

¿Qué evidencia respalda la validez de la evaluación docente en Chile?

Jorge Manzi

Daniela Jiménez

Centro de Medición MIDE UC

Partamos por precisar el concepto de validez

- “Validez se refiere al grado en que evidencia y teoría respaldan las interpretaciones de puntajes de pruebas para sus usos propuestos” (AERA, APA & NCME, 2014)

¿Cómo se puede establecer un programa de validez?

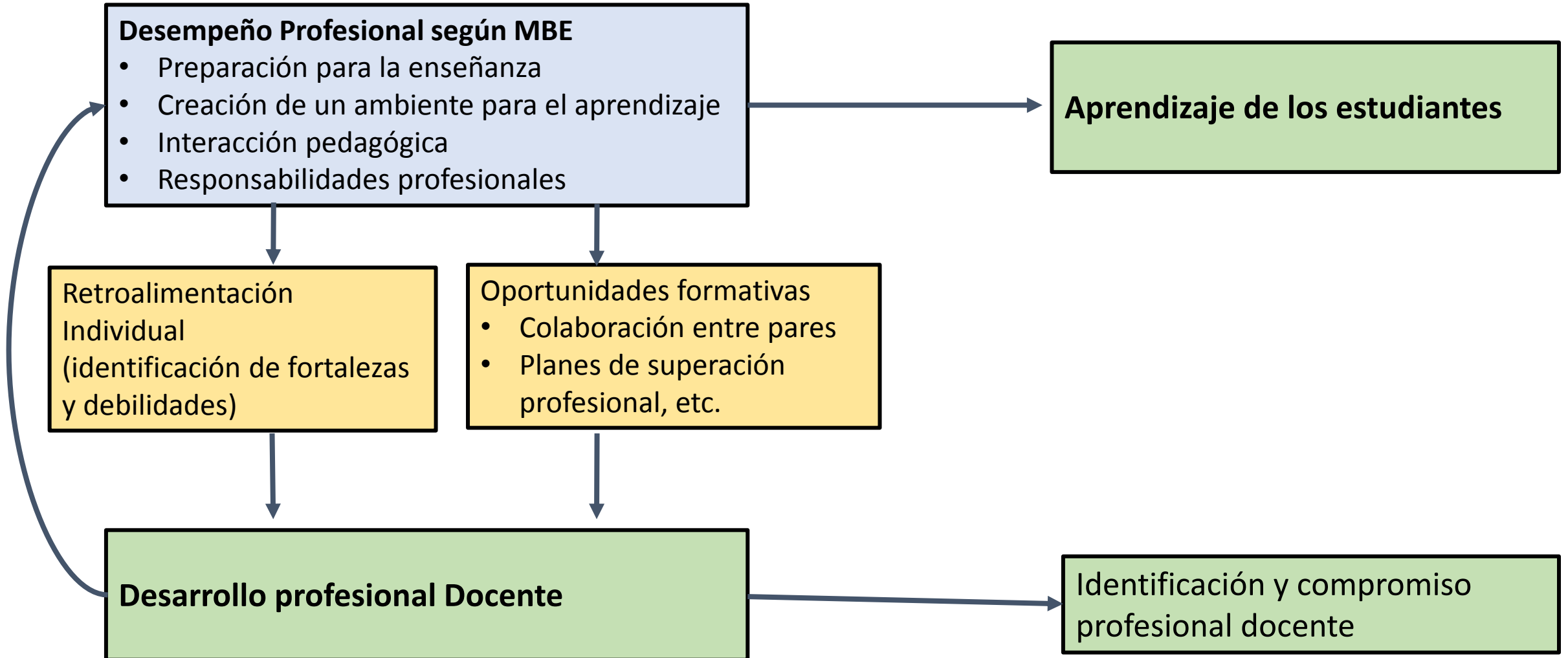
- *Construcción de una argumentación*
 - » *La validación conlleva un proceso de construcción y evaluación de argumentos a favor y en contra de la interpretación que desea hacerse de los puntajes. Esta argumentación debe establecer su relevancia o pertinencia para los usos propuestos.*
 - » *La argumentación debe contemplar hipótesis alternativas a la interpretación prevista*

¿Cómo se puede establecer un programa de validez?

- *Obtención de evidencia*

- » *La interpretación propuesta, y el análisis de hipótesis alternativas, debe orientar la selección de evidencia que permita establecer si hay sustento para la misma.*
- » *La evidencia también debe considerar los usos y consecuencias intencionados (y no intencionados) de los instrumentos.*

El Marco de Referencia de la Evaluación Docente



El programa de validación de la Evaluación Docente

- Evidencia basada en el proceso de desarrollo de los instrumentos
 - » Alineamiento con estándares (Marco de la Buena Enseñanza)
 - » Verificación del alineamiento por parte de comisiones técnicas
 - » Estudios acerca de la estructura factorial de los instrumentos (Análisis Factorial)

El programa de validación de la Evaluación Docente

- Evidencia acerca de la diferenciación entre docentes calificados en los distintos niveles de desempeño
 - » Destacados vs Insatisfactorios
 - » Competentes vs Básicos

El programa de validación de la Evaluación Docente

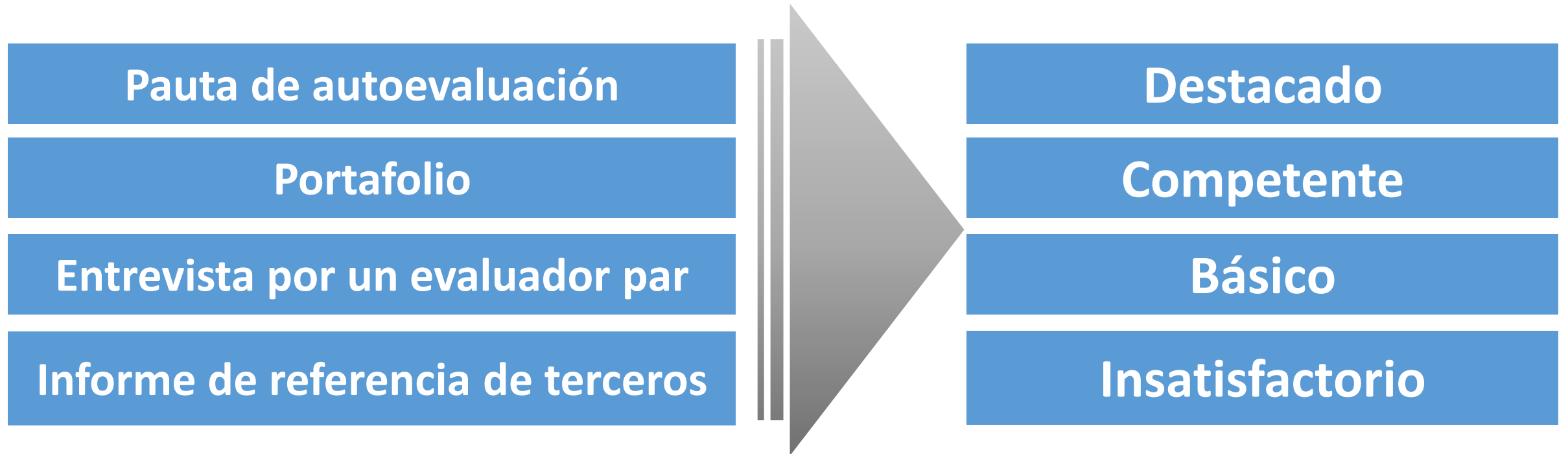
- Evidencia acerca del mejoramiento de los desempeños en docentes previamente evaluados

El programa de validación de la Evaluación Docente

- Evidencia acerca de la relación entre el desempeño docente evaluado y el logro escolar de los estudiantes
 - » Comparación del logro escolar según número de profesores bien evaluados
 - » Modelamiento multinivel
 - » Modelamiento con metodología de valor agregado

Validación de las categorías de desempeño

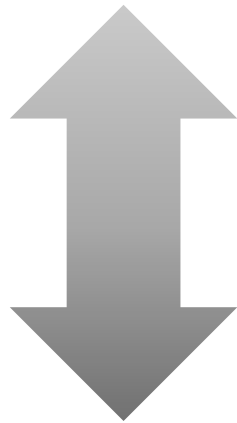
Validez asociada a niveles de desempeño



Estudio de comparación de categorías extremas

Destacado

n=32

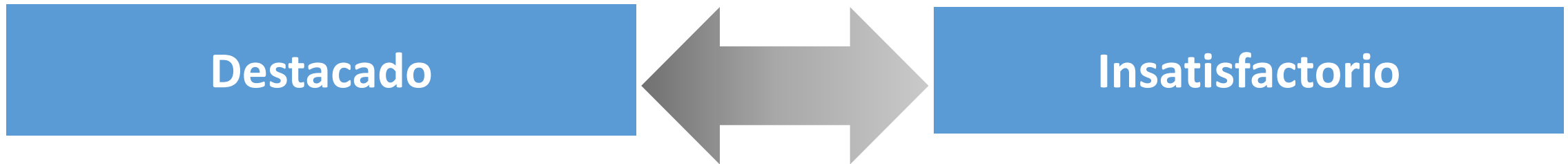


Insatisfactorio

n=26

- ¿Por qué usar categorías extremas?
 - consecuencias asociadas
 - inicio lógico de la agenda de validación
- Se investigaron prácticas pedagógicas de profesores generalistas (1º a 4º) durante un año escolar

Estudio de comparación de categorías extremas



- Evidencia de Validez convergente (basada en relaciones entre variables)
 - Registro de observación de 3 clases
 - Logro académico de los estudiantes (inicio y fin del año escolar)
 - Material de trabajo correspondiente a una unidad de aprendizaje
 - Cuestionario de contexto escolar y reflexión docente
 - Pruebas de conocimientos disciplinarios y pedagógicos

Principales Resultados

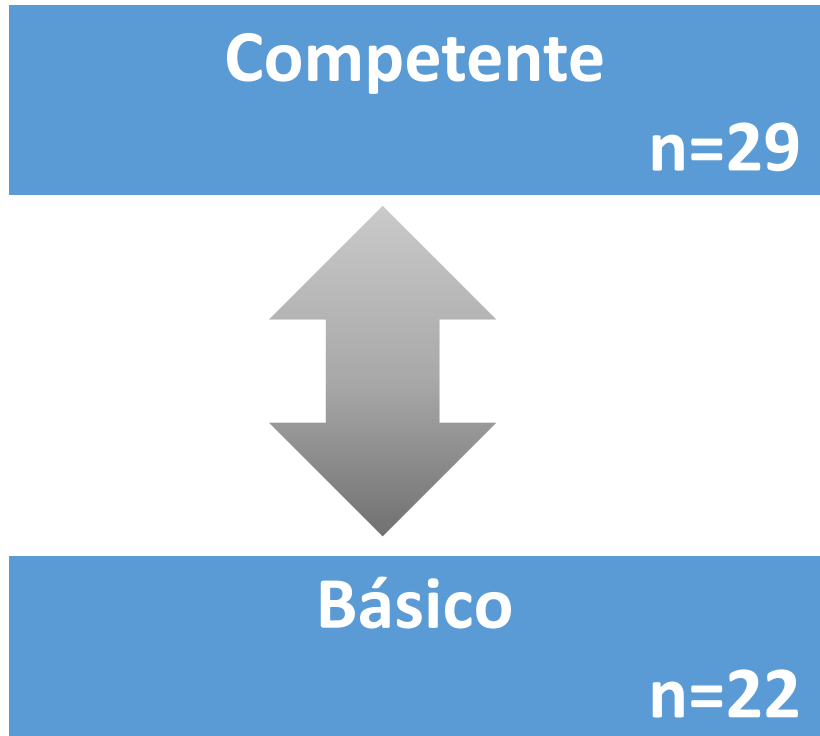
- Diferencias en la dirección esperada en todos los indicadores
- Significativa estadísticamente en la mitad de ellos, para los que se calculó el tamaño del efecto

Indicador	<i>D</i> de Cohen
Calidad de materiales instruccionales (ev. holística)	0,46
Diseño de evaluación	1,21
Desempeño de los alumnos	0,58
Diseño de evaluación y desempeño de alumnos (ev. holística)	0,95
Reflexión de la propia práctica (ev. holística)	0,75

Principales Resultados

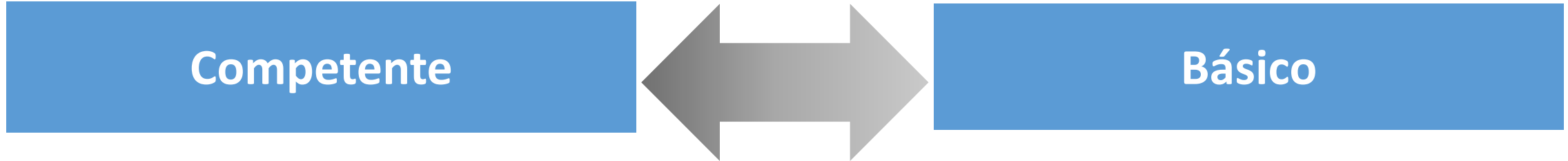
Indicador	<i>D</i> de Cohen
Prueba de conocimientos (selección múltiple)	0,58
Prueba de conocimientos (pregunta abierta)	0,48
Tiempo con menos de 75% de alumnos enfocados en la tarea	0,98
Tiempo con menos de 95% de alumnos enfocados en la tarea	0,86
Estructura de la clase	1,11
Enseñanza especialmente estimulante	0,62
Conducta adecuada de los estudiantes	1,05

Estudio de comparación de categorías intermedias



- ¿Por qué comparar categorías intermedias?
 - consecuencias asociadas
 - progresión en la agenda de validación
- Se investigaron prácticas pedagógicas de profesores de matemáticas de 7º básico durante un año escolar

Estudio de comparación de categorías intermedias



- Evidencia de Validez convergente (basada en relaciones entre variables)
 - Grabación de 2 clases
 - Logro académico de los estudiantes (inicio y fin del año escolar) + en prueba de unidad de potencias o geometría
 - Cuestionario de estudiantes acerca de la percepción de prácticas docentes
 - Pruebas de conocimientos disciplinarios y pedagógicos

Resultados: aprendizaje de los estudiantes

Pruebas de conocimientos	<i>Media B (DS)</i>	<i>Media C (DS)</i>
Inicio de año escolar*	594 (10,4)	603 (19,8)
Fin de año escolar	615 (12,4)	622 (23,4)
Prueba de Unidad geometría	7,6 (1,1)	8,6 (1,2)
Prueba de Unidad potencias*	9,2 (1,6)	10,7 (2,1)

* = nivel de significancia marginal $p < 0,1$

Principales Resultados: prácticas docentes

- No hay diferencias estadísticamente significativas en los indicadores descriptivos

Indicadores descriptivos	<i>B</i>	<i>C</i>
Tiempo instruccional efectivo	81,2%	84,5%
Tiempo de organización	13,4%	9,4%
Tiempo de clases curso completo	61,6%	61,2%
Tiempo dedicado a trabajo grupal	9,8%	4,2%
Tiempo dedicado a trabajo individual	20,2%	26,8%
Tiempo dedicado a repaso de contenidos	53,5%	49,7%

Principales Resultados: prácticas docentes

Dimensiones CLASS-S	<i>B</i>	<i>C</i>	<i>D de Cohen</i>
Clima positivo	3,6	3,8	0,4
Sensibilidad del profesor**	3,5	4,0	0,7
Consideración por la perspectiva del adolescente*	1,8	2,0	0,6
Manejo de Conducta	6,2	6,3	0,2
Productividad	5,8	6,1	0,5
Ausencia de clima negativo	6,8	6,8	0,2

Principales Resultados: prácticas docentes

Dimensiones CLASS-S	<i>B</i>	<i>C</i>	<i>D de Cohen</i>
Formatos de enseñanza	3,7	4,0	0,6
Comprensión de contenidos*	2,9	3,3	0,7
Análisis e Indagación	1,2	1,4	0,3
Calidad de la retroalimentación*	2,1	2,4	0,6
Diálogo Pedagógico*	1,8	2,1	0,6

Principales Resultados: prácticas docentes

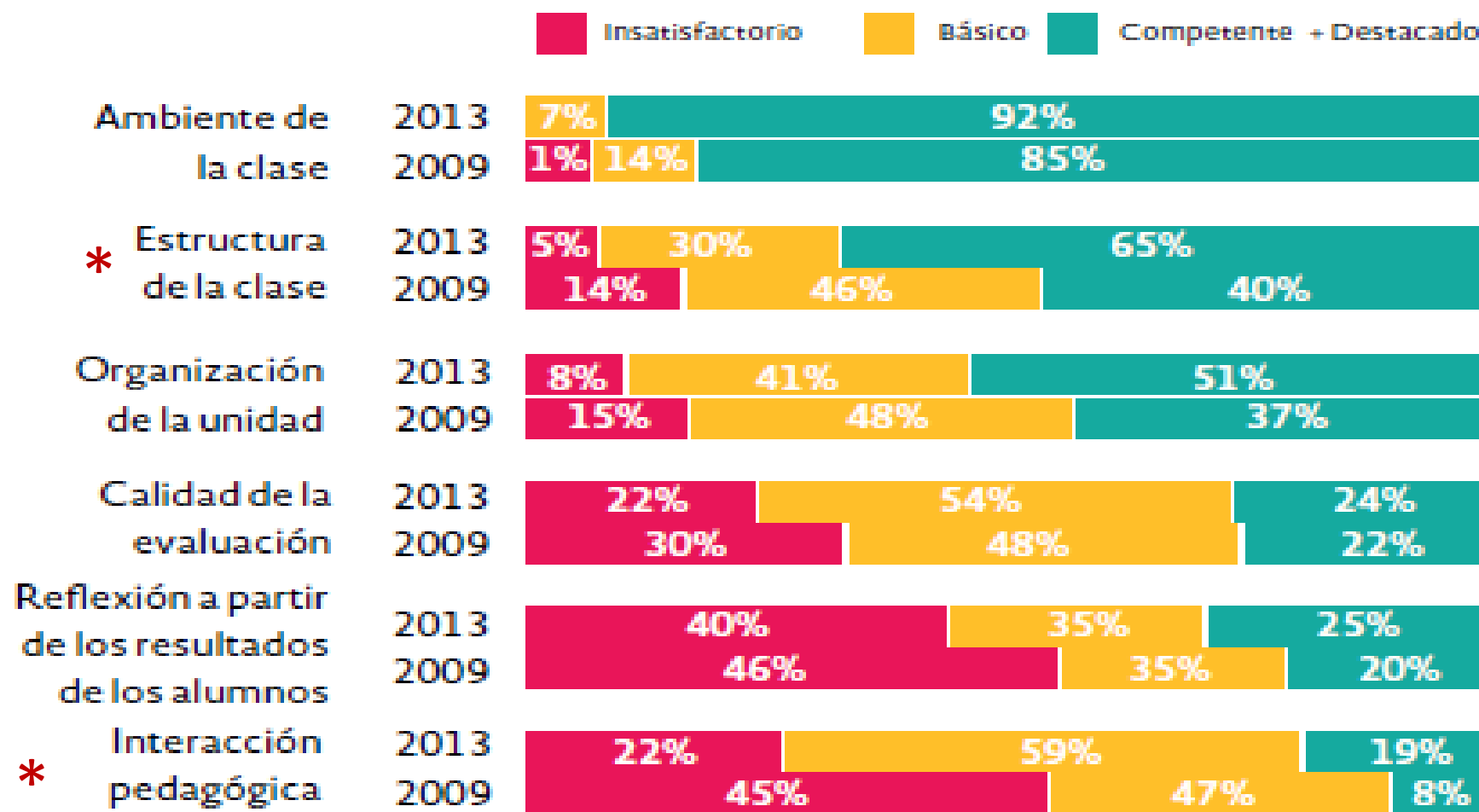
Percepción de los estudiantes	<i>Media B (DS)</i>	<i>Media C (DS)</i>
Constructivismo**	3,2 (0,30)	3,4 (0,24)
Instrucción directa	3,9 (0,35)	4 (0,36)
Apoyo pedagógico**	4 (0,38)	4,3 (0,35)

- Al controlar por el rendimiento de los estudiantes, el nivel de significancia disminuye a marginal ($p < 0,1$)

Evolución del desempeño de profesores previamente evaluados

Distribución de docentes que se evaluaron en 2009 y 2013 (N=5528)

Fuente: Herrada y Zapata, 2015

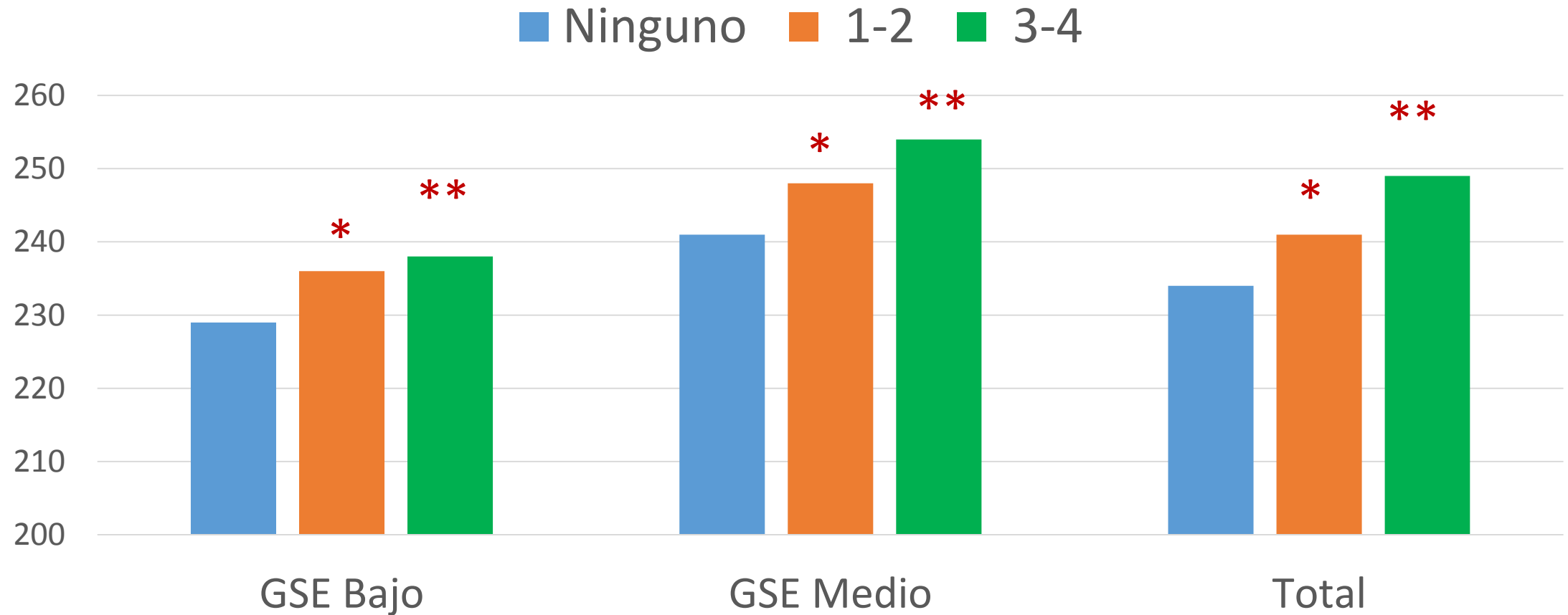


Distribución de docentes reevaluados

Aspecto	Porcentaje de docentes que cambia su categoría de desempeño			Diferencia promedio	Tamaño del efecto
	Baja	Mantiene	Sube		
Interacción pedagógica	16%	41%	43%	,26	,45
Estructura de la clase	16%	42%	43%	,24	,42
Organización de la unidad	22%	38%	40%	,17	,26
Ambiente de la clase	7%	76%	17%	,05	,16
Reflexión a partir de resultados de los alumnos	26%	39%	35%	,12	,11
Calidad de la evaluación	26%	41%	33%	,06	,06

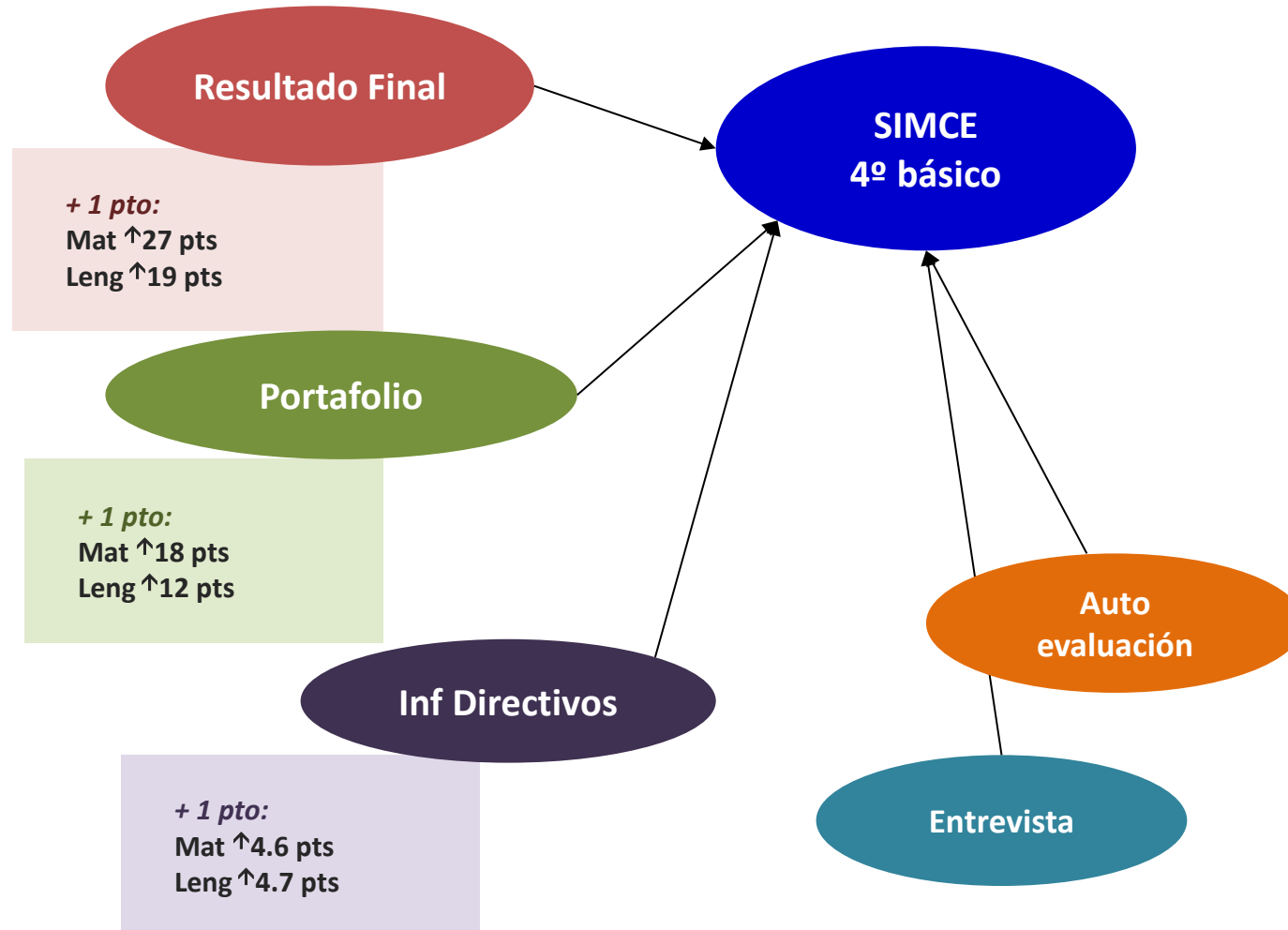
Relación entre evaluación docente y logro escolar en SIMCE

Puntaje SIMCE Promedio según el número de profesores Competentes o Destacados de los estudiantes. Fuente: informe SIMCE 4º básico 2008



Relación entre evaluación docente y SIMCE.

Análisis multinivel. Fuente: Alvarado et al, 2012



Alvarado et al (2012):

- 1 punto más en resultado final Docentemás se asocia a +27 ptos. en SIMCE Matemática y 19 en Lenguaje.
- El instrumento que muestra mayor relación con SIMCE es el Portafolio.

Correlaciones entre instrumentos de Evaluación Docente y logro de estudiantes con modelos de Valor Agregado. Fuente: Taut et al, 2016

Tabla 1. Correlaciones entre puntajes ED y resultados de aprendizaje (Modelo de Valor Agregado, MVA⁴) en Matemática (N=150) y Lenguaje (N=162).



	Valor agregado Lenguaje	Valor agregado Matemática
Auto-evaluación	-0.08	0.07
Evaluación Par	0.07	0.08
Evaluación Directivos	0.05	0.14
Portafolio	0.19*	0.33**
Resultado final	0.16*	0.28**
** Los asteriscos indican que estas correlaciones son estadísticamente significativas (diferentes de cero).		

Evidencia acerca de validez
consecuencial

Validez Consecuencial (Taut y otros, 2010)

Efectos intencionados

Ordenar a profesores de acuerdo a la calidad de su enseñanza	+
Diagnosticar fortalezas y debilidades de los docentes	+
Apoyar la colaboración entre docentes	+
Proveer información para toma de decisiones comunal	(+)
Promover reconocimiento social de los buenos profesores	(+)
Mejorar la docencia como profesión al dar acceso a incentivos monetarios vinculados a calidad	?
Apoyar desarrollo profesional a través de los Planes de Superación Profesional	?

Validez Consecuencial

Efectos no intencionados

Positivos:

- Impacto formativo del proceso gatillado por el desarrollo del Portafolio, familiarización con estándares (MBE), capacitación de evaluadores pares y correctores de portafolio.
- Apoyo de municipios y escuelas a profesores durante su evaluación.

Negativos:

- Sobrecarga de trabajo, reacciones emocionales negativas (miedo, ansiedad, inseguridad laboral)
- Evasión de la evaluación (subterfugios legales)
- Prácticas poco éticas (copiar o “comprar” portafolios)

Conclusiones acerca de la evidencia acumulada

- En cuanto a las diferencias entre las categorías de desempeño, se encuentran diferencias sustanciales entre las categorías extremas y más atenuadas entre las categorías intermedias (lo que es esperable)
- En los últimos años se han observado mejoramientos relevantes en docentes que se han evaluado en más de una ocasión
- Clara consistencia entre evaluación docente y logro de estudiantes, especialmente asociada al portafolio
- Respaldo parcial para efectos positivos intencionados del programa de evaluación. También se observan efectos no intencionados, algunos positivos (como apoyo en escuelas a profesores durante su evaluación) y otros negativos (como sobrecarga laboral y emocionalidad negativa)

Recomendaciones

- Diseñar un programa de validación para los instrumentos en el marco de la carrera docente (dado el cambio de uso de los instrumentos y el incremento de las consecuencias de los mismos)
 - Cambios en los instrumentos: ahora el portafolio tiene 3 módulos
 - Pruebas se aplican a todos los docentes para su progresión en la carrera (antes se rendían en forma voluntaria en AEP o AVDI)
 - Validar las reglas que combinan los instrumentos para la progresión en la carrera
 - Estudios longitudinales a través de las etapas de la carrera docente
 - Monitorear de manera sistemática las consecuencias de la evaluación
 - Analizar el impacto de la evaluación en el sector particular subvencionado cuando se incorpore a la carrera docente