



Fondo de Investigación y Desarrollo en Educación - FONIDE
Departamento de Estudios y Desarrollo.
División de Planificación y Presupuesto.
Ministerio de Educación.

Validación de proceso de Assessment Center para la selección de Directivos Escolares

Investigador Principal:

Paulo Volante B.

Investigadores Secundarios:

Dr. Antonio Mladinic
M. Ed. Magdalena Fernández
Ing. Cristian Lincovil
Dr. Michael Johaneck
Dr. Andrew Porter
MBA, Ms, Sergio Valenzuela

Institución Adjudicataria:

Pontificia Universidad Católica de Chile

Proyecto FONIDE N°:

F711295, año 2012

Informe Final realizado en el marco de proyecto FONIDE ganador del Séptimo Concurso 2012, ejecutado durante el año 2013. No citar.

Noviembre 2013

Información: Secretaría Técnica FONIDE. Departamento de Estudios y Desarrollo - DIPLAP. Alameda 1371, Piso 8, MINEDUC. Fono: 2 406 6073. E-mail: fonide@mineduc.cl

ÍNDICE

Abstract (máximo 500 palabras y debe incluir palabras claves)	3
Contextualización/Antecedentes	4
Preguntas de Investigación.	5
Hipótesis y Objetivos.....	6
Marco Teórico/ Conceptual.	7
Factibilidad económica de los Assessment Center	10
Metodología	11
Instrumentos de recolección de información	15
Resultados de Investigación	16
Resultados de validez de constructo.....	16
Cuestionario ILI y test FIX.....	20
Validación de criterio	21
Análisis de correlación.....	21
Análisis de diferencia de medias	26
Construcción y distribución de puntajes.....	30
Puntajes de corte	32
Resultados de confiabilidad	33
Conclusiones	38
Conclusiones respecto a la validez del Assessment Center	38
Conclusiones acerca de la siguiente etapa de desarrollo de Assessment Center	39
Bibliografía	42
Otros anexos	45

Abstract

Este estudio busca validar un proceso de Assessment Center 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 e 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 la influencia en la enseñanza y el aprendizaje. Adicionalmente la investigación internacional avala la creación de Assessment Centers para la selección de candidatos a un cargo debido a su capacidad 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 realizados entre mayo y julio de 2013 a una muestra de 164 participantes. Del análisis se concluye alto grado de confiabilidad del Assessment Center y validez 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.

Contextualización/Antecedentes.

La validación de instrumentos de un Assessment Center (AC) para la selección de equipos directivos en establecimientos educacionales está directamente vinculado 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 de los establecimientos educacionales. Secundariamente la validación de instrumentos busca evaluar la factibilidad de un proceso de selección enmarcado en el nuevo marco legal promovido en 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 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 a través 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 y basado 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.

Se espera que esta investigación contribuya a la política pública para la selección de directores mediante:

1. Un método de Assessment Center de selección de directores válido, basado en estándares, que predice efectivamente el futuro desempeño del postulante seleccionado.
2. Un método a disposición de la Alta Dirección Pública, de las empresas consultoras, municipios, expertos y ciudadanía.
3. Facilitar la implementación de la ley 20.501.

Preguntas de Investigación.

Esta investigación gira 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 correlacionar 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?

Hipótesis y Objetivos.

El estudio realizado propuso un **modelo de competencias de directivos escolares**, basado en la literatura internacional sobre efectos del liderazgo escolar en calidad de la enseñanza y resultados de aprendizaje de los estudiantes. Para ello se priorizaron los siguientes focos de desempeño directivo, inspirados en los Estándares ILLSC (1996, 2008):

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

Este modelo ha sido testeado utilizando un set de instrumentos que fueron diseñados para detectar diferencias en el desempeño de estas competencias, a través de un proceso de AC aplicado a directores en ejercicio y 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 comprobarse a partir de los datos recolectados en el proceso AC. Al respecto, se trabajó con tres hipótesis:

1. Hipótesis 1 de validez de constructo: La estructura de variables de AC se ajusta a los datos empíricos.
2. Hipótesis 2 de validez de criterio: El resultado de los Directores participantes en el AC correlaciona con el mejoramiento de los aprendizajes en el establecimiento del director.
3. 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 el sector de administración municipal de la educación. Los objetivos específicos son:

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

Marco Teórico/ Conceptual.

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, y 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 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 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 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 N°19.979 se define la función principal del director como “dirigir y liderar el Proyecto Educativo Institucional”, definiendo también un marco de acción y un perfil de competencias directivas. El 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 basados en estándares de desempeño.

Durante el 2011 se realizó el diseño de un modelo de Assessment Center (Centro de Evaluación, "AC" en sus siglas en inglés), que permitiera detectar y evaluar el desempeño actual y potencial de los directivos escolares (actuales y futuros), así como sus necesidades de apoyo y 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 a través 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 de director(a):

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

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

1. Entrevista en base a competencias
2. Ejercicio de análisis y presentación estratégicos
3. Ejercicio de gestión (manejo) de personal
4. 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. Adicionalmente, durante el pre-piloto se aplicaron encuestas a evaluadores y directivos participantes para evaluar 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, validez de criterio concurrente y 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 rendimiento futuro (Gómez & Stephenson, 1987; Schmitt & Schechtman, 1990; Bobrow & Leonards, 1997). El modelo de competencias en 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 como 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 en base a competencias están basados 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 Assessment Center. 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 aun mejor las evaluaciones de potencial para la gestión, es decir el talento de las personas para ser buenos directivos. En estos casos los Assessment Centers 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 validación de un Assessment Center refiere a la importancia de establecer la validez de contenido y validez de constructo de los procesos de evaluación (Sackett,1987; Chan 1996; Fleenor, 1996). Por ello el proyecto descrito a continuación, 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 Assessment Centers para directivos escolares, ya que su costo suele ser mayor al que se está dispuesto a invertir en procesos de selección de directivos escolares.

Factibilidad económica de los Assessment Center

La ley de Calidad y Equidad de la Educación (ley 20.501) señala que el proceso de selección de directores se desarrollará de manera análoga a los procesos de selección de cargos de grado 2 del sistema de Alta Dirección Pública (ADP), lo que supone la presencia de Asesoría Técnica externa al Municipio a la hora de establecer y seleccionar candidatos aptos. En este sentido, quienes estarían facultados por la ley para participar de este proceso son instituciones que se encuentren dentro del registro del Servicio Civil para el proceso de Alta Dirección Pública, utilizando para el financiamiento de esta asesoría los montos que la ley destine para ello.

El costo del actual sistema de Alta Dirección Pública es entre 12 y 14 millones de pesos por cargo de primer nivel, y 7 millones por cargos de segundo nivel (Blanco, s.f.). Dadas las características del sistema de selección de directores definido en la Ley de Calidad y Equidad de la Educación (Ley 20.501), el costo de la selección de un cargo directivo debiese bordear los 7 millones de pesos correspondientes a la selección de un cargo de segundo nivel mediante el sistema de Alta Dirección Pública. Sin embargo, este valor excede por mucho los costos del actual sistema de selección de directores el que puede llegar a un costo cercano a los 2 millones de pesos cuando se utiliza asesoría externa en la selección. Por ello, el proceso de evaluación de desempeño esperado mediante Assessment Center, previamente establecido debe tener un costo de operación menor a ese monto.

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 miden 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 para 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 las más relevantes para un proceso de *Assessment Center* para 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 el 2012 a través de un panel de expertos y de la aplicación de un prepiloto de *Assessment Center*.

Para la recolección de evidencia relacionada con la validez de constructo se han realizado dos métodos. El primero corresponde a análisis factorial confirmatorio, en el cual se evaluó si el modelo conceptual previamente establecido en relación a las competencias y sus indicadores es apoyado por los datos empíricos que se obtienen una vez aplicados los instrumentos del *Assesment 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 nuestro 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 a la literatura, la muestra apropiada para realizar el análisis factorial confirmatorio es de 10 casos por cada dimensión (Arrindell & van der Ende, 1985, p. 166; Kunce, 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 personas 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 es a través de la aplicación de dos tests psicológicos que miden constructos relacionados a los que el *Assessment Center* pretende medir, es decir, tests con los cuales los resultados del *Assessment Center* 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. Cabe mencionar que este test ha sido estandarizado en Chile por el Centro de Desarrollo de Tecnologías de Inclusión (CEDETI) en el marco del proyecto FONDEF D09I1238. La misma muestra descrita en el párrafo anterior será utilizada para los análisis de validez de constructo convergente.

Tanto el test ILI como el test FIX se aplicaron de manera conjunta al *Assessment Center*. El objetivo fue correlacionar el Puntaje General del *Assessment Center* (Overall Assessment Rating OAR) con el puntaje y/o percentil de cada uno de estos tests. Los argumentos para utilizar estos tests se basan en el supuesto teórico de que el despliegue de las competencias requiere tanto habilidades cognitivas como sociales o relacionales. Es importante mencionar que estos tests no fueron utilizados para predecir el desempeño exitoso de un directivo escolar, sino que fueron utilizados para obtener evidencia de validez de constructo. En otras palabras, solo se analizó la correlación entre estos tests 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 de 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 por fin 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(a) en ejercicio de a lo menos 3 años de antigüedad en el cargo. 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(as) de establecimientos de bajo y alto nivel socio-económico y bajo y alto tamaño, y se observaron tendencias en términos 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 sólo 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 2do. año como una forma de aislar el efecto del director anterior.

Para la aplicación del AC se seleccionaron 8 evaluadores psicólogos y se los ha entrenó en la aplicación del proceso de evaluación en *Assessment Centers* 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 *Assesment 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" significaba la ausencia del 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 sólo en los casos en los cuáles 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.

Síntesis de objetivos, actividades y dimensiones del estudio:

Tabla 1. Objetivos, actividades y muestra.

Objetivos	Actividades	Muestra
1. Evaluar la validez de constructo de los instrumentos del Assessment Center para la selección de directivos escolares.	<p>Pilotaje de Assessment Center.</p> <p>Aplicación de test FIX e inventario ILI.</p> <p>Análisis factorial confirmatorio.</p> <p>Cálculo de validez de constructo convergente.</p>	164 directores, directivos y docentes.
2. Evaluar la validez de criterio concurrente del Assessment Center en la predicción de indicadores de gestión escolar.	<p>Levantamiento de indicadores de gestión escolar.</p> <p>Segmentación de la muestra.</p> <p>Cálculo de correlaciones entre puntajes AC e indicadores de gestión escolar.</p>	Sub-muestra 67 directores en ejercicio con tres o más años de experiencia en el cargo.
3. Evaluar la confiabilidad entre evaluadores al utilizar los instrumentos de evaluación del Assessment Center.	<p>Análisis de confiabilidad inter evaluadores (índice Kappa)</p>	164 directores, directivos y docentes.

Cabe mencionar que el Assessment Center ofrece como incentivo a los participantes la entrega de un informe de desempeño que incluye recomendaciones para mejorar sus competencias directivas.

Instrumentos de recolección de información

Se han utilizado 7 instrumentos para recolectar información:

1. 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.
2. Entrevista por competencias: Los participantes son evaluados mediante una entrevista en base a competencias guiada aplicada por un psicólogo. La entrevista dura 60 minutos.
3. Ejercicio de análisis y presentación estratégica: Un evaluador entrega instrucciones al postulante. El postulante lee un caso de una escuela ficticia y debe preparar un plan estratégico para implementar a 3 años. Se retira a preparar su presentación durante 2 horas y luego presenta su análisis y estrategia los evaluadores. Los evaluadores pueden realizar preguntas al candidato en base a una pauta. El ejercicio completo dura 2 horas y 35 minutos.
4. Ejercicio de manejo de personal: En este juego de rol el participante interactúa con un actor, quien en base a una pauta estándar interpreta 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.
5. Ejercicio de Observación de clases: En este ejercicio el participante ve 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 aspectos no discutidos en la retroalimentación, especialmente, para evaluar aspectos metacognitivos.
6. 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 a 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.
7. 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.

Cada instrumento incluye sus respectivas rúbricas. En los otros anexos se incluyen detalles de cada uno de estos instrumentos.

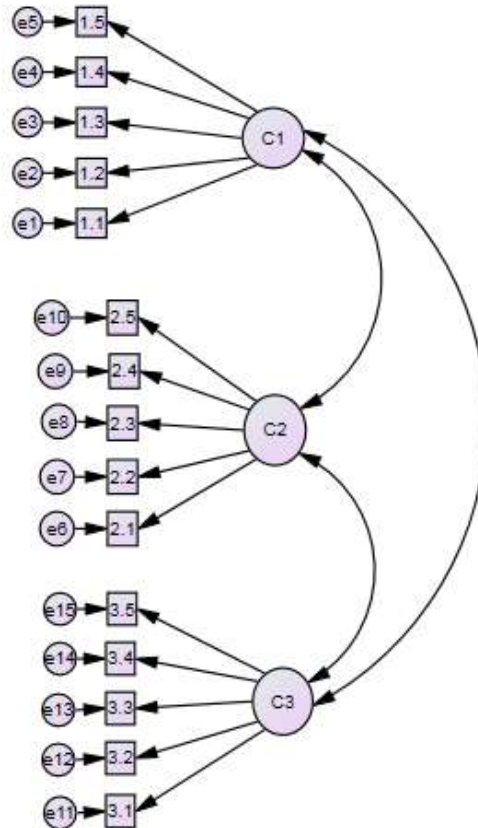
Resultados de Investigación

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.

Resultados de validez de constructo

El modelo teórico que está a la base del 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 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 (Structural Equation Modelling). Para ello se establecieron tres variables latentes principales (que no se miden directamente), que corresponden a las tres competencias del modelo, cada variable latente se asume tiene una influencia causal en un grupo de indicadores conductuales, que son las variables que se midieron directamente. Adicionalmente a 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 e1, e2,.....e15). Por último, se especificaron correlaciones entre las competencias dado que miden capacidades de liderazgo que se espera tengan relación entre sí. La siguiente figura muestra la especificación general del modelo:

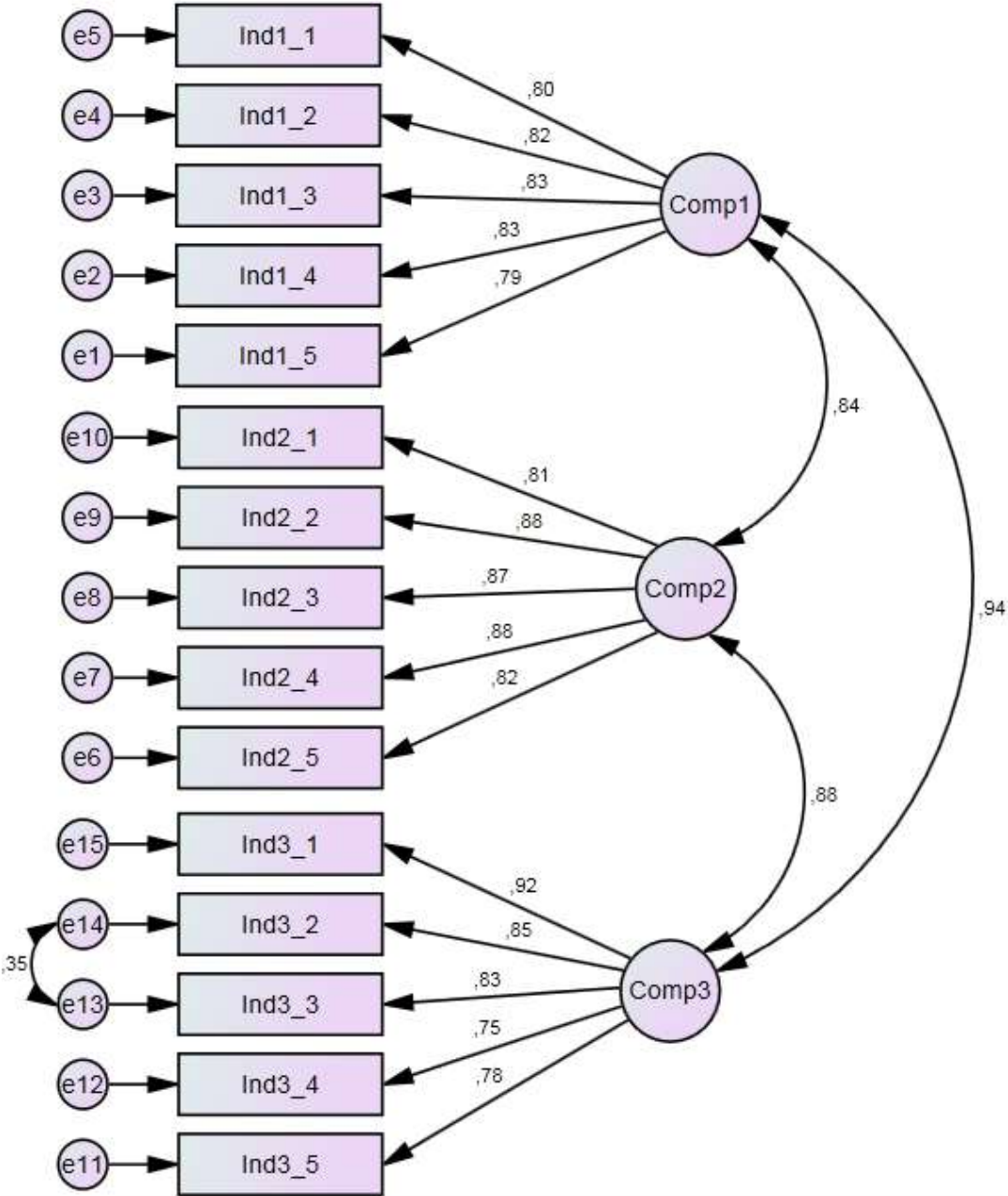
Figura 1. Especificación general modelo CFA SEM



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.

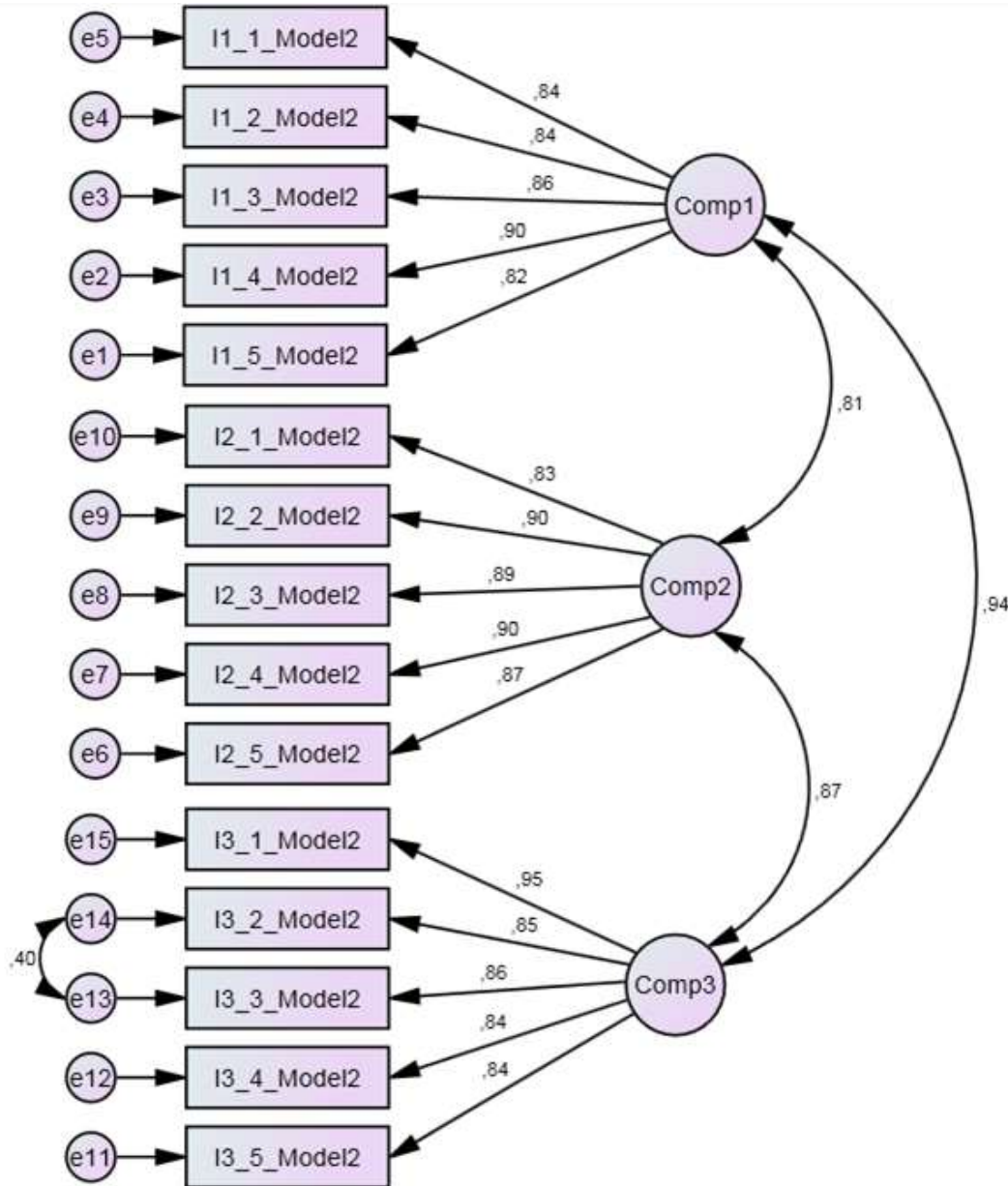
Figura 2. Modelo 1



De acuerdo a 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.

En relación con los índices de ajuste para ambos modelos, las siguientes tablas resumen los resultados encontrados.

	Índices de ajuste general					Índices de ajuste incremental			
	χ^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	χ^2/df	AIC
Modelo 1	0,819	0,669	1,133	165,477
Modelo 2	0,786	0,663	1,242	174,792

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 (χ^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.

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 al modelo 2 (OAR2).

Cuestionario ILI y test FIX

Además de la validación de constructo realizada a través del análisis factorial confirmatorio, los puntajes del AC tanto del modelo 1 como del modelo 2 fueron correlacionado 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, constructos relacionados con lo que el AC mide.

Para el caso del ILI, los argumentos para sustentar la probable asociación de este instrumento con el AC se relacionan con la similitud de las competencias evaluadas en el AC con los ítems del ILI que miden liderazgo instruccional. Por lo tanto era esperable una correlación positiva. Sin embargo, los resultados mostraron que no hay correlación entre el puntaje en el AC y el puntaje en el cuestionario ILI. La principal explicación de este resultado tiene que ver con la forma en que el cuestionario ILI se implementó, ya que se trató de un auto-reporte. Esto quiere decir que cada participante contestó el cuestionario y se autoevaluó en las conductas que cada ítem preguntaba. Es posible suponer que la situación de evaluación del AC, las expectativas y ganas de los participantes por ser bien evaluados, haya incidido en la honestidad de sus respuestas, fenómeno que se conoce como deseabilidad social. Este argumento insiste en que debiesen correlacionar ambos constructos y se atribuye al diseño de la investigación el resultado encontrado.

Correlations

		ILI	OAR_Model1	OAR_Model2
ILI	Pearson Correlation	1	-.037	-.052
	Sig. (2-tailed)		.647	.519
	N	158	158	158
OAR_Model1	Pearson Correlation	-.037	1	.960**
	Sig. (2-tailed)	.647		.000
	N	158	164	164
OAR_Model2	Pearson Correlation	-.052	.960**	1
	Sig. (2-tailed)	.519	.000	
	N	158	164	164

** . Correlation is significant at the 0.01 level (2-tailed).

En el caso del test FIX, la posible correlación se sustentaba en los datos de validación del test FIX que mostraban una relación entre este constructo y la inteligencia general. Sin embargo, los resultados tampoco mostraron una correlación entre el puntaje en AC y el puntaje en el test FIX. La explicación de la no correlación con el AC se fundamenta en el concepto de inteligencia fluida. Este constructo tiene que ver con la capacidad de las personas para resolver problemas nuevos, razonar lógicamente sobre situaciones abstractas, y establecer asociaciones. Este tipo de inteligencia es de carácter genético y no tiene su base en la experiencia y los conocimientos adquiridos como si la tiene la inteligencia cristalizada que tiene que ver con el desarrollo cognitivo a partir de las experiencias de aprendizajes. En conclusión, la no correlación del test FIX con el AC puede interpretarse como que el estar dotado genéticamente con inteligencia fluida, no es suficiente para puntuar bien en el AC.

Correlations

		FIX Percentil	OAR_Model1	OAR_Model2
FIX Percentil	Pearson Correlation	1	.070	.058
	Sig. (2-tailed)		.374	.458
	N	164	164	164
OAR_Model1	Pearson Correlation	.070	1	.960**
	Sig. (2-tailed)	.374		.000
	N	164	164	164
OAR_Model2	Pearson Correlation	.058	.960**	1
	Sig. (2-tailed)	.458	.000	
	N	164	164	164

** . Correlation is significant at the 0.01 level (2-tailed).

Validación de criterio

Análisis de correlación

Adicionalmente a 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/a, la muestra para estos cálculos fue menor (N = 67). El criterio de inclusión de estos directores/as para el procesamiento fue que tuvieran al menos tres años de experiencia en la cargo. El principal indicador utilizado fue el puntaje SIMCE del establecimiento educacional al cual cada director/a 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. Las siguientes tablas resumen estas correlaciones.

Correlations

		OSR_Model1	OSR_Model2	Simce_Lect_2m_2012	Simce_Mat_2m_2012	Simce_Lect_4b_2012	Simce_Mat_4b_2012
OSR_Model1	Pearson Correlation	1	.965**	.417*	.205	.280*	.267*
	Sig. (2-tailed)		.000	.027	.295	.025	.033
	N	67	67	28	28	64	64
OSR_Model2	Pearson Correlation	.965**	1	.376*	.160	.231	.212
	Sig. (2-tailed)	.000		.049	.416	.066	.093
	N	67	67	28	28	64	64
Simce_Lect_2m_2012	Pearson Correlation	.417*	.376*	1	.901**	.747**	.653**
	Sig. (2-tailed)	.027	.049		.000	.000	.000
	N	28	28	28	28	26	26
Simce_Mat_2m_2012	Pearson Correlation	.205	.160	.901**	1	.842**	.763**
	Sig. (2-tailed)	.295	.416	.000		.000	.000
	N	28	28	28	28	26	26
Simce_Lect_4b_2012	Pearson Correlation	.280*	.231	.747**	.842**	1	.909**
	Sig. (2-tailed)	.025	.066	.000	.000		.000
	N	64	64	26	26	64	64
Simce_Mat_4b_2012	Pearson Correlation	.267*	.212	.653**	.763**	.909**	1
	Sig. (2-tailed)	.033	.093	.000	.000	.000	
	N	64	64	26	26	64	64

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Con respecto al SIMCE del año 2012, el puntaje en el AC correlaciona significativamente con el puntaje SIMCE de lectura de 2do. Medio, y los puntajes de lectura y matemáticas de 4to. Básico.

Correlations

		OSR_Model1	OSR_Model2	Simce_Lect_4b_2011	Simce_Mat_4b_2011	Simce_Len_8vo_2011	Simce_Mat_8vo_2011
OSR_Model1	Pearson Correlation	1	.965**	.260*	.366**	-.074	-.081
	Sig. (2-tailed)		.000	.038	.003	.565	.529
	N	67	67	64	62	63	63
OSR_Model2	Pearson Correlation	.965**	1	.231	.321*	-.076	-.079
	Sig. (2-tailed)	.000		.066	.011	.554	.540
	N	67	67	64	62	63	63
Simce_Lect_4b_2011	Pearson Correlation	.260*	.231	1	.900**	-.011	.010
	Sig. (2-tailed)	.038	.066		.000	.935	.940
	N	64	64	64	62	60	60
Simce_Mat_4b_2011	Pearson Correlation	.366**	.321*	.900**	1	-.077	-.062
	Sig. (2-tailed)	.003	.011	.000		.564	.646
	N	62	62	62	62	58	58
Simce_Len_8vo_2011	Pearson Correlation	-.074	-.076	-.011	-.077	1	.981**
	Sig. (2-tailed)	.565	.554	.935	.564		.000
	N	63	63	60	58	63	63
Simce_Mat_8vo_2011	Pearson Correlation	-.081	-.079	.010	-.062	.981**	1
	Sig. (2-tailed)	.529	.540	.940	.646	.000	
	N	63	63	60	58	63	63

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Con respecto a los resultados SIMCE del año 2011, el puntaje del AC correlaciona significativamente con el puntaje de lectura y matemáticas de 4to. Básico.

Correlations

		OSR_Model1	OSR_Model2	Simce_Lect_4b_2010	Simce_Mat_4b_2010	Simce_Lect_2m_2010	Simce_Mat_2m_2010
OSR_Model1	Pearson Correlation	1	.965**	.360**	.415**	.249	.241
	Sig. (2-tailed)		.000	.004	.001	.210	.236
	N	67	67	63	63	27	26
OSR_Model2	Pearson Correlation	.965**	1	.338**	.387**	.189	.191
	Sig. (2-tailed)	.000		.007	.002	.346	.351
	N	67	67	63	63	27	26
Simce_Lect_4b_2010	Pearson Correlation	.360**	.338**	1	.926**	.853**	.855**
	Sig. (2-tailed)	.004	.007		.000	.000	.000
	N	63	63	63	63	25	24
Simce_Mat_4b_2010	Pearson Correlation	.415**	.387**	.926**	1	.794**	.817**
	Sig. (2-tailed)	.001	.002	.000		.000	.000
	N	63	63	63	63	25	24
Simce_Lect_2m_2010	Pearson Correlation	.249	.189	.853**	.794**	1	.947**
	Sig. (2-tailed)	.210	.346	.000	.000		.000
	N	27	27	25	25	27	26
Simce_Mat_2m_2010	Pearson Correlation	.241	.191	.855**	.817**	.947**	1
	Sig. (2-tailed)	.236	.351	.000	.000	.000	
	N	26	26	24	24	26	26

** . Correlation is significant at the 0.01 level (2-tailed).

Finalmente, los puntajes del AC correlacionaron significativamente con puntajes SIMCE del año 2010 de Lectura y Matemáticas de 4to. Básico. Es importante mencionar que en la mayoría de los resultados, el modelo 1 presenta mejores correlaciones con los puntajes SIMCE que el modelo 2.

El análisis de correlación no es concluyente respecto a que el desempeño de los participantes en el Assessment Center esté relacionado con puntaje SIMCE del establecimiento del director

No se observaron correlaciones significativas entre la evaluación del director realizada por los profesores en los resultados SIMCE y los resultados del AC. Esto se puede deber a las inconsistencias de las evaluaciones realizadas por los profesores, ya que dentro de un mismo año, las correlaciones de las evaluaciones de los profesores de diferentes ciclos son bajas. La única correlación significativa se produjo entre la evaluación del director realizada por los profesores de 2do. medio del SIMCE de 2012 y el percentil OAR del modelo 2 ($R = 0,39$; $p < 0,05$). Se trata de una correlación moderada positiva, que significa que a mayor percentil del puntaje OAR, mayor es la percepción de desempeño efectivo por parte de los profesores.

Tampoco se observaron diferencias significativas en la evaluación de los directores (realizada por los profesores el día de la prueba SIMCE) y las distintas agrupaciones de los resultados del AC

Correlations

		Ev_Director_2 M_2012	Ev_Director_4 B_2012	Ev_Director_8 B_2011	Ev_Director_4 B_2011	Ev_Director_2 M_2010	Ev_Director_4 B_2010	PercOAR_Mo del1	PercOAR_Mo del2
Ev_Director_2M_2012	Pearson Correlation	1	.199	.321	.118	.427*	.413*	.353	.387*
	Sig. (2-tailed)		.340	.118	.584	.030	.040	.071	.046
	N	27	25	25	24	26	25	27	27
Ev_Director_4B_2012	Pearson Correlation	.199	1	.294*	.265*	.225	.304*	.009	-.024
	Sig. (2-tailed)	.340		.020	.037	.280	.015	.944	.854
	N	25	64	62	62	25	63	64	64
Ev_Director_8B_2011	Pearson Correlation	.321	.294*	1	.198	.295	.283*	.187	.184
	Sig. (2-tailed)	.118	.020		.130	.152	.027	.146	.153
	N	25	62	62	60	25	61	62	62
Ev_Director_4B_2011	Pearson Correlation	.118	.265*	.198	1	.192	.269*	.009	.008
	Sig. (2-tailed)	.584	.037	.130		.380	.036	.944	.948
	N	24	62	60	62	23	61	62	62
Ev_Director_2M_2010	Pearson Correlation	.427*	.225	.295	.192	1	.259	.099	.036
	Sig. (2-tailed)	.030	.280	.152	.380		.211	.624	.860
	N	26	25	25	23	27	25	27	27
Ev_Director_4B_2010	Pearson Correlation	.413*	.304*	.283*	.269*	.259	1	.180	.163
	Sig. (2-tailed)	.040	.015	.027	.036	.211		.159	.201
	N	25	63	61	61	25	63	63	63
PercOAR_Model1	Pearson Correlation	.353	.009	.187	.009	.099	.180	1	.969**
	Sig. (2-tailed)	.071	.944	.146	.944	.624	.159		.000
	N	27	64	62	62	27	63	67	67
PercOAR_Model2	Pearson Correlation	.387*	-.024	.184	.008	.036	.163	.969**	1
	Sig. (2-tailed)	.046	.854	.153	.948	.860	.201	.000	
	N	27	64	62	62	27	63	67	67

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

según el percentil del director (50% superior versus 50%), aunque en la mayoría de los casos el promedio (de la evaluación del director realizada por los profesores) del 50% superior de los puntajes OAR es mayor al promedio de los directores que se encuentran en el 50% inferior de los puntajes OAR. Una de las posibles explicaciones para la falta de diferencias significativas, además de la esbozada en el párrafo anterior, tiene que ver con los pocos casos que están siendo comparados (grupos de 12 directores en algunos casos). Por la cantidad de casos no es pertinente comparar los grupos 20% superior versus 20% inferior, ya que la comparación sería entre grupos de aún menos casos que los mencionados anteriormente. La siguiente tabla muestra la correlación de OAR con la percepción del desempeño del director según profesores

A continuación se muestran las comparaciones de los promedios de las evaluaciones de los directores (realizadas por los profesores tanto de 2do. Medio como de 4to. Básico durante la aplicación del SIMCE del año 2012) agrupados según el percentil en el AC (50% superior versus 50%) tanto para el OAR del modelo 1 como el OAR del modelo 2.

Comparación de promedios de la evaluación del director en los grupos 50% superior (1) versus 50% inferior (2) de resultados en el AC (según percentil OAR modelo 1). *Evaluación realizada por los profesores de 2do.Medio durante el SIMCE 2012*

Descriptives

Ev_Director_2M_2012

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00	14	3.291031304	.4515510522	.1206820950	3.030313489	3.551749120	2.666666667	4.000000000
2.00	13	3.116737963	.4826152429	.1338533850	2.825096490	3.408379435	2.255555556	4.000000000
Total	27	3.207112288	.4661816919	.0897167084	3.022696953	3.391527623	2.255555556	4.000000000

ANOVA

Ev_Director_2M_2012

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.205	1	.205	.940	.342
Within Groups	5.446	25	.218		
Total	5.650	26			

Comparación de promedios de la evaluación del director en los grupos 50% superior (1) versus 50% inferior (2) de resultados en el AC (según percentil OAR modelo 2). *Evaluación realizada por los profesores de 2do.Medio durante el SIMCE 2012*

Descriptives

Ev_Director_2M_2012

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00	15	3.319834346	.4491975205	.1159823011	3.071077050	3.568591641	2.666666667	4.000000000
2.00	12	3.066209716	.4667806077	.1347479547	2.769631467	3.362787965	2.255555556	4.000000000
Total	27	3.207112288	.4661816919	.0897167084	3.022696953	3.391527623	2.255555556	4.000000000

ANOVA

Ev_Director_2M_2012

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.429	1	.429	2.053	.164
Within Groups	5.222	25	.209		
Total	5.650	26			

Comparación de promedios de la evaluación del director en los grupos 50% superior (1) versus 50% inferior (2) de resultados en el AC (según percentil OAR modelo 1). *Evaluación realizada por los profesores de 4to. Básico durante el SIMCE 2012*

Descriptives

Ev_Director_4B_2012

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00	28	3.402134572	.6330087112	.1196274020	3.156679418	3.647589726	1.066666667	4.000000000
2.00	36	3.305014584	.6719676941	.1119946157	3.077653427	3.532375741	1.800000000	4.000000000
Total	64	3.347504579	.6518761658	.0814845207	3.184670757	3.510338400	1.066666667	4.000000000

ANOVA

Ev_Director_4B_2012

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.149	1	.149	.346	.559
Within Groups	26.623	62	.429		
Total	26.771	63			

Comparación de promedios de la evaluación del director en los grupos 50% superior (1) versus 50% inferior (2) de resultados en el AC (según percentil OAR modelo 2). *Evaluación realizada por los profesores de 4to. Básico durante el SIMCE 2012*

Descriptives

Ev_Director_4B_2012

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00	30	3.373103378	.6207643585	.1133355474	3.141306157	3.604900599	1.066666667	4.000000000
2.00	34	3.324917403	.6866535310	.1177601091	3.085332660	3.564502146	1.800000000	4.000000000
Total	64	3.347504579	.6518761658	.0814845207	3.184670757	3.510338400	1.066666667	4.000000000

ANOVA

Ev_Director_4B_2012

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.037	1	.037	.086	.771
Within Groups	26.734	62	.431		
Total	26.771	63			

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 en el AC versus los participantes con menor puntaje en AC se asocia con 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

Descriptive Statistics

	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	37.270
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

Descriptive Statistics

	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

Descriptive Statistics

	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

Descriptive Statistics

	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	27.598
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

Modelo 2: 50% Superior

Descriptive Statistics

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	29	268.76	30.073
Simce_Mat_4b_2009	29	263.31	34.575
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		

Descriptive Statistics

	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

Descriptive Statistics

	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	35	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	56.376
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

Descriptive Statistics

	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		

En las siguientes tablas se demuestra que el Assessment Center no favorece ningún tipo de dependencia en particular.

Modelo 1: 50% Superior por dependencia

Descriptive Statistics^a

	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		

a. Dependencia = Municipal

Modelo 1: 50% Inferior por dependencia

Descriptive Statistics^a

	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		

a. Dependencia = Municipal

Descriptive Statistics^a

	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		

a. Dependencia = Particular Subvencionado

Descriptive Statistics^a

	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	16	264.00	27.585
Simce_Len_8vo_2009	15	256.47	27.129
Simce_Mat_8vo_2009	15	264.27	27.442
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		

a. Dependencia = Particular Subvencionado

Descriptive Statistics^a

	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	2	326.00	18.385
Simce_Lect_4b_2012	3	305.33	10.504
Simce_Mat_4b_2012	3	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	16.000
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		

a. Dependencia = Particular Pagado

Descriptive Statistics^a

	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		

a. Dependencia = Particular Pagado

Modelo 2: 50% Superior por dependencia

Descriptive Statistics^a

	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	23.591
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		

a. Dependencia = Municipal

Modelo 2: 50% Inferior por dependencia

Descriptive Statistics^a

	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		

a. Dependencia = Municipal

Descriptive Statistics^a

	N	Mean	Std. Deviation
Simce_Lect_4b_2009	13	279.77	20.356
Simce_Mat_4b_2009	13	276.92	27.753
Simce_Lect_4b_2010	13	286.15	20.169
Simce_Mat_4b_2010	13	267.77	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		

a. Dependencia = Particular Subvencionado

Descriptive Statistics^a

	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	254.14	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		

a. Dependencia = Particular Subvencionado

Descriptive Statistics^a

	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	2	326.00	18.385
Simce_Lect_4b_2012	3	305.33	10.504
Simce_Mat_4b_2012	3	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	16.000
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		

a. Dependencia = Particular Pagado

Descriptive Statistics^a

	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		

a. Dependencia = Particular Pagado

Construcción y distribución de puntajes

Por definición, un AC es un conjunto de instrumentos que sólo adquieren sentido cuando el puntaje es considerado en su totalidad. En otras palabras, la interpretación por ejercicio no tiene sentido desde el punto de vista teórico puesto que el puntaje agregado explica mayor varianza del fenómeno que se pretende predecir que la varianza explicada por cada ejercicio en forma individual. El término para denominar a este puntaje agregado es conocido como *OAR (Overall Assessment Ratings)*.

Para la construcción de los *OAR* de cada participante se realizó el siguiente procedimiento. El primer paso fue promediar la evaluación del observador 1 y el observador 2 por cada indicador conductual en cada ejercicio. En un segundo paso se promedió el resultado obtenido del paso anterior por cada indicador a través de los distintos ejercicios. Por ejemplo, se promediaron todos los indicadores 1.1 de los distintos ejercicios, y así sucesivamente. El resultado obtenido de los pasos anteriores en un puntaje por cada indicador conductual que en total son 15. El paso final consistió en sumar estos 15 indicadores, lo que resulta en una escala teórica de *OARs* que varía entre 15 puntos (si todos los indicadores están en nivel 1) y 60 puntos (si todos los indicadores están en nivel 4).

Como se describió anteriormente, dos modelos de construcción del *OAR* se construyeron, uno donde se considera el ejercicio de manejo de personal y otro donde se excluye. A continuación se muestran las estadísticas descriptivas para ambos modelos.

Descriptives

		Statistic	Std. Error	
OAR_Model1	Mean	37.0939	.41027	
	95% Confidence Interval for Mean	Lower Bound	36.2838	
		Upper Bound	37.9040	
	5% Trimmed Mean	37.2362		
	Median	37.3021		
	Variance	27.605		
	Std. Deviation	5.25405		
	Minimum	22.17		
	Maximum	47.96		
	Range	25.79		
	Interquartile Range	7.94		
	Skewness	-.261	.190	
	Kurtosis	-.270	.377	
OAR_Model2	Mean	39.6343	.54653	
	95% Confidence Interval for Mean	Lower Bound	38.5551	
		Upper Bound	40.7135	
	5% Trimmed Mean	39.7063		
	Median	39.2292		
	Variance	48.987		
	Std. Deviation	6.99906		
	Minimum	21.54		
	Maximum	55.75		
	Range	34.21		
	Interquartile Range	10.68		
	Skewness	-.071	.190	
	Kurtosis	-.496	.377	

Como se observa en la tabla presentada, los puntajes para el modelo 1 (OAR1) variaron entre 22,17 y 47,96, cuyo promedio es 37,09 y la desviación estándar es 5,25. Para el modelo 2 (OAR2) los puntajes variaron entre 21,54 y 55,75, con un promedio de 39,63 y una desviación estándar de 7.

Cabe señalar que la distribución para ambos modelos es normal de acuerdo al test de normalidad Shapiro-Wilk como muestra el siguiente cuadro:

Tests of Normality

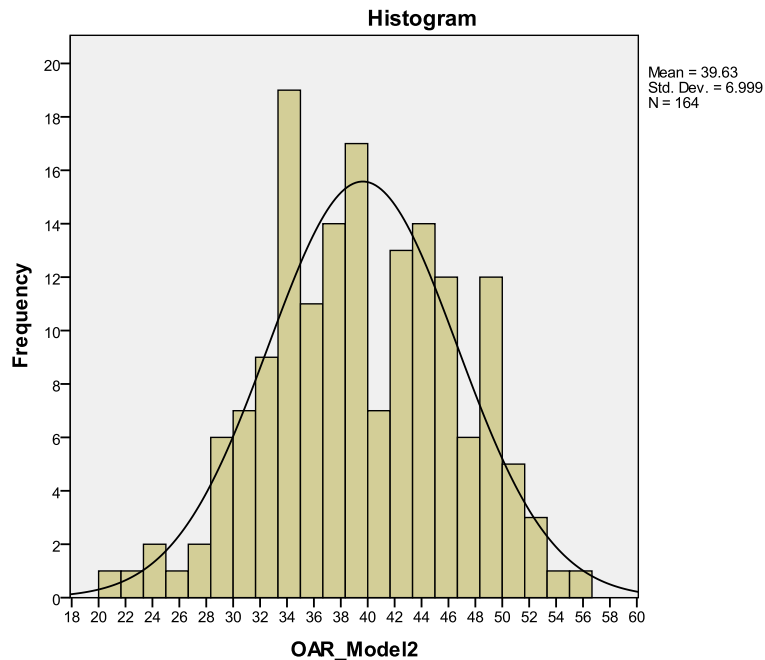
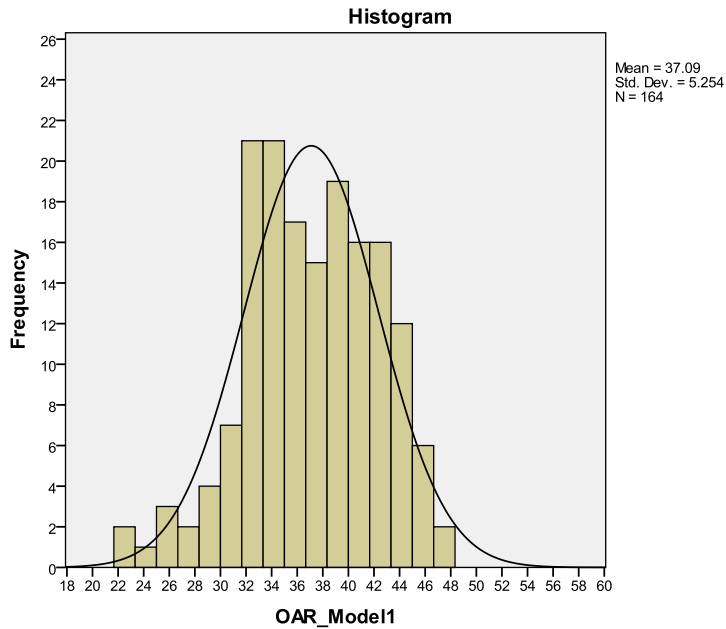
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
OAR_Model1	.059	164	.200*	.985	164	.076
OAR_Model2	.058	164	.200*	.990	164	.283

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

En este test se espera fallar en rechazar la hipótesis nula de que los datos están normalmente distribuidos. En otras palabras, se espera que el test de significancia estadística sea no significativo (n.s.), es decir, sea mayor a 0,05. A continuación se presentan los histogramas para cada modelo, donde se puede constatar visualmente que la distribución de los datos se aproxima

a la curva normal. Del mismo modo se puede observar que en el modelo 1 los puntajes se alejan más del puntaje máximo teórico (60 puntos).



Puntajes de corte

De acuerdo a la literatura de los modelos de competencias (e.g., Spencer & Spencer, 1993), el desempeño superior es definido por aquellas personas que se encuentran en el 20% superior de los rendimientos, es decir, aquellas personas que obtienen puntajes que los sitúan sobre el

percentil 80. De acuerdo con los datos obtenidos para ambos modelos, el puntaje de corte (definido por el percentil 80) es de 41,92 puntos para el modelo 1 (OAR1) y 46,08 puntos para el modelo 2 (OAR2).

La siguiente tabla muestra los puntajes que corresponden a cada percentil para ambos modelos.

		OAR_Model1	OAR_Model2
N	Valid	164	164
	Missing	0	0
Percentiles	10	30.7917	30.6875
	20	32.4792	33.8750
	30	34.0417	35.0625
	40	35.3542	37.0000
	50	37.3021	39.2292
	60	39.0208	41.9167
	70	40.4375	44.2917
	80	41.9167	46.0833
	90	43.9479	49.0417

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

Discriminación por género y edad

Como se muestra en las siguientes tablas, los instrumentos de AC no favorecen a un género en particular. Tanto en el modelo 1 como en el modelo 2, los porcentajes de aceptados y rechazados son muy similares entre hombres y mujeres. Se puede concluir que la aceptación o rechazo (de acuerdo al criterio del percentil 80), no depende del género, por lo que tanto hombres como mujeres tienen las mismas oportunidades de ser aceptados independiente de su sexo.

Modelo 1

Género * AceptacionModelo1 Crosstabulation

			AceptacionModelo1		Total
			1.00	2.00	
Género	F	Count	7	20	27
		% within Género	25.9%	74.1%	100.0%
	M	Count	10	30	40
		% within Género	25.0%	75.0%	100.0%
Total		Count	17	50	67
		% within Género	25.4%	74.6%	100.0%

Modelo 2

Género * AceptacionModelo2 Crosstabulation

			AceptacionModelo2		Total
			1.00	2.00	
Género	F	Count	6	21	27
		% within Género	22.2%	77.8%	100.0%
	M	Count	10	30	40
		% within Género	25.0%	75.0%	100.0%
Total		Count	16	51	67
		% within Género	23.9%	76.1%	100.0%

Asimismo, las siguientes tablas muestran que tener mayor o menor edad no implica mejor o peor desempeño en el AC. Tanto para el modelo 1 como el modelo 2, las diferencias entre las medias de edad de los aceptados y rechazados no son significativas. Se puede concluir que el AC no discrimina por edad.

Modelo 1

Descriptives

AceptacionModelo1			Statistic	Std. Error		
Edad	1.00	Mean	49.18	1.596		
		95% Confidence Interval for Mean	Lower Bound	45.79		
			Upper Bound	52.56		
		5% Trimmed Mean	49.53			
		Median	50.00			
		Variance	43.279			
		Std. Deviation	6.579			
		Minimum	33			
		Maximum	59			
		Range	26			
		Interquartile Range	9			
		Skewness	-.871	.550		
		Kurtosis	.880	1.063		
		2.00	2.00	Mean	53.08	1.021
				95% Confidence Interval for Mean	Lower Bound	51.03
Upper Bound	55.13					
5% Trimmed Mean	53.53					
Median	54.00					
Variance	52.116					
Std. Deviation	7.219					
Minimum	30					
Maximum	67					
Range	37					
Interquartile Range	10					
Skewness	-.929			.337		
Kurtosis	1.660			.662		

Descriptives

Edad

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00	17	49.18	6.579	1.596	45.79	52.56	33	59
2.00	50	53.08	7.219	1.021	51.03	55.13	30	67
Total	67	52.09	7.219	.882	50.33	53.85	30	67

ANOVA

Edad

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	193.312	1	193.312	3.871	.053
Within Groups	3246.151	65	49.941		
Total	3439.463	66			

Modelo 2

Descriptives

AceptacionModelo2			Statistic	Std. Error	
Edad	1.00	Mean	50.31	1.347	
		95% Confidence Interval for Mean	Lower Bound	47.44	
			Upper Bound	53.18	
		5% Trimmed Mean	50.35		
		Median	50.00		
		Variance	29.029		
		Std. Deviation	5.388		
		Minimum	41		
		Maximum	59		
		Range	18		
		Interquartile Range	9		
		Skewness	-.293	.564	
		Kurtosis	-.722	1.091	
		2.00	2.00	Mean	52.65
95% Confidence Interval for Mean	Lower Bound			50.49	
	Upper Bound			54.80	
5% Trimmed Mean	53.15				
Median	54.00				
Variance	58.753				
Std. Deviation	7.665				
Minimum	30				
Maximum	67				
Range	37				
Interquartile Range	11				
Skewness	-.957			.333	
Kurtosis	1.349			.656	

Descriptives

Edad

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00	16	50.31	5.388	1.347	47.44	53.18	41	59
2.00	51	52.65	7.665	1.073	50.49	54.80	30	67
Total	67	52.09	7.219	.882	50.33	53.85	30	67

ANOVA

Edad

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	66.378	1	66.378	1.279	.262
Within Groups	3373.085	65	51.894		
Total	3439.463	66			

Conclusiones

Conclusiones respecto a la validez del Assessment Center

Una de las principales expectativas de los sistemas de evaluación tipo Assessment Center, consiste en su poder predictivo en términos de estimación del desempeño futuro, tanto de candidatos en procesos de selección, como de aspectos a mejorar o monitorear en procesos de inducción o capacitación. En consecuencia es clave contar con diferentes fuentes de validación tanto para los instrumentos como de 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 confirmó 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 de Assessment Center propuesto lo asumió. Los resultados mostraron que no hay correlación entre el puntaje en el AC y el puntaje en el cuestionario ILI. Los resultados tampoco mostraron una correlación entre el puntaje en AC y el puntaje en el test FIX.

En segundo lugar, a través 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 en AC, se asocia con diferencias en el SIMCE de sus establecimientos de al menos 10 puntos, en todas las pruebas y en todos los años. Esta relación entre puntaje AC y SIMCE se presenta como un resultado exploratorio y no asume una relación causal. En futuras investigaciones con muestra más amplias se pueden profundizar 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 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 obtiene 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.

Conclusiones acerca de la siguiente etapa de desarrollo de Assessment Center

Teniendo en cuenta los resultados del proceso de validación y la experiencia ganada por parte del equipo, es posible delinear algunas mejoras posibles de desarrollar:

- En lo logístico, es clave resolver el problema de las inasistencias. Si bien la mayoría de los participantes inscritos en la actividad participan con entusiasmo en el AC, una minoría no se presentó a la actividad. El pago a los evaluadores se comprometió por instrumento aplicado, por lo que las ausencias no afectaron el presupuesto, sin embargo desgastaron el equipo. Esto se solucionó en gran medida mediante conversaciones telefónicas con cada participante antes de confirmar su inscripción en el AC, reconfirmaciones telefónicas y la inscripción de más interesados en una lista de espera.
- Reducir el costo de aplicación del set de instrumentos. En este proyecto el costo directo de aplicar el set de instrumentos a cada participante fue de \$450.000. A esto se debe agregar el costo de horas, traslado, alojamiento y alimentación que asumieron muchos participantes. Además de costos de infraestructura y gestión de servicio que en esta aplicación fueron aportados por la institución ejecutora. Para ser competitivo en el marco de los procesos de Alta Dirección Pública, se deben hacer esfuerzos para reducir los costos a 3 UF por participante (\$70.305) en los futuros AC.
- Dado que se trata de un proceso de aplicación entre personas, es clave asegurar mayor uniformidad de los estímulos. Los pilotos de Assessment Center presencial incluyeron ejercicios de juego de rol en el que participaron actores profesionales. Si bien los juegos de rol demostraron ser útiles para observar competencias directivas en los participantes, el estímulo presente en el ejercicio tiende a variar en cada performance del actor a cargo.
- Respecto a la validez de criterio del Assessment Center se debe continuar explorando otras correlaciones con variables que puedan estar vinculadas más directamente al desempeño del director. Los próximos análisis podrían explorar correlaciones entre el desempeño en el Assessment Center y el desempeño observado del director por parte de los docentes. En este sentido, son especialmente útiles los resultados de cuestionarios SIMCE aplicados a profesores, los cuales incluyen ítems de evaluación a los directores y al equipo directivo. Otra alternativa consistiría en aplicar a docentes y supervisores del establecimiento de los participantes, cuestionarios que incluyeran ítems que midieran específicamente las competencias directivas que incluye el Assessment Center. Adicionalmente se puede considerar estudiar relaciones entre resultados de OAR de AC con otros test de inteligencia cristalizada y con test de inteligencia emocional. Sin embargo, estos test no son tests diseñados para contextos educativos, es decir, se tendrían que validar para estos propósitos.¹

¹ Test como el Rorschach, el Lusher, y el Zullinger son clínicos, por lo que no se recomiendan para selección de personas. Están incluso prohibidos para la selección de personas en algunos países, mientras en otros,

- Solucionar el problema de “efecto halo”. Posibles mejoras podrían incluir capacitación específica sobre este efecto a los evaluadores, modificaciones a los valores de las rúbricas - desde valores absolutos a valores relativos del tipo “desempeño bajo el promedio”, “desempeño promedio” o “desempeño sobre el promedio”-, o medir este efecto pidiendo a los evaluadores que asignen un puntaje general a los postulantes.
- Virtualizar el Assessment Center. El desarrollo de una plataforma online con los instrumentos de assessment Center en formato digital y con evaluaciones asincrónicas –es decir sin interacción simultánea entre participante y evaluador- permitiría reducir los costos de aplicación, dar uniformidad a los estímulos y resolver el problema de las inasistencias. No obstante, en un futuro piloto se deben sondear las limitaciones y validez de este formato virtual de AC, considerando especialmente las dificultades que pueden tener las personas de edad avanzada o con poco manejo de herramientas TIC.
- Considerar la posibilidad de capacitar e incluir directivos y docentes como evaluadores, de manera de incluir su experiencia en la evaluación y facilitar la validación del AC por parte de los directivos y docentes.
- Mientras no se uniforme en mayor medida los estímulos el ejercicio de Manejo de Personal, se puede considerar eliminar este ejercicio del Assessment Center. El Assessment Center sin este ejercicio fue validado mediante el modelo 2. Esta medida implicaría reducir los costos y tiempos de aplicación.

simplemente no se usan para ningún propósito (ni clínico ni laboral). Se busca desincentivar su uso, por lo que no se recomienda correlacionar el AC con este tipo de test (incluidos el dibujo de la persona bajo la lluvia y los test grafológicos). Se propone adaptar y validar test de competencias para contexto educativos. Un test de esta naturaleza podría correlacionar con el puntaje del AC, sin que necesariamente sea más predictivo del desempeño que el AC. Un test de inteligencia emocional general, también podría ser una alternativa. Por otra parte, la no existencia de correlación no constituye evidencia de validez. Es importante recalcar que la principal evidencia de validez que proporcionaron los resultados, fue la de validez de constructo factorial. La evidencia de criterio es muy preliminar y debe ser mayormente explorada en futuros estudios.

Recomendaciones para la formulación de políticas públicas

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

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

- Como un instrumento complementario a los instrumentos 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 un instrumento para el desarrollo de 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 un instrumento para la formación y evaluación. Ministerio de Educación, CPEIP, universidades y organizaciones formadoras pueden utilizar este Assessment Center 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>

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.

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.

Gauger,B.B., Rosenthal D.B., Thornton, G.C., Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72: 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.

Firestone, W. y Riehl, C., 2005. *A new agenda: Directions for research on educational leadership* (Eds.). New York: Teachers College Press.

Fleener,J. 1996. Constructs and developmental assessment centers: further troubling empirical findings. *Journal of business and psychology*. 10(3), pp.319-335.

Second International Handbook of Educational Leadership and Administration, Part 2, Dordrecht: Kluwer. pp.1071-1099.

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
- Kunze, 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. Standford, CA: Standford University, Standford 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):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.M; 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 interventios. *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.
- Tziner, A., Ronen, S. & Hacohen, 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 (1e), 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, 231-251.

Volante, P., 2010. Influencia instruccional de la organización escolar en logros académicos. Tesis Grado de Doctor, Escuela de Psicología Universidad Católica de Chile.

Otros anexos

- Entrevista por competencias.
- Ejercicio de Observación de clases.
- Ejercicio de manejo de personal.
- Ejercicio de análisis y presentación estratégica.
- Test ILI de liderazgo instruccional.
- Rúbricas generales.
- Rúbricas especiales para ejercicios de observación de clases y ejercicio de análisis estratégico.
- Base completa de datos recolectados.
- Base filtrada sólo de directores.
- Difusión del proyecto en prensa.
- Ficha de inscripción para participar.
- Bases de dato SIMCE.
- Base de datos de cruce Assessment Center y SIMCE (Sólo directores participantes con 3 o más años de experiencia)