

LA VALIDEZ DE LA EVALUACIÓN DOCENTE EN CHILE

SANDY TAUT, PH.D.
P. UNIVERSIDAD CATOLICA DE CHILE

3er Encuentro de Políticas Públicas, 19-1-2012, Santiago, Chile

Características clave de la ED en Chile



La validez: Un constructo clave de la medición educativa

- Evolución del concepto hacia modelo unificado de validez de constructo (Messick, 1989; Kane, 2006)
- Varias fuentes de evidencia, necesidad de integrarlas y llegar a un juicio complejo
- Foco en las interpretaciones y usos de resultados de un test, siempre en el contexto de sus usos propuestos
- La validez se tiene que estudiar

Pregunta central

- Hay suficiente evidencia para interpretar los resultados de la ED para los usos propuestos?
- Usos propuestos de la ED en Chile:
 - Ordenar docentes según su desempeño (insat<bás<comp<dest)
 - Diagnosticar fortalezas y debilidades de docentes evaluados para motivar cambios en prácticas débiles y mantener buenas prácticas
 - Enfocar el desarrollo de capacidades a través de Planes de Superación Profesional (PSP)
 - Tomar decisiones informadas a nivel comunal (contratación, despido, asignación de docentes)
 - Mejorar trayectoria laboral a través de vínculo entre desempeño y incentivos monetarios
 - Invitar el reconocimiento social de buenos docentes
 - Incentivar la colaboración entre pares respecto de buenas prácticas

La agenda de validación de la ED

1. Evidencia de validez basada en el contenido
2. Evidencia de validez basada en la estructura interna de los instrumentos
3. Evidencia de relaciones con otras variables
4. Evidencia sobre consecuencias
5. Evidencia de confiabilidad y generalizabilidad
6. Evidencia de equidad del proceso

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.

Métodos



- **Análisis psicométricos:**
 - Análisis factorial exploratorio y confirmatorio
 - Estudios de generalizabilidad
 - Confiabilidad y acuerdo entre correctores
- **Análisis estadísticos:**
 - Descriptivo
 - Correlacional
 - Comparación de medias
 - Regresión multinivel
- **Estudios cualitativos:**
 - Grupos focales
 - Entrevistas personales
 - Análisis de Contenido y Teoría Fundada usando ATLAS.ti

Resultados: Contenido

- Estudio de alineamiento ED – Marco para la Buena Enseñanza (2005)
- Específicamente, se analizó rúbrica de puntuación del portafolio, y los ítemes de los otros tres instrumentos
- Portafolio cubre Dominios B y C, parcialmente A
- Se excluyen temas de prácticas específicas de la disciplina y conocimientos disciplinarios (estos últimos se miden a través de prueba AVDI)
- Otros instrumentos cubren D
- Pendiente: estudio con expertos externos

Resultados: Estructura interna

- Consistencia interna (Alpha de Cronbach) de los instrumentos:
 - > 0.7 para todos los instrumentos en todos los años
- Análisis factorial exploratorio (2005-2009): Representa el portafolio diferentes dimensiones (factores) del desempeño docente? Cuáles son?
- Resultados varían entre años, pero en general se encuentran 5 o 6 factores: 3 o 4 de la parte escrita, 1 o 2 de la parte filmada.
- Reflejan habilidades como a) planificar y diseñar la enseñanza, b) construir evaluaciones, c) capacidad reflexiva. Habilidades observadas en video aparecen correlacionadas.
- Análisis factorial confirmatorio (2010): confirma estructura teórica.

Ejemplo: Resultados EFA 2009

Factor	Indicador (loading >0.35)
F1	A1 A2 A3 Planificación
F2	B1 B2 B3 D1 D2 E1 E2 E3 Reflexión
F3	C1 C2 C3 Evaluación
F4	F1 F3 G3
F5	F2 G1 G2 H1 H2 H3
F6	H4 Pedagogía especif. asignatura

Estructura teórica del portafolio ED 2010

DimA: Organización de los elementos de la unidad

DimB: Análisis de las actividades de las clases

DimC: Calidad de la evaluación de la unidad

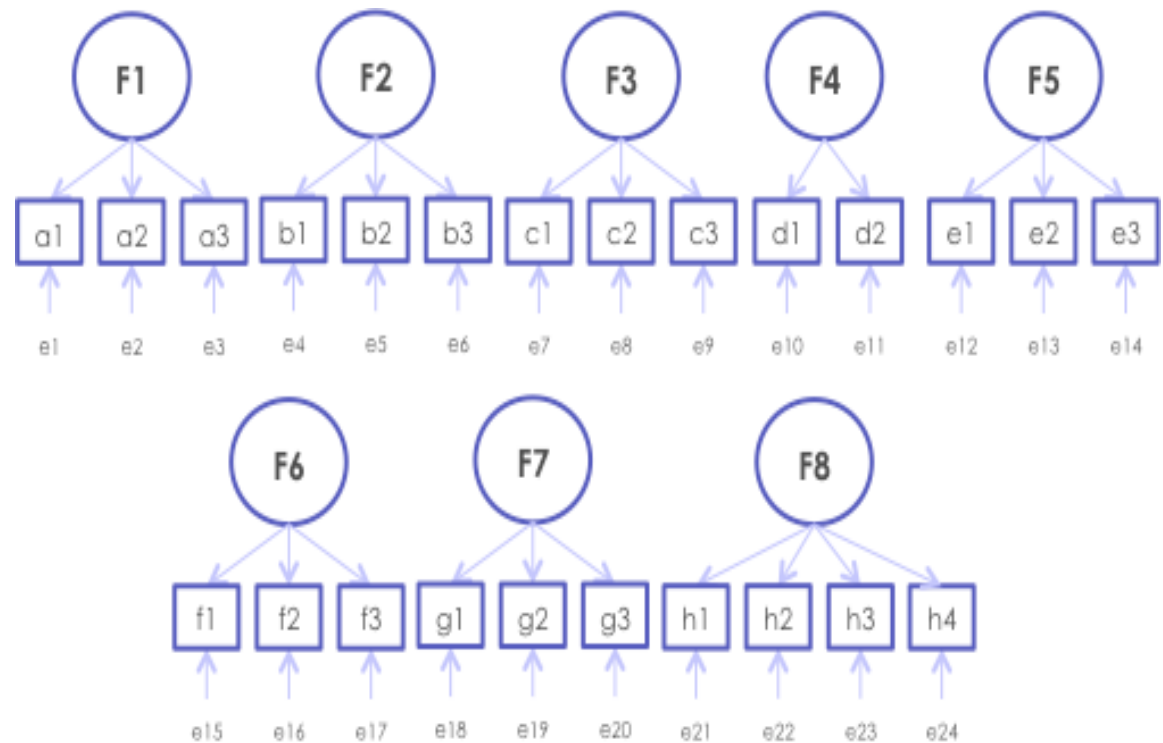
DimD: Reflexión a partir de los resultados de la evaluación

DimE: Reflexión pedagógica

DimF: Ambiente de la clase

DimG: Estructura de la clase

DimH: Interacción pedagógica



Resultados: Relación con otras variables

- Relación entre ED y prácticas pedagógicas estudiados en profundidad (Santelices & Taut, 2011)
 - ▣ Destacados (N=32) mostraron mejor desempeño que insatisfactorios (N=26) en todos los indicadores, en la mitad estadísticamente significativo, en algunos de tamaño de efecto grande (ver tabla)
 - ▣ Correlaciones significativas con instrumentos ED (salvo auto-evaluación)
 - ▣ Pendiente: estudio contrastando prácticas de aula de docentes básicos versus competentes
- Relación entre ED y resultados SIMCE (Manzi et al., 2007; Bravo et al., 2008; Ministerio de Educación, 2008, 2009)
 - ▣ Relación positiva entre ED y SIMCE
 - ▣ Pendiente: estudio usando datos longitudinales de aprendizaje de los estudiantes de docentes evaluados

Tamaños de efecto t-test: Insat vs. Dest

Instrumento	Indicador / subescala	Cohen <i>d</i>
Prueba de conocimiento pedagógico y disciplinario	Preguntas de alternativas múltiples	0,58
	Preguntas abiertas	0,48
Observaciones de clase	Proporción de tiempo con menos de 75% de alumno en la tarea	0,98
	Proporción de tiempo con más de 95% de alumno en la tarea	0,86
	Estructura de la clase	1,11
	Enseñanza especialmente estimulante	0,62
	Conducta de los alumnos adecuada	1,05
Carpeta con materiales de enseñanza	Materiales instruccionales (evaluación holística)	0,46
	Diseño de evaluación de los alumnos	1,21
	Desempeño de los alumnos (indicador continuo)	0,58
	Diseño de evaluación y desempeño de los alumnos (evaluación holística)	0,95
	Reflexión de la propia práctica (evaluación holística)	0,75

Discusión: Integración de evidencia

- “... la validez es, a menudo, presentada como una dicotomía: la conclusión en base al desempeño en un instrumento es válida o inválida. Desafortunadamente, la situación (en la vida real) es menos clara. La validación es un continuo (...) Algunas conclusiones están mejor respaldadas que otras, pero la evidencia con la cual generalmente se cuenta es limitada...” (Koretz, 2008, p. 219).
- Es una cuestión de valores y perspectivas (Herman & Baker, 2009), de los cuales hay varios en el caso de la ED.
- Se necesita convergencia entre distintos cuerpos de evidencia, pero no todos tienen la misma relevancia.
- Efectos y usos *no* intencionados también se reportan.
- Nuestra visión: evidencia inicial lleva a un juicio positivo respecto de la validez de la ED para sus principales usos propuestos (pero para algunos más que otros, ver tabla).
- Hay que seguir investigando la validez de la ED (Cronbach, 1989).

Tabla de integración de evidencia

Uso propuesto de la ED	Evidencia inicial
Ordenar docentes según su desempeño	+
Diagnosticar fortalezas y debilidades de docentes	+
Informar toma de decisiones a nivel comunal	(+)
Mejorar trayectoria laboral a través de incentivos	0
Apoyar el desarrollo de capacidades (PSP)	0
Invitar el reconocimiento social de buenos docentes	(+)
Incentivar la colaboración entre pares	+

Publicaciones relacionadas

- Santelices, V., Taut, S. & Valencia, E. (2008). Estudio exploratorio-comparativo de los resultados de los programas AVDI, AEP y Docentemás entre 2002 y 2006 [Exploratory comparative quantitative analysis of the results of the national teacher evaluation program, the individual teacher incentive program and the accreditation of excellence in teaching program between 2002 and 2006]. Internal document MIDE UC.
- Santelices, V., Taut, S. & Valencia, E. (2009). La Evaluación Docente y los Planes de Superación Profesional: Estudio descriptivo. Internal document MIDE UC.
- Santelices, V., Taut, S., Araya, C., & Manzi, J. (2009). Consequential Validity of Chile's Teacher Evaluation System: Consequences at the Municipal (Local) Level. *Paper presented at the Annual Meeting of the American Educational Research Association, April 13-19, 2009*, in San Diego, USA.
- Taut, S., Santelices, V. & Valencia, E. (2010). Resultado de re-evaluaciones y situación laboral de los docentes evaluados por el Sistema de Evaluación de Desempeño Docente entre 2003 y 2008. Informe técnico MIDE UC IT1007. Disponible en <http://www.mideuc.cl/docs/informes/it1007.pdf>
- Taut, S., Santelices, V., Araya, C. & Manzi, J. (2010). Theory underlying a national teacher evaluation program. *Evaluation and Program Planning*, 33, 477-489. <http://dx.doi.org/10.1016/j.evalprogplan.2010.01.002>
- Tornero, B. & Taut, S. (2010). A mandatory, high-stakes national teacher evaluation system: Perceptions and attributions of teachers who actively refuse to participate. *Studies in Educational Evaluation*, 36, 132-142.
- Taut, S., Santelices, V., Araya, C. & Manzi, J. (2011). Effects and uses of the national teacher evaluation system in Chilean elementary schools. *Paper presented at the Annual Conference of the American Educational Research Association, April 8-12, 2011*, New Orleans, USA.
- Cortés, F., Taut, S., Santelices, V. & Lagos, M.J. (2011, January). Formación continua en profesores y la experiencia de los Planes de Superación Profesional (PSP) en Chile: Fortalezas y debilidades a la luz de la evidencia internacional [Teacher professional development and the experience with the program Planes de Superación Profesional (PSP) in Chile: Strengths and weaknesses in light of the international evidence]. *Paper presented at the second meeting of the Sociedad Chilena de Políticas Públicas, January 19, 2011*, Santiago, Chile.
- Taut, S., Santelices, V. & Manzi, J. (2011). Estudios de validez de la Evaluación Docente [Validation studies of the Chilean National Teacher Evaluation]. In: Jorge Manzi, Roberto Gonzalez, Yulan Sun (Eds.), *La Evaluación Docente en Chile* [The Chilean national teacher evaluation system]. Pontificia Universidad Católica de Chile, Santiago, Chile.