.1 Áreas disciplinares</b>	4.1 Comprensión lectora y Lenguaje y comunicación, Matemática, Ciencias, Inglés, y Tecnologías de la Información y de las Comunicaciones.	4.1 Comprensión lectora y Lenguaje y comunicación, Escritura, Matemática, Biología, Física, y Química.	4.1 Comprensión lectora y Lenguaje y comunicación, Matemática, Biología, Física, Química, Historia, Filosofía, Música, y Educación Física.	4.1 Comprensión lectora y Lenguaje y comunicación, Escritura, Historia, Inglés, y Educación Física	4.1 Matemática y Escritura.
<b>4.2 Áreas no disciplinares</b>	4.2 Estilos de aprendizaje.	4.2 Estrategias de aprendizaje.	4.2 Capacidades pedagógicas.	4.2 Estrategias metacognitivas y disposición o vocación de profesor.	4.2 No reportaron instrumentos de áreas no disciplinares.
<b>5. Propósitos y usos</b>	Obtener información sobre las competencias y las habilidades de los estudiantes que reciben. Conocer las áreas en que los estudiantes requieren apoyo y, por ende, planificar y ajustar las clases según las brechas identificadas.	Caracterizar el grado de dominio de los contenidos escolares por parte de los estudiantes al egreso del sistema escolar, con el fin de generar acciones institucionales de apoyo. Usos específicos: diseño e implementación de acciones de nivelación; construcción del perfil de los estudiantes de primer año diferenciados por curso para los docentes, y conformación del modelo de evaluación de cada carrera.	Medir las habilidades básicas que un profesor requiere desarrollar. Estas habilidades serían diagnosticadas para establecer mecanismos de apoyo que permitan su desarrollo. Otros usos anticipados para las evaluaciones diagnósticas refirieron a informar procesos formativos y sensibilizar a los docentes respecto de la necesidad de considerar resultados en la práctica docente. No se realizan acompañamientos o tutorías focalizadas en el apoyo académico vinculadas con las evaluaciones diagnósticas. Se dispone de talleres de inducción a la vida universitaria para todos los estudiantes de primer año.	Generar planes de mejora en las carreras de pedagogía. Usos: potenciar los planes de mejora, conocer el perfil de ingreso de los estudiantes, apoyar los procesos de innovación curricular y aseguramiento de la calidad, y monitorear a los estudiantes a lo largo del proceso formativo.	Contribuir a la construcción de un perfil académico de los nuevos estudiantes que permitiera diseñar políticas institucionales. La institución externa encargada del diseño de las evaluaciones diagnósticas utilizó instrumentos que ya había elaborado con anterioridad. Los entrevistados no hacen referencia a los propósitos originales de aquellos instrumentos. Los diagnósticos no son considerados en el diseño de nivelaciones debido a la extensión del periodo de revisión de resultados exigido por la institución externa.

Dimensión	Caso 1	Caso 2	Caso 3	Caso 4	Caso 5
<b>6. Proceso de construcción de los instrumentos y 7. Aplicación de las evaluaciones</b>	<p>6. La institución construye evaluaciones y adapta evaluaciones de proyectos previos.</p> <p>7. Aplicación en línea. Cuentan con un soporte tecnológico que apoya el proceso evaluativo y permite realizar un seguimiento a los estudiantes.</p>	<p>6. El diseño y la elaboración de evaluaciones diagnósticas se basa en instrumentos utilizados en años anteriores. Existe un proceso de rediseño de cada instrumento cada dos años, de forma intercalada, según lineamiento institucional. Se llevan a cabo aplicaciones piloto como proceso de validación.</p> <p>7. Aplicación en línea. Articulan distintas unidades (aplicación de evaluaciones, soporte, etc.). Procesos institucionalizados y desarrollo de materiales de apoyo.</p>	<p>6. El proceso de elaboración, aplicación, análisis y reporte de las evaluaciones diagnósticas fue llevado a cabo internamente en las escuelas de pedagogía por un equipo liderado por un profesor experto en evaluación. Se elaboró un perfil de ingreso común para todas las pedagogías, el que estableció capacidades pedagógicas y disciplinares como base conceptual para la construcción de las evaluaciones.</p> <p>7. Aplicación en línea.</p>	<p>6. El diseño, aplicación y evaluación de las evaluaciones diagnósticas es responsabilidad de una unidad que opera de manera más o menos centralizada. Pruebas generales se adaptaron para las pedagogías. Los referentes de las pruebas son variados y se conducen procesos complementarios para definir los contenidos: revisión bibliográfica; ajustes de ítems del banco de preguntas, y uso de ítems liberados de otras evaluaciones.</p> <p>7. Aplicación en lápiz y papel, donde los profesores/docentes de la Facultad juegan un rol fundamental. La universidad cuenta con un coordinador de aplicación por sede, protocolos de aplicación y procesos sistematizados.</p>	<p>6. Los instrumentos son contruidos por una institución externa. La dirección institucional encargada de asuntos académicos organiza la implementación de evaluaciones diagnósticas a todos los estudiantes de primer año.</p> <p>7. Aplicación en lápiz y papel que es responsabilidad de la institución externa a cargo de las evaluaciones. Se valora la decisión de externalizar este proceso, ya que la universidad no tiene aún las capacidades necesarias.</p>

Dimensión	Caso 1	Caso 2	Caso 3	Caso 4	Caso 5
<b>8. Análisis de contenido y psicométrico</b>	<p>Los ítems eran adecuados para los instrumentos, sin embargo, las pruebas disciplinares incluían muy pocos ítems, lo cual impidió que se obtuviera una confiabilidad aceptable (Alfa de Cronbach entre 0,2 a 0,49).</p>	<p>Los instrumentos estaban correctamente contruidos y se observa una referencia a la PSU en la estructura y contenido de los ítems. Las evaluaciones de matemáticas y de lenguaje presentaron buenos índices de confiabilidad (mayores a 0,7).</p>	<p>Las pruebas de lenguaje eran adecuadas; sin embargo, existen problemas de construcción en la prueba de matemáticas y la prueba de aspectos no disciplinares. Por ejemplo, esta última contenía muy pocos ítems por dimensión. Los resultados psicométricos indicaron que las pruebas de lenguaje y matemáticas tenían bajas confiabilidades (rango entre 0,10 y 0,45).</p>	<p>Se identificaron problemas de construcción; por ejemplo, combinar disciplinas en una misma evaluación o utilizar muy pocos ítems en cada constructo evaluado. Los resultados psicométricos mostraron bajas confiabilidades en el conjunto de ítems de lenguaje y matemáticas (rango entre 0,09 y 0,33).</p>	<p>En general, los instrumentos eran adecuados en su construcción. Respecto del análisis psicométrico, la evaluación analizada presentó una confiabilidad baja de 0,52.</p>

*Nota:* la información presentada en esta tabla proviene del análisis de entrevistas y de la documentación asociada (dimensiones 1 a 7) y del análisis de contenido y psicométrico de los instrumentos (dimensión 8).

*Fuente: elaboración propia.*

## Síntesis y discusión final

A partir del análisis de estos resultados, es posible dar cuenta de las temáticas asociadas con las preguntas de investigación inicialmente planteadas. Respecto de la primera pregunta, referida al desarrollo de las evaluaciones diagnósticas de pedagogía, se observan niveles variados de desarrollo de los sistemas de evaluación diagnóstica de cada institución. El Caso 2 tenía un sistema establecido y amplio institucionalmente, al cual fueron incorporadas las pedagogías. Su objetivo era medir conocimientos escolares. El sistema era robusto y articulado entre las unidades centrales de la universidad. Al ser institucional, las evaluaciones no toman especificidades para las pedagogías. Es la única institución cuyas pruebas tuvieron índices de confiabilidad sobre 0,7. Por otro lado, el Caso 3 implementó un sistema independiente y específico para las pedagogías con base en un marco de evaluación teórico propio. El proceso fue desarrollado basado en un equipo que tenía conocimiento técnico para realizar la tarea, pero que se vio muy sobrecargado. En el Caso 1, las evaluaciones también se aplicaban a la mayoría de la universidad, lo que permitiría eventualmente usar las capacidades institucionales; sin embargo, se observó un desarrollo menos avanzado que en el Caso 2. En el Caso 4, se optó por adaptar un sistema existente previamente que se aplicaba a nivel de toda la institución. Este sistema contaba con procesos sistematizados y protocolos de seguimiento y registro de proceso. No obstante, había importantes fallas en el diseño de las evaluaciones y en sus indicadores psicométricos. Finalmente, en el Caso 5 se decidió externalizar el diseño, implementación y análisis del sistema. La evaluación utilizada corresponde a un instrumento construido con propósitos más generales, vinculados con el diagnóstico de estudiantes de primer año, no específicos para pedagogía. Esta decisión se corresponde con la inexistencia de organismos internos que pudiesen hacerse cargo de la tarea, restándole capacidad de influir en el proceso.

Respecto de la segunda pregunta, los casos muestran que el nivel de desarrollo de los sistemas de evaluación y su calidad varían de un contexto institucional a otro y dependen, en primer lugar, de las capacidades instaladas previamente en cada institución; en segundo lugar, de la definición de propósitos y usos asociados a las evaluaciones, y, en tercer lugar, de la decisión de aplicar pruebas generales para toda la universidad o específicas para las pedagogías.

En relación con el primer punto, estas capacidades son altamente complejas y de lento desarrollo, ya que implican contar con procesos formales, sistema de seguimiento, equipos consolidados y estables, y conocimiento técnico de evaluación. Por ejemplo, la única universidad cuyas evaluaciones tenían confiabilidades altas tenía un sistema diagnóstico implementado durante varios años y lo aprovechó para seguir evaluando a las pedagogías (Caso 2). El resto de las universidades implementa sistemas nuevos, adaptados o externalizados y ninguna de ellas pudo tener un sistema totalmente satisfactorio técnicamente.

En lo concerniente al segundo punto, se destaca como un aspecto clave el establecimiento de propósitos y usos de la evaluación y la subsecuente articulación de las evaluaciones con estos propósitos y usos. El caso en el que esta articulación fue mejor desarrollada también corresponde al Caso 2, en donde los propósitos estaban definidos con mayor precisión y especificidad (caracterizar el grado de dominio de los contenidos escolares), lo que permite hacer pruebas acordes a estos propósitos. En los otros casos no se observó una articulación explícita y los propósitos no estaban claramente definidos.

Finalmente, y en relación con el tercer punto, entre los casos estudiados se distinguen dos formas de abordar la implementación: pruebas generales utilizadas para múltiples carreras o pruebas específicas para las pedagogías. Esta decisión estaba relacionada con el contexto institucional, ya que aplicar pruebas generales permite utilizar toda la estructura y capacidades de una universidad, mientras que aplicar pruebas específicas suele ser una responsabilidad solo de las escuelas de pedagogía, que tienden a estar constituidas por equipos más pequeños. Esto podría explicar por qué las instituciones que desarrollaron evaluaciones específicas para las pedagogías tuvieron problemas en la definición de propósitos y conceptualización de las pruebas. Un ejemplo es el Caso 4,

que intentó adaptar el sistema general para pedagogía, y sus evaluaciones presentaron problemas de diseño. Otro ejemplo es el Caso 3, en donde la escuela de pedagogía asumió la responsabilidad de llevar a cabo el proceso, asumiendo el costo de sobrecargar a un equipo pequeño.

Un último resultado a destacar respecto del desarrollo de los sistemas de evaluación es que no se observa una articulación entre estas evaluaciones y los modelos formativos; tampoco se observa una relación con los perfiles del estudiantado. Los entrevistados identifican los modelos formativos institucionales y el perfil del estudiantado, pero estos elementos no son mencionados al explicar los sistemas de evaluación diagnóstica.

La Ley N° 20.903 determina la existencia de evaluaciones diagnósticas para las carreras de pedagogía, así como su obligatoriedad. Ello deja en manos de los organismos estatales correspondientes el desarrollo de políticas y lineamientos que faciliten y orienten su implementación. Al igual que las universidades, los organismos estatales se encuentran en proceso de establecer estos lineamientos, lo cual deja abiertas varias interrogantes. Por ejemplo, la ley indica que la Comisión Nacional de Acreditación tiene que verificar el desarrollo de estas evaluaciones diagnósticas (Ley N° 20.903), pero no se aclara hasta qué punto esta evaluación diagnóstica deberían alinearse con los modelos formativos, apuntar a problemas como retención y rendimiento académico o al desarrollo curricular. Estas interrogantes son clave, ya que las evaluaciones diagnósticas definen, en parte, lo que los estudiantes de pedagogía deben saber antes de comenzar su formación (Klassen et al., 2017). Asimismo, en el marco del sistema de desarrollo profesional docente (a cargo del CPEIP) hay interrogantes abiertas respecto de su posible coordinación con evaluaciones diagnósticas posteriores en la formación como la ENDFID. Por otra parte, la organización y desarrollo de las evaluaciones diagnósticas es responsabilidad de cada institución. En consecuencia, son también las instituciones las que pueden definir los propósitos de sus propios sistemas de forma independiente. Es por ello que se genera una tensión entre esta autonomía y las posibles orientaciones y propósitos establecidos a nivel del sistema profesional docente.

La literatura muestra que evaluar correctamente es difícil y costoso en términos de tiempo y recursos, y además depende de muchos actores dentro de una comunidad (Banta & Palomba, 2014; Richmond et al., 2019) y se requiere que las instituciones cuenten con condiciones operacionales y conocimiento técnico específico. Por ello, junto con las acciones a desarrollar por parte de las universidades, es esencial que el Estado provea recursos humanos y materiales para sostener esta política. Nuestro análisis confirma lo que indica la literatura respecto de los desafíos asociados con el desarrollo de sistemas de evaluación, destacando la importancia de que las universidades cuenten con equipos que se sostengan en el tiempo, lo que permitirá la implementación y evaluación continua de las evaluaciones realizadas. Asimismo, falta que las universidades definan propósitos más específicos y que articulen las evaluaciones para cumplir estos propósitos. En este sentido, es esperable que un sistema de evaluación no funcione satisfactoriamente en sus primeros años de implementación.

Los resultados de este estudio ilustran cómo las tensiones entre la ley, los organismos públicos y las universidades se materializan en sistemas de evaluación incipientes y no consolidados, lo cual debe ser considerado en los procesos de rendición de cuentas que las comisiones de acreditación de calidad piden a las universidades. En este sentido, es necesario que las instituciones puedan evaluar y mejorar sus sistemas de evaluación antes de tomar decisiones. También es fundamental que las universidades reciban los apoyos técnicos y materiales necesarios para el logro de estos objetivos. Para esto, es clave promover la asociación y colaboración entre instituciones. Considerando lo complejo de evaluar y la necesidad de testear evaluaciones en muestras grandes, es necesario optimizar recursos y promover reflexiones frente a necesidades comunes.

Por último, en relación con las limitaciones del estudio, la primera es que los datos corresponden al primer año de implementación de la ley. Este momento es importante de mirar ya que ilustra la puesta en marcha; sin embargo, es necesario realizar un seguimiento, ya que estos sistemas de evaluación pueden haber evolucionado.

Una segunda limitación es que los propósitos y usos de las evaluaciones no estaban claramente definidos en la mayoría de las instituciones, lo que complica poder evaluar la pertinencia de los instrumentos aplicados. Considerando el rol central que tiene el uso de estas evaluaciones, creemos que una próxima investigación tiene que poner el foco en este tema. Por último, una tercera limitante es que el estudio no permite entender el rol de los organismos públicos relevantes, los cuales son clave para analizar la implementación de políticas evaluativas (Flórez Petour, 2015; Stobart, 2006). En esta línea, sería interesante incluir entre los informantes de un futuro estudio a representantes de estos organismos públicos.

---

**Financiamiento:** el proyecto de investigación del que emana el artículo fue financiado por el Ministerio de Educación de Chile, mediante el Fondo de Investigación y Desarrollo en Educación (código proyecto: FON170009, año 2017).

**Agradecimientos:** este trabajo fue apoyado por el Centro de Modelamiento Matemático (CMM), ACE210010 y FB210005 y el CIAE FB0003, financiamiento BASAL para centros de excelencia de ANID-Chile.

---

El artículo original fue recibido el 20 de abril de 2020  
 El artículo revisado fue recibido el 8 de marzo de 2021  
 El artículo fue aceptado el 9 de agosto de 2021

## Referencias

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2018). *Estándares para pruebas educativas y psicológicas*. American Educational Research Association.
- Ávalos, B. (2014). La formación inicial docente en Chile: Tensiones entre políticas de apoyo y control. *Estudios Pedagógicos (Valdivia)*, 40(Especial), 11–28. <https://doi.org/10.4067/S0718-07052014000200002>
- Banta, T. W., & Palomba, C. A. (2014). *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education*. John Wiley & Sons.
- Biggs, J. (1996). Assessing Learning Quality: reconciling institutional, staff and educational demands. *Assessment and evaluation in Higher Education*, 21(1), 5-16. <https://doi.org/10.1080/0260293960210101>
- Black, P., & Wiliam, D. (2006). The Reliability of Assessments. En J. Gardner (Ed.), *Assessment and Learning* (pp. 149-167). SAGE.
- Brookhart, S. (2017). Formative Assessment in Teacher Education. En D. J. Clandini, & J. Husu (Eds.), *The SAGE Handbook of Research on Teacher Education* (Vol. 2) (pp. 927-943). SAGE.
- Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas. (2018). *Resultados Nacionales. Evaluación Nacional Diagnóstica de la Formación Inicial Docente 2018*. CPEIP.
- Cosi, A., Voltas, N., Lázaro-Cantabrana, J. L., Morales, P., Calvo, M., Molina, S., & Quiroga, M. Á. (2020). Formative assessment at university through digital technology tools. *Profesorado, revista de currículum y formación del profesorado*, 24(1), 164-183. <https://doi.org/10.30827/profesorado.v24i1.9314>
- Daugherty, R., & Ecclestone, K. (2006). Constructing Assessment for Learning in the UK Policy Environment. En J. R. Gardner (Ed.), *Assessment and Learning* (pp. 149-167). SAGE.
- Fisher, R., Cavanagh, J., & Bowles, A. (2011). Assisting transition to university: Using assessment as a formative learning tool. *Assessment & Evaluation in Higher Education*, 36(2), 225-237. <https://doi.org/10.1080/02602930903308241>

- Flórez Petour, M. T. (2015). Systems, ideologies and history: a three-dimensional absence in the study of assessment reform processes. *Assessment in Education: Principles, Policy & Practice*, 22(1), 3-26. <https://doi.org/10.1080/0969594X.2014.943153>
- Giaconi, V., Varas, M. L., Ravest, J., Martin, A., Gómez, G., Quepil, J. P., & Díaz, K. (2019). *Fortaleciendo la Formación Inicial Docente: experiencia universitaria en la implementación de la Evaluación Diagnóstica Inicial para Pedagogías. Informe FONIDE 170009*. Ministerio de Educación de Chile.
- Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From Evidence to Action: A Seamless Process in Formative Assessment? *Educational Measurement: Issues and Practice*, 28(3), 24-31. <https://doi.org/10.1111/j.1745-3992.2009.00151.x>
- Ingvarson, L., & Rowley, G. (2017). Quality Assurance in Teacher Education and Outcomes: A Study of 17 Countries. *Educational Researcher*, 46(4), 177-193. <https://doi.org/10.3102/0013189X17711900>
- Joffe, H. (2011). Thematic Analysis. En D. Harper, & A. R. Thompson (Eds.), *Qualitative Research Methods in Mental Health and Psychotherapy* (pp. 209-223). John Wiley & Sons. <https://doi.org/10.1002/9781119973249.ch15>
- Klassen, R., Durksen, T., Patterson, F., & Rowett, E. (2017). Filtering functions of assessment for selection into initial teacher education programs. En D. J. Clandini, & J. Husu (Eds.), *The SAGE Handbook of Research on Teacher Education* (Vol 2) (pp. 893-909). SAGE.
- Klassen, R. M., & Kim, L. E. (2018). Selecting teachers and prospective teachers: A meta-analysis. *Educational Research Review*, 26, 32-51. <https://doi.org/10.1016/j.edurev.2018.12.003>
- Kohoutek, J. (2014). European standards for quality assurance and institutional practices of student assessment in the UK, the Netherlands and the Czech Republic. *Assessment & Evaluation in Higher Education*, 39(3), 310-325. <https://doi.org/10.1080/02602938.2013.830694>
- Ley N° 20.903. Crea el sistema de desarrollo profesional docente y modifica otras normas. (2016, 1 de abril). <https://www.leychile.cl/Navegar?idNorma=1087343&idParte=>
- Martínez-Arias, R. (1996). *Psicometría: Teoría de los tests psicológicos y educativos*. Síntesis.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2013). *Qualitative Data Analysis: A Methods Sourcebook*. SAGE Publications.
- Richmond, G., Salazar, M. D. C., & Jones, N. (2019). Assessment and the Future of Teacher Education. *Journal of Teacher Education*, 70(2), 86-89. <https://doi.org/10.1177/0022487118824331>
- Raaper, R. (2017). Tracing assessment policy discourses in neoliberalised higher education settings. *Journal of Education Policy*, 32(3), 322-339. <https://doi.org/10.1080/02680939.2016.1257160>
- Santelices, M. V., Catalán, X., & Horn, C. (2019). University Admission Criteria in Chile. En M. V. Santelices, C. Horn, & X. Catalán (Eds.), *The quest for equity in Chile's higher education. Decades of Continued Efforts* (pp. 81-90). Lexington Books.
- Sato, M., & Kemper, S. (2017). Teacher Assessment from Pre-service through In-service Teaching. En D. J. Clandini, & J. Husu (Eds.), *The SAGE Handbook of Research on Teacher Education* (Vol 2) (pp. 944-962). SAGE.
- Sireci, S., & Faulkner-Bond, M. (2014). Validity Evidence Based on Test Content. *Psicothema*, 26(1), 100-107. <https://doi.org/10.7334/psicothema2013.256>
- Stobart, G. (2006). The Validity of Formative Assessment. En J. Gardner (Ed.), *Assessment and Learning* (pp. 149-167). SAGE.
- Stull, J., Varnum, S. J., Ducette, J., & Schiller, J. (2011). The Many Faces of Formative Assessment. *International Journal of Teaching and Learning in Higher Education*, 23(1), 30-39. <https://www.isetl.org/ijtlhe/pdf/IJTLHE851.pdf>
- Yin, R. K. (2011). *Case Study Research and Applications: Design and Methods* (6a Ed.). SAGE.
- Yorke, M. (1998). The Management of Assessment in Higher Education. *Assessment & Evaluation in Higher Education*, 23(2), 101-116. <https://doi.org/10.1080/0260293980230201>