

Capítulo 4

Validación del proceso de Assessment Center para la selección de directivos escolares¹

INVESTIGADOR PRINCIPAL: PAULO VOLANTE

Equipo de investigación: Alejandro Díaz, Antonio Mladinic, Magdalena Fernández, Cristian Lincovil, Michael Johaneck, Andrew Porter
Institución adjudicataria: Pontificia Universidad Católica de Chile
Proyecto FONIDE N° F711295

Resumen

Este estudio busca validar un proceso de Assessment Center (Centro de Evaluación, AC en sus siglas en inglés) para la selección de directivos escolares. El proceso de evaluación e instrumentos que forman parte del Assessment Center fueron diseñados el 2011 y piloteados durante el 2012 en una muestra de 12 directivos. Basándose en una selección teórica y empírica de las prácticas y competencias asociadas a buenos resultados en las escuelas (estándares de liderazgo escolar, ILSCC-2008), y en virtud de los requerimientos de la Ley 20.501, se propone un proceso de evaluación de postulantes a cargos directivos que enfatiza en la influencia de la enseñanza y del aprendizaje. Adicionalmente, la investigación internacional avala la creación de Assessment Centers para la selección de candidatos a un cargo, debido a su capacidad

¹ Siempre que es posible, el presente Informe intenta usar un lenguaje inclusivo y no discriminatorio, sin embargo, con el fin de respetar la ley lingüística de la economía expresiva y así facilitar la lectura y comprensión del texto, en algunos casos se usará el masculino genérico que, según la Real Academia de la Lengua Española, se acepta como representante de hombres y mujeres en igual medida.



predictiva por sobre los métodos tradicionales de selección (basados únicamente en entrevistas, pruebas proyectivas, y/o inventarios de personalidad). Este estudio iniciará la validación de constructo y de criterio concurrente, además de la confiabilidad entre evaluadores de los instrumentos generados para el Assessment Center. Con este fin, se analizan los resultados del desempeño en Assessment Center obtenidos entre mayo y julio de 2013, a una muestra de 164 participantes. Del análisis se concluye un alto grado de confiabilidad y de validez del instrumento en términos de constructo y criterio. El desempeño en el Assessment Center de los directores con 3 o más años de experiencia en su cargo, se asocia con los resultados SIMCE de su establecimiento.

Palabras claves: *directores, liderazgo, liderazgo instruccional, liderazgo directivo, Assessment Center, validación, efectividad escolar, selección de directivos, formación de directivos.*

1. ANTECEDENTES Y PLANTEAMIENTO DEL PROBLEMA

La validación de instrumentos de un Assessment Center (AC) para la selección de equipos directivos en establecimientos educacionales está directamente vinculada al ámbito de gestión escolar y liderazgo educativo; puesto que apunta a establecer e identificar aquellas competencias de liderazgo que ayudan a mejorar las condiciones de aprendizaje de los alumnos al interior de la escuela, es decir, las competencias de liderazgo instruccional en el equipo de profesionales que lidera la gestión al interior de los establecimientos educacionales. Con esto, la validación de instrumentos busca evaluar la factibilidad de un proceso de selección contenido en el nuevo marco legal de la Ley 20.501, lo que tiene valor para guiar las decisiones en políticas públicas.

En el contexto de la reciente Ley sobre Calidad y Equidad de la Educación (Ley 20.501), vigente desde febrero de 2011, se establecen tres líneas de acción en el campo de directivos escolares: mejorar la selección directiva, otorgar mayores atribuciones para el ejercicio de su cargo y aumentar las remuneraciones con el fin de atraer mejores postulantes.

En el terreno de la selección de equipos directivos, la ley considera un proceso análogo al implementado en la Alta Dirección Pública, agregando al proceso actual el apoyo de asesorías externas en la preselección y la presencia de un miembro del Consejo de Alta Dirección Pública en la Comisión Evaluadora. Por lo tanto, es pertinente explorar modelos y aplicaciones que puedan ser transferidos por medio de políticas y programas públicos de desarrollo profesional.

Una línea de acción relevante frente a este problema es la aplicación de un proceso de evaluación y selección de equipos directivos, válido, que prediga efectivamente el futuro desempeño de los seleccionados, basándose en estándares de desempeño. Por esta razón, la validación de un modelo de AC, que permita detectar y evaluar el desempeño actual y potencial de los directivos escolares, se constituye en una herramienta que aspira a facilitar la implementación de la ley, y así contribuir a dar más transparencia y equidad en la distribución de capital humano en el sistema escolar público.

Esta investigación giró en torno a 3 preguntas:

- ¿Los datos empíricos recolectados mediante Assessment Center se estructuran y comportan como el modelo lo asume? En otras palabras, ¿es válido el set de instrumentos de Assessment Center en términos de constructo?
- ¿Los datos del desempeño de los participantes en el AC están relacionados con el desempeño del director en su establecimiento? En otras palabras, ¿el resultado de los participantes en el AC puede correlacionarse con otras medidas?
- ¿En qué medida las mediciones tomadas mediante el Assessment Center son confiables? ¿En qué medida los ejercicios, instrumentos y rúbricas permiten una evaluación consistente entre distintos evaluadores?

Este modelo ha sido testeado utilizando un set de instrumentos que fueron diseñados para detectar diferencias en el desempeño de estas competencias, por medio de un proceso de AC aplicado a directores en ejercicio, y a directivos y docentes interesados en postular a este tipo de cargos. Por tanto, se hipotetizó que el modelo de competencias seleccionado tendría validez empírica, y sería factible de comprobar a partir de los datos recolectados en el proceso AC. Al respecto, se trabajó con tres hipótesis:

- Hipótesis 1 de validez de constructo: la estructura de variables de AC se ajusta a los datos empíricos.
- Hipótesis 2 de validez de criterio: el resultado de los directores participantes en el AC se correlaciona con el mejoramiento de los aprendizajes en el establecimiento del director.
- Hipótesis 3 de confiabilidad: las evaluaciones de AC son consistentes.

En consecuencia, el objetivo general propuesto inicialmente para este estudio es validar un proceso de Assessment Center para la selección de directivos escolares que sea factible de ser implementado en la administración municipal de la educación. Los objetivos específicos son:

- Evaluar la validez de constructo de los instrumentos del AC para la selección de directivos escolares.
- Evaluar la validez de criterio concurrente del AC en la predicción de indicadores de gestión escolar.
- Evaluar la confiabilidad de los instrumentos de AC.

2. MARCO TEÓRICO

La eficacia de los equipos directivos en las escuelas es muy relevante para el logro de resultados educativos. En efecto, existe evidencia de que el liderazgo directivo impacta de manera positiva en los aprendizajes de los estudiantes, sobre todo en las escuelas más vulnerables (Leithwood, 2009). Por otro lado, el mayor impacto se produciría cuando el director centra su gestión en lo relativo a la enseñanza y el aprendizaje (Robinson, Honepa & Lloyd, 2009).

En Chile, 7.735 directores administran la enseñanza para 3,6 millones de estudiantes. De ellos, el 97% tiene título de pedagogo y más del 80% ha realizado estudios de postítulo y postgrado, preferentemente en el área de administración educacional. Además, se ha detectado que dedican un 28% de su tiempo a la gestión pedagógica, un 56% a labores de administración internas y un 16% a relaciones inter-institucionales. Por otro lado, un 53% de ellos tiene más de 50 años de edad (OCDE, 2008), por lo que es evidente la urgente necesidad de proveer buenos directivos en el sistema escolar en el mediano y corto plazo (Huber & West, 2002; Firestone & Riehl, 2005).

Hoy existe cada vez más evidencia sobre el impacto de la formación de directores escolares en su desempeño dentro de los establecimientos y el efecto de este en los resultados de aprendizajes

(Leithwood, 2009; Robinson, Honepa & Lloyd, 2009; Uribe, 2009). Sin embargo, en Chile no se cuenta con evidencia del impacto de los programas de capacitación en resultados educativos, aunque sí de las características de su oferta (Muñoz & Farfan, 2011). Por otra parte, existe una falta de programas de inducción y desarrollo que apoyen, de manera dirigida, la generación y el fortalecimiento de competencias de gestión, lo que se contrapone a la necesidad de garantizar la calidad del desempeño de nuevos directores en el ejercicio de sus cargos. De este modo, la hipótesis que guía las políticas y los programas de formación de directivos afirma que si se garantizara un buen director para cada establecimiento del país, los indicadores de calidad educativa mejorarían sustantivamente. Por lo tanto, la formación, selección, inducción y evaluación de los futuros directores escolares se constituye en una prioridad nacional, y la experiencia internacional sugiere elementos comunes para un cambio de perspectiva en su diseño, implementación y evaluación de impacto (La Pointe, et al., 2007).

En Chile, las políticas educativas se han centrado en la gestión escolar y especialmente en los directores. En modificaciones introducidas al Estatuto Docente en 1996 por la Ley 19.979, se define la función principal del director como “dirigir y liderar el Proyecto Educativo Institucional”, estableciendo también un marco de acción y un perfil de competencias directivas. En el año 2005, por otro lado, se crea la Asignación de Desempeño Colectivo, como incentivo al cumplimiento de metas por parte de los equipos directivos de establecimientos escolares. La Ley 20.370 (Ley General de Educación), por su parte, describe nuevamente las funciones del director, centrándose en lo pedagógico. La reciente Ley sobre Calidad y Equidad de la Educación (Ley 20.501), vigente desde febrero de 2011, establece la mejora de los procesos de selección de directores. En el terreno de selección de equipos directivos, la ley considera un proceso análogo al implementado en la Alta Dirección Pública. Por lo tanto, es pertinente explorar modelos de selección de personal que puedan ser implementados en este marco. Una línea de acción relevante frente a este problema, es la aplicación de un proceso de evaluación y selección de equipos directivos basado en estándares de desempeño.

Durante el año 2011, se realizó el diseño de un modelo de Assessment Center, que permitiera detectar y evaluar el desempeño actual y potencial de los directivos escolares (actuales y futuros), así como sus necesidades de apoyo y el potencial de desarrollo en el contexto del sistema escolar público. En la primera etapa de esta investigación, se logró definir un set de competencias claves de dirección escolar, validadas teóricamente, testeadas con expertos y priorizadas por medio de una consulta a más de 200 profesores y directivos en ejercicio. El set de competencias claves de dirección escolar propuesto está centrado en tres aspectos globales en los que se basaría el proceso de evaluación de postulantes al cargo del director:

- Definir una misión y una visión de aprendizaje compartidas por los miembros de la organización escolar.
- Gestionar una cultura y un programa escolar conducente al aprendizaje de los estudiantes y al desarrollo profesional del staff docente.
- Gestionar la organización, la operación y los recursos.

Adicionalmente, se diseñaron prototipos de instrumentos, rúbricas y un flujo del proceso de selección, de acuerdo con los supuestos del modelo AC. Este proceso de AC cuenta con cuatro instrumentos o procesos de evaluación:

- Entrevista en base a competencias
- Ejercicio de análisis y presentación estratégicos
- Ejercicio de gestión (manejo) de personal
- Ejercicio de observación y retroalimentación de clases

Durante el 2012 se implementó un pre-piloto para evaluar en la práctica los instrumentos y el proceso de Assessment Center, desde la perspectiva de los evaluadores y participantes. Para desarrollar este piloto se capacitó a evaluadores en los instrumentos y en el proceso de evaluación. En el piloto participaron 24 evaluadores y 12 directivos. Además, durante el pre-piloto se aplicaron encuestas a evaluadores y directivos participantes para medir la calidad y pertinencia de los instrumentos. Posteriormente, se aplicó un focus group a evaluadores que participaron del pre-piloto para obtener información cualitativa que permitiera mejorar tanto el sistema de evaluación como sus instrumentos. Adicionalmente se realizó un panel de expertos para recolectar evidencia en torno a la validez de contenido.

El resultado de este pre-piloto son una serie de recomendaciones y mejoramientos al proceso de aplicación y a los instrumentos. El índice de consistencia entre los evaluadores medido por el índice Kappa, varió entre 0,16 (leve) y 0,7 (considerable). Se espera que con el mejoramiento de las rúbricas, los instrumentos y el proceso de evaluación realizados en este pre-piloto, los índices de confiabilidad en la siguiente aplicación sean aún mayores a los que se alcanzaron.

El presente proyecto buscó determinar la validez de constructo, la validez de criterio concurrente y la confiabilidad del Assessment Center para la selección de directivos escolares. En la literatura existe un amplio consenso respecto a que las experiencias de AC son una estrategia efectiva para seleccionar y promover cargos laborales (Rupp et al., 2006), y son atractivos, puesto que su mayor potencial consiste en predecir el rendimiento futuro (Gómez & Stephenson, 1987; Schmitt & Schechtman, 1990; Bobrow & Leonards, 1997). El modelo de competencias en la selección de personal presenta una ventaja por sobre los métodos tradicionales de selección de personal, como pruebas proyectivas e inventarios de personalidad de tipo test de Rorschach, test Z de Zulliger, test de Lusher o test de los colores, MBTI, 16 PF Catell, Edwards, y Big Five. Esto se debe a que las pruebas de personalidad indican cómo las personas son, pero no indican lo que hacen en la práctica (Arribas y Pereña, 2009). En otras palabras, los métodos tradicionales de selección de personal permiten observar lo que la persona podría ser capaz de hacer y no lo que efectivamente hace. En contraste, los modelos de selección se fundamentan en la conducta y lo que la persona sabe hacer.

Es necesario avanzar en el desarrollo de nuevas herramientas para la selección de personal que midan competencias en la línea de los Assessment Centers, pues son más predictivos que los instrumentos tradicionales de selección de personal (Gaugler et al., 1987). La validez de un Assessment Center en la predicción del desempeño es de 0,36. Esto significa que un 36% de la varianza en la evaluación de desempeño se explicaría por el puntaje en el AC. Esto es mucho más de lo que predice una entrevista semi-estructurada de selección de personal (0,25) (Gaugler et al., 1987). Los Assessment Centers predicen aún mejor las evaluaciones de potencial para la gestión, es decir, el talento de las personas para ser buenos directivos. En estos casos, predicen las evaluaciones de potencial para la gestión en 0,53 (Gaugler et al., 1987). Es decir, un 53% de la varianza en la evaluación de potencial para la gestión es explicada por el puntaje en el Assessment Center, lo que convierte a esta herramienta en la mejor alternativa para la selección de directivos escolares desde el punto de vista de su capacidad para predecir el éxito en la gestión directiva.

Por lo mismo, los instrumentos que son utilizados en los Assessment Centers son construidos en base al análisis del cargo y las dimensiones más relevantes que debe encontrarse en los candidatos para cumplir con las exigencias de este. La investigación internacional referida a la validación de un instrumento como este, señala la importancia de establecer la validez de contenido y de constructo de los procesos de evaluación (Sackett, 1987; Chan, 1996; Fleenor, 1996). Por ello, el presente proyecto busca principalmente, responder a la necesidad de validar los instrumentos de un Assessment Center para la selección de directivos escolares, basándose en la experiencia internacional sobre validación de contenido, constructo y criterio (concurrente) disponible hasta la fecha (Bobrow & Leonards, 1997; Sackett, 1987; Fleenor, 1996; Norton, 1977; Gómez & Stephenson, 1987; Schmitt & Schechtman, 1990; Bobrow & Leonards, 1997; Van Iddekinge & Ployhart, 2008). Un aspecto original de esta propuesta se aprecia en el hecho de que no existen experiencias sistemáticas de AC para directivos escolares, ya que su costo suele ser mayor al que se está dispuesto a invertir en procesos de selección de directivos escolares.

3. METODOLOGÍA

El proceso de validación de un Assessment Center requiere la recolección de distintos tipos de evidencia para establecer que esta herramienta y sus instrumentos midan aquello que pretenden medir, y con un grado aceptable de confiabilidad. La concepción de validez adoptada por este proyecto está alineada con los Standards for educational and psychological testing desarrollados por la American Educational Research Association (AERA), la American Psychological Association (APA), y el National Council on Measurement in Education (NCME). De acuerdo con estos estándares, la validez es un concepto unitario y no dividido en tipos de validez. En este sentido, para evaluar la validez de un instrumento se deben recolectar distintos tipos de evidencia, los que dependerán del propósito por el cual el test es construido. En el caso de este proyecto, la evidencia en torno a la validez de contenido, de constructo y de criterio, se definen como los más relevantes para un proceso de AC en la selección de directivos escolares.

En este estudio se realizó la evaluación y validación de un proceso de Assessment Center para la selección de directivos escolares cuyos instrumentos fueron pre-diseñados en el contexto de una investigación patrocinada por el Concurso de Políticas Públicas UC 2011. La validez de contenido de los instrumentos se realizó en el año 2012, por medio de un panel de expertos y de la aplicación de un pre-piloto de Assessment Center.

Para la recolección de evidencia relacionada con la validez de constructo se han realizado dos métodos. El primero corresponde a un análisis factorial confirmatorio, en el cual se evaluó si el modelo conceptual previamente establecido en relación con las competencias y sus indicadores es apoyado por los datos empíricos que se obtienen una vez aplicados los instrumentos del Assessment Center. Para realizar este análisis se utilizó la técnica estadística Structural Equation Modeling (SEM), la que puede ser ejecutada por una aplicación del programa estadístico SPSS llamada AMOS (Analysis of Moment Structures).

El análisis factorial confirmatorio es una herramienta que se utiliza cuando se cuenta previamente con un modelo de medición o estructura que establece, a modo de hipótesis, los ítems que miden un factor determinado. En el caso de este estudio, los factores corresponden a las 3 competencias que mide el Assessment Center, y los ítems son los indicadores conductuales que dan cuenta de estas competencias (5 indicadores por cada competencia). Es importante mencionar que no se requiere estrictamente que las variables sean independientes para realizar este análisis, ya que el modelo puede ser especificado con correlaciones entre los factores. En el caso de las competencias medidas por el Assessment Center existe redundancia, ya que se trata de habilidades en el ámbito de la gestión escolar. No se contempló la utilización de esta herramienta para la eliminación o reducción de variables, sino que, a partir de este análisis, se determinó si el modelo teórico que vincula las competencias con sus respectivos indicadores conductuales es consistente con los datos empíricos.

De acuerdo con la literatura, la muestra apropiada para realizar el análisis factorial confirmatorio es de 10 casos por cada dimensión (Arrindell & Van der Ende, 1985; Kuncze, Cook, & Miller, 1975; Velicer & Fava, 1998). En el Assessment Center existen 15 dimensiones que evalúan 3 competencias, por lo que la muestra se estima en 150 personas. El criterio general de selección es que sean individuos que ejerzan o pretendan ejercer cargos directivos en establecimientos educacionales. Se difundió la invitación a participar en el sitio web de Liderazgo Escolar de la Pontificia Universidad Católica y participaron personas de Santiago y otras regiones.

Un segundo método que se utilizó para la recolección de evidencia relativa a la validez de constructo, fue por medio de la aplicación de dos test psicológicos que miden constructos relacionados a los que el Assessment Center pretende medir, es decir, test con los cuales sus resultados podrían correlacionar positivamente. Este tipo de método es referido como validez de constructo convergente. El primer test consiste en el Inventario de Liderazgo Instruccional (ILI) (Alig-Mielcarek, 2003) previamente validado en Chile (Volante, 2010), que mide competencias de liderazgo instruccional. El segundo test consiste en el FIX, que está diseñado para la evaluación de

la inteligencia fluida y para obtener una rápida estimación del CI. La inteligencia fluida corresponde a la capacidad para pensar lógicamente y resolver problemas en situaciones novedosas. El test está compuesto de 20 ítems, en los que el examinado debe seleccionar la respuesta correcta entre 5 alternativas. La aplicación de este test dura 15 minutos (este test ha sido estandarizado en Chile por el Centro de Desarrollo de Tecnologías de Inclusión (CEDETI) en el marco del proyecto FONDEF D0911238). La misma muestra descrita en el párrafo anterior será utilizada para los análisis de validez de constructo convergente.

Tanto el test ILLI como el test FIX se aplicaron de manera conjunta al Assessment Center. El objetivo fue correlacionar el Puntaje General del AC (Overall Assessment Rating OAR) con el puntaje y/o percentil de cada uno de estos test. Los argumentos para utilizar estos test se basan en el supuesto teórico de que el despliegue de las competencias requieren tanto habilidades cognitivas como sociales o relacionales. Es importante mencionar que estos no fueron utilizados para predecir el desempeño exitoso de un directivo escolar, sino para obtener evidencia de validez de constructo. En otras palabras, solo se analizó la correlación entre estos test y el Assessment Center, no con indicadores de desempeño o gestión.

Finalmente, para la recolección de evidencia en torno a la validez de criterio concurrente, se recolectaron datos SIMCE de los establecimientos a los que pertenecen los participantes, del tipo directores con tres o más años de experiencia en el cargo. Luego se analizó la diferencia de medias considerando el desempeño en el AC de los participantes y el puntaje SIMCE de sus establecimientos.

Es importante mencionar que los indicadores de gestión que se recolectaron tienen como objetivo la exploración de posibles correlaciones con el Assessment Center, sin que ello implique relaciones causales entre el comportamiento del director y los indicadores de gestión. La literatura ilustra que entre todas las variables que intervienen en los resultados escolares, la gestión directiva tiene una influencia indirecta importante. Sin embargo, es necesario desarrollar argumentos lógicos para relacionar la labor del director con los indicadores de gestión, con el fin de evitar correlaciones espurias. Por otra parte, variables como el nivel socioeconómico de los estudiantes, que ha demostrado ser la más determinante en predecir los resultados escolares, y el tamaño del establecimiento educacional, se controlaron en este estudio para aislar estos efectos del Assessment Center.

Para realizar estos análisis, se utilizó una muestra de las 164 personas. El criterio de selección para la sub-muestra utilizada para el análisis de validez concurrente es que el evaluado fuera un director en ejercicio de a lo menos 3 años de antigüedad en el cargo; así, 67 personas cumplieron con este requisito. Por otra parte, se han controlado las características de los establecimientos, de tal modo de asegurar contextos organizacionales reales y diferenciados. De esta manera, se distingue entre directores de establecimientos de bajo y alto nivel socioeconómico, y bajo y alto tamaño, y se observaron tendencias en términos de disminución, mantenimiento o aumento en el rendimiento SIMCE en los últimos tres años.

Es importante recalcar que las correlaciones con variables como SIMCE son exploratorias, y no se asume un efecto directo de la labor del director sobre dichos puntajes. En otras palabras, el SIMCE es solo una variable más a explorar. Sin perjuicio de lo anterior, en el caso del puntaje SIMCE se puede observar la tendencia bajo la gestión del director a partir de su segundo año como una forma de aislar el efecto del director anterior.

Para la aplicación del AC se seleccionaron 8 evaluadores psicólogos y se los entrenó en la aplicación del proceso de evaluación en general, y en los instrumentos y rúbricas en particular. Para calcular la confiabilidad de las mediciones se calcularon índices Kappa entre los evaluadores, estableciéndose la consistencia inter-raters (Cohen & Swerdlik, 2006).

Como se explicó anteriormente, las competencias a evaluar fueron determinadas a partir de un estudio previo, donde expertos en educación y directores jerarquizaron las competencias prioritarias para la gestión directiva con foco en el aprendizaje. Dado lo anterior, las competencias evaluadas por el Assessment Center no son todas las competencias que un director debiese desarrollar, pero sí las más importantes para impactar en los aprendizajes. En este sentido, dichas competencias aplican para cualquier tipo de establecimiento educacional. Sin perjuicio de lo anterior, igualmente se han controlado variables como el nivel socioeconómico de los estudiantes y el tamaño del establecimiento para explorar cómo funciona el Assessment Center en la evaluación de directores que se desempeñan en diferentes contextos.

En el análisis, la función utilizada para la agregación de los puntajes asignados por los evaluadores fue el promedio. Cada evaluador debía determinar el nivel de presencia de los indicadores de conducta en una escala operacionalizada de "1 a 4", donde "4" implicaba una mayor presencia del indicador conductual, y "1", la ausencia o nivel mínimo del indicador conductual. Con el objetivo de evitar la pérdida de poder estadístico, se realizó un proceso de imputación para aquellos indicadores que no fueron completados. El procedimiento para la imputación fue la asignación del promedio del resto de los indicadores conductuales que están midiendo la misma competencia que el indicador faltante. Esta imputación se realizó solo en los casos en los cuales había a lo menos tres indicadores de los cinco evaluados por cada competencia (todos los datos faltantes cumplían este criterio). De esta manera se evitó eliminar alrededor de 15 casos.

Se utilizaron 7 instrumentos para recolectar información:

- **Ficha de inscripción para participar:** los interesados en participar de una sesión de Assessment Center completan esta ficha de inscripción. La ficha contiene preguntas que permiten la caracterización de la muestra como: cargo, tamaño del establecimiento, índice de vulnerabilidad del establecimiento, ubicación rural o urbana, años de experiencia del director, entre otras variables.
- **Entrevista por competencias:** los participantes son evaluados mediante una entrevista en base a competencias guiadas, aplicada por un psicólogo. La entrevista dura 60 minutos.
- **Ejercicio de análisis y presentación estratégica:** un evaluador entrega instrucciones al postulante.

El postulante lee un caso de una escuela ficticia y debe reparar un plan estratégico para implementarla a 3 años. Se retira a preparar su presentación durante 2 horas y luego presenta su análisis y estrategia a los evaluadores, quienes pueden realizar preguntas al candidato en base a una pauta. El ejercicio completo dura 2 horas y 35 minutos.

- **Ejercicio de manejo de personal:** en este juego de rol, el participante interactúa con un actor, quien, a partir de una pauta estándar, interpreta a una jefa de UTP en una situación difícil de resolver para el postulante. El evaluador califica la interacción haciendo observación no participante. La actividad dura 30 minutos.
- **Ejercicio de observación de clases:** el participante observa el video de una clase; luego, en un juego de rol, debe retroalimentar a un actor que interpreta al docente que realiza la misma clase, mientras el evaluador participa como observador no participante. Finalmente se realiza una post-entrevista con el postulante para evaluar los aspectos no discutidos en la retroalimentación, especialmente, para evaluar los aspectos metacognitivos.
- **Test FIX de inteligencia fluida:** el participante realiza un test de razonamiento abstracto no verbal e inteligencia fluida. En el test, el examinado debe inferir el elemento que falta en una serie de estímulos de acuerdo con un patrón. El test está diseñado para que el participante no pueda apoyarse en conocimientos adquiridos durante su educación formal. El tiempo de aplicación es de 10 minutos.
- **Test ILI de liderazgo instruccional:** el participante debe declarar su grado de acuerdo con 28 afirmaciones relacionadas con su estilo de liderazgo. La aplicación dura 10 minutos.

Un total de 174 personas participaron en el AC, de las cuales 164 completaron el proceso, es decir, fueron evaluadas con todos los ejercicios del AC. Sobre la base de este número (N=164) se efectuaron los análisis que se presentan a continuación.

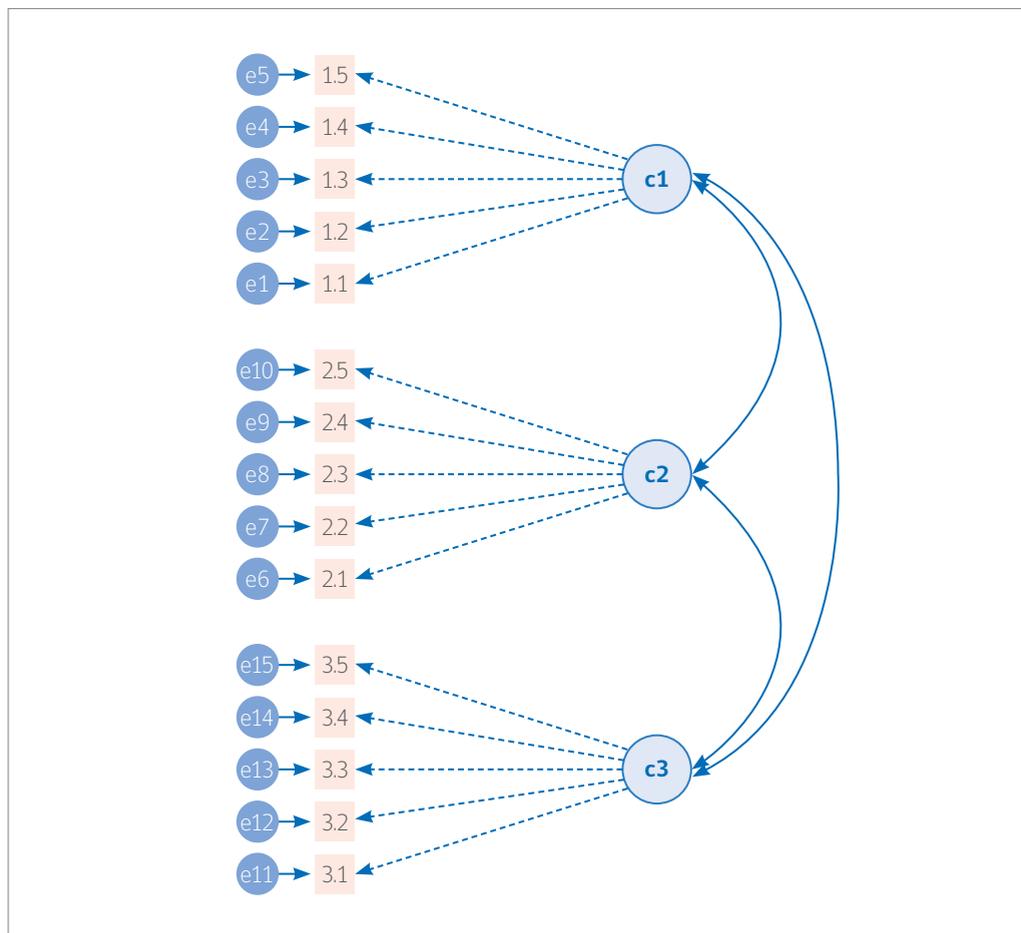
4. PRINCIPALES RESULTADOS

4.1 Resultados de validez de constructo

El modelo teórico sobre el cual se basa el AC es que los 15 indicadores conductuales miden tres competencias. Como se cuenta con este modelo previo, es decir, se sabe exactamente qué indicador carga en qué competencia, el método apropiado para testear la validez de constructo corresponde a un análisis factorial confirmatorio o CFA (Confirmatory Factor Analysis). Para realizar los análisis se utilizó el módulo AMOS del programa estadístico SPSS, que permite realizar SEM (Estructural Equation Modelling). Para ello se establecieron tres variables latentes principales (que no se miden directamente) que corresponden a las tres competencias del modelo. Se asume que cada variable latente tiene una influencia causal en un grupo de indicadores conductuales, que son las variables que se midieron directamente. Con esto, para cada variable medida (indicador conductual), se asume que además de ser influenciada por una competencia, es influenciada por un término de error (simbolizados como e_1, e_2, \dots, e_{15}). Por último, se especificaron correlaciones

entre las competencias, ya que miden capacidades de liderazgo que se espera que tengan relación entre sí. La siguiente figura muestra la especificación general del modelo:

FIGURA 1: ESPECIFICACIÓN GENERAL DEL MODELO CFA SEM

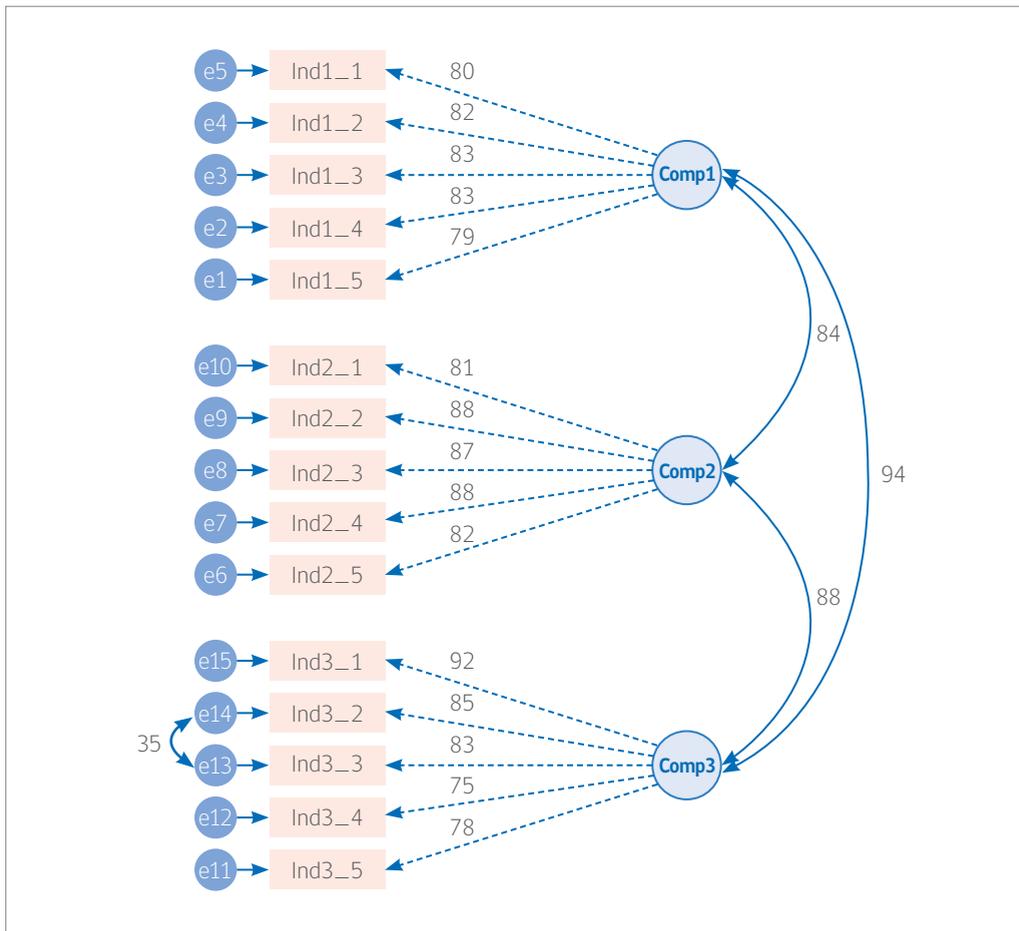


Es importante mencionar que una vez que se introduce el modelo teórico en el programa estadístico y se cargan los datos en él, nuevas relaciones pueden ser especificadas para mejorar los índices de ajuste del modelo.

Para testear la validez de constructo del AC, se probaron dos modelos. En el modelo 1, el puntaje obtenido por los participantes en cada indicador fue promediado en los distintos ejercicios. Por ejemplo, el indicador 1.1 en el ejercicio de entrevista fue promediado con el indicador 1.1 del ejercicio de manejo de personal, el indicador 1.1 del ejercicio de análisis estratégico, y el indicador

1.1 del ejercicio de observación de clases. El mismo procedimiento se utilizó con el resto de los indicadores. La siguiente figura muestra la especificación del modelo 1.

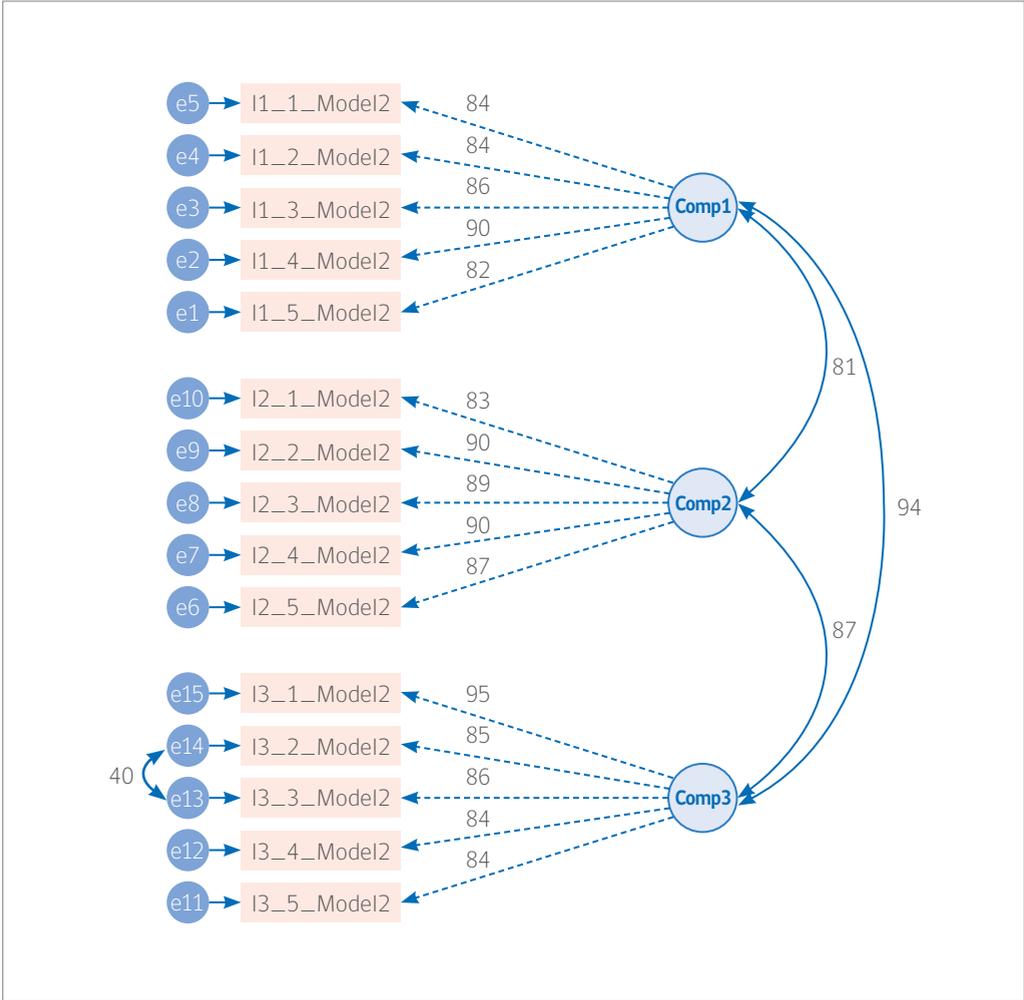
FIGURA 2: MODELO 1



De acuerdo con los datos obtenidos, se observó que la especificación de una correlación entre los términos de error e13 y e14, mejoraba los índices de ajuste del modelo 1, por lo que se procedió a especificar dicha correlación como se observa en la figura ($R = 0,35$). En general, las cargas de los indicadores en las competencias son altas, variando entre 0,75 y 0,92. Adicionalmente, se observa que las correlaciones entre las competencias son también altas (entre 0,84 y 0,94).

En el modelo 2 los indicadores fueron computados de la misma forma que en el modelo 1, pero se excluyó del análisis el ejercicio de manejo de personal.

FIGURA 3: MODELO 2



Los datos mostraron, al igual que para el modelo 1, que al agregar la especificación de correlación entre los errores e13 y e14, se obtienen mejores índices de ajuste para el modelo 2, por lo que dicha relación fue especificada como muestra la figura ($R = 0,40$). Las cargas de los indicadores conductuales en las competencias son también altas para el modelo 2 y levemente superiores que las del modelo 1, variando entre 0,82 y 0,95. Por otra parte, al igual que en el modelo 1, las correlaciones entre las competencias son bastante altas (entre 0,81 y 0,94). Dicha situación puede ser interpretada por la relación de dominio que existe entre las competencias, ya que si bien remiten a diferentes ámbitos de la gestión escolar, comparten un dominio común o general

que está dado por la gestión de establecimientos educacionales. En otras palabras, estos datos indican que una persona que es competente en uno de estos ámbitos de la gestión escolar, es muy probable que también sea competente en los otros ámbitos. También es importante destacar que las competencias fueron construidas poniendo un fuerte énfasis en la capacidad de los directores y directivos de impactar en los aprendizajes de los estudiantes, por lo que el dominio en común de las tres competencias evaluadas puede ser la gestión de los aprendizajes.

Como se exhibe en la siguiente tabla, ambos modelos demostraron un buen ajuste con los datos, es decir, los modelos especificados fueron sustentados por los datos, proporcionando evidencia para la validez de constructo del AC. Como índice de ajuste general se puede mencionar el RMSEA que se espera sea inferior a 0,05, lo que se cumple para ambos modelos (0,029 y 0,039 para los modelos 1 y 2 respectivamente). Se espera también que el Chi-cuadrado (X^2) no sea significativo, lo que también se cumple en ambos modelos. El resto de los índices están también dentro de lo esperado para ambos modelos.

TABLA 1

	Índices de ajuste general					Índices de ajuste incremental			
	X^2	GFI	RMR	RMSEA	ECVI	AGFI	TLI	NFI	CFI
Modelo 1	97,477 n.s.	0,933	0,004	0,029	1,015	0,907	0,993	0,956	0,995
Modelo 2	106,792 n.s.	0,925	0,007	0,039	1,072	0,895	0,990	0,960	0,992

Índices de ajuste de parsimonia				
	PNFI	PGFI	X^2/df	AIC
Modelo 1	0,819	0,669	1,133	165,477
Modelo 2	0,786	0,663	1,242	174,792

A la luz de estos resultados y dada la economía que significa el realizar el AC con un ejercicio menos, ambos modelos serán utilizados para testear su capacidad predictiva de otras variables. Por lo tanto, se computarán dos tipos de puntajes: uno que representa el OAR del modelo 1 (OAR1) y otro OAR que representa el puntaje en el AC de acuerdo con el modelo 2 (OAR2).

Además de la validación de constructo realizada por medio del análisis factorial confirmatorio, los puntajes del AC, tanto del modelo 1 como del modelo 2, fueron correlacionados con dos instrumentos de evaluación: un cuestionario de liderazgo instruccional (ILI) y un test de inteligencia fluida (FIX). De acuerdo con las hipótesis previas al estudio, se esperaba que los puntajes del AC correlacionaran con estos dos instrumentos, lo que sería considerado como evidencia de validez de constructo al tratarse, teóricamente, de constructos relacionados con lo que el AC mide. Sin

embargo, los resultados mostraron que no hay correlación entre el puntaje en el AC y el puntaje en el cuestionario ILLI. Los resultados tampoco mostraron una correlación entre el puntaje en AC y el puntaje en el test FIX.

Validación de criterio

Análisis de correlación

Además de la validez de constructo factorial y concurrente, se correlacionó el puntaje del AC con indicadores de gestión del establecimiento educacional del participante. Como no todos los participantes ocupaban el cargo de director, la muestra para estos cálculos fue menor (N = 67). El criterio de inclusión de estos directores para el procesamiento fue que tuvieran al menos tres años de experiencia en el cargo. El principal indicador utilizado fue el puntaje SIMCE del establecimiento educacional al cual cada director participante del AC pertenecía. De esta manera, se solicitaron al Ministerio de Educación los resultados SIMCE del año 2010, 2011 y 2012. De acuerdo con esta información, los resultados mostraron correlaciones positivas que variaron entre pequeñas a moderadas con los resultados de los distintos años, pero que no son concluyentes respecto a que el desempeño de los participantes en el Assessment Center esté relacionado con el puntaje SIMCE del establecimiento del director.

Análisis de diferencia de medias

A continuación se muestra para el modelo 1 y modelo 2 el puntaje promedio SIMCE de los establecimientos de los directores pertenecientes a los grupos de mayor y menor desempeño en el Assessment Center. El análisis de diferencia de medias revela que las diferencias entre los participantes con mayor puntaje versus los participantes con menor puntaje en AC, se asocia a diferencias en el SIMCE de sus establecimientos de al menos 10 puntos, en todas las pruebas y en todos los años.

MODELO 1: 20% SUPERIOR

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	16	261.31	31.519
Simce_Mat_4b_2009	16	253.44	33.383
Simce_Lect_4b_2010	15	274.53	20.410
Simce_Mat_4b_2010	15	259.60	28.933
Simce_Lect_4b_2011	16	267.31	28.735
Simce_Mat_4b_2011	14	269.64	32.270
Simce_Lect_4b_2012	16	268.88	27.907
Simce_Mat_4b_2012	16	264.75	33.601
Simce_Len_8vo_2009	15	255.07	28.042
Simce_Mat_8vo_2009	15	265.07	31.585
Simce_Len_8vo_2011	16	228.63	68.493
Simce_Mat_8vo_2011	16	233.44	71.005
Simce_Lect_2m_2010	8	274.88	32.202
Simce_Mat_2m_2010	8	271.13	52.127
Simce_Lect_2m_2012	8	277.25	34.932
Simce_Mat_2m_2012	8	282.00	51.873
Valid N (listwise)	6		

MODELO 2: 20% SUPERIOR

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	15	265.47	34.165
Simce_Mat_4b_2009	15	259.60	39.659
Simce_Lect_4b_2010	14	279.36	24.963
Simce_Mat_4b_2010	14	265.93	35.409
Simce_Lect_4b_2011	15	270.20	31.333
Simce_Mat_4b_2011	15	273.60	38.866
Simce_Lect_4b_2012	15	272.00	30.905
Simce_Mat_4b_2012	15	267.87	35.597
Simce_Len_8vo_2009	14	258.71	32.662
Simce_Mat_8vo_2009	14	268.21	34.508
Simce_Len_8vo_2011	15	235.40	70.693
Simce_Mat_8vo_2011	15	241.40	73.040
Simce_Lect_2m_2010	7	283.43	31.669
Simce_Mat_2m_2010	7	286.71	53.559
Simce_Lect_2m_2012	7	283.86	40.499
Simce_Mat_2m_2012	7	296.14	60.204
Valid N (listwise)	7		

MODELO 1: 20% INFERIOR

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	12	246.58	30.213
Simce_Mat_4b_2009	12	237.17	34.504
Simce_Lect_4b_2010	14	255.36	24.434
Simce_Mat_4b_2010	14	227.36	27.083
Simce_Lect_4b_2011	14	246.93	28.848
Simce_Mat_4b_2011	14	239.50	25.913
Simce_Lect_4b_2012	14	254.79	28.011
Simce_Mat_4b_2012	14	249.07	32.267
Simce_Len_8vo_2009	14	236.00	22.975
Simce_Mat_8vo_2009	14	248.07	20.915
Simce_Len_8vo_2011	12	225.00	78.918
Simce_Mat_8vo_2011	12	229.50	80.519
Simce_Lect_2m_2010	4	248.25	27.072
Simce_Mat_2m_2010	4	230.75	28.300
Simce_Lect_2m_2012	4	247.25	16.070
Simce_Mat_2m_2012	4	259.25	34.384
Valid N (listwise)	3		

MODELO 2: 20% INFERIOR

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	12	249.92	22.371
Simce_Mat_4b_2009	12	241.67	28.741
Simce_Lect_4b_2010	14	258.43	24.289
Simce_Mat_4b_2010	14	230.43	27.298
Simce_Lect_4b_2011	14	252.50	27.134
Simce_Mat_4b_2011	14	245.86	24.642
Simce_Lect_4b_2012	14	254.14	28.169
Simce_Mat_4b_2012	14	252.00	30.795
Simce_Len_8vo_2009	14	242.50	23.104
Simce_Mat_8vo_2009	14	254.57	24.566
Simce_Len_8vo_2011	12	230.75	78.435
Simce_Mat_8vo_2011	12	234.25	81.581
Simce_Lect_2m_2010	2	269.50	16.263
Simce_Mat_2m_2010	2	253.00	15.556
Simce_Lect_2m_2012	2	252.00	8.485
Simce_Mat_2m_2012	2	286.50	23.335
Valid N (listwise)	1		

MODELO 1: 50% SUPERIOR

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	29	268.76	30.073
Simce_Mat_4b_2009	29	263.31	34.757
Simce_Lect_4b_2010	28	279.43	22.720
Simce_Mat_4b_2010	28	264.00	31.108
Simce_Lect_4b_2011	29	271.21	26.426
Simce_Mat_4b_2011	27	272.04	31.936
Simce_Lect_4b_2012	29	274.48	26.364
Simce_Mat_4b_2012	29	271.72	31.187
Simce_Len_8vo_2009	27	258.81	29.824
Simce_Mat_8vo_2009	27	267.96	32.283
Simce_Len_8vo_2011	31	241.58	51.925
Simce_Mat_8vo_2011	31	247.74	54.918
Simce_Lect_2m_2010	16	275.88	32.900
Simce_Mat_2m_2010	16	277.25	47.355
Simce_Lect_2m_2012	16	277.25	34.609
Simce_Mat_2m_2012	16	288.81	50.802
Valid N (listwise)	12		

MODELO 2: 50% SUPERIOR

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	31	268.35	29.304
Simce_Mat_4b_2009	31	263.42	33.574
Simce_Lect_4b_2010	30	279.37	21.943
Simce_Mat_4b_2010	30	263.20	30.199
Simce_Lect_4b_2011	31	271.16	25.546
Simce_Mat_4b_2011	29	270.55	31.274
Simce_Lect_4b_2012	31	273.94	25.559
Simce_Mat_4b_2012	31	270.42	30.921
Simce_Len_8vo_2009	29	260.34	29.381
Simce_Mat_8vo_2009	29	267.86	31.422
Simce_Len_8vo_2011	33	240.91	50.430
Simce_Mat_8vo_2011	33	247.09	53.259
Simce_Lect_2m_2010	17	276.24	31.890
Simce_Mat_2m_2010	16	277.25	47.355
Simce_Lect_2m_2012	17	278.59	33.961
Simce_Mat_2m_2012	17	290.47	49.662
Valid N (listwise)	12		

MODELO 1: 50% INFERIOR

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	32	251.16	28.425
Simce_Mat_4b_2009	32	243.31	33.081
Simce_Lect_4b_2010	35	262.03	23.942
Simce_Mat_4b_2010	35	238.77	28.700
Simce_Lect_4b_2011	35	258.74	23.728
Simce_Mat_4b_2011	35	249.63	23.993
Simce_Lect_4b_2012	35	259.91	22.545
Simce_Mat_4b_2012	34	253.14	25.803
Simce_Len_8vo_2009	34	246.62	26.575
Simce_Mat_8vo_2009	34	255.06	27.030
Simce_Len_8vo_2011	32	243.09	53.931
Simce_Mat_8vo_2011	32	250.22	56.576
Simce_Lect_2m_2010	11	269.91	30.111
Simce_Mat_2m_2010	10	267.30	46.462
Simce_Lect_2m_2012	12	258.42	27.698
Simce_Mat_2m_2012	12	278.92	43.196
Valid N (listwise)	9		

MODELO 2: 50% INFERIOR

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	30	250.40	29.012
Simce_Mat_4b_2009	30	241.87	33.520
Simce_Lect_4b_2010	33	261.03	24.296
Simce_Mat_4b_2010	33	237.97	29.360
Simce_Lect_4b_2011	33	258.03	24.254
Simce_Mat_4b_2011	33	249.58	24.727
Simce_Lect_4b_2012	33	259.55	23.184
Simce_Mat_4b_2012	33	253.24	26.188
Simce_Len_8vo_2009	32	244.47	25.819
Simce_Mat_8vo_2009	32	254.34	27.408
Simce_Len_8vo_2011	30	243.93	55.573
Simce_Mat_8vo_2011	30	251.10	58.156
Simce_Lect_2m_2010	10	268.70	31.457
Simce_Mat_2m_2010	10	267.30	46.462
Simce_Lect_2m_2012	11	254.64	25.598
Simce_Mat_2m_2012	11	275.45	43.523
Valid N (listwise)	9		

MODELO 1: 50% SUPERIOR POR DEPENDENCIA

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	14	250.57	26.331
Simce_Mat_4b_2009	14	243.86	28.985
Simce_Lect_4b_2010	13	268.92	20.512
Simce_Mat_4b_2010	13	253.00	31.909
Simce_Lect_4b_2011	14	258.07	27.706
Simce_Mat_4b_2011	13	257.08	33.969
Simce_Lect_4b_2012	14	263.14	26.712
Simce_Mat_4b_2012	14	257.71	33.011
Simce_Len_8vo_2009	13	239.92	21.926
Simce_Mat_8vo_2009	13	248.62	24.551
Simce_Len_8vo_2011	16	246.06	29.733
Simce_Mat_8vo_2011	16	252.94	34.013
Simce_Lect_2m_2010	3	230.33	10.693
Simce_Mat_2m_2010	3	218.00	12.166
Simce_Lect_2m_2012	3	242.67	21.221
Simce_Mat_2m_2012	3	230.00	13.892
Valid N (listwise)	0		

MODELO 1: 50% INFERIOR POR DEPENDENCIA

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	18	240.17	23.458
Simce_Mat_4b_2009	18	229.22	29.309
Simce_Lect_4b_2010	18	251.50	19.675
Simce_Mat_4b_2010	18	227.11	23.676
Simce_Lect_4b_2011	18	251.17	22.833
Simce_Mat_4b_2011	18	241.11	20.745
Simce_Lect_4b_2012	18	249.56	18.718
Simce_Mat_4b_2012	18	242.50	20.261
Simce_Len_8vo_2009	18	237.61	24.108
Simce_Mat_8vo_2009	18	246.06	24.570
Simce_Len_8vo_2011	16	240.38	68.672
Simce_Mat_8vo_2011	16	243.81	70.766
Simce_Lect_2m_2010	0		
Simce_Mat_2m_2010	0		
Simce_Lect_2m_2012	0		
Simce_Mat_2m_2012	0		
Valid N (listwise)	0		

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	12	280.08	21.228
Simce_Mat_4b_2009	12	276.83	28.985
Simce_Lect_4b_2010	12	286.50	21.026
Simce_Mat_4b_2010	12	268.67	26.040
Simce_Lect_4b_2011	12	278.25	16.410
Simce_Mat_4b_2011	12	279.25	16.443
Simce_Lect_4b_2012	12	280.00	21.200
Simce_Mat_4b_2012	12	280.08	23.701
Simce_Len_8vo_2009	11	270.27	25.251
Simce_Mat_8vo_2009	11	278.09	26.872
Simce_Len_8vo_2011	12	254.42	24.912
Simce_Mat_8vo_2011	12	261.00	28.626
Simce_Lect_2m_2010	10	282.60	28.316
Simce_Mat_2m_2010	10	282.70	40.912
Simce_Lect_2m_2012	10	277.60	31.031
Simce_Mat_2m_2012	10	293.10	46.422
Valid N (listwise)	10		

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	13	264.08	29.522
Simce_Mat_4b_2009	13	261.31	30.483
Simce_Lect_4b_2010	16	269.94	19.911
Simce_Mat_4b_2010	16	246.81	23.659
Simce_Lect_4b_2011	16	264.06	20.276
Simce_Mat_4b_2011	16	255.25	20.703
Simce_Lect_4b_2012	16	270.56	22.091
Simce_Mat_4b_2012	15	264.00	27.585
Simce_Len_8vo_2009	15	256.47	27.129
Simce_Mat_8vo_2009	15	264.27	27.422
Simce_Len_8vo_2011	15	247.87	36.035
Simce_Mat_8vo_2011	15	258.47	38.989
Simce_Lect_2m_2010	10	266.60	29.557
Simce_Mat_2m_2010	9	261.56	45.357
Simce_Lect_2m_2012	11	257.64	28.911
Simce_Mat_2m_2012	11	275.64	43.709
Valid N (listwise)	8		

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	3	308.33	15.567
Simce_Mat_4b_2009	3	300.00	29.816
Simce_Lect_4b_2010	3	296.67	23.438
Simce_Mat_4b_2010	3	293.00	32.357
Simce_Lect_4b_2011	3	304.33	12.423
Simce_Mat_4b_2011	3	326.00	18.385
Simce_Lect_4b_2012	3	305.33	10.504
Simce_Mat_4b_2012	2	303.67	10.970
Simce_Len_8vo_2009	3	298.67	8.145
Simce_Mat_8vo_2009	3	314.67	7.234
Simce_Len_8vo_2011	3	166.33	144.071
Simce_Mat_8vo_2011	3	167.00	144.938
Simce_Lect_2m_2010	3	299.00	16.00
Simce_Mat_2m_2010	3	318.33	34.530
Simce_Lect_2m_2012	3	310.67	27.025
Simce_Mat_2m_2012	3	333.33	34.819
Valid N (listwise)	2		

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	1	281.00	-
Simce_Mat_4b_2009	1	263.00	-
Simce_Lect_4b_2010	1	325.00	-
Simce_Mat_4b_2010	1	320.00	-
Simce_Lect_4b_2011	1	310.00	-
Simce_Mat_4b_2011	1	313.00	-
Simce_Lect_4b_2012	1	276.00	-
Simce_Mat_4b_2012	1	271.00	-
Simce_Len_8vo_2009	1	261.00	-
Simce_Mat_8vo_2009	1	279.00	-
Simce_Len_8vo_2011	1	215.00	-
Simce_Mat_8vo_2011	1	229.00	-
Simce_Lect_2m_2010	1	303.00	-
Simce_Mat_2m_2010	1	319.00	-
Simce_Lect_2m_2012	1	267.00	-
Simce_Mat_2m_2012	1	315.00	-
Valid N (listwise)	1		

MODELO 2: 50% SUPERIOR POR DEPENDENCIA

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	15	250.47	25.377
Simce_Mat_4b_2009	15	244.40	28.010
Simce_Lect_4b_2010	14	269.36	19.774
Simce_Mat_4b_2010	14	252.57	30.699
Simce_Lect_4b_2011	15	259.13	27.013
Simce_Mat_4b_2011	14	256.50	32.708
Simce_Lect_4b_2012	15	263.27	25.745
Simce_Mat_4b_2012	15	256.07	32.444
Simce_Len_8vo_2009	14	242.29	22.845
Simce_Mat_8vo_2009	14	248.71	22.591
Simce_Len_8vo_2011	17	245.82	28.806
Simce_Mat_8vo_2011	17	252.35	33.021
Simce_Lect_2m_2010	3	230.33	10.693
Simce_Mat_2m_2010	3	218.00	12.166
Simce_Lect_2m_2012	3	242.67	21.221
Simce_Mat_2m_2012	3	230.00	13.892
Valid N (listwise)	0		

MODELO 2: 50% INFERIOR POR DEPENDENCIA

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	17	239.65	24.073
Simce_Mat_4b_2009	17	227.88	29.637
Simce_Lect_4b_2010	17	250.12	19.358
Simce_Mat_4b_2010	17	225.94	23.863
Simce_Lect_4b_2011	17	249.82	22.790
Simce_Mat_4b_2011	17	240.65	21.287
Simce_Lect_4b_2012	17	248.65	18.881
Simce_Mat_4b_2012	17	243.06	20.741
Simce_Len_8vo_2009	17	235.53	23.122
Simce_Mat_8vo_2009	17	245.82	25.306
Simce_Len_8vo_2011	15	240.27	71.081
Simce_Mat_8vo_2011	15	243.87	73.250
Simce_Lect_2m_2010	0		
Simce_Mat_2m_2010	0		
Simce_Lect_2m_2012	0		
Simce_Mat_2m_2012	0		
Valid N (listwise)	0		

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	13	279.77	20.356
Simce_Mat_4b_2009	13	276.92	27.752
Simce_Lect_4b_2010	13	286.15	20.169
Simce_Mat_4b_2010	13	167.77	25.140
Simce_Lect_4b_2011	13	277.38	16.018
Simce_Mat_4b_2011	13	277.15	17.464
Simce_Lect_4b_2012	13	279.00	20.616
Simce_Mat_4b_2012	13	279.31	22.863
Simce_Len_8vo_2009	12	271.83	24.675
Simce_Mat_8vo_2009	12	278.50	25.660
Simce_Len_8vo_2011	13	251.69	25.795
Simce_Mat_8vo_2011	13	258.69	28.643
Simce_Lect_2m_2010	11	282.55	26.864
Simce_Mat_2m_2010	10	282.70	40.912
Simce_Lect_2m_2012	11	279.64	30.204
Simce_Mat_2m_2012	11	295.27	44.625
Valid N (listwise)	10		

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	12	263.08	30.607
Simce_Mat_4b_2009	12	259.92	31.405
Simce_Lect_4b_2010	15	269.13	20.339
Simce_Mat_4b_2010	15	246.13	24.328
Simce_Lect_4b_2011	15	263.87	20.972
Simce_Mat_4b_2011	15	255.47	21.411
Simce_Lect_4b_2012	15	270.80	22.845
Simce_Mat_4b_2012	15	263.60	28.505
Simce_Len_8vo_2009	14	254.12	26.559
Simce_Mat_8vo_2009	14	262.93	27.966
Simce_Len_8vo_2011	14	249.93	36.466
Simce_Mat_8vo_2011	14	260.43	39.685
Simce_Lect_2m_2010	9	264.89	30.820
Simce_Mat_2m_2010	9	261.56	45.357
Simce_Lect_2m_2012	10	253.40	26.634
Simce_Mat_2m_2012	10	271.50	43.745
Valid N (listwise)	8		

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	3	308.33	15.567
Simce_Mat_4b_2009	3	300.00	29.816
Simce_Lect_4b_2010	3	296.67	23.438
Simce_Mat_4b_2010	3	293.00	32.357
Simce_Lect_4b_2011	3	304.33	12.412
Simce_Mat_4b_2011	3	326.00	18.385
Simce_Lect_4b_2012	3	305.53	10.504
Simce_Mat_4b_2012	2	303.67	10.970
Simce_Len_8vo_2009	3	298.67	8.145
Simce_Mat_8vo_2009	3	314.67	7.234
Simce_Len_8vo_2011	3	166.33	144.071
Simce_Mat_8vo_2011	3	167.00	144.938
Simce_Lect_2m_2010	3	299.00	16.00
Simce_Mat_2m_2010	3	318.33	34.530
Simce_Lect_2m_2012	3	310.67	27.025
Simce_Mat_2m_2012	3	333.33	34.819
Valid N (listwise)	2		

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	1	281.00	-
Simce_Mat_4b_2009	1	263.00	-
Simce_Lect_4b_2010	1	325.00	-
Simce_Mat_4b_2010	1	320.00	-
Simce_Lect_4b_2011	1	310.00	-
Simce_Mat_4b_2011	1	313.00	-
Simce_Lect_4b_2012	1	276.00	-
Simce_Mat_4b_2012	1	271.00	-
Simce_Len_8vo_2009	1	261.00	-
Simce_Mat_8vo_2009	1	279.00	-
Simce_Len_8vo_2011	1	215.00	-
Simce_Mat_8vo_2011	1	229.00	-
Simce_Lect_2m_2010	1	303.00	-
Simce_Mat_2m_2010	1	319.00	-
Simce_Lect_2m_2012	1	267.00	-
Simce_Mat_2m_2012	1	315.00	-
Valid N (listwise)	1		

4.2 Resultados de confiabilidad

Para testear la confiabilidad de los modelo 1 y 2, se procedió a calcular el Alpha de Cronbach. Para el modelo 1, este indicador fue de 0,96. En el caso del modelo 2, la confiabilidad fue de 0,97. Ambos índices son bastante altos, lo que refuerza la consistencia interna de las respuestas. Es importante señalar que este índice no mejora eliminando algún ítem, por lo que para ambos modelo se mantuvieron los 15 indicadores conductuales.

TABLA 2: CONFIABILIDAD MODELO 1

Cronbach's Alpha	Cronbach's Alpha Based on Standardized items	N of items
962	963	15

TABLA 3: CONFIABILIDAD MODELO 2

Cronbach's Alpha	Cronbach's Alpha Based on Standardized items	N of items
969	970	15

5. CONCLUSIONES

Una de las principales expectativas de los sistemas de evaluación de tipo Assessment Center, consiste en su poder predictivo en términos de estimación del desempeño futuro, tanto de candidatos en procesos de selección, como los aspectos a mejorar o monitorear en los procesos de inducción o capacitación. En consecuencia, es clave contar con diferentes fuentes de validación tanto para los instrumentos como para los procedimientos a utilizar.

Para el caso de competencias de directores escolares, es especialmente necesario contar con herramientas de selección, inducción, entrenamiento y evaluación que alineen su desempeño observado con las prioridades del marco político institucional vigente. Este énfasis es particularmente importante en contextos de desempeño directivos de vulnerabilidad, donde las demandas exceden muchas veces las capacidades y es muy difícil mantener el foco en variables claves de la gestión escolar.

De este modo, este estudio busca ofrecer medidas de validación del proceso de Assessment Center para directivos escolares, probando empíricamente sus hipótesis, instrumentos y procedimientos en 164 sujetos que se desempeñan en este segmento profesional.

En primer lugar, el análisis factorial confirmatorio de los datos ratificó la hipótesis 1 de validez de constructo del Assessment Center para la medición de competencias directivas. Los datos empíricos recolectados se estructuran y comportan como el modelo propuesto lo asumió. Los resultados mostraron que no hay correlación entre el puntaje en el AC y el puntaje en el cuestionario ILI, ni tampoco entre el primero y el puntaje en el test FIX.

En segundo lugar, por medio de un análisis de diferencia de medias se logró detectar evidencia robusta que apoya la hipótesis 2 de validez de criterio. Hasta el momento los datos indican que las diferencias entre los participantes con mayor puntaje en el AC versus los participantes con menor puntaje, se asocian con diferencias en el SIMCE de sus establecimientos en al menos 10 puntos, en todas las pruebas y en todos los años. Esta relación entre el puntaje de AC y SIMCE se presenta como un resultado exploratorio y no asume una relación causal. En futuras investigaciones con muestras más amplias se pueden profundizar en estos análisis. No se encontraron correlaciones con el test FIX de inteligencia fluida y con el test ILI.

En tercer lugar, y a nivel de procedimientos, se utilizó el análisis de confiabilidad con Alpha de Cronbach para evaluar la hipótesis 3 de confiabilidad entre evaluadores, donde se obtuvo que los instrumentos de Assessment Center demuestran un alto grado de confiabilidad y consistencia interna. Por otra parte, se detectó una alta correlación entre los puntajes de diferentes competencias evaluadas en un mismo ejercicio, más alta que la correlación entre puntajes de una misma competencia evaluada en diferentes ejercicios. Este fenómeno debe ser analizado en futuras aplicaciones para

controlar un eventual "efecto halo", que pueda afectar la sensibilidad de los evaluadores, y que ha sido observado en procesos de AC con recursos reducidos.

Finalmente, el proceso de validación realizado representa una fase clave para continuar perfeccionando la técnica y factibilidad del modelo, ya que al calibrar instrumentos y entrenar un equipo más amplio de evaluadores, se obtienen capacidades que no existen en el sector educativo y es posible buscar aplicaciones tecnológicas con mayor rigurosidad. En lo sustantivo, los indicadores de validez de criterio muestran lo importante que es el dominio específico del contexto escolar a la hora de demostrar desempeño en tareas de gestión escolar claves.

6. RECOMENDACIONES PARA LA FORMULACIÓN DE POLÍTICAS PÚBLICAS

En primer término, es necesario destacar que el desempeño de los directivos escolares contempla muchas demandas y variables de distinto tipo. Esto hace necesario enfatizar en focos de acción claves, por lo que los sistemas de desarrollo profesional de estos agentes deben estar especialmente alineados con objetivos prioritarios. En este caso se han privilegiado variables de gestión académica (instruccional) y en futuras aplicaciones se incluirían otros aspectos claves.

En consecuencia, a la hora de la implementación de Assessment Center por parte de organismos públicos, se recomienda que se consideren los siguientes aspectos para su utilización:

- Como instrumento complementario a los exigidos por ley en el proceso de selección de directores para establecimientos de educación municipal. Municipios, corporaciones municipales y servicio de alta dirección pública, pueden difundir la existencia de los instrumentos de Assessment Center entre las empresas consultoras que ejecutan los procesos de selección.
- Como instrumento para el desarrollo de una carrera directiva, específicamente:
Municipios y corporaciones municipales pueden utilizar los instrumentos de Assessment Center para conocer las necesidades de formación de directores y directivos de los establecimientos de la comuna, mediante un reporte general no sensible a diferencias individuales y que permita formular mejores planes de formación.

Directores y directivos de establecimientos municipales, municipios y corporaciones municipales pueden utilizar los instrumentos de Assessment Center para conocer las necesidades individuales de formación directiva, de manera de diseñar e implementar planes de coaching personalizados.

- Como instrumento para la formación y evaluación. El Ministerio de Educación, la CPEIP, las universidades y las organizaciones formadoras pueden utilizarlo para medir el efecto en el aprendizaje de los programas de formación aplicando los instrumentos en modalidades ex-ante y ex-post.

BIBLIOGRAFÍA

- Alig-Mielkarek, J. M. (2003). A model of school success: Instructional leadership, academic press and student achievement. Dissertation for Degree of doctor of philosophy in the Graduate School of the Ohio State University.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). Standards for educational and psychological testing (2nd ed.). Washington, DC: American Educational Research Association.
- Arribas, D. y Pereña, J. (2009). CompeTEA. Evaluación de competencias. Madrid: TEA Ediciones.
- Arrindell, W. A., & Van der Ende, J. (1985). An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement*, 9, 165-178.
- Blanco, R. (s.f). Presentación de Rafael Blanco, vocero del Sistema de Alta Dirección Pública en debate CDC-KAS: Origen, logros y desafíos de la institución. Recuperado de <http://www.cdc.cl/columna/sistema-de-alta-direccion-publica>
- Bobrow, W. & Leonards, J. (1997). Development and Validation of an Assessment Center During Organizational Change. *Journal of Social Behavior and Personality*, (12)5, pp. 217-236.
- Chan, D. (1996). Criterion and construct validation of an assessment center. *Journal of Occupational and Organizational Psychology*, 69, pp.167-181.
- Cohen, R. J. & Swerdlik, M. E. (2006). Pruebas y evaluación psicológicas: Introducción a las pruebas y a la medición (6ta Ed.). McGraw-Hill.
- Firestone, W. y Riehl, C. (2005). A new agenda: Directions for research on educational leadership. New York: Teachers College Press.
- Fleenor, J. (1996). Constructs and developmental assessment centers: further troubling empirical findings. *Journal of business and psychology*, 10(3), pp. 319-335.
- Gauger, B. B., Rosenthal D. B., Thornton, G. C., y Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72, pp. 493-511.
- Goldring, E., Porter, A., Murphy, J., Elliot, S. N. y Cravens, X. (2009). Assessing Learning-Centered Leadership: Connections to Research, Professional Standards and Current Practices. *Leadership and Policy in Schools*, (8)1, pp. 1-36.
- Gómez, J. & Stephenson, R. (1987). Validity of an Assessment Center for the Selection of School-Level Administrators. *Educational evaluation & policy analysis*, 9(1), pp. 1-7.

- Interstate School Leaders Licensure Consortium (ISLLC). (1996). Standards for School Leaders. Washington, DC: Council of Chief State School Officers.
- Krüger, M. (2009). The big five of school leadership competence in the Netherland. En *School leadership and management*, (29) pp. 109-127.
- Kuncze, J. T., Cook, W. D., & Miller, D. E. (1975). Random variables and correlational overkill. *Educational and Psychological Measurement*, 35, 529-534.
- LaPointe, M., Darling-Hammond, L., Meyerson, D. y Orr, M. (2007). Preparing school leaders for a changing world: Executive summary. Stanford, CA: Stanford University, Stanford Educational Leadership Institute (SELI).
- Muñoz, G., Farfan, J. (2011). Formación y Entrenamiento de los Directores Escolares en Chile: situación actual, desafíos y propuesta política. Fundación Chile, financiado por Ministerio de Educación: Fondo de Investigación y Desarrollo en Educación (FONIDE).
- Norton, S. (1977). The empirical and content validity of assessment centers vs. traditional methods for predicting managerial success. *The academy of management review*, 2(3), pp. 442-453.
- Organization for Economic Co-Operation and Development. (2008). Improving School Leadership: Country Background Report for Chile, Paris.
- Porter, A., Polikoff, M., Goldring, E., Murphy, J., Elliot, S. & May, H. (2010). Developing a psychometrically sound assessment of school leadership: The VAL-ED as a case study. *Educational Administration Quarterly*, 46(2), pp. 135-173.
- Robinson, V., Lloyd, C., Rowe, K. (2008). The Impact of Leadership on Student Outcomes: an Analysis of the Differential Effects of Leadership Types. *Educational Administration Quarterly*, 44(5), pp. 635-674.
- Robinson, V., Hohepa, M. y Lloyd, C. (2009). School Leadership and Student Outcomes: Identifying What Works and Why. Best Evidence Synthesis Iteration [BES]. Wellington: New Zealand Ministry of Education.
- Rupp, D., Gibbons, A., Baldwin, A., Anderson, L., Spain, S., Woo, S., Brummel, B. J., Sims, C., Kim, M. (2006). An initial validation of developmental assessment centers as accurate assessments and effective training interventions. *The psychologist-manager journal*, 9(2), pp. 171-200.
- Sackett, P. (1987). Assessment centers and content validity: some neglected issues. *Personnel psychology*, 40, pp. 13-25.
- Schmitt & Schechtman. (1990). The selection of school administrators. *Journal of personnel selection and evaluation*, 3(3), 231-238.

Second International Handbook of Educational Leadership and Administration. Dordrecht: Kluwer, pp. 1071-1099.

- Tziner, A., Ronen, S. & Hacoheh, D. (1993). A four-year validation study of an assessment center in a financial corporation. *Journal of organizational behavior*, 14, pp. 225-237.
- Uribe, M. (2010). Profesionalizar la dirección escolar potenciando el liderazgo: Una clave ineludible en la mejora escolar. *Revista Iberoamericana de Evaluación Educativa*, 3 (1º ed.), pp. 303-322.
- Van Iddekinge, C. & Ployhart, R. (2008). Developments in the criterion-related validation of selection procedures: A critical review and recommendations for practice. *Personnel psychology*, 61, pp. 871-925.
- Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, 3, pp. 231-251.
- Volante, P. (2010). Influencia instruccional de la organización escolar en logros académicos. Tesis Grado de Doctor, Escuela de Psicología U. Católica de Chile.

