

FONDO DE INVESTIGACIÓN Y DESARROLLO EN EDUCACIÓN
DEPARTAMENTO DE ESTUDIOS Y DESARROLLO
DIVISIÓN DE PLANIFICACIÓN Y PRESUPUESTO
MINISTERIO DE EDUCACIÓN

Fortaleciendo la Formación Inicial Docente: experiencia universitaria en la implementación de la Evaluación Diagnóstica Inicial para Pedagogías.

INSTITUCION ADJUDICATARIA: Departamento de Evaluación, Medición y Registro
Educativo. Universidad de Chile
Universidad de O'Higgins
Pontificia Universidad Católica de Valparaíso

INVESTIGADORA PRINCIPAL: Valentina Giaconi

EQUIPO DE INVESTIGACIÓN: María Leonor Varas
Javiera Ravest
Angel Martin
Gabriela Gómez
Juan Pablo Queupil
Karina Díaz

PROYECTO FONIDE N°: 170009

Versión resumida del proyecto señalado, en el marco de UNDÉCIMO CONCURSO FONIDE
Santiago/2019

Secretaría Técnica FONIDE, Departamento de Estudios y Desarrollo. División de Planificación
y Presupuesto, MINEDUC.

Alameda N° 1371, Piso 8. E-mail: fonide@mineduc.cl www.fonide.cl

Resumen:

Entre los esfuerzos institucionales por mejorar la calidad en la Formación Inicial Docente, la Ley 20.903 establece la obligatoriedad de una evaluación diagnóstica al comenzar los estudios universitarios en pedagogía, siendo requisito para la acreditación de las instituciones formadoras. Bajo ese contexto, el presente proyecto estudia las experiencias de implementación de la evaluación diagnóstica en 5 universidades. La metodología del estudio fue mixta y centró su análisis en tres focos principales, a saber, propósitos y usos de la evaluación, administración en la aplicación, y contenidos y comportamiento psicométrico de los instrumentos aplicados. Los hallazgos del proyecto dan cuenta de una falta de claridad sobre el propósito de la evaluación entre las instituciones. Pese a que todas las instituciones valoran realizar diagnósticos en primer año, su especificidad y pertinencia para el campo de las pedagogías ha sido escasamente discutido. De momento, los resultados de estas evaluaciones no aportan mayormente al diseño o ajuste de planes de acompañamiento y nivelación estudiantil. En cuanto a los instrumentos, es necesario definir sus objetivos y los referentes teóricos que los articulan, además de mejorar aspectos de diseño y construcción, lo que ayudaría a mejorar su confiabilidad y discriminación, favoreciendo así la toma de decisiones pertinentes y contextualizadas. En cuanto a las universidades, se deben robustecer las capacidades técnicas para implementar las evaluaciones y generar mayores instancias de colaboración. Se concluye la necesidad de esclarecer los propósitos de estas evaluaciones en las pedagogías y su lugar en un sistema integral de Formación Inicial Docente, considerando la diversidad de contextos institucionales.

Palabras claves: Formación Inicial Docente – Evaluación Diagnóstica – Fuentes de evidencia de validez – Competencias Docentes - Educación Superior

Introducción

En la última década el Estado chileno ha realizado sostenidos esfuerzos por mejorar la formación de profesores y las condiciones laborales en que se desarrolla el ejercicio docente. El año 2016 se aprobó la Ley 20.903, conocida como la Nueva Política Nacional Docente, que entre sus medidas incluye la obligatoriedad de realizar una evaluación diagnóstica al ingreso de todas las carreras de pedagogía. En esta primera fase - en vigencia desde 2017- hubo una deliberada ausencia de orientaciones nacionales que, si bien han complicado en parte los procesos de implementación, también presentan un gran potencial para mejorar la calidad de la formación docente aprovechando al máximo la experiencia de las instituciones formadoras.

Entre las medidas específicas de la Ley en el ámbito de la Formación Inicial Docente (en adelante, FID), se exige la acreditación de los programas de pedagogía bajo criterios formativos y de excelencia, que administra la Comisión Nacional de Acreditación (CNA) (Ley 20.903 de 2016. Art. 2º, BCN). Dentro de estos requisitos, el nuevo marco regulador establece que las y los estudiantes de pedagogía rindan evaluaciones al inicio y al finalizar su carrera. La primera de estas evaluaciones es aplicada por cada una de las universidades y la Ley señala que su propósito es que las universidades cuenten con más información referencial, esperando con ello que se desarrollen acciones de nivelación y acompañamiento, si corresponde (BCN, 2016).

Este estudio es un primer esfuerzo por sistematizar y analizar los resultados de las experiencias diagnósticas realizadas por un conjunto diversos de instituciones en el año 2017. En esta sistematización se incluye evidencia sobre las características técnicas de los instrumentos aplicados por las universidades. Además, se presentan recomendaciones a la política pública para promover el logro de los objetivos de la Ley 20.903.

Este informe se estructura de la siguiente forma: en primer lugar, se contextualiza la discusión revisando iniciativas nacionales e internacionales por mejorar la FID -en particular en la fase de ingreso a los estudios de pedagogía-, y se caracterizan los cambios en la matrícula del estudiantado de pedagogía en la última década en Chile. En segundo lugar, se discute teóricamente sobre las competencias docentes fundamentales referidas en la literatura especializada, y se reflexiona sobre la necesidad de contar con evidencias de validez para las interpretaciones de los puntajes de los instrumentos utilizados en estos contextos educativos, además de los requerimientos operacionales que deben cumplir para alinearse con los estándares de evaluación y medición educacional (AERA; APA; NCME, 2014)¹. En tercer lugar, se presentan las preguntas de investigación, hipótesis y objetivos. En cuarto lugar, se muestra el enfoque metodológico utilizado, que incluye el uso de fuentes primarias y secundarias, además de estrategias cualitativas y cuantitativas para la recolección y análisis de los datos.

Finalmente, se presentan los resultados de la investigación, organizados en tres dimensiones: (1) propósito de la evaluación, (2) aspectos operativos y procedimentales, y (3) dominios evaluados y fuentes de evidencias de validez. Se incluyen al final, conclusiones y recomendaciones para la formulación de políticas públicas en esta materia. Estas últimas apuntan a definir los propósitos de la evaluación en el marco de un proceso integral de FID, empoderar a las universidades para que logren procesos exitosos de implementación, y con ello asegurar que las evaluaciones sean situadas y respondan tanto al sello formativo de la institución como a las características particulares de su estudiantado.

Los esfuerzos recientes de la política pública para mejorar la FID

Diversos estudios han sostenido que mejorar el desempeño docente es una de las variables que guarda mayor impacto sobre la calidad de la enseñanza (Hattie, 2008; Hanushek & Rivkin, 2012; OECD, 2015). Además, los países mejor posicionados en las comparaciones internacionales demuestran alto compromiso con la profesionalización de la carrera docente y cuentan con rigurosos procedimientos para el aseguramiento de la calidad en la formación de profesores (Lauzano & Moriconi, 2014; Ingvarson L. , 2013). Sin embargo, en América Latina persisten ciertas dificultades para avanzar hacia un programa de mejora de la FID realmente efectivo. La UNESCO (2013) sintetiza estos problemas en seis “nudos críticos”: 1) el bajo nivel

¹ *American Psychological Association, American Educational Research Association y National Council on Measurement in Education.*

de formación con que ingresan las y los jóvenes a los estudios pedagógicos; 2) la baja calidad de algunos programas y procesos formativos; 3) la escasa preparación de quienes forman docentes; 4) la estandarización de las estrategias de enseñanza que termina excluyendo a minorías y grupos sociales vulnerables; 5) la tensión entre lógicas escolarizantes y lógicas académicas en la formación de docentes; y, finalmente, 6) la insuficiente regulación existente en los programas de formación.

Para atender a estos problemas los países han emprendido diversas estrategias de intervención. En esa línea, el trabajo de Ingvarson y Rowley (2017) realiza un análisis internacional y distingue tres ámbitos de acción de las políticas de fortalecimiento a la FID: (1) políticas enfocadas en el aseguramiento de la calidad en la fase de reclutamiento y selección de las y los estudiantes que ingresan a los programas de pedagogía, cuyo objetivo es atraer a estudiantes de alto rendimiento académico a través de medidas de equidad y acceso, filtros de excelencia académica, y potenciamiento de la carrera docente como una alternativa profesional atractiva; (2) políticas abocadas al monitoreo y aseguramiento de la calidad de la actividad propia de los programas de FID, comúnmente desarrollada a través de procesos de acreditación institucional y de programas; y (3) políticas de certificación, enfocadas en garantizar calidad en los conocimientos y habilidades de las y los docentes que egresan de los programas de pedagogía e ingresan al ámbito profesional.

En Chile, la política orientada al fortalecimiento y aseguramiento de la calidad de la FID comienza a establecerse con fuerza desde la década de 1990. Ante la débil situación de los recursos académicos de los programas y del bajo rendimiento académico de sus estudiantes (Ávalos, 2014), el Estado implementó el Programa de Fortalecimiento a la FID, cuyos componentes apuntaban a la asignación de recursos competitivos a instituciones que adquirieran compromisos de mejoramiento y un robusto sistema de apoyo a los proyectos de auto-mejora derivados del MINEDUC (Ávalos, 2002). Hacia la década del 2000 se establece una nueva política de control y mejoramiento de la formación de profesores, dada la creciente demanda de las instituciones por orientaciones curriculares y nuevos mecanismos de auto-evaluación, sumado a un explosivo y desregulado incremento en la oferta de programas de FID (Ávalos, 2002). Se constituye así el Sistema de Aseguramiento de la Calidad de la Educación Superior, que posteriormente se refuerza con otras políticas desagregadas que apuntan a su perfeccionamiento mediante mayores exigencias e incentivos (Ávalos, 2014).

La Tabla 1 ofrece una síntesis de las políticas de la última década diseñadas para la mejora de la FID en Chile considerando las tres dimensiones de análisis de Ingvarson y Rowley (2017). El primer ámbito corresponde a la promoción del ingreso de mejores candidatos a la profesión docente, desafío especialmente urgente en Chile según la OECD (2004) dado el bajo rendimiento de las y los postulantes a pedagogía en mediciones estandarizadas. En el segundo ámbito, el principal mecanismo fue la exigencia de acreditación obligatoria para todos los programas de pedagogía, a través del establecimiento del Sistema de Aseguramiento de la Calidad en la Educación Superior (OECD, 2017). Sobre el tercer ámbito, referido al egreso, destaca la Prueba INICIA (2008-2015), que informa sobre la preparación, formación y calidad profesional de futuros docentes, aplicada a estudiantes egresados y titulados. Esta prueba actualmente lleva por nombre Evaluación Nacional Diagnóstica de la Formación Inicial

Docente (ENDFID) y a diferencia de INICIA, constituye un requisito para obtener el título profesional (artículo 27 bis de la Ley N°20.129).

Tabla 1. Ámbitos de intervención de las políticas para el fortalecimiento de la FID en Chile.

Ámbitos de sistemas nacionales de aseguramiento de la calidad en FID (Ingvarson & Rowley, 2017)	Medidas contempladas en la Nueva Carrera Docente
Aseguramiento de la calidad del reclutamiento y selección de quienes ingresan a los programas de pedagogía.	Beca Vocación de Profesor: desde 2010 financia los estudios de pedagogía en programas acreditados a jóvenes con altos puntajes en la Prueba de Selección Universitaria. Cambios en los requisitos de ingreso a programas de pedagogía.
Monitoreo y aseguramiento de la calidad de la actividad propia de los programas de FID (acreditación institucional y de los programas).	Acreditación obligatoria para los programas de pedagogía. Desde 2017 esta acreditación incluye como requisito la aplicación de la evaluación diagnóstica en cada universidad.
Políticas de certificación para docentes que egresan de los programas de pedagogía.	INICIA (2008-2015) y actualmente ENDFID, la cual es requisito para obtener el título profesional.

Fuente: Elaboración propia.

En este marco general, la evaluación diagnóstica parece responder simultáneamente a dos de los ‘nudos críticos’ que puntualiza la UNESCO: el bajo nivel de formación de quienes ingresan a estudiar pedagogía y la calidad de los programas impartidos por las universidades. Lo primero se debe a que la evaluación diagnóstica se piensa fundamentalmente para que las universidades dispongan de información útil que facilite realizar adecuaciones curriculares y ajustar sus estrategias de enseñanza a las necesidades de las y los nuevos estudiantes de primer año. Lo segundo, porque la Ley vincula el proceso de acreditación de las carreras de pedagogía con la aplicación efectiva de la ED.

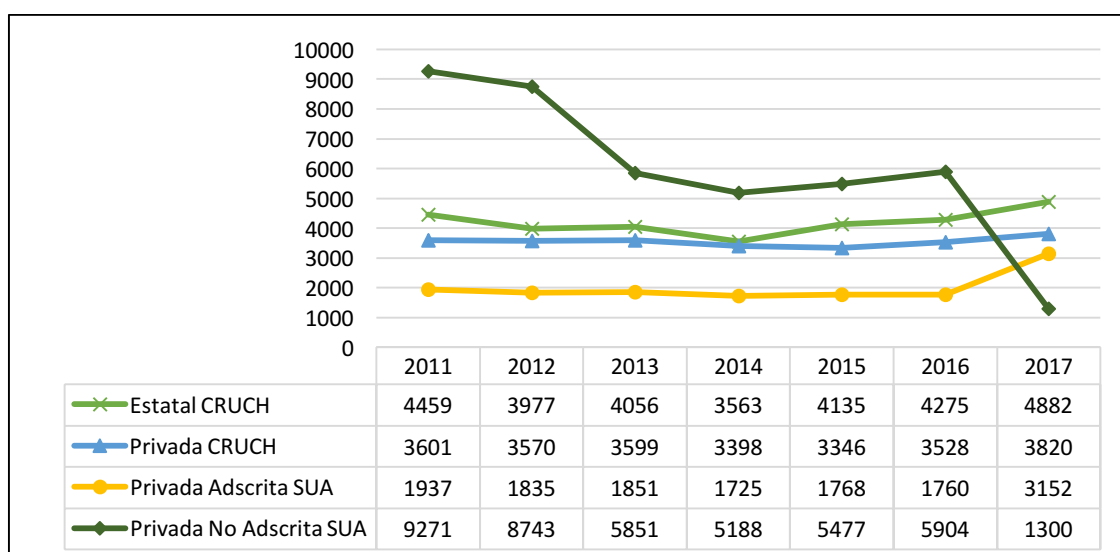
El nuevo panorama de las pedagogías en Chile

Este conjunto de medidas y acciones elaboradas en la última década también ha tenido un impacto sobre la oferta de programas de pedagogía y en la composición de la matrícula de estudiantes. Ambos aspectos han cambiado considerablemente en el último lustro.

Un estudio de Pedraja *et al.* (2012) identifica una relación positiva y directa entre el porcentaje de logro promedio de la Prueba INICIA 2010 (como indicador de la FID) y los años de acreditación de las universidades. Por ello, los esfuerzos de la política pública se han centrado en mejorar la calidad de las instituciones formadoras. La oferta de programas ha aumentado y diversificado en la última década, pero su crecimiento se ha visto supeditado al cumplimiento de estándares de calidad más rigurosos. Uno de los indicadores más llamativos de este cambio regulatorio es que, para el período 2011-2017, la matrícula de primer año descendió en un 31,7%, considerando la totalidad de universidades que imparten pedagogía (SIES, 2017).

Este descenso de la matrícula estudiantil de primer año también ha sido acompañado de una redistribución favorable a las instituciones de calidad certificada, tal como muestra el Gráfico 1², siendo notorio el incremento que han experimentado las universidades estatales y privadas del Consejo de Rectores (CRUCH) desde 2015. Por otra parte, entre 2016 y 2017 se observa una caída en la matrícula de las universidades privadas no adscritas al Sistema Único de Admisión (SUA) a la vez que incrementa la matrícula en las privadas adscritas al SUA. Así, este abrupto descenso de la matrícula en las privadas no adscritas al SUA puede explicarse, al menos en parte, por el aumento de exigencias para el acceso a la pedagogía y la obligatoriedad de la acreditación.

Gráfico 1. Evolución de la matrícula de primer año por tipo de universidad (2011-2017)



Fuente: Elaboración propia a partir de los datos del Servicio de Información de Educación Superior (SIES). 2017. Base de datos Matrícula Histórica 2007-2017 (actualizada el 19 de julio de 2017).

Particularmente en el año 2017, de las 43 universidades acreditadas, 15 son estatales del CRUCH, 8 son privadas del CRUCH, 11 son privadas adscritas al SUA y 9 son privadas no adscritas al SUA (ver Tabla 2).

Tabla 2. Matrícula total y de primer año según tipo de universidad en pedagogía (2017)

Tipo de institución	N° universidades	Media años acredit.	Matrícula Total	Matrícula 1er año	1er año Mujeres	1er año Hombres
Estatad CRUCH	15 (todas con gratuidad)	4.6	18274 (31.4%)	4718 (38.1%)	2918 (61.8%)	1800 (38.2%)
Privada CRUCH	8 (todas con gratuidad)	4.5	16237 (27.9%)	3820 (30.9%)	2570 (67.3%)	1250 (32.7%)
Privada Adscrita SUA	11 (5 con gratuidad)	5.7	16453 (28.3%)	3152 (25.5%)	2157 (68.4%)	995 (31.6%)
Privada No Adscrita SUA	9 (sin gratuidad)	2.8	7270 (12.5%)	685 (5.5%)	439 (64.1%)	246 (35.9%)

2 Se consideran aquellas carreras que cumplen con los siguientes criterios: jornada: diurna; modalidad: presencial; nivel: pregrado; área de conocimiento: educación; área de carrera genérica: pedagogías.

Total	43	58234	12375	8084	4291
-------	----	-------	-------	------	------

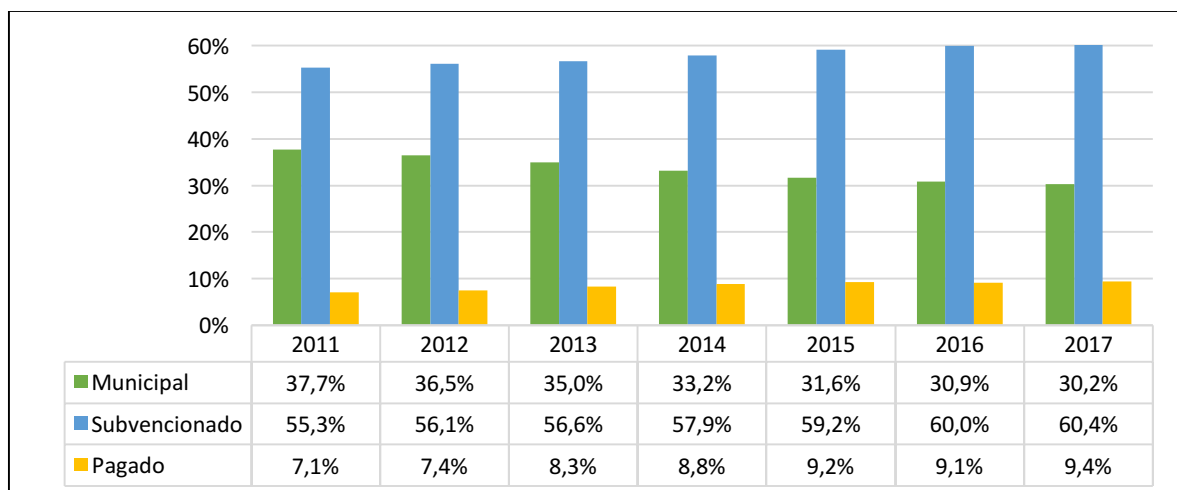
Fuente: Elaboración propia a partir de los datos del Servicio de Información de Educación Superior (SIES). 2017. Base de datos Matrícula Histórica 2007-2017 (actualizada el 19 de julio de 2017).

La tabla también da cuenta que la matrícula de mujeres en primer año prácticamente duplica a la matrícula de hombres, lo que coincide con lo señalado por otras mediciones acerca del carácter altamente feminizado de la profesión docente (CNED, 2016; SIES, 2017). Hablar de pedagogías es referirse a uno de los principales campos desde los cuales las mujeres se han incorporado a los estudios superiores (Ávalos, 2002). Esta tendencia se ha mantenido históricamente: los datos del Consejo Nacional de Educación muestran que las mujeres, en las carreras asociadas a la educación, tenían una tasa de participación (matrícula) cercana al 70% para el año 2005, y luego en el año 2016 este porcentaje aumenta a 73%.

Por otra parte, en cuanto a la oferta de programas de pedagogía, según SIES (2017) al año 2017 existen 419 programas de pedagogía acreditados en las universidades del país, equivalentes al 85,8% de la oferta total en esta área. Además, se observa una disminución del 26,3% respecto de la oferta disponible en 2011. En 2017, el 32,4% de los programas corresponde a universidades estatales del CRUCH, el 25,3% a universidades privadas adscritas al SUA, el 22,8% a universidades privadas no adscritas al SUA, y el 19,6% a privadas pertenecientes al CRUCH.

En relación con la procedencia de quienes ingresan a carreras de pedagogía, desde 2011, la proporción de la matrícula de estudiantes de establecimientos subvencionados se mantiene sobre el 55% en cada año, llegando a superar el 60% en 2016 y 2017. La proporción de estudiantes procedentes de establecimientos particulares es inferior al 10% en el período 2011-2017 (ver Gráfico 2).

Gráfico 2. Proporción de estudiantes que ingresa a pedagogía según dependencia de su establecimiento de origen (2011-2017)



Fuente: Elaboración propia a partir de los datos del Servicio de Información de Educación Superior (SIES). 2017. Base de datos Matrícula Histórica 2007-2017 (actualizada el 19 de julio de 2017).

Con respecto a los puntajes de la Prueba de Selección Universitaria (en adelante, PSU), para el ingreso en 2017, el 61,8% de los programas de pedagogía tuvo promedios PSU en Lenguaje y Comunicación y Matemática superiores a 500 puntos, debido a la exigencia de obtener este puntaje como mínimo para el ingreso. Los programas de pedagogía de las universidades privadas del CRUCH tienen medias

más altas en puntajes PSU, Ranking y Notas de Enseñanza Media (NEM). Luego se encuentran las universidades estatales del CRUCH, seguidas por las privadas adscritas al SUA y las privadas no adscritas (ver Tabla 3).

Tabla 3. Media de los promedios PSU, Ranking, NEM de los programas de pedagogía (2017)

Tipo de universidad	PSU (Prom. Lenguaje y Matemática)	Ranking	NEM
Estatal CRUCH	562.09	595.50	574.42
Privada CRUCH	571.74	609.16	585.36
Privada Adscrita SUA	545.74	534.18	521.08
Privada No Adscrita SUA	526.89	543.67	530.92

Fuente: Elaboración propia a partir de los datos del Consejo Nacional de Educación (CNEC). 2017. Base de datos INDICES Matrícula Instituciones/Programa Educación Superior 2005-2017 (actualizada el 3 de noviembre de 2017).

En relación con las características de los estudiantes que ingresan a las carreras de pedagogía –dada la ampliación de la matrícula y la oferta de carreras durante la década 2000-2010–, el ingreso a programas de pedagogía se asocia al ascenso, en términos de movilidad educativa y social, a través de la educación superior. Así, la mayor parte de la matrícula del estudiantado de pedagogía se caracteriza por su composición concentrada en las “primeras generaciones” universitarias y tener comparativamente puntajes PSU bajos (Cox, Meckes, & Bascopé, 2010).

En definitiva, la implementación reciente de diversas políticas que buscan hacerse cargo de las características de la juventud que se forman para ser docentes, y de la calidad de las universidades que imparten los programas, ha tenido un impacto significativo en la composición de la matrícula del estudiantado. Sin embargo, no han tenido impacto en otras características como la composición por sexo y género. La nueva evaluación diagnóstica por tanto se implementa en un conjunto de instituciones más acotado y selectivo que el existente en la primera década del 2000, y debe caracterizar a una nueva población de estudiantes que cuenta con un mejor desempeño relativo en pruebas estandarizadas que antes, pero esto poco informa sobre los perfiles concretos del estudiantado.

Se debe tener en cuenta además que la evaluación propuesta tiene consecuencias, en la medida en que se convierte en un referente importante para la toma de decisiones en materias relativas al acompañamiento, las estrategias de enseñanza y las mejoras curriculares de las carreras de pedagogía. En ese sentido, es fundamental que estas evaluaciones eviten producir o reproducir sesgos de diversa índole, sobre todo aquellos de orden socioeconómico y de género. En cuanto a este último, algunas investigaciones han demostrado que, en contextos de examen o evaluación de matemáticas, las mujeres tienden a tener un rendimiento diferencial más bajo que el de los hombres aun cuando su habilidad sea la misma (Niederle & Vesterlund, 2010), y la evidencia para Chile confirma esta tendencia en evaluaciones altamente competitivas, como la PSU (Oscar, Mizala, & Meneses, 2016).

Menos atención han recibido en la política pública nacional aquellos aspectos que dicen relación con las competencias a evaluar entre los docentes en formación y en ejercicio. Parece necesario entonces dar cuenta de los enfoques actuales que ganan

popularidad y que pueden orientar a las instituciones en la elaboración de los marcos de evaluación.

Tendencias actuales sobre competencias docentes

En los últimos 20 años distintas investigaciones se han centrado en comprender y conceptualizar los conocimientos y competencias necesarias para el desempeño de la profesión docente (Guerrero, 2017; Schleicher, 2016; Townsend & Bates, 2007). El estudio y la elaboración de marcos teóricos bien delimitados que conceptualicen los saberes, habilidades y disposiciones implicadas en el ejercicio del profesorado resulta crucial como insumo para la formación de docentes y para la implementación de nuevos currículos (Remillard, 2005; Lloyd, 2008).

Lo que se entiende por competencia docente refiere a un constructo multidimensional que comprende conocimientos, habilidades y creencias, que se manifiestan durante el desempeño de la enseñanza (Yang, Kaiser, G, König, J, & Blömeke, 2018). Su delimitación conceptual ha significado un desafío puesto que está atravesada por las situaciones específicas del aula y por las expectativas que la sociedad pone en sus docentes, las que difieren según los distintos contextos (Johansson, Strietholt, Rosén, & Myrberg, 2014; Blömeke & Kaiser, 2017). Pese a lo anterior, la investigación muestra un avance en los marcos conceptuales que permiten identificar una base común respecto a los elementos presentes en el desempeño profesional docente (Yang, Kaiser, G, König, J, & Blömeke, 2018).

Hasta hace algunos años las principales propuestas teóricas estaban centradas en describir los conocimientos docentes desde un enfoque cognitivo, es decir, a partir de una detallada especificación de los tipos de saberes involucrados en el ejercicio de la enseñanza. Destaca el trabajo de Shulman (1986) del que se desprenden otros trabajos posteriores. Este autor describe la competencia docente a partir de tres dimensiones: a) conocimiento del contenido, que se centra en las áreas disciplinares a enseñar; b) conocimiento pedagógico general, que incluye nociones de evaluación y enseñanza, y c) conocimiento pedagógico del contenido, referido a aspectos didácticos que integran lo disciplinar con su enseñanza. Esta categorización ha sido reelaborada por Weinert (2001) en el estudio sobre competencias docentes a gran escala TEDS-M. Otras investigaciones han reelaborado y complejizado las dimensiones de Shulman para la enseñanza de distintas disciplinas, por ejemplo, para la enseñanza de matemáticas (Ball, 2008) y de lengua inglesa (Freeman, Katz, Gomez, & Burns, 2015). Otros trabajos han conceptualizado la competencia docente centrándose en los aspectos afectivos y motivacionales implicados en el ejercicio de la enseñanza (Richardson, 1996; Thompson, 1992), Atendiendo a elementos como creencias epistemológicas sobre la materia a enseñar, creencias asociadas a la enseñanza de tal materia, motivación y habilidades como la autorregulación.

Una de las limitaciones de estos marcos conceptuales es que la competencia docente se entiende como un conjunto de recursos cognitivos y motivacionales que subyacen al desempeño en el aula. Sin embargo, en la práctica, dichos recursos no se manifiestan simple y directamente en el desempeño, sino que estarán mediados por otras habilidades más relacionadas con la capacidad de las y los docentes para

mediar situaciones complejas en el aula (Blömeke, Gustafsson, & Shavelson, 2015; Blömeke & Kaiser, 2017).

Las conceptualizaciones actuales de competencia docente han considerado este aspecto problemático y han elaborado un enfoque que definen como “situado”, es decir, han ampliado los marcos tradicionales enfocados en descripciones cognitivas incorporando una concepción que toma en cuenta la situación específica del aula (Mulder, 2014). Bajo este enfoque los procesos de percepción e interpretación de una situación de aula específica, junto con la toma de decisiones, influyen en la forma en que un docente despliega sus recursos de conocimiento (Blömeke, Gustafsson, & Shavelson, 2015). Esta línea teórica ha sido denominada bajo el nombre de *Noticing*, que involucra tres fases: (a) la percepción, es decir, la forma en el maestro selecciona eventos particulares de un entorno educativo para situar la atención y por cuánto tiempo, (b) la interpretación, es decir, la forma en que interpreta las actividades percibidas y las deducciones que genera a partir de las mismas, y cómo relaciona eventos particulares del aula con categorías más abstractas para darles sentido, y (c) la toma de decisiones: ya sea como anticipando respuestas a las actividades de los estudiantes o proponiendo estrategias de instrucción alternativas. De esta forma los recursos cognitivos y emocionales-motivacionales de las y los docentes son aprovechados en el curso de la práctica docente a través de la puesta en escena de estas tres fases del *noticing*, las que se integran para dar forma a la *performance* del docente en la sala de clases.

Considerando lo anterior, los marcos teóricos actuales de competencias docentes que abarcan aspectos cognitivos y afectivo-motivacionales así como su puesta en práctica en la situación del aula escolar, brindan un punto de partida para orientar las evaluaciones que tienen lugar en distintos momentos de la formación inicial de profesores.

Fuentes de evidencia de validez y procedimientos de aplicación de la Evaluación Diagnóstica

Habiendo considerado los antecedentes que marcan la trayectoria de reformas a la FID, la actual composición de la oferta y del estudiantado de pedagogía, y los marcos teóricos de competencias docentes, la pregunta que se abre tanto para las universidades como para el sistema en general es cómo asegurar que esta evaluación cumpla con el espíritu que persigue la Ley al tiempo que contribuya a fortalecer un sistema integral de FID.

Los instrumentos de evaluación diseñados para tal propósito tienen un rol gravitante en la producción de evidencia sobre habilidades y conocimientos claves para el éxito de los estudiantes que ingresan a pedagogía. Es importante que estos instrumentos cumplan con ciertas características fundamentales que permitan entregar evidencias de la validez de las interpretaciones de sus resultados, entre las cuales destaca la calidad técnica, pertinencia e imparcialidad, y la definición de un referente o marco teórico (Domínguez, Bascopé, Carrillo, Lorca, Olave, & Pozo, 2012). En ese sentido, es fundamental que el primer paso en la construcción sea establecer claramente los propósitos y usos específicos que se darán a los resultados que informen estas

evaluaciones, los cuales deben ser conocidos y compartidos por todos quienes participen de la evaluación. Para cumplir con lo anterior, a nivel mundial se han considerado como referente los estándares psicométricos y técnicos definidos por APA, AERA, y NCME (2014), que proveen los criterios necesarios para asegurar el ajuste de la evaluación a sus objetivos. Adicionalmente, establecen criterios sobre la confiabilidad, validez, discriminación, uso y administración de los instrumentos.

Dentro de todo sistema o proceso que considera evaluaciones, la información producida a partir de estas, y las decisiones que se tomen, deben guardar pertinencia y alineación con sus usos y propósitos. En esa discusión destaca el concepto de validez, la cual es definida como “el grado en que la evidencia y la teoría sostienen las interpretaciones de los puntajes de una evaluación, respecto de sus usos propuestos” (AERA; APA; NCME, 2014, pág. 11). Es por tanto una consideración fundamental al momento de desarrollar o analizar instrumentos de evaluación. Provee evidencia que establece la base científica necesaria para una correcta interpretación de los resultados, considerando los propósitos de la evaluación misma. Dicho de otra forma, las evidencias de validez ofrecen información para un juicio evaluativo integrado, que sostiene la adecuación y pertinencia de las inferencias o acciones basadas en los resultados de una evaluación. En ese sentido, y según Stufflebeam y Coryn (2014), la validez no es una proposición de “todo o nada”, sino que una gradación. Es incorrecto generalizar que un instrumento o procedimiento es válido o no válido, deben juzgarse como válidas o no válidas las distintas inferencias o conclusiones que emanan de un uso particular de tal instrumento.

Respecto de lo anterior, los estándares internacionales establecen cinco fuentes de evidencia que pueden ser utilizadas para evaluar la validez de las interpretaciones sobre un instrumento, según sus usos particulares (AERA; APA; NCME, 2014), las que se describirán a continuación. Sobre la calidad de las evidencias de validez, Chan (2014), sugiere que el foco no solo debe estar puesto en evaluar si existe evidencia de validez, sino también en los enfoques metodológicos para obtener tal evidencia.

1. **La evidencia de validez basada en el contenido** habla sobre la relación entre el contenido del instrumento y el constructo teórico³ que se pretende medir. Esto permite garantizar que el conjunto de ítems que componen la evaluación es relevante y representativo del constructo (Sireci & Faulkner-Bond, 2014). Esta evidencia puede incluir análisis lógicos o empíricos sobre qué tan adecuadamente el contenido de la evaluación representa el dominio definido y su relevancia al momento de interpretar los resultados. En general, este análisis se realiza por expertos disciplinares, quienes analizan la relación entre las partes que componen la evaluación y el constructo definido.
2. **La evidencia de validez basada en los procesos de respuesta** muestra el ajuste entre el constructo, la naturaleza del desempeño y las formas de responder de quienes son evaluados. En otras palabras, se observan los procesos cognitivos implicados en la resolución de los ítems (Elosua, 2003). Este tipo de evidencia incluye análisis teóricos y empíricos que permiten conocer el ajuste entre el

³ Un constructo es el atributo teórico (Sireci & Faulkner-Bond, 2014), concepto o característica (AERA; APA; NCME, 2014), que el instrumento pretende medir.

constructo y las diversas estrategias cognitivas. Generalmente se obtiene a partir de estudios cualitativos a tareas realizadas por sujetos particulares: así, amplias diferencias individuales en una misma tarea pueden revelar la necesidad de modificar o reconsiderar ciertos formatos de evaluación para resguardar la medición efectiva del dominio propuesto. Además, esta evidencia puede ayudar a tomar mejores decisiones de interpretación de resultados considerando las características de los distintos subgrupos sometidos a evaluación.

3. **La evidencia de validez basada en la estructura interna** indica el grado en que las relaciones entre los ítems y componentes de la evaluación conforman el constructo, del cual se generan los puntajes y sus interpretaciones de puntajes (AERA; APA; NCME, 2014). Todo esto, bajo el presupuesto de que una evaluación debe medir el dominio unidimensionalmente. Así, en general los estudios sobre la estructura interna son diseñados para revelar cuando los ítems funcionan diferenciadamente en subgrupos de la población evaluada o no son capaces de evaluar unificadamente el dominio.
4. **La evidencia de validez basada en relaciones con otras variables** se obtiene mediante el análisis de la relación entre las interpretaciones de puntaje y variables externas, para conocer el grado en que esas relaciones son consistentes con el constructo subyacente (McCoach, Gable, & Madura, 2013). En ese contexto, es esencial que las mediciones de un instrumento demuestren un patrón de relaciones externas con variables específicas, escalas de otros instrumentos conocidos, u otros criterios externos de modo consistente con las expectativas teóricas.
5. **La evidencia de validez basada en las consecuencias de la evaluación** refiere a los usos de los instrumentos y las consecuencias emanadas desde la información que proveen. Como los sistemas de evaluación están destinados a tener un impacto tanto en la instrucción como en el aprendizaje, es importante incluir las consecuencias, sobre todo aquellas con potencial impacto en la población de interés, particularmente impacto adverso o sistémicos (Kane, 2013). A este respecto, las instituciones que mandatan el uso de los instrumentos tienen la responsabilidad de monitorear su impacto, además de identificar y minimizar potenciales consecuencias negativas (AERA; APA; NCME, 2014).

Dicho lo anterior, debe recalarse que, para iniciar cualquier estrategia de validación, siempre es necesario previamente establecer los usos y propósitos de las evaluaciones, pues el objetivo final de la medición es mejorar la toma de decisiones. Como argumentan Newton y Shaw (2014), si se mide para tomar decisiones, entonces lo primero que debe considerarse, tanto al diseñar como al evaluar pruebas, es el uso que se dará a sus puntajes: es decir, las decisiones que se tomarán en su base.

El resguardo de la validez de un instrumento de evaluación se juega en las distintas etapas de su diseño, desarrollo, aplicación y análisis (AERA; APA; NCME, 2014). En otras palabras, las características de las condiciones operacionales de un instrumento definirán en buena parte las evidencias de validez que éste demostrará. Esta preocupación por la validez debe considerar en un sentido amplio la etapa de diseño y desarrollo del instrumento; la definición de los propósitos de la evaluación, y su subsecuente marco de evaluación, junto con el formato, extensión y modo de

corrección del instrumento, y el reporte de resultados. Luego debe considerar los procesos de interpretación y vinculación de puntajes, y el establecimiento de puntos de corte. Por último, la aplicación y monitoreo de condiciones de administración, calificación, presentación e interpretación de informes de las evaluaciones.

1. **Sobre el diseño y desarrollo de la evaluación:** considera el diseño y evaluación de las especificaciones de la prueba, desarrollo y evaluación de sus ítems, reunión de nuevos formularios de prueba, y procedimientos y materiales para la administración y calificación. La etapa de diseño y desarrollo tiene en su núcleo la determinación y documentación de la finalidad y usos previstos que guiarán la estructura de todo el instrumento, y por tanto orientarán los criterios de validez. Con ello, derivan las determinaciones sobre contenido, formato, extensión, características psicométricas de los ítems y de la prueba, modo de ejecución y administración, calificación y reporte de puntajes.
2. **Sobre la asignación de puntajes y resultados:** la asignación de puntajes a través de una escala corresponde a una herramienta para facilitar la interpretación de resultados. Por esta razón, los usos previstos para el instrumento juegan un rol fundamental en la determinación del tipo de escala de puntaje. Por último, debe prestarse especial atención a la definición de puntajes de corte, debido a que la determinación de rangos de rendimiento en categorías estructura la base de la interpretación de resultados, y por tanto se relaciona directamente con la validez del instrumento.
3. **Sobre la administración, calificación, reporte e interpretación de la evaluación:** la estandarización de procedimientos detallados en cada uno de estos ámbitos permite elevar la posibilidad de que todos quienes se someten a evaluación tengan la misma posibilidad de mostrar su rendimiento, aumentando la precisión y comparabilidad de los resultados. Por este motivo, es relevante la estandarización de instrucciones de evaluación, protocolos de acción y capacitación de examinadores, monitoreo del proceso de asignación de puntajes, y la administración de información contextual para la interpretación de los resultados.

Preguntas de la investigación y objetivos

Pregunta: ¿Qué características tienen las evaluaciones diagnósticas y qué desafíos técnicos deben enfrentar las universidades para implementarlas?

Objetivo General: Describir el estado actual de la implementación de evaluaciones diagnósticas de estudiantes que ingresan a carreras de pedagogía en un conjunto diverso de Instituciones de Educación Superior.

Objetivos Específicos:

1. **(OE1)** Describir los fundamentos teóricos y los aspectos técnicos y operativos de las evaluaciones diagnósticas para estudiantes de primer año de pedagogía realizadas por una muestra diversa de Instituciones de Educación Superior.
2. **(OE2)** Analizar los aspectos cualitativos (de contenido) y cuantitativos (psicométricos) de los instrumentos evaluativos utilizados por las Instituciones

de Educación Superior incluidas en el estudio.

3. **(OE3)** Identificar los intereses y necesidades comunes de las Instituciones de Educación Superior incluidas en el estudio, en torno al diagnóstico de estudiantes de pedagogía y la pertinencia de los instrumentos utilizados respecto de estas mismas necesidades.
4. **(OE4)**⁴ Entregar orientaciones técnicas para la mejora del diseño de la evaluación diagnóstica ajustadas colaborativamente con las Instituciones de Educación Superior incluidas en el estudio.

Metodología

La muestra se compone de cinco universidades públicas y privadas que abarcan parte importante de la matrícula de pedagogía en Chile, con diversas características respecto de: ubicación geográfica, acreditación institucional, variedad de programas y adscripción al SUA. La conformación de la muestra correspondió a un muestreo cualitativo por conveniencia y las conclusiones no se pueden extrapolar a todas las instituciones formadoras de docentes.

La metodología del proyecto es mixta, dado que combina enfoques cualitativos y cuantitativos para responder a sus objetivos, a partir de fuentes de información primarias y secundarias. La Tabla 4 resume las estrategias metodológicas para analizar tal información.

Tabla 4. Síntesis de estrategia metodológica para el análisis de fuentes de información

	Descripción	Objetivos a los que aporta
Análisis Procedimental	Sistematización de instructivos, operativos y manuales según el uso de cada instrumento. Análisis de contenido a entrevistas y documentación (Hernández Sampieri, et al., 2010).	Cumplimiento del objetivo OE1
Entrevistas	Entrevistas semiestructuradas (Flick, 2007; Canales, 2006) a actores clave de las universidades.	Cumplimiento de los objetivos OE1 y OE3.
Análisis Cualitativo de los instrumentos	Análisis de fuentes de evidencias de validez basada en el contenido, a partir del juicio de expertos (Sireci & Faulkner-Bond, 2014; McCoach, et al., 2013).	Cumplimiento del objetivo OE2.
Análisis Cuantitativo de los instrumentos	Análisis estadístico descriptivo y psicométrico de los instrumentos. El análisis psicométrico se basó en la Teoría Clásica del Test (Martínez-Arias, 1996).	Cumplimiento del objetivo OE2.

Fuente: Elaboración propia.

⁴ El cumplimiento del OE4 se presenta en un documento anexo titulado "Orientaciones Técnicas para el desarrollo de la Evaluación Diagnóstica inicial en pedagogías".

El análisis holístico de la información obtenida a partir de los análisis descritos en la Tabla 4, permitió cumplir el objetivo específico 4 (OE4), el documento de Orientaciones Técnicas que responden a este objetivo se encuentra en los anexos de investigación.

Es importante señalar que todos estos procesos de recolección de información, tanto primaria como secundaria, cuentan con estrictos protocolos de seguridad y protección de información. En ese sentido, la documentación sometida a revisión se anonimizó y se entregaron declaraciones de confidencialidad y consentimientos informados a cada participante, a fin de asegurar el resguardo de los procesos internos.

Como ha de notarse, un elemento relevante en el análisis son los instrumentos aplicados, su contenido y resultados. Considerando la diversidad de baterías de instrumentos utilizadas por las universidades, se decidió analizar solo los instrumentos que tienen aspectos o dominios comunes entre baterías. En ese sentido, se determinó que los instrumentos de Lenguaje y Matemática son dominios comunes para todas las universidades y a estos instrumentos se les analizó su contenido basándose en juicio experto y se les realizó análisis psicométricos. Por su parte, se encuentran los instrumentos “no disciplinares” que también son aplicados por prácticamente todas las universidades. Estos instrumentos no evalúan una disciplina específica, sino que más bien perfilan a las y los estudiantes respecto de sus estrategias de aprendizaje, creencias, percepciones, entre otros aspectos. Estos instrumentos evalúan temáticas muy diversas, por lo que únicamente se analizó su contenido basándose en juicio de expertos. No fueron analizados cuantitativamente, ya que en muchos casos no tenían una conceptualización clara que permitiera interpretar estos análisis.

Principales resultados de la investigación

Sobre los propósitos de la Evaluación Diagnóstica

En esta sección se presentan los resultados correspondientes al objetivo 3, que se articulan en torno a la falta de propósitos de las evaluaciones, ya que ésta corresponde a la principal necesidad detectada en el estudio.

En términos generales, las visiones institucionales recogidas a partir de las entrevistas dan cuenta de una especial valoración por la evaluación diagnóstica, pues permite identificar aspectos o dominios a reforzar para el desempeño en la educación superior. En ese sentido, estos procesos son parte de un conjunto de iniciativas que las universidades ya venían desarrollando.

Si bien en todas las universidades existe una necesidad por diagnosticar a sus estudiantes, no hay mayor claridad sobre el posible uso de estas evaluaciones, ni sobre la especificidad de aquel diagnóstico en el campo de las pedagogías. Antes de que la Ley 20.903 entrara en vigor, la mayoría de las universidades disponían de diversas baterías de instrumentos que aplicaban a toda la cohorte sin distinciones por carrera. Solo a contar del año 2017 se introducen nuevos instrumentos de caracterización sobre el quehacer docente, pero se mantienen la mayoría de los otros instrumentos.

[EL CPEIP] “(...) han planteado que es deseable que avancemos hacia pruebas

*específicas de acceso, que tengan que ver con la carrera de pedagogía. No pruebas [de habilidades] generales, como las que estamos tomando ahora (...) Han planteado que sería deseable que las instituciones avancen a pruebas que tienen más que ver con, por ejemplo, la disposición en la pedagogía, la vocación pedagógica, o habilidades cognitivas que tienen que ver con el quehacer del profesor.” - **Informante 2, U1***

*“Lo que pasa es que nosotros comenzamos los procesos diagnósticos de manera incipiente el año 2014, entonces esta ley nos aparece después. (...) pero la aplicamos a todos los alumnos que ingresan a la universidad, y ahí están incluidos los alumnos de pedagogía. Entonces nosotros no tenemos esto especialmente dedicado a pedagogía, sino que es para todos.” - **Informante 1, U2***

Debido a lo anterior, las universidades esperan estrechar lazos de trabajo y establecer un marco de colaboración con el CPEIP (Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas), en tanto representante del MINEDUC. Los representantes institucionales ven como natural esta relación dado que este organismo administra la ENDFID, realizada al finalizar los estudios de pregrado. Considerando que la Ley establece que son las universidades quienes administran la evaluación diagnóstica inicial, el rol esperado es más bien de una guía que oriente la realización de las evaluaciones diagnósticas.

Como manera de profundizar en los dominios específicos a contemplar por la evaluación diagnóstica, destacan dos líneas de trabajo. Una es la ya mencionada con el CPEIP, mientras que la otra es la generación de instancias de socialización de experiencias con diversas universidades, así como también la construcción de redes de trabajo y colaboración permanentes.

*“Conseguimos en una oportunidad que vinieran desde el Ministerio a entregar resultados, para compartir los resultados, pero la verdad es que... muy buena disposición, pero poco espacio de concreción. Entonces más que deliberativo o consultivo, son reuniones que uno va a escuchar la resolución que se tomó.” - **Informante 3, U4***

*“Nosotros hemos estado en conversaciones con otras universidades, [una de la región del Maule y otra de Valparaíso], como para ver las pruebas que aplican ellos, y reconsiderar lo nuestro. (...) Ahora, dado que la cosa salió como salió, en el sentido de que claro, tenemos que tener evaluaciones diagnósticas, ustedes se arreglan sobre qué y en cómo, si sería agradable, no sé que lo haga el Ministerio directamente pero sí que propicie instancias donde las distintas universidades pudiesen compartir lo que se ha estado haciendo en este ámbito (...)” - **Informante 1, U5***

Se destaca la importancia de resguardar la autonomía de las universidades en la elaboración de las evaluaciones diagnósticas, lo que necesariamente debe complementarse con la necesidad de mayores orientaciones generales. Se entiende esta necesidad como un marco general en el cual situar las evaluaciones diagnósticas, pero ello en ningún caso debe perder de vista la necesidad de un diagnóstico situado y en sintonía con el contexto local. Este último aspecto es relevado especialmente en el caso de las universidades regionales, quienes cumplen un papel importante en el desarrollo de sus respectivas zonas del país.

“...yo pensaría que lo deseable no sería estandarizar todos los procesos, (...) pienso que las universidades que tengan la libertad de construir sus instrumentos, ojalá, así como lo estábamos haciendo en conjunto con otras, (...) me parece que esa libertad es favorable, porque nos permite recoger realidades que son propias de nuestros

contextos.”- **Informante 1, U5**

A grandes rasgos, se puede inferir que todas las universidades estudiadas han presentado dificultades para la aplicación de estas evaluaciones, principalmente porque las disposiciones de la Ley no entregan orientaciones a este respecto. Dada la premura las universidades respondieron a la contingencia y, en algunos casos, adecuaron instrumentos que ya ocupaban para otros fines, postergando con ello procesos de reflexión y discusión más profundos sobre los fines de la evaluación específica para pedagogías.

Junto con lo anterior, existe un problema de falta de conexión entre los resultados de la evaluación diagnóstica y los usos para el acompañamiento y nivelación de las y los estudiantes. La Ley establece que la evaluación diagnóstica debe informar y reforzar los planes remediales de las universidades, pero de momento eso no se ha logrado. Solo en contadas situaciones, con docentes específicos de ciertas universidades, los resultados de la evaluación diagnóstica han alimentado estrategias para el afianzamiento de aprendizajes. Además, debido al tiempo de procesamiento de resultados, en muchos casos los planes de acompañamiento no pueden ejecutarse en sintonía con los resultados de la evaluación diagnóstica durante el primer semestre.

*“(...) en realidad la nivelación no se hace en base a los resultados de la evaluación diagnóstica, si no que se nivela a todos por igual, todos están obligados a asistir a esa nivelación, independientemente de su nivel en ese diagnóstico. Así que yo imagino que esos resultados, todavía no han llegado, que sería lo ideal que fuese así, que la nivelación sea más bien pertinente al nivel de logro que alcanzaron en la nivelación, pero entiendo que eso no ocurrió, porque entre el momento en que se tomó la prueba, y la nivelación, no pasaron más de dos días.” - **Informante 1, U5***

En ese sentido, la creación o adaptación de mecanismos de nivelación y acompañamiento a partir de las evaluaciones diagnósticas se encuentran en diversas etapas de desarrollo en las universidades participantes. La mayoría de las universidades reporta la implementación de talleres de fortalecimiento académico (generalmente en Lenguaje y Matemática), orientado a estudiantes de todas las unidades académicas. Dado que estos talleres se implementan al inicio del periodo académico, ninguno de éstos se diseña en articulación con los resultados de las evaluaciones diagnósticas. Además, al ser algunos de estos talleres dirigidos a toda la cohorte de estudiantes, no permiten discernir claramente acciones específicas para las y los estudiantes de pedagogía, considerando sus características particulares.

Además del diseño de talleres, charlas, y otras actividades específicas, las universidades participantes discuten la posibilidad de entregar los resultados de las evaluaciones diagnósticas a las unidades académicas y docentes como insumo para el diseño o ajuste curricular y de las estrategias pedagógicas de cada curso. La limitante identificada para lograr este proceso es justamente la dificultad de obtener los resultados de forma oportuna.

En síntesis, en las universidades analizadas no se explicitan los usos de los instrumentos. Dada esta falta de claridad sobre los usos para las evaluaciones diagnósticas, es difícil determinar si estas evaluaciones, ya sea en relación con factores institucionales, operativos o técnicos (e.g. confiabilidad de los instrumentos), contribuyen al cumplimiento del propósito de la Ley, que es mejorar y fortalecer la FID.

Además, esta indeterminación sobre los usos explícitos y acciones derivadas a partir de los resultados de las evaluaciones diagnósticas dificulta elaborar argumentos consistentes respecto de su validez y una retroalimentación apropiada de los instrumentos mismos.

Sobre los procedimientos

En esta sección se presentan los resultados relacionados con el objetivo 1, respecto a los fundamentos teóricos y los aspectos técnicos y operativos de las evaluaciones diagnósticas.

Proceso de elaboración de ítems

Las evaluaciones diagnósticas se implementaron con relativa premura, lo que dada la diversidad de experiencias previas de diagnóstico resultó en procesos marcadamente disímiles. Mientras que las universidades que ya aplicaban evaluaciones diagnósticas a estudiantes de primer año asumieron las nuevas exigencias con relativa fluidez, adaptando los procesos que ya desarrollaban para cumplir con las demandas de la Ley, otras universidades debieron diseñar, elaborar, aplicar y analizar instrumentos de evaluación diagnóstica a través de procesos totalmente nuevos.

Como aspecto general, se observa que la mayoría de las universidades no presentó marcos de evaluación claramente constituidos, entendiéndose estos últimos como el núcleo articulador de un proceso de evaluación. A su vez, la referencia a marcos teóricos o conceptuales resulta general y no siempre coherente, en algunos casos mezclando diversas perspectivas bajo los mismos instrumentos. La definición de los dominios a evaluar se mantiene en descripciones generales. Lo mismo ocurrió en el caso de las tablas de especificaciones de cada instrumento, ya que no todas las universidades contaban con esta desagregación de la información, y en aquellos casos donde estaba presente, de igual forma no era exhaustiva si se considera el set completo de la batería de instrumentos.

El proceso de elaboración de ítems difiere entre las universidades si se consideran los diversos modos de distribuir tareas entre las y los profesionales asignados, el grado de centralización de los procesos y la formalización de los protocolos para la elaboración de los instrumentos. En cuanto a quienes construyen los ítems, todas las instituciones remarcan el perfil experto de quienes se encargan de la elaboración de instrumentos.

En términos generar evidencias de validez, las universidades dan cuenta de diversos criterios y distintos formatos de pilotaje. Los más frecuentes son el criterio de validación externa⁵, panel de expertos, pruebas estadísticas y psicométricas posteriores y aplicaciones piloto previas a la aplicación oficial. Destaca el caso de la U2 como una institución que, fruto de la experiencia acumulada, ha logrado consolidar procedimientos de validación rigurosos basado en pilotajes y juicio de expertos.

5 Cabe hacer un llamado de atención: la validez, dentro de sus elementos fundantes, contiene ciertas condiciones como las características de la población y el uso posterior que se le quiere dar al instrumento. En ese contexto, un instrumento con evidencias de validez generadas a nivel externo puede no haber considerado los factores contextuales que dan forma a la evaluación actual, por lo tanto, las acciones y conclusiones deben interpretarse con cautela.

Batería de instrumentos de Evaluación Diagnóstica

Una primera observación general es que las universidades de mayor antigüedad en procesos de diagnóstico cuentan con un número mayor de evaluaciones, enfocadas a temáticas más amplias. Por el contrario, las universidades de proyectos más recientes tienden a focalizarse en evaluaciones de Lenguaje, Matemáticas, y evaluaciones no disciplinares. Tras la revisión preliminar de los instrumentos de evaluación diagnóstica, estos se pueden clasificar en tres grupos y se presentan con detalle en la Tabla 5. **Error! No se encuentra el origen de la referencia.** Como se comentó, el foco de análisis del Proyecto está en los instrumentos de Matemática, Lenguaje y “No disciplinares” al ser dominios comunes.

Tabla 5. Total de instrumentos aplicados por cada Institución de Educación Superior.

U.	Tº de respuesta (mins.)	Instr. de Leng. y/o Mate.	Instr. “no disciplinares”	Instr. de otras disciplinas	Total de instr. aplicados
1	No definido debido a la corta extensión de las pruebas.	2	1	0	3
2	60', excepto Comp. Lectora: 90'	3	1	3 (Biología, Física y Química)	7
3	90'	3	1	7 (Biología, Física, Química, Historia, Filosofía, Música y Educación Física)	11
4	90', excepto Escritura: 30'	6	2	3 (Historia, inglés y Educación Física)	11
5	120' para Matemática y 90' para Escritura	2	0	0	2

Fuente: Elaboración propia.

En cuanto a los referentes conceptuales que las universidades tomaron para el desarrollo de las evaluaciones diagnósticas, estas incluyeron experiencias internacionales como, por ejemplo, PISA y líneas académicas específicas de la Universidad de Queen y Sherbrooke, y nacionales, como el currículum escolar y los estándares de formación específicos de cada disciplina.

En cuanto a la corrección, la U2, U4 y U5 describen brevemente el formato de la evaluación de Producción Escrita, a través de una rúbrica. Destaca el caso de la U2, cuya rúbrica fue elaborada por un equipo especializado en lingüística y la corrección estuvo a cargo de estudiantes del área de cuarto año. Respecto a la asignación de puntajes, todas las universidades reportan utilizar formatos de puntuación que otorgan un punto a cada respuesta correcta para los ítems de selección múltiple. Los tipos de ítems de selección múltiple se muestran en la

Tabla 6 a continuación.

Tabla 6. Tipos de ítems utilizados

U.	Matemática	Lenguaje	"No disciplinares"
1	Selección múltiple de cinco alternativas	Selección múltiple de cuatro alternativas	Escalas de acuerdo graduando entre "+" y "-"
2	Selección múltiple de cinco alternativas	Selección múltiple de cinco alternativas	Verdadero/Falso
3	Selección múltiple de cuatro alternativas, donde la cuarta es "No puedo responder"	Selección múltiple de cuatro alternativas	<ul style="list-style-type: none"> Preguntas de cinco puntos en escala de acuerdo Selección múltiple de cuatro alternativas, donde la cuarta es "No puedo responder"
4	Selección múltiple de cuatro alternativas	Selección múltiple de cuatro alternativas	Enunciados en escala Likert de cuatro puntos, en grado de frecuencia, y en grado de acuerdo, con la opción "No tengo opinión"
5	<ul style="list-style-type: none"> Selección múltiple de cuatro alternativas Verdadero/Falso Sí/No Preguntas de respuesta breve 		

Fuente: Elaboración propia.

Procesos de aplicación

En general, los tiempos determinados para la elaboración de los instrumentos varían ampliamente entre las universidades, mientras que la planificación logística y acciones de preparación del material para las evaluaciones se concentran en los últimos meses del año académico anterior a su aplicación (Tabla 7).

Tabla 7. Organización logística de las aplicaciones

U1	U2	U3	U4	U5
<ul style="list-style-type: none"> Noviembre-diciembre: Inicio proceso logístico para habilitar laboratorios y plataforma próximo año Marzo: Aplicación 	<ul style="list-style-type: none"> Abril-mayo: Inicio proceso: validación, prueba de constructo e implementación en plataforma digital Noviembre-diciembre: Planificación logística para aplicación próximo año Marzo: Aplicación 	<ul style="list-style-type: none"> Marzo-abril-mayo: Elaboración pruebas y carga a la plataforma Junio: Aplicación 	<ul style="list-style-type: none"> Enero: Revisión y validación Febrero: Impresión, mecanización y envío a sedes Marzo: Aplicación 	<ul style="list-style-type: none"> Enero: Preparación logística por institución externa Marzo: Aplicación

Fuente: Elaboración propia.

La estructura organizacional sobre la que las universidades llevaron a cabo el desafío de implementar las evaluaciones diagnósticas permite comprender el contexto en el cual se desarrollaron los instrumentos en cada universidad. En este sentido, la marcada diferencia entre la experiencia de las universidades según su tradición en evaluación diagnóstica, vinculada a estructuras organizacionales más o menos estables, manifiesta diferencias en las características generales y grado de desarrollo técnico de los instrumentos de cada universidad.

Por un lado, a las universidades que ya contaban con procesos de evaluación diagnóstica antes de la promulgación de la Ley, les bastó con adaptar los procesos institucionales que se encontraban alojados en ciertas unidades específicas de la universidad o de responsabilidad de los mismos actores. A su vez, y por la misma razón, estas universidades cuentan con departamentos centralizados con la responsabilidad de implementar pruebas de diagnóstico, como es el caso de la U1 y U2. Por el contrario, las universidades que no contaban con procesos de diagnóstico tomaron diversas estrategias para enfrentar la contingencia. Para estas universidades las evaluaciones diagnósticas se desarrollaron a través de equipos pequeños, alojados dentro de las facultades de educación, con una capacidad menor de implementación en comparación a las universidades con experiencia. El caso excepcional es el de la U5 que externalizó parte de la administración de la evaluación diagnóstica.

Con respecto a la formalización de los protocolos para la aplicación y análisis de las evaluaciones diagnósticas, ninguno de los casos cuenta con un documento único. En su mayoría, las universidades trabajan con documentos internos que describen algunas áreas concretas del proceso, los cuales no siempre son socializados con las personas relevantes en la aplicación de las evaluaciones. Por esta razón, la coordinación se basa en instrucciones y comunicaciones específicas entre los involucrados dentro de las unidades institucionales encargadas.

Información a estudiantes

Otro aspecto importante para las universidades es que las y los estudiantes tengan información respecto de las etapas de aplicación de los instrumentos. Un buen proceso informativo permite comprometer a la población evaluada con la instancia facilitando que las acciones derivadas a partir de los resultados en la evaluación diagnóstica sean consensuadas y beneficien a todas las partes interesadas (estudiantes, docentes, universidades). En este sentido, para todos los casos analizados, las y los estudiantes son citados para rendir las evaluaciones diagnósticas e informados acerca de sus características y propósitos durante los procesos de matrícula y/o durante la primera semana del calendario académico, el cual es el periodo definido por todas las universidades para su realización.

En general las universidades informan personalmente o vía correo electrónico sobre el objetivo de las evaluaciones y sus períodos de aplicación. Además, se emplea la estrategia de enmarcar la semana de bienvenida y la aplicación de pruebas diagnósticas como las primeras actividades académicas para aumentar la participación y compromiso de las y los estudiantes.

Equipo, formato y lugares de aplicación

En general, y como se observa en la Tabla 8 el personal de aplicación de las evaluaciones diagnósticas tiene un perfil similar en todas las universidades, siendo académicos, directivos, o estudiantes. En cuanto a las tareas asignadas, estas se relacionan con el registro de información, resolución de dudas, mantenimiento del orden, y ser capaces de resolver situaciones que se presenten en el momento. También existen diversas instancias de capacitación del personal de aplicación.

Tabla 8. Aspectos relativos a los equipos de aplicación

U.	Personal de aplicación	Funciones	Capacitación
1	Miembros del equipo encargado del proceso de evaluación diagnóstica, apoyados por 2 o 3 estudiantes colaboradores por sala	Los profesionales del equipo están a cargo de entre 2 y 4 laboratorios, mientras que cada laboratorio está a cargo de un estudiante	Sólo estudiantes colaboradores(as) reciben una capacitación una semana antes sobre el proceso general y el uso de la plataforma virtual
2	Examinadores definidos autónomamente por cada Jefe de Carrera	Los examinadores deben hacerse cargo de eventualidades relacionadas con el material de aplicación	Se entregan instrucciones de la aplicación a través de un correo electrónico
3	Examinadores definidos por la Dirección de Asuntos Estudiantiles, que contrata estudiantes de último año	Los examinadores preparan los laboratorios, monitorean el avance, evitan que se dejen preguntas sin contestar, controlar los tiempos de evaluación, y resolver problemas de conectividad	Las y los estudiantes contratados reciben una capacitación por la Dirección de Asuntos Estudiantiles
4	Examinadores son docentes de cada Unidad Académica, privilegiando aquellos con jornada completa	Los encargados deben resolver dudas y asegurar el correcto registro de los datos de identificación	Se capacita a las y los académicos en los Consejos de Facultad, Escuela y carrera
5	Coordinadores de la institución externa y examinadores son académicos(as) de las distintas Unidades	Los docentes encargados de la sala deben llevar un control del tiempo y registrar en formularios la asistencia de las y los estudiantes	La institución externa se encarga de entregar las instrucciones operativas al personal académico

Fuente: Elaboración propia.

En cuanto al formato y lugares de aplicación, U1, U2 y U3 utilizaron un formato digital, argumentando la facilidad que este formato permite para manejar datos de una gran cantidad de estudiantes, junto con la velocidad que permite otorgar a los procesos de análisis. Las dos primeras dispusieron de laboratorios de computación y la U3 habilitó una sala de clases con tabletas digitales.

Por otro lado, las U4 y U5 utilizaron un formato de lápiz y papel, debido a las posibilidades permitidas por su infraestructura. La U4 aplicó las pruebas en horario regular de clases en las salas correspondientes. De manera diferente, la U5 agrupó a las y los estudiantes por número de RUT y no por criterios de carrera, disponiendo de un horario simultáneo para su rendición, durante la semana de bienvenida.

Finalmente, las aplicaciones de pruebas deben considerar las diversas características de sus evaluados, en particular de quienes requieren adecuaciones especiales, en virtud de una evaluación equitativa (AERA; APA; NCME, 2014). En ese sentido, las U1 y U2 reportan adecuaciones de los instrumentos para estudiantes con necesidades educativas especiales o en situación de discapacidad, a partir de políticas institucionales hacia personas sordas y/o con dificultades visuales. Las universidades restantes han realizado ciertos ajustes formulados a partir de contingencias.

Análisis de resultados

Quienes son responsables de aplicar las evaluaciones diagnósticas y analizar sus resultados son, en tres casos (U3, U4 y U5), las mismas unidades institucionales que se encargan del diseño y elaboración de los instrumentos. En el caso de las U1 y U2, el análisis de algunos instrumentos depende de organismos distintos dentro de la misma universidad, pudiendo ser facultades u otros departamentos institucionales.

Respecto del análisis y producción de reportes, estos varían de acuerdo con cada universidad. Destaca la U2, la cual desarrolló procesos de análisis y cruces de información automática, lo que ha agilizado la producción de informes. Esta prontitud en reportar la información puede, eventualmente, favorecer la elaboración de planes de nivelación adecuados y oportunos durante el primer semestre de estudios, pero de momento esto no ocurre en ninguna de las universidades analizadas. En otros casos, la información llega a mediados de año, lo que eventualmente dificulta la vinculación de los resultados con las estrategias de nivelación y acompañamiento, que se realizan usualmente al inicio del año académico.

Sobre los dominios evaluados y las fuentes de evidencia de validez

Como se indicaba en el apartado anterior, se observaron tres “dominios comunes” por evaluar dentro de los procesos analizados, cada uno con instrumentos particulares. Estos se analizaron considerando las diversas fuentes de evidencia de validez, en particular aspectos de contenido y de estructura interna, según fuera el caso. Estos análisis corresponden al cumplimiento del objetivo 2.

Instrumentos no disciplinares

En primer lugar, están los instrumentos que miden los aspectos “**no disciplinares**”, presentes en cuatro universidades de este Proyecto. Si bien son importantes, sus resultados deben tratarse con cautela, debido a la complejidad de su medición (Zwick, 2017). En efecto, desde la evidencia analizada surge principalmente un inconveniente en torno a su diseño: la deseabilidad social, o un sesgo en los ítems que son respondidos desde lo socialmente esperado y aceptado (An & Carr, 2017). En ese sentido, el juicio experto recomienda a quienes utilizan estos instrumentos tener en consideración la deseabilidad social, midiéndola y controlando su influencia. No obstante, respecto del diseño de estos instrumentos, se valora positivamente la inclusión de temáticas y contextos cercanos, relacionados con temas académicos.

*“En términos generales, me parece que los ítems son claros (..) me parece que se usan oraciones relacionadas con contextos cercanos para los estudiantes y que se relacionan de manera simple con temas académicos. Sólo me preocupa que igual varios de ellos se podrían responder desde la deseabilidad social.” - **Revisora caracterización 2***

La revisión del contenido de estos instrumentos muestra dos grandes temáticas sometidas a medición, las cuales deben utilizarse complementariamente, pues la información que entregan cada una por sí sola es insuficiente para la toma de decisiones y las acciones propuestas por la Ley. Por un lado, un área temática evaluada incluye metacognición, estilos y estrategias de aprendizaje (que llamamos aquí “habilidades transversales relacionadas con el aprendizaje”); permitiendo conocer como estudian y aprenden las y los estudiantes. Sin embargo, es difícil tomar decisiones en torno a estilos y estrategias de aprendizaje, ya que no siguen trayectorias comunes y no necesariamente se relacionan con un aprendizaje eficiente (Kirschner, 2017; Delgado, Martínez, & Ponce, 2017). Por otro lado, se observa la medición de las temáticas docentes, cuya evaluación se justifica en la necesidad de

que el profesorado se adapte a las transformaciones del sistema educacional y desarrolle competencias profesionales fuertes (Selvi, 2010) y que evalúa aspectos afectivos y motivacionales implicados en el ejercicio de la enseñanza. Bajo ese contexto, destaca la medición de creencias pedagógicas, que influyen sobre las prácticas docentes (Tagle, Díaz, Alarcón, Quintana, Ramos, & Etchegaray, 2017) e impactan en el desempeño y aprendizaje de sus estudiantes (Solis, 2015). Respecto a la construcción de estos instrumentos, se observaron como espacios de mejora: el desarrollo de marcos teóricos más actualizados y precisos, mejorar la traducción de ítems de instrumentos extranjeros y utilizar varios ítems por dimensión.

Lectura y Escritura

En segundo lugar, se encuentra la evaluación en **Lenguaje y Comunicación**, materializada en instrumentos de Producción Escrita y Comprensión Lectora. A nivel general, se observa PISA, -en parte- la PSU y las Bases Curriculares del subsector de Lenguaje y Comunicación, como los referentes que articulan estas evaluaciones. Tras la conversación con los representantes institucionales, queda en cuestión si efectivamente las Bases son su referente o, si estas exponen el estado del arte más actualizado y consensuado en torno a la lectura y escritura, lo que también es seguido por las universidades que desarrollan instrumentos de este tipo.

Respecto a los resultados descriptivos en las pruebas de producción escrita, se contó con datos para dos instituciones. Los resultados se presentan en la Tabla 9. En ambas pruebas las y los estudiantes se concentran en los niveles intermedios de desempeño y la proporción de estudiantes en el nivel de desempeño más alto, es superior a la del nivel más bajo. Para la U5, los resultados según sexo muestran que las mujeres se concentran en los dos niveles de desempeño más altos, mientras que los hombres se concentran en los niveles 2 y 3.

Tabla 9. Síntesis de resultados en Producción Escrita, por sexo.

U.	Instrumento	Estudiantes que rinden	Nivel Desempeño	General	Sexo	
					Mujeres	Hombres
4	Instrumento General	407	1	1.2%	No disponible	
			2	28.3%		
			3	59.2%		
			4	11.3%		
5	Instrumento General	143	1	4.1%	3.1%	6.4%
			2	32.2%	34.4%	29.8%
			3	48.6%	46.9%	55.3%
			4	13.0%	15.6%	8.5%

Fuente: Elaboración propia.

Del análisis de contenido realizado por expertos, los cuales revisaron los instrumentos de las tres instituciones que aplicaban evaluaciones de escritura, se sostiene que la escritura, en tanto constructo, aborda una variedad de propósitos y audiencias, como también una diversidad de habilidades cognitivas y estructuras de conocimiento, por lo que no es de extrañar que las pruebas de escritura no evalúen la gama completa del constructo (Attali & Power, 2008). En ese contexto, dentro de los instrumentos

analizados los revisores destacan la inclusión de la temática del quehacer docente, enmarcada en la evaluación de la escritura. A su vez, los instrumentos analizados buscan la construcción de textos altamente relacionados con la vida académica: argumentativos y expositivos. Los textos argumentativos son propios de la vida académica, donde se requiere la toma de posición dentro de una negociación colaborativa de significados o puntos de vista en los procesos de construcción de conocimientos disciplinares (Bañales, Vega, Araujo, & Reyna, 2015). Por su parte, los textos expositivos muestran como la escritura se despliega en función de los géneros discursivos y, a la vez, se realiza en situaciones comunicativas organizadas de acuerdo con propósitos comunicativos específicos y contextualizados (Veal & Hudson, 1983).

Las tareas de escritura revisadas presentaban rasgos claros de contextualización, lo cual es clave para conceptualizar la escritura como un proceso de resolución de problemas comunicativos y sociales. Dentro de este aspecto, la revisión destaca las temáticas evaluadas, principalmente las que se relacionan con el quehacer docente, en tanto permiten conocer por parte de las y los estudiantes las creencias sobre la pedagogía, su rol en sociedad y las actitudes que deben tener el profesorado. En ese sentido, las tareas de escritura son afines al modelo situado: cuando un tema despierta interés, el hecho de escribir se convierte en un proceso de reflexión que permite transformar el conocimiento.

La escritura, al tener un formato de respuesta abierta, requiere que los criterios de evaluación sean claros y explícitos. La revisión experta destacó la influencia de la buena calidad de las instrucciones y su estrecha relación con la rúbrica de corrección, que contiene los criterios y requisitos formales de evaluación. Las experiencias analizadas presentan dos rúbricas analíticas y una holística para la corrección, donde la elección de una u otra estrategia dependía de los propósitos definidos por cada institución. Desde las experiencias analizadas, las y los revisores advierten que hay elementos generales de las rúbricas que aseguran la calidad de los resultados, permitiendo trabajar elementos más específicos de las habilidades de escritura: criterios graduados con precisión, la descripción objetiva de los niveles de desempeño y rasgos descriptivos para diferenciar correctamente los niveles. No obstante, y tal como plantea Marinkovich (2002) debe mantenerse una estructura que resguarde la flexibilidad, ya que las etapas de composición de la escritura son interactivas y simultáneas.

“En general los niveles pueden distinguirse de manera objetiva, ya que describen y ejemplifican detalladamente las evidencias que debieran aparecer en el texto creado por los estudiantes evaluados. No obstante, existen estas ambigüedades que generan que un evaluador tenga que tomar decisiones subjetivas, como es el caso de un texto que se salga del género solicitado (...) pero que cumple de una forma lateral tanto las características de niveles altos como el objetivo discursivo declarado. Como dijimos, esto podría suscitar decisiones arbitrarias.” - Revisor lenguaje 4

Por otro lado, los resultados descriptivos de las pruebas de Comprensión Lectora se presentan en la Tabla 10. Respecto a los resultados descriptivos en las pruebas de comprensión lectora, en ninguno de los instrumentos se alcanza el puntaje máximo posible. Además, en la mayoría de los instrumentos el promedio de ítems correctos se encuentra bajo la mitad del total de preguntas. Respecto a los patrones por sexo, en 6

de 9 instrumentos se observa que las mujeres obtienen, en promedio, más preguntas correctas que los hombres.

Tabla 10. Síntesis de resultados en Comprensión Lectora, por sexo.

U.	Instrumento	Cantidad de ítems	Estudiantes que rinden	Media General	D. E. General	Sexo		Diferencia Mujeres – Hombres
						Mujeres	Hombres	
1	Instrumento General	12	852	5.8	1.7	64.0%	36.0%	0.2
2	Instrumento General	50	660	21.3	7.4	63.1%	36.9%	-0.6
3	Instrumento Específico (PEL, PEI)	30	143	19.6	3.5	70.7%	29.3%	0.2
	Instrumento Específico (PEB, PEP, PED)	30	141	14.2	2.6	94.7%	5.3%	-1.0
4	Instrumento General	24	405	12.8	2.6	77.4%	22.6%	0.1
	Instrumento Específico (PEL)	36	13	16.9	2.1	76.9%	23.1%	1.6
	Instrumento específico (PEP)	13	14	5.1	1.7	100.0%	0.0%	--
	Instrumento específico (PED)	19	177	8.9	2.3	96.8%	3.2%	2.1
	Instrumento específico (PEB)	20	27	10.2	2.0	96.0%	4.0%	2.2

(*) La diferencia es estadísticamente significativa ($p < 0.05$).

Fuente: Elaboración propia.

Del análisis de contenido realizado por expertos se identifican distintas estrategias de evaluación, las que se corresponden con los propósitos que tienen las universidades en torno a la interpretación de la información. En ese contexto, se presentan instrumentos aplicados a toda la cohorte, según nivel (educación de párvulos, diferencial y básica) y especialidad (pedagogía en Lenguaje).

En este dominio, las conclusiones emanan de la integración del análisis de contenido y el psicométrico. A partir del análisis de contenido realizado por las y los expertos se destacan los instrumentos que incorporan textos diversos y no solo literarios, pues la diversidad de temáticas, tipologías textuales y selecciones léxicas, entre otros aspectos, influyen en la complejidad de la lectura y, por tanto, contribuyen a una mayor gradación de las tareas a evaluar. Sin embargo, se enfatizó desde la revisión disciplinar, que en general faltaban ítems para evaluar habilidades de orden superior, pues en los instrumentos analizados predominaban las preguntas inferenciales y de comprensión literal. Se sugiere la inclusión de preguntas de evaluación y transferencia de información (Neira, Reyes, & Riffo, 2015), donde se observe un sentido crítico en torno a la forma y el contenido de los textos, o la capacidad de re-contextualizar

información, utilizar la información para resolver problemas y realizar otras actividades de naturaleza reflexiva.

“(...) destaco la inclusión de una pregunta de inferencia global, cuestión relevante para la obtención de información y que también debiera estar más equilibrada dentro del instrumento. En comprensión lectora se acepta que existen 3 niveles: explícito, implícito (local y global) y evaluativo. El tercero de estos no está representado en éste ni en los instrumentos anteriores.” - Revisor lenguaje 4

Respecto del análisis psicométrico, este mostró que 5 de los 6 instrumentos analizados no cumplían con el supuesto de unidimensionalidad y además presentaban bajas confiabilidades⁶, a excepción del instrumento general de la U2 (ver Tabla 11).

Tabla 11. Confiabilidades de los instrumentos de Comprensión Lectora

U.	Instrumento	Ítems	Confiabilidad (α)	Intervalo de Confianza
1	General Lenguaje	12	0.204	(-0.59, 0.20)
2	General Lenguaje	50	0.810	(0.73, 0.81)
3	Específico Lenguaje	30	0.450	(0.13, 0.45)
	Específico Lenguaje	30	0.100	(-0.46, 0.10)
4	General Lenguaje	24	0.331	(-0.11, 0.33)
	Específico Lenguaje	19	0.280	(-0.26, 0.28)

Fuente: Elaboración propia

Esto puede explicarse por una representación inadecuada del constructo, entre otros problemas puntuales de construcción, como falta de claridad en la redacción. Por ejemplo, una universidad incluyó ítems en los que además de comprensión lectora evaluaba el desarrollo de procedimientos matemáticos, los cuales interferían en la medición del constructo. Otra universidad presentaba ítems que se podían resolver con conocimientos previos, que dependían fuertemente de la cobertura curricular de sus establecimientos de origen y/o de contenidos que deberían ser impartidos por la universidad en la FID. Además, el único instrumento que tuvo buena confiabilidad fue el que tenía mayor cantidad de ítems.

Además, se observan varios ítems con discriminación nula, es decir, que no entregan mucha información sobre los diferentes niveles de habilidades del estudiantado (Tabla 12). Estos ítems disminuyen la confiabilidad de las pruebas y en la mayoría de los instrumentos se observó una relación entre los ítems con muy baja discriminación, con problemas importantes de construcción (e.g. clave errónea o redacción confusa).

Tabla 12. Dificultades y discriminaciones de instrumentos de Comprensión Lectora

U.	Instrumento	Dificultad muy Baja*	Dificultad muy Alta**	Disc. Nula***	Disc. Negativa
1	General Lenguaje	0	1	6	5
2	General Lenguaje	0	1	7	5
3	Específico Lenguaje	4	0	17	3
	Específico Lenguaje	1	6	13	12
4	General Lenguaje	1	1	16	2
	Específico Lenguaje	0	2	14	3

⁶ La confiabilidad se estimó con el coeficiente alfa de Cronbach. Cabe señalar que cuanto más cerca se encuentre el valor del α a 1, mayor es la consistencia interna de los ítems analizados (el mínimo aceptable para este coeficiente es 0.7) (Celina & Campo, 2005).

*Dificultad muy baja: sobre 90% de respuestas correctas

**Dificultad muy alta: bajo 10% de respuestas correctas

***Discriminación nula: $-0.1 < d < 0.1$

Fuente: *Elaboración propia*

Matemática

Los estadísticos descriptivos de los instrumentos de matemática se presentan en la Tabla 13. En ninguno de los instrumentos se alcanza el puntaje máximo posible y en la totalidad de los instrumentos el promedio de ítems correctos se encuentra bajo la mitad del total de preguntas. Respecto a diferencias por sexo, en 5 de 7 instrumentos se observa que los hombres obtienen, en promedio más preguntas correctas que las mujeres. En este caso las diferencias de rendimiento entre hombres y mujeres son estadísticamente significativas en los instrumentos de la U2 y la U3, en desmedro de las estudiantes mujeres.

Tabla 13. *Síntesis de resultados en Matemática, por sexo.*

U.	Matemática	Cantidad de ítems	Estudiantes que rinden	Media General	D.E. General	Sexo		Diferencia Mujeres – Hombres
						Mujeres	Hombres	
1	Instrumento General	12	852	5.6	2.3	64.0%	36.0%	-1.2
2	Instrumento General	30	617	9.6	4.7	60.9%	39.1%	-1.2*
3	Instrumento Específico (PEM)	30	61	14.7	3.5	48.2%	51.8%	-2.0*
4	Instrumento Específico (PEP)	oct-40	14	4.2	1.5	100.0%	0.0%	--
	Instrumento Específico (PED)	oct-53	177	3.1	1.5	96.8%	3.2%	0.9
	Instrumento Específico (PEB)	dic-60	27	5.2	2.5	96.0%	4.0%	-0.8
5	Instrumento General	40	155	14.3	3.7	69.0%	31.0%	-0.7

(*) *La diferencia es estadísticamente significativa ($p < 0.05$).*

Fuente: *Elaboración propia.*

En cuanto al análisis integrado de contenido y psicométricos en el caso de los instrumentos de matemática, se observan diversas estrategias de evaluación según las interpretaciones y los usos dados por las universidades. En ese contexto, se presentan instrumentos aplicados a toda la cohorte, según nivel (educación de párvulos, diferencial y básica) y especialidad (pedagogía en Matemática). Se observa la medición de contenidos y competencias matemáticas y la referencia a las “grandes áreas del currículum nacional”, (MINEDUC, 2015) como comunes y relevantes de evaluar para las universidades analizadas. Otro referente observado es la PSU, que es discutible, puesto que sus ítems fueron diseñados con propósitos de selección universitaria. Si las universidades plantean la necesidad de levantar sus propios datos,

sería provechoso que estos generaran evidencia adicional o complementaria a la producida por los actuales instrumentos de evaluación.

En los casos que corresponde, se destaca la inclusión de indicadores observables de los estándares de aprendizaje de educación básica y las habilidades disciplinares propuestas en el currículum. No obstante, en general, en las evaluaciones hay temas fundamentales de la matemática elemental que no están incorporados. También se observa claramente el referente de PISA en la formulación de los objetivos de evaluación de los instrumentos que declaran evaluar competencias y habilidades, como también en ítems de resolución de problemas y razonamiento matemático.

“En general, en las evaluaciones no están incorporados los temas relevantes del currículum hasta sexto año básico. Hay temas fundamentales que no son evaluados como la relación entre la multiplicación y la división; vistas desde distintas perspectivas; el proceso de medir y el uso de distintas unidades de medida; argumentación y justificación de fórmulas como el área y perímetro.” - Revisor matemática 3

En la construcción, en términos generales los revisores señalan el cumplimiento del dominio a evaluar, con ciertas dificultades compartidas que limitan la obtención de información útil para diagnóstico y acompañamiento. En ese sentido, existen problemas de construcción relacionados con las alternativas de respuesta (preguntas con la clave errónea y distractores poco plausibles) y la distribución de ítems en las distintas dimensiones de la prueba.

“(…) debe quedar más claro qué tipo de aproximación se está considerando, esto porque si del gráfico obtengo un valor y luego aproximo por redondeo, o truncamiento, o exceso, etc. pueden dar otros valores.” - Revisor matemática 4

Respecto de los distractores -y sobre todo en procesos de diagnóstico-, se sugiere no incluir opciones como “No sé la respuesta”, y que éstos se determinen por errores comunes en el planteamiento y/o la resolución de los problemas. Esto requiere un trabajo arduo en la construcción respecto de la plausibilidad de los distractores; de lo contrario, no es posible evaluar errores conceptuales o de razonamiento, sino resultados que solo se pueden obtener producto del azar. Además, se sugiere resguardar el balance en los dominios efectivamente reconocibles, mediante suficientes oportunidades para demostrar el conocimiento en cada uno. En algunos casos, hay dimensiones evaluadas (sucede en general con el eje de ‘Probabilidad y Estadística’) con muy pocos reactivos asociados. Para efectos de resultados y toma de decisiones, la información que reflejen estos ítems puede ser incompleta o poco precisa acerca del saber que poseen las y los estudiantes en tal ámbito. Por su parte, el análisis psicométrico confirma empíricamente estos problemas evidenciando bajas confiabilidades. Sin embargo, se destaca que la mayoría de los instrumentos (salvo el de la U4) cumple con el supuesto de unidimensionalidad (ver Tabla 14)

Tabla 14. Confiabilidades de los instrumentos de Matemática

U.	Instrumento	N ítems	Confiabilidad (α)	Intervalo de Confianza
1	General Matemática	12	0.490	(-0.02, 0.49)
2	General Matemática	30	0.740	(0.59, 0.74)
4	Específico Matemática	10	0.091	(-0.92, 0.09)
5	General Matemática	40	0.520	(0.27, 0.52)

Fuente: Elaboración propia

En la revisión de contenido, se evidencia una tensión importante: el desacople entre los instrumentos y las diversas poblaciones que los enfrentan, pudiendo explicar en parte los problemas psicométricos identificados. Se observan preguntas muy difíciles y poco pertinentes para pedagogías no relacionadas con el área científica y matemática, y otras muy sencillas y rutinarias para pedagogías del área.

A nivel psicométrico, la mayoría de los instrumentos de Matemática tiene dificultad apropiada de acuerdo con los criterios estadísticos establecidos. La discriminación, sin embargo, presenta un panorama más complejo, con varios instrumentos con una alta proporción de ítems con discriminación nula o negativa. Como espacio de mejora, se sugiere, definir precisamente el constructo a medir, además de resguardar la calidad en la construcción de ítems (ver Tabla 15).

Tabla 15. Dificultades y discriminaciones de instrumentos de Matemática

U.	Instrumento	Dificultad muy baja*	Dificultad muy alta**	Disc. Nula***	Disc. Negativa
1	General Matemática	0	0	2	1
2	General Matemática	0	0	3	1
4	Específico Matemática	0	0	2	5
5	General Matemática	1	1	7	4

*Dificultad muy baja: sobre 90% de respuestas correctas

**Dificultad muy alta: bajo 10% de respuestas correctas

***Discriminación nula: $-0.1 < d < 0.1$

Fuente: Elaboración propia

Hay acuerdo en la importancia de la evaluación diagnóstica de habilidades y competencias matemáticas, lo que desafía a los programas a concentrar esfuerzos y recursos por involucrar sistemáticamente este trabajo durante la FID. Esta es la forma más significativa para modelar la enseñanza en torno al desarrollo de habilidades: las actuales Bases Curriculares establecen es una parte fundamental en el aprendizaje, por lo tanto, la formación docente debe contemplarla.

Conclusiones generales

Los hallazgos de esta investigación aportan evidencia para la mejora de las evaluaciones diagnósticas en distintos ámbitos y niveles del proceso. Las conclusiones de este trabajo buscan sintetizar y hacer dialogar parte de los análisis aplicados sobre fuentes de información diversas (entrevistas, documentación, instrumentos, bases de datos, entre otras). Estas conclusiones se organizan en torno a los objetivos de investigación de la siguiente manera. Primero, se concluye en relación al objetivo 3 que ofrece aspectos clave para la comprensión del resto de las conclusiones. Luego, se concluye respecto al objetivo 1, que considera los fundamentos teóricos y los aspectos técnico y operativos de las evaluaciones. Luego, sobre el objetivo 2 de manera conjunta para los distintos dominios evaluados. Las conclusiones respecto al objetivo 4⁷, que se refieren a orientaciones técnicas para la mejora de la evaluación diagnóstica, aparecen transversalmente a lo largo de esta sección según

⁷ Este objetivo se cumple a cabalidad en el documento anexo "Orientaciones Técnicas para el desarrollo de la Evaluación Diagnóstica inicial en pedagogías".

correspondan: al propósito de la evaluación (objetivo 3), aspectos técnicos y operativos (objetivo 1) o aspectos psicométricos y de contenido (objetivo 2).

Sobre el propósito de la Evaluación Diagnóstica

En esta sección, se presentan las conclusiones relacionadas con el objetivo 3 respecto a los intereses y necesidades comunes de las universidades y la pertinencia de los instrumentos. Como la principal necesidad identificada fue la falta de un propósito claro para las evaluaciones, las conclusiones se centran en este tema.

Las recientes políticas para mejorar la FID han tenido un impacto positivo porque aseguran la calidad de las instituciones que imparten programas de pedagogía, como también porque hay una mayor preocupación por el perfil del estudiantado que se interesa e ingresa a estos programas. Los antecedentes de este estudio muestran cómo parte de estos cambios ocurridos en la última década se reflejan en mayores restricciones y requisitos para la acreditación de programas y para el ingreso de los candidatos, dando lugar a un panorama cambiante y que requiere de estudios contextualizados sobre aspectos más puntuales de la FID, como es el caso de las nuevas pruebas diagnósticas. En pocas palabras, evaluar a este nuevo estudiantado debe partir por reconocer sus perfiles, necesidades, intereses y capacidades, en sintonía con el desarrollo de competencias docentes que involucren recursos cognitivos, pero que a la vez los pongan en práctica en el desempeño profesional (en el aula), y de esta forma tengan un impacto significativo en el entorno escolar (Blömeke, Gustafsson, & Shavelson, 2015; Blömeke & Kaiser, 2017; Yang, Kaiser, G, König, J, & Blömeke, 2018). Este tipo de competencias ha sido evaluada al principio de la formación inicial en otros países (König, Ligtoet, Klemenz, & Rothland, 2017).

Una de las principales lecciones que se desprenden del estudio, es que resulta necesario establecer objetivos más precisos sobre lo que se busca evaluar con estas pruebas diagnósticas en pedagogía, siendo esta la principal necesidad común a las instituciones identificada en este estudio. En varios casos, estos objetivos fueron establecidos con anterioridad a la promulgación de la Ley y apuntan a la totalidad de la matrícula de la institución y no a estudiantes de pedagogía en particular. Esta falta de claridad sobre los objetivos puede afectar el ejercicio de delimitación de las dimensiones o dominios a evaluar en cada instrumento. Este problema es identificado en los análisis de contenido sobre los instrumentos “no disciplinares”, de Escritura, Comprensión Lectora y Matemática.

Un problema derivado de la falta de objetivos claros es que tampoco se observan procedimientos o políticas de intervención sobre la cohorte de estudiantes a partir de los resultados que informan los instrumentos. Mejorar las evaluaciones diagnósticas es necesario no solamente para asegurar su calidad sino también para que los resultados sirvan de insumo para el acompañamiento y nivelación estudiantil en los planteles. De momento, existe una desconexión evidente entre los resultados de la evaluación diagnóstica y los usos de la información, lo que se presenta como uno de los aspectos más débiles en el conjunto de las universidades consideradas en este estudio. Si bien se puede estar cumpliendo formalmente con la Ley, la poca claridad que hay respecto de los objetivos y los usos de la información pueden limitar los avances esperados en la mejora de la FID a través de la ED.

En este proceso de mejora, se espera un papel más activo por parte de las instituciones estatales, particularmente del CPEIP, órgano especializado en FID y representante del MINEDUC. La Ley en principio no define la forma específica de participación del organismo en la ED, sino que delega aquella responsabilidad de elaboración y administración de la evaluación en las instituciones formadoras. Si bien esto parece una decisión adecuada dado que resguarda la autonomía universitaria y da cabida a los distintos sellos formativos según institución, aparece la necesidad de orientación y de promoción de la colaboración entre universidades.

Sobre las capacidades técnicas para administrar la Evaluación Diagnóstica

En esta sección, se concluye sobre el objetivo 1, que considera los fundamentos teóricos y los aspectos técnico y operativos de las evaluaciones. Esos aspectos son afectados por la falta de definición del propósito de evaluación. Esto se observó en los instrumentos estudiados. De hecho, en la mayoría de los casos no es claro cuáles son los referentes teóricos, ni cómo estos unifican las dimensiones evaluadas por cada instrumento. Establecer estas definiciones constituye uno de los principales desafíos para los equipos constructores de los instrumentos en las universidades. Además, es necesario sofisticar los procesos de validación de los instrumentos. En la mayoría de los casos estudiados estos procesos son aún muy débiles o inexistentes.

Igual de relevante es instalar las capacidades técnicas que permitan disponer de los resultados de la evaluación de forma oportuna. Esto es importante porque informa a todas las partes interesadas a tiempo, pero también porque así se pueden ajustar las estrategias de enseñanza y las estructuras curriculares del primer año en base al diagnóstico efectivo. Para alcanzar este objetivo es deseable que se automaticen los procesos de corrección y producción de reportes de los resultados.

Entre aquellas capacidades técnicas que parecen más necesarias de incorporar en las universidades, destacan 1) la capacidad de generar evidencias de validez para los instrumentos en función de sus propiedades psicométricas y los marcos de evaluación previamente establecidos; 2) la capacidad de procurar y asegurar que los usos previstos para el instrumento se vinculen con los resultados efectivos de la evaluación; 3) la capacidad de administrar las aplicaciones de los instrumentos al tiempo que se aseguran las condiciones de equidad en su rendición para todo el estudiantado; 4) la capacidad de reportar los resultados de forma que resulten funcionales a las acciones de acompañamiento y nivelación; y 5) la capacidad de pilotear periódicamente distintos ítems que contribuyan a la validez y vigencia de los instrumentos en el tiempo.

Lo anterior puede ayudar a evitar que se inviertan recursos en pruebas que poco informan sobre las verdaderas capacidades y dificultades de las y los estudiantes al iniciar sus estudios. Contar con evidencias de validez sobre estos instrumentos no sólo representa una tarea para responder satisfactoriamente a los estándares evaluativos, sino que además vela por la obtención de resultados confiables y que contribuyan a la elaboración de políticas de acompañamiento robustas.

Sobre los instrumentos y las decisiones de evaluación

En esta sección se concluye sobre el objetivo 2, relacionado con el contenido y las características psicométricas de los instrumentos. Respecto a los contenidos evaluados, los instrumentos de Matemática, Lenguaje (Comprensión Lectora y Escritura) y “no disciplinares” se encuentran presente en la mayoría de las universidades y representan los dominios comunes que son evaluados en el contexto de la ED. Tanto los instrumentos de Escritura como los “no disciplinares” son transversales a la matrícula de pedagogía, mientras que los de Matemática y Comprensión Lectora se ajustan dependiendo de las especialidades de los programas. De acuerdo con el alto interés de las universidades y el juicio positivo de los expertos consultados, estos dominios son pertinentes y relevantes de ser evaluados al inicio de la carrera.

En el caso de los instrumentos “no disciplinares”, la experiencia de las universidades indica que su propósito es la caracterización estudiantil, tanto en lo relativo a sus procesos de aprendizaje (habilidades transversales) como en las disposiciones y creencias sobre la docencia y su rol profesional (aspectos afectivos y motivacionales). En estos casos es recomendable disponer de un constructo teórico sólido, además de controlar la deseabilidad social en la medición. Esta información no es suficiente por sí misma para desarrollar planes de mejora. Sin embargo, si puede ser un insumo relevante para diseñar mejores planes de estudio, abordar dificultades asociadas a disciplinas específicas y aportar información para los perfiles de egreso.

Como estos instrumentos apuntan a la caracterización del estudiantado en su ámbito social y emocional, y considerando que los docentes en ejercicio son un modelo que seguir por sus estudiantes, sería relevante diagnosticar y trabajar posibles sesgos desde el inicio de la formación. Un sesgo fundamental a considerar es el de género en la educación. Existen estereotipos de género que permean la sociedad, y en particular afectan a los futuros docentes (Mendoza-Mendoza, & Sanhueza, 2018). Mendoza-Mendoza y Sanhueza (2018) investigaron las percepciones de género de estudiantes de último año de pedagogía chilenos y encontraron que los estudiantes entrevistados conocían sobre las desigualdades de género presentes en la sociedad, sin embargo, reproducían estereotipos y roles de género en sus discursos y no tenían referentes teóricos para explicar los fenómenos relacionados con género. Las autoras mencionan la necesidad de implementar en la formación docente acciones para educar en estos aspectos. De hecho, los estereotipos se reflejan en las prácticas y creencias de los docentes, por ejemplo, en el aula de matemáticas los niños participan más y reciben preguntas más complejas cognitivamente que las niñas (Espinoza, & Taut, 2016) y los docentes chilenos de matemáticas tienen expectativas sobre el rendimiento de sus estudiantes que están sesgadas por el género (Mizala, Martínez, & Martínez, 2015). La necesidad de abordar temáticas de género y diversidad sexual en la formación inicial es reconocida por otros autores (Barrientos, Andrade, y Montenegro 2018, Abett de la Torre, 2014). Evaluar la presencia de sesgos de género al inicio de la formación inicial, para informar a los programas de formación de profesores y para retroalimentar a los estudiantes podría ser una herramienta muy útil para trabajar temas de sesgo de género en la formación inicial docente.

Sobre la evaluación en Lenguaje, en los casos estudiados, se destaca el hecho de que las tareas estén relacionadas con la vida académica, tanto en Comprensión Lectora como en Escritura. En el primer dominio, las tareas se centran en la comprensión de tipologías textuales, estructuras y géneros diversos. Por su parte, en Escritura, se evaluaba la construcción de textos argumentativos y expositivos, indispensables en la construcción de conocimiento disciplinar. En ambos dominios, evaluar con diversidad de textos tiene potencial de mejorar las habilidades escriturales y lectoras de las y los estudiantes, fundamentales para el ejercicio profesional. Se requiere, sin embargo, que los instrumentos aumenten su confiabilidad y capacidad de discriminación, lo que se encuentra estrechamente relacionado al establecimiento de procedimientos rigurosos de construcción y validación.

Finalmente, se encuentran los instrumentos de Matemática. En varias universidades se evaluaron competencias o habilidades matemáticas, lo cual es positivo dado el consenso entre los revisores sobre la importancia de una evaluación que considere tanto habilidades como contenidos. Para eso también se sugiere mejorar el diseño y la construcción del instrumento pensando en la población que lo rinde, dado que distintos programas de pedagogía precisan de conocimientos de distinto nivel. En particular, hay temas curriculares fundamentales hasta sexto básico que no están siendo considerados y que son relevantes en el ejercicio docente en el aula. Vale la pena destacar aquí también el hecho de que estos instrumentos, al menos en dos de los casos aquí analizados, muestran diferencias de rendimiento significativas entre hombres y mujeres, en desmedro de estas últimas. Es muy importante que no haya sesgos de género en los conocimientos y competencias matemáticas porque, desde un punto de vista del desarrollo individual, se espera una equidad en el desarrollo de las habilidades. Además, el nivel de los estudiantes en conocimientos y competencias matemáticas pueden relacionarse con otras variables, que pueden finalmente afectar su aprendizaje durante la formación docente y su futuro desempeño en el aula. Por ejemplo, la ansiedad matemática se relaciona negativamente con el nivel de conocimiento y competencias matemáticas (Ashcraft, 2002). A su vez, el nivel de ansiedad de las profesoras puede afectar el nivel de desempeño de las niñas (Beilock, Gunderson, Ramirez, & Levine, 2010). Este punto supone un gran desafío para las universidades en miras de favorecer la equidad y mitigar los sesgos de género que aún prevalecen en la formación de profesores.

Un referente observado en algunos instrumentos de Comprensión Lectora y Matemática es la PSU, lo que es discutible, pues sus ítems fueron diseñados con propósitos de selección universitaria y no de diagnóstico en pedagogías. Si las universidades plantean la necesidad de levantar sus propios datos, sería provechoso que generaran evidencia adicional o complementaria a la ya producida por los actuales instrumentos estandarizados. Esto claramente se alinea con lo dicho particularmente en la sección del marco conceptual referido a las competencias docentes.

El logro de evaluaciones pertinentes que cumplan con criterios técnicos requiere del desarrollo de capacidades institucionales y un uso considerable de recursos económicos, materiales y humanos. En particular este es un proceso iterativo que debe orientarse al desarrollo de un sistema de diagnóstico coherente con los programas de estudio de pedagogía. De esta manera, las evaluaciones podrán aportar al cumplimiento de las metas de la formación docente. Esto tanto desde el punto de

vista de la normativa nacional como de los perfiles específicos del FID en cada institución.

Los resultados presentados aquí corresponden a la primera aplicación diagnóstica obligatoria por ley. Su principal mérito es que permitió a las instituciones entender mejor los instrumentos que estaban usando, tomar conciencia de la necesidad de perfeccionar sus instrumentos y procesos diagnósticos, generar redes de colaboración entre instituciones y abrirse al diálogo con los organismos estatales involucrados en este proceso. Si bien, en sus versiones actuales, la mayor parte de los instrumentos carecen de suficiente rigurosidad técnica, es necesario relevar el potencial de mejora de estos instrumentos. Una muestra de ello es la institución que lleva al menos 4 años trabajando en el desarrollo de evaluaciones diagnósticas y perfeccionando su sistema. Con un adecuado apoyo de los organismos estatales involucrados y fomentando el trabajo en red, es posible visualizar avances consistentes en el perfeccionamiento de estos sistemas de evaluación diagnóstica.

Limitantes de la investigación

Se identifican tres grandes limitantes en esta investigación. En primer lugar, la muestra de universidades que aportan información no es representativa del sistema de educación superior completo, únicamente caracteriza parte importante de la matrícula en instituciones formadoras que son especialmente relevantes para el país. En ese sentido, sus resultados dan cuenta de una tendencia importante, pero no son generalizables. En segundo lugar, hay otros procesos diagnósticos que estas universidades realizan, que no se reportan al MINEDUC para efectos de acreditación de las pedagogías y no se analizan en este estudio. Por lo tanto, estas conclusiones no hablan de la totalidad de las iniciativas institucionales de la muestra en este ámbito. Finalmente, debido al periodo de ejecución del proyecto, solo se analizan las evaluaciones diagnósticas del año 2017. Cabe destacar que las instituciones participantes han modificado considerablemente algunos procesos e instrumentos de diagnóstico entre 2017 y 2018.

Recomendaciones para la formulación de políticas públicas

Las recomendaciones al sistema pueden elaborarse en tres ámbitos. Por un lado, es necesario que el conjunto de actores e instituciones involucradas definan el rol de la evaluación diagnóstica en la Formación Inicial Docente, de modo tal que la evaluación inicial no sea un requisito aislado, sino que parte de un sistema integral que apunte a mejora en la calidad de la formación de las y los futuros profesores de Chile. Se debe clarificar cuál es el propósito de diagnosticar a las y los estudiantes de pedagogía en particular, más allá de las iniciativas transversales de diagnóstico que actualmente implementan las universidades. Se sugiere en esta línea revisar las contribuciones de distintos académicos abocados al estudio sobre las competencias docentes, algunos ya referenciados en el marco conceptual de esta investigación. Además, se sugiere considerar las necesidades educativas que tiene que abordar el país en términos de calidad y equidad. En particular, un desafío urgente es abordar temas de género y

diversidad sexual en la educación. En este contexto la evaluación diagnóstica podría jugar un rol muy útil, ya que podría informar a los programas sobre las percepciones y conocimiento que tienen los estudiantes de temáticas de género y también informar a los mismos estudiantes sobre sus propias creencias. También, permitiría identificar brechas de género entre el estudiantado que se pueden abordar durante la formación. Un posible camino para avanzar en la definición de los propósitos es que en los criterios que establece la CNA para la acreditación de las pedagogías, se contemple que las universidades expliciten los propósitos de las evaluaciones diagnósticas y las acciones llevadas a cabo a partir de sus resultados.

Un segundo ámbito de interés es la generación y fortalecimiento de las capacidades internas de las universidades para el desarrollo y uso de estas evaluaciones. Por un lado, se necesitan personas expertas en diseñar, implementar y evaluar instrumentos diagnósticos. Por otro, instalar capacidades de procesamiento de datos eficaz, lo que implica gestión y posiblemente automatización, para que los resultados de las evaluaciones se puedan usar a tiempo. Todas las universidades analizadas cuentan con equipos para las acciones de acompañamiento o apoyo, sin embargo, es imperante articular el trabajo de estos equipos con las evaluaciones. Además de esto, hay que procurar que las universidades cuenten con tiempo y recursos para la implementación de las evaluaciones diagnósticas. Esto se evidencia en que, en la muestra considerada, las instituciones con mayor experiencia en evaluación diagnóstica mostraban, en consecuencia, mayor desarrollo y formalización de estos procesos e instrumentos más robustos y articulados con las acciones posteriores.

Finalmente, es también relevante que en esta discusión se considere la diversidad de realidades y contextos de las universidades formadoras de docentes. La evaluación diagnóstica debe alinearse con el sello formativo y el perfil de egreso de cada institución, pero además debe procurar que se resguarde la autonomía universitaria en el diseño y aplicación de estas evaluaciones. Las directrices u orientaciones técnicas que se elaboren desde los organismos responsables deben considerar las visiones de cada una de las instituciones involucradas. Una evaluación situada ayuda a la calidad de los instrumentos y mejores interpretaciones de sus resultados, dado que se ajusta a las características de su población.

Bibliografía

- Abett de la Torre, P. (2014) Educación y género: una reflexión sobre los principios de igualdad y diversidad en la educación chilena en el escenario de la reforma educativa actual. Paulo Freire. *Revista de Pedagogía Crítica*, 16, pp. 35-47.
- AERA; APA; NCME. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- An, D., & Carr, M. (2017). Learning styles theory fails to explain learning and achievement: Recommendations for alternative approaches. *Personality and Individual Differences*, 116, 410-416.
- Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current directions in psychological science*, 11(5), 181-185.

- Attali, Y., & Power, D. (2008). *A Developmental Writing Scale*. Princeton, NJ: Educational Testing Service.
- Ávalos, B. (2002). *Profesores para Chile. Historia de un Proyecto*. Santiago de Chile: Ministerio de Educación.
- Ávalos, B. (2014). La formación inicial docente en Chile: Tensiones entre políticas de apoyo y control. *Estudios Pedagógicos (Valdivia)*, 4(Especial), 11–28.
- Ball, D. L. (2008). Content knowledge for teaching: What makes it special? . *Journal of Teacher Education*, 59(5), 389–407.
- Bañales, G., Vega, N., Araujo, N., & Reyna, A. (2015). La enseñanza de la argumentación escrita en la universidad: una experiencia de intervención con estudiantes de lingüística aplicada. *Revista Mexicana de Investigación Educativa*, 20(66), 879-910.
- Barrientos, P., Andrade, D., y Montenegro C. (2018). La formación docente en género y diversidad sexual: Tareas pendientes. *Cuaderno de Educación. Desarrollo Profesional. Facultad de Educación Universidad Alberto Hurtado*. 81, pp. 1-13. Recuperado de http://mailing.uahurtado.cl/cuadernosdeeducacion/81/documentos/desarrollo_profesional_81.pdf
- BCN. (2016, abril 1). Retrieved marzo 26, 2018, from Ley 20.903 CREA EL SISTEMA DE DESARROLLO PROFESIONAL DOCENTE Y MODIFICA OTRAS NORMAS: <http://bcn.cl/1uzzn>
- Beilock, S., Gunderson, E., Ramírez, G., & Levine, S. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences*, 107(5), 1860-1863.
- Blömeke, S., & Kaiser, G. (2017). Understanding the development of teachers' professional competencies as personally, situationally, and societally determined. In D. J. Husu, *International handbook of research on teacher education* (pp. 783–802). Thousand Oakes, CA: SAGE.
- Blömeke, S., Gustafsson, J. E., & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie*, 25, 3-13.
- Canales, M. (2006). *Metodologías de investigación social. Introducción a los oficios*. Santiago: LOM.
- Celina, H., & Campo, A. (2005). Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría*, 34(4), 572-580.
- Chan, E. (2014). Standards and Guidelines for Validation Practices: Development and Evaluation of Measurement Instruments. In B. Zumbo, & E. Chan (Eds.), *Validity and Validation in Social, Behavioral, and Health Sciences* (pp. 9-27). Springer International Publishing Switzerland.
- CNED. (2016). *Índices 2016. Tendencias Matrícula Educación Superior*. Retrieved from https://www.cned.cl/sites/default/files/presentacion_indices_junio_2016.pdf
- Cox, C., Meckes, L., & Bascopé, M. (2010). La institucionalidad formadora de profesores en Chile en la década del 2000: Velocidad del mercado y parsimonia de las políticas. *Revista Pensamiento Educativo*, 46-47, 205-245.

- Delgado, U., Martínez, F., & Ponce, D. (2017). Análisis crítico de los estilos de aprendizaje desde una perspectiva naturalista. *Revista Iberoamericana de Psicología: Ciencia y Tecnología*, 9(1), 45-52.
- Domínguez, M., Bascopé, M., Carrillo, C., Lorca, E., Olave, G., & Pozo, M. d. (2012). Procesos de acreditación de pedagogías: un estudio del quehacer de las agencias. *Calidad en la Educación*(36), 53-85.
- Elosua, P. (2003). Sobre la validez de los tests. *Psicothema*, 15(2), 315-321.
- Espinoza, A. M., & Taut, S. (2016). El rol del género en las interacciones pedagógicas de aulas de matemática chilenas. *Psykhé (Santiago)*, 25(2), 1-18.
- Flick, U. (2007). *Introducción a la investigación cualitativa*. Madrid: Morata Paideia.
- Freeman, D., Katz, A., Gomez, P. A., & Burns, A. (2015). English-for-Teaching: Rethinking teacher proficiency in the classroom. *ELT Journal*, 69(1), 129–139.
- Guerriero, S. (2017). *Pedagogical knowledge and the changing nature of the teaching profession*. París: OECD.
- Hanushek, E., & Rivkin, S. (2012). The distribution of teacher quality and implications for policy. *Annual Review of Economics*, 4, 131-157.
- Hattie, J. (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. New York: Routledge.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2010). *Metodología de la investigación* (Quinta ed.). México D.F: Mc Graw Hill Educación .
- Ingvarson, L. (2013). Estándares de egreso y certificación inicial docente: la experiencia internacional. *Calidad en la educación n° 38*, 21-77.
- Ingvarson, L., & Rowley, G. (2017). Quality Assurance in Teacher Education and Outcomes: A Study of 17 Countries. *Educational Researcher*, 46(4), 177–193.
- Johansson, S., Strietholt, R., Rosén, M., & Myrberg, E. (2014). Valid Inferences of Teachers' Judgements of Pupils' Reading Literacy: Does Formal Teacher Competence Matter? *School Effectiveness and School Improvement*, 25(3), 394-407.
- Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Kirschner, P. (2017). Stop propagating the learning styles myth. *Computers & Education*, 106, 166-171.
- König, J., Ligtoet, R., Klemenz, S., & Rothland, M. (2017). Effects of opportunities to learn in teacher preparation on future teachers' general pedagogical knowledge: Analyzing program characteristics and outcomes. *Studies in Educational Evaluation*(53), 122-133.
- Lauzano, P., & Moriconi, G. (2014). Visión de la docencia y características de los sistemas de formación docente. In U. OREALC, *Temas críticos para formular nuevas políticas docentes en América Latina y el Caribe: el debate actual*. (pp. 10-52). Santiago: Centro de Estudios de Políticas y Prácticas en Educación (CEPPE), © UNESCO.

- Lloyd, G. M. (2008). Curriculum use while learning to teach: One student teacher's appropriation of mathematics curriculum materials. *Journal for Research in Mathematics Education*, 39(1), 63-94.
- Marinkovich, J. (2002). Enfoques de proceso en la producción de textos escritos. *Revista Signos*, 35(51-52), 217-230.
- Martínez-Arias, R. (1996). *Psicometría: Teoría de los tests psicológicos y educativos*. Madrid: Síntesis.
- McCoach, B., Gable, R., & Madura, J. (2013). *Instrument Development in the Affective Domain School and Corporate Applications* (Third ed.). New York: Springer Science Business Media.
- Mendoza-Mendoza, I., y Sanhueza S. 2018. Percepciones de equidad de género en las/os futuras/os profesoras/es. *Ex aequo*, n.o 37, pp. 129-142. DOI: <https://doi.org/10.22355/exaequo.2018.37.09>
- MINEDUC. (2015). *Bases Curriculares. 7º básico a 2º medio*. Retrieved julio 10, 2018, from Ministerio de Educación: <https://media.mineduc.cl/wp-content/uploads/sites/28/2017/07/Bases-Curriculares-7%C2%BA-b%C3%A1sico-a-2%C2%BA-medio.pdf>
- Mizala, A., Martínez, F., & Martínez, S. (2015). Pre-service elementary school teachers' expectations about student performance: How their beliefs are affected by their mathematics anxiety and student's gender. *Teaching and Teacher Education*, 50, 70-78.
- Mulder, M. (2014). Conceptions of professional competence. In S. Billett, C. Harteis, & H. Gruber, *International handbook on research into professional and practice-based learning* (pp. 107-137). Dordrecht: Springer.
- Neira, A., Reyes, F., & Riffo, B. (2015). Experiencia académica y estrategias de comprensión lectora en estudiantes universitarios de primer año. *Literatura y lingüística*, 31, 221-244.
- Newton, P., & Shaw, S. (2014). *Validity in educational & psychological assessment*. Los Angeles: SAGE Publications.
- Niederle, M., & Vesterlund, L. (2010). Explaining the Gender Gap in Math Test Scores: The Role of Competition. *Journal of Economics Perspectives*, 24(2), 129-144.
- OECD. (2004). *Reviews of National Policies for Education. Chile*.
- OECD. (2015). *Reformulando la Carrera Docente en Chile. Evidencia internacional seleccionada*. Retrieved enero 11, 2018, from <http://www.oecd.org/chile/Reformulando-la-carrera-docente-en-chile.pdf>
- OECD. (2017). *Educación en Chile*. París -Ciudad de México: OECD Publishing-Fundación SM.
- Oscar, A., Mizala, A., & Meneses, F. (2016). *Brecha de género en matemáticas: el sesgo de las pruebas competitivas (evidencia para Chile)*.
- Pedraja, L., Araneda-Guirriman, C., Rodríguez-Ponce, E., & Rodríguez-Ponce, J. (2012). Calidad en la Formación Inicial Docente: Evidencia Empírica en las Universidades Chilenas. *Form. Univ. [online]*, 15-26.

- Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, 75(2), 211-246.
- Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula, *Handbook of research on teacher education* (pp. 102–119). New York, NY: Macmillan.
- Schleicher, A. (2016). *Teaching excellence through professional learning and policy reform*. París: OECD.
- Selvi, K. (2010). Teachers' Competencies. *Cultura. International Journal of Philosophy of Culture and Axiology*, 7(1).
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- SIES. (2017, septiembre 11). Retrieved from Servicio de Información de Educación Superior: <http://www.mifuturo.cl/index.php/estudios/estudios-recientes>
- Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content . *Psicothema*, 26(1), 100-107.
- Solis, C. (2015). Beliefs about teaching and learning in university teachers: Revision of some studies. *Propósitos y Representaciones* , 3(2), 227-260.
- Stufflebeam, D., & Coryn, C. (2014). *Evaluation theory, models, and applications* (Second edition ed.). San Francisco: Jossey-Bass.
- Tagle, T., Díaz, C., Alarcón, P., Quintana, M., Ramos, L., & Etchegaray, P. (2017). Aprendizaje y enseñanza: creencias de estudiantes en formación de pedagogía en inglés. *Páginas de Educación*, 10(1), 64-90.
- Thompson, A. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws , *Handbook of research on mathematics teaching and learning* (pp. 127-146). New York: MacMillan.
- Townsend, T., & Bates, R. (2007). *Handbook of teacher education: Globalisation, standards and professionalism in times of change*. Heidelberg: Springer.
- UNESCO. (2013). *Antecedentes y Criterios para la Elaboración de Políticas Docentes en América Latina y el Caribe*. Santiago: Centro de Estudios de Políticas en Educación (CEPPE).
- Veal, L., & Hudson, S. (1983). Direct and indirect measures for large-scale evaluation of writing. *Research in the Teaching of English*, 17(3), 290-296.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification . In D. S. Rychen , & L. H. Salganik, *Defining and selecting key competencies* (pp. 45–66.). Göttingen: Hogrefe.
- Yang, X., Kaiser, G. G., König, J. J., & Blömeke, S. (2018). Measuring Chinese teacher professional competence: adapting and validating a German framework in China. *Journal of Curriculum Studies*, 50(5), 638-653.
- Zwick, R. (2017). *Who Gets In?: Strategies for Fair and Effective College Admissions* . Harvard University Press.