



INFORMES TÉCNICOS MIDE UC / TECHNICAL REPORTS MIDE UC

Centro de Medición MIDE UC / Measurement Center MIDE UC

IT1202

La validez de la Evaluación Docente en Chile usando como criterio estimaciones de Valor Agregado de profesores de Enseñanza Media

SANDY TAUT, EDGAR VALENCIA & JORGE ESCOBAR ⁽¹⁾

Diciembre 2012



PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE
ESCUELA DE PSICOLOGÍA

⁽¹⁾ Agradecemos el financiamiento Fondecyt Regular No. 1120441. También agradecemos la valiosa colaboración de Verónica Santelices, Ph.D., Facultad de Educación, y Jorge González, Ph.D., Facultad de Matemática, en la fase inicial de este estudio.

Resumen

Esta investigación analiza la validez de la Evaluación Docente en Chile (*Sistema de Evaluación de Desempeño Profesional Docente*, o *Docentemás*) usando el aprendizaje de estudiantes chilenos de enseñanza media como criterio de validación. Hemos combinado datos de los profesores en la Evaluación Docente con datos de los estudiantes en la prueba SIMCE (*Sistema de Medición de la Calidad de la Educación*). Para estos efectos usamos datos de una cohorte de estudiantes que se evaluaron dos veces en el tiempo, primero en 8° básico (2004) y luego en 2° año medio (2006). Para el análisis sólo se incluyeron docentes que enseñaron al mismo grupo de estudiantes durante 1° y 2° medio y que tenían datos completos para un mínimo de 10 estudiantes. Aplicamos la metodología de valor agregado (MVA) para obtener estimaciones de valor agregado (VA) para cada profesor a través de regresiones multinivel, contrastando diferentes especificaciones para los modelos. En un segundo paso correlacionamos estos índice de VA de los docentes con los resultados que ellos habían obtenido en la Evaluación Docente (tomando el resultado más reciente registrado para ellos). Nuestro estudio encuentra una relación positiva entre ambas medidas de efectividad docente; los resultados se pueden interpretar como evidencia positiva respecto de la validez de la Evaluación Docente. Las relaciones son más fuertes entre VA y resultados en el instrumento del portafolio, algo más débiles con la evaluación que entrega el director, y cercanas a cero con la entrevista del evaluador par y la autoevaluación. Las relaciones se observan más fuertemente en matemática, y también son más fuertes para una sub-muestra de docentes evaluados por ambas medidas en el mismo año.

Introducción

El presente estudio se enmarca en una agenda de investigación respecto de la validez de la Evaluación Docente (ED) en Chile. Específicamente, el estudio examina la relación entre el desempeño docente medido en un grupo de profesores a través del Sistema de Evaluación de Desempeño Docente y el aprendizaje de los alumnos de estos profesores. La característica particular de nuestro estudio es que utilizamos datos que reflejan el progreso en el aprendizaje ya que contamos con dos mediciones en el tiempo del mismo grupo de estudiantes (resultados SIMCE de 8° básico en 2004 y 2° medio en 2006). La disponibilidad de estos datos longitudinales sobre el aprendizaje de los alumnos abre la posibilidad de aplicar modelos de valor agregado (MVA) para estimar el “efecto profesor” (McCaffrey, Lockwood, Hamilton, & Koretz, 2003; Lissitz, 2005). Después nos preguntamos cómo se relaciona ese “efecto profesor” con los resultados de los profesores en la Evaluación Docente.

¿Por qué es relevante examinar la validez de un sistema de medición tal como el Sistema de Evaluación de Desempeño Docente? En el campo de la medición educacional la validez es un concepto clave que permite llegar a un juicio respecto de si decisiones basadas en los resultados de una medición son defendibles o más bien inadecuadas (AERA, APA & NCME, 1999; Kane, 2006). Esta validación es especialmente crucial en contextos de decisiones que traen importantes consecuencias (*high stakes*) para los evaluados, como es el caso de la Evaluación Docente en Chile. Ya existen varios estudios de validación de la Evaluación Docente y sus resultados se resumen en Taut, Santelices y Stecher (2012).

Sin embargo, todavía faltan estudios respecto de la relación entre los resultados de la Evaluación Docente y el aprendizaje de los estudiantes como un criterio de éxito relevante de la labor docente. Existen estudios que han relacionado los resultados de la Evaluación

Docente con los puntajes SIMCE tanto a nivel curso como escuela (Bravo, Falck, González, Manzi & Peirano, 2008; Manzi, Strasser, San Martín & Contreras, 2008; Alvarado, Cabezas, Falck & Ortega, 2012) pero desde esta investigación el aprendizaje se entiende como el progreso de los alumnos a través del tiempo, utilizando datos longitudinales (también llamados panel). No existen estudios que hayan utilizado datos panel en ese contexto.

Modelos de valor agregado intentan capturar la contribución específica de un profesor al aprendizaje de sus alumnos usando al menos dos mediciones de desempeño de los mismos alumnos en el tiempo. Los modelos de valor agregado controlan al menos el punto de partida de los alumnos, y en muchos casos, incluyen variables adicionales de características de los alumnos y del contexto, todos los cuales describen aspectos que influyen en la labor docente pero que están fuera del control del profesor (ver Braun, 2005). El VA es relativo al grupo de profesores incluidos en el análisis. Controlando por el punto de partida de los alumnos se puede estimar la contribución del docente al aprendizaje de forma más adecuada (Raudenbush, 2004; Bryk & Raudenbush, 1992; McCaffrey et al., 2003; Lissitz, 2005; Kane, Taylor, Tyler, & Wooten, 2010). Discutimos las ventajas y limitaciones de estos MVA en más detalle en la sección de antecedentes.

En un sentido metodológico, la contribución del presente estudio recae en la comparación de distintos modelos de regresión (con y sin controlar por el rendimiento previo de los alumnos), además de examinar el impacto de controlar por características del estudiante y de la sala de clase en las estimaciones.

Nuestras preguntas de investigación son las siguientes:

1. ¿Existe evidencia que apoye la validez de la Evaluación Docente usando aprendizaje de los estudiantes como criterio de desempeño docente?

2. ¿Cuál es la correspondencia de los resultados provenientes de distintos modelos estadísticos que se usan para determinar el efecto del profesor?

La primera hipótesis es que los resultados de la Evaluación Docente están positivamente asociado con estimaciones de VA del profesor. ¿Qué esperamos en términos de la fuerza de la relación entre ambas medidas? No esperamos una relación perfecta, y ni siquiera fuerte, ya que la Evaluación Docente se enfoca en prácticas de enseñanza del docente y no pretende reflejar el aprendizaje de los alumnos. Sin embargo, debe haber alguna relación entre la calidad de las prácticas del docente y el aprendizaje de sus alumnos, por lo que parece uno de los criterios de validez relevantes de estudiar. En este estudio, el aprendizaje de los alumnos en pruebas estandarizadas es el criterio para determinar la validez convergente de los resultados de la Evaluación Docente. Los expertos están de acuerdo que el impacto de un buen profesor en el aprendizaje de sus alumnos es considerable (la variable más fuerte después de características sociodemográficas de los alumnos) (Kupermintz, 2003; Wenglinsky, 2002; Sanders & Horn, 1998; Nye, Konstantopoulous & Hedges, 2004). Esperamos entonces una relación de tamaño pequeño a mediano (correspondiente a un coeficiente de correlación de Spearman de alrededor de 0.3).

La segunda hipótesis es que los resultados de distintos modelos tienen una alta correlación entre sí, y que MVA se relacionan más fuertemente con resultados de la Evaluación Docente que otros tipos de modelos, ofreciendo evidencia de validez más relevante.

El informe tiene la siguiente estructura. Empezamos con una revisión de la literatura respecto de la metodología de valor agregado y estudios con propósitos similares del contexto de EE.UU. Seguimos con una descripción de nuestra metodología (muestra, bases de datos, variables, especificación de modelos). La sección de resultados muestra resultados descriptivos, resultados de las estimaciones de VA, la correspondencia entre distintos tipos de modelos, y las correlaciones entre el “efecto profesor” y los resultados de la ED. En la

discusión presentamos nuestra interpretación de los resultados y ofrecemos conclusiones y próximos pasos de la investigación en el tema.

Las fortalezas y limitantes de la metodología de Valor Agregado para estimar el efecto del profesor en el aprendizaje de sus alumnos

Fortalezas de los Modelos de Valor Agregado (MVA)

Los MVA tienen utilidad en el ámbito de la educación ya que pueden ayudar de aislar el efecto de instituciones y actores educativos en el aprendizaje de los estudiantes, descontando la influencia de factores no controlables por ellos, como por ejemplo el nivel de logro al entrar o las características socioculturales de los estudiantes (Braun, Chudowsky & Koenig, 2010; McCaffrey, Lockwood, Koretz, & Hamilton, 2003). Por ende, tienen utilidad en sistemas de responsabilización educacional, cuando uno de sus propósitos fundamentales es monitorear la efectividad de las instituciones y actores educacionales en términos del aprendizaje que logran en los estudiantes a su cargo, pero sin penalizar el trabajo con estudiantes de vulnerabilidad social y de un bajo rendimiento inicial. En ese sentido, el uso de MVA constituye un enfoque metodológico más justo a la hora de comparar la efectividad de actores escolares, siempre y cuando los consideren variables contextuales y el rendimiento previo de los estudiantes (San Martín & Carrasco, 2012).

Los MVA centran la atención de la labor educacional en el incremento del aprendizaje de los alumnos, específicamente en el progreso individual de los estudiantes (Braun, 2005). Los MVA presentan similar confiabilidad y validez que otros métodos de evaluación docente (Glazerman, Loeb, Goldhaber, Staiger, Raudenbush, & Whitehurst, 2010). Si se utilizan con precaución, pueden añadir información útil para el análisis comprensivo y minucioso del progreso en el aprendizaje de los estudiantes, apoyando inferencias sobre la influencia de las prácticas de profesores y colegios en el aprendizaje (Baker et al., 2010). Asimismo, pueden ser una de las fuentes de información para identificar profesores que requieren desarrollo profesional y colegios con baja efectividad (Braun, 2005).

Dado que su confiabilidad está limitada, actualmente se recomienda ampliamente utilizar los MVA sólo para la investigación y para decisiones de bajo impacto (*low stakes*) (Braun et al., 2010; Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2011). Asimismo, los MVA pueden ayudar a validar otras medidas utilizadas para la evaluación docente (Darling-Hammond et al., 2011).

Debilidades de los MVA

Como en el caso de cualquier método de evaluación docente, la asignación no aleatoria de estudiantes a profesores constituye un problema en el caso de los MVA (Baker et al., 2010). Aunque los MVA intentan controlar la falta de aleatoriedad para diferencias pre-existentes, este intento de control puede subestimar o sobreestimar el efecto del colegio o profesor (Braun et al., 2010). Existe evidencia indicando que la falta de aleatoriedad podría variar como una función del rendimiento más reciente de los estudiantes y que puede resultar en **estimaciones inestables de año a año** (Steele et al., 2010; Newton, Darling-Hammond, Haertel & Thomas, 2010). Del mismo modo, los MVA no resuelven completamente el problema del **sesgo de las variables omitidas**. Dado que el efecto real del profesor puede estar correlacionado con las características de los estudiantes que enseña, los enfoques actuales de MVA no pueden separar los efectos contextuales de estos efectos reales (Braun et al., 2010). Si bien los ajustes de covarianza se utilizan para dar cuenta de los factores que podrían confundir las estimaciones del efecto profesor o colegio, controlar por puntajes de años anteriores y otras co-variables disponibles **no garantiza que los efectos estimados de los profesores sean efectos causales** (Ballou, Sanders, & Wright, 2004; McCaffrey et al., 2003).

Otras posibles amenazas a la robustez de los MVA tienen que ver con las características de las pruebas de aprendizaje. Los MVA se basan en puntajes de pruebas estandarizadas, pero éstas **reflejan medidas incompletas de los objetivos educativos** (Braun et al., 2010; Baker et al., 2010; McCaffrey et al., 2003). Además, sólo pueden aplicarse los VAM a **profesores cuyas asignaturas y grados pueden ser evaluados a través de pruebas estandarizadas** (Steele et al., 2010). Adicionalmente, el **escalamiento**

vertical año a año de las pruebas es usualmente inadecuado (Baker et al., 2010; Braun et al., 2010; McCaffrey et al., 2003). Finalmente, para proporcionar una clasificación consistente del valor agregado de colegios, profesores o programas, un supuesto importante que subyace al análisis de valor agregado utilizando modelos de regresión, es que las pruebas utilizadas son reportadas en una **escala de intervalo**. La mayor parte de las pruebas no cumple este requisito, al menos no de forma exacta, pero influye en la validez de las interpretaciones de valor agregado (Braun et al., 2010). Aún cuando la calidad de los datos es importante para cualquier sistema de evaluación, los requisitos para los MVA tienden a ser mayores porque se necesitan datos longitudinales con al menos dos puntos de medición en el tiempo (Braun et al., 2010).

Como en todo sistema de responsabilización con consecuencias de alto impacto, el uso de MVA en estos sistemas puede incentivar a que los profesores distorsionen su enseñanza para evitar las consecuencias negativas que se asocian a bajos índices de valor agregado (Baker et al., 2010), llevando tanto a **consecuencias deseadas como no deseadas** (Braun et al., 2010). De hecho, existe gran preocupación de los expertos sobre su uso exclusivo al tomar decisiones de alto impacto (Braun, et al., 2010; Darling-Hammond et al., 2011). Se menciona reiteradamente la **necesidad de más investigación** antes que los MVA puedan ser utilizados para apoyar decisiones importantes sobre profesores (Braun, 2005) o colegios (McCaffrey et al., 2003).

Estudios relevantes del contexto norteamericano

Un estudio de Milanowski (2004) presenta los resultados de un análisis relacional entre los resultados de evaluación docente basado en el Framework for Teaching de Charlotte Danielson (1996) y el logro de los estudiantes en pruebas estandarizadas en las escuelas públicas de Cincinnati, EE.UU. Usando MVA, el autor correlacionó la diferencia entre el logro predicho y el obtenido en estudiantes de tercer a octavo grado y la clasificación de evaluación docente (N=212). Se encontraron correlaciones pequeñas a moderadas positivas en cada asignatura testada. Cuando estas se combinaron a través de grados dentro de las asignaturas, el promedio fue de 0.27 para ciencia, 0.32 para lectura, y

0.43 para matemáticas. Muchos estudiantes fueron excluidos del análisis por no haber participado en ambas evaluaciones, o por no haber podido ser asociado a un profesor.

Se realizó un análisis de dos pasos. El primero se estimó la efectividad docente modelando el rendimiento de los estudiantes al finalizar el año, controlando por logro previo en la asignatura y características del estudiante que pudieran influenciar los puntajes (educación especial, antigüedad en la escuela, información socioeconómica, género y raza, incluidas como control en el nivel 1 del modelo lineal jerárquico). De este modelo se obtuvieron los residuos de los interceptos empíricos de Bayes (EB) para el nivel 2 (profesor). Estos residuos fueron tomados como la medida del logro promedio de los estudiantes para cada profesor. El segundo paso incluyó la correlación del residuo del intercepto EB con el puntaje de evaluación de los profesores, pudiendo calcular las correlaciones sólo para aquellos docentes (23% del total) cuyos puntajes de evaluación estaban disponibles.

El autor concluyó que los puntajes de un riguroso sistema de evaluación docente puede estar sustancialmente relacionado al logro de los estudiantes y proporcionar evidencia de validez para este sistema de evaluación docente.

En el mismo contexto del distrito escolar de Cincinnati, Kane et al. (2010) investigaron la relación entre los resultados del mismo sistema de evaluación docente (basado en observaciones de clase y un portafolio) y el logro de los estudiantes utilizando MVA. Los datos contienen registros de la evaluación docente para 2.071 profesores en un periodo de 8 años y datos de panel de los estudiantes para los 8 años. Las observaciones por año de los estudiantes incluyen género, raza, nivel de inglés, necesidades educativas especiales, curso, asignación de profesores por asignatura, y cuando aplicaba, puntajes de pruebas estandarizadas. Se utilizaron todos los datos disponibles para construir estimaciones de VA para los profesores, incluyendo el logro de los estudiantes en años anteriores. Luego se dividieron los profesores en cuartiles según su puntaje de valor agregado.

Se encontró que las observaciones en el aula pueden capturar elementos de la enseñanza que se relacionan con el rendimiento escolar. Las estimaciones muestran una

relación positiva y no trivial entre las puntuaciones de la evaluación docente y el crecimiento del rendimiento de los estudiantes. Así, pasar de una calificación global de “Basico” a “Competente” o de “Competente” a “Distinguido” se asocia con aumentos en el logro de los estudiantes desde un sexto a un quinto de una desviación estándar. Relacionar las prácticas observadas en el aula con el crecimiento del logro también ofrece una idea respecto a qué tipos de prácticas pueden ser importantes en el aumento del rendimiento de los estudiantes. Estos resultados apoyan la idea de que varias medidas alternativas de la eficacia de los profesores pueden predecir efectos futuros en el logro estudiantil por sobre una única medida. Esto es verdadero cuando puntajes de observación de aula se llevan a un modelo que solo considera medidas de logro estudiantil. Lo mismo sucede, y tal vez más aún, cuando las medidas del logro estudiantil se incluyen en un modelo que sólo considera observaciones de aula. El aumento en el logro de “estudiantes previos” de un profesor, son un buen predictor de futuros aumentos en el logro, pero es probable que al medir las practicas de aula mejore dicha predicción.

Hill, Kapitula y Umland (2011) estudiaron en qué medida los índices de VA son consistentes con otros indicadores de calidad docente, tales como observaciones, entrevistas, y pruebas de conocimiento. Se tomó una muestra de profesores de matemática (n=24) en segundo ciclo a quienes se les realizaron seis observaciones codificadas utilizando el instrumento *Mathematical Quality of Instruction* (MQI), dos entrevistas, una encuesta de conocimiento *Mathematical Knowledge for Teaching* (MKT), y una evaluación realizada por los calificadores a los docentes llamada *Lesson-based MKT* (LBMKT) basada en el conocimiento matemático manifestado por los mismos al momento de hacer la clase. Los estudiantes entre 3° y 8° grado fueron evaluados a través del *State Mathematics Assessment* (SMA). Se construyeron índices de VA para todos los profesores de matemática en el distrito con información completa de 9 ó más estudiantes para el año del estudio.

Todos los modelos explorados fueron *covariate adjustment models*, los efectos de la escuela fueron tratados como fijo o aleatorio, mientras que el del profesor fue generalmente aleatorio. Se reportaron resultados de tres modelos: (1) Modelo sencillo, ajustado por el

resultado de los estudiantes en años anteriores más el efecto aleatorio del profesor; (2) Modelo efecto fijo de la escuela, ajustado por el resultado de los estudiantes en años anteriores, el efecto aleatorio del profesor y adicionalmente el efecto fijo del colegio; (3) Modelo ajustado por los antecedentes de los estudiantes, se ajustó el resultado de los estudiantes en años anteriores, el efecto aleatorio del profesor, y los antecedentes del estudiante, pero sin el efecto fijo del colegio. Aunque los tres modelos tuvieron una alta correlación, hubo algunas variaciones. La correlación entre el modelo simple y el modelo ajustado a los antecedentes del estudiante fue de 0.93, y la correlación del modelo simple y el modelo con el colegio como efecto fijo fue de 0.90.

El análisis encontró una fuerte correlación positiva entre los recursos docentes, medidos por la encuesta de MKT, y la calidad de su instrucción matemática. Las correlaciones de Spearman muestran que las estimaciones de la encuesta y LBMKT de los profesores están correlacionados a 0.72, LBMKT y MQI están correlacionados a 0.90, y MQI se correlaciona con la encuesta en 0.58. Las correlaciones parciales de Spearman entre las Medidas de Observación del profesor y los puntajes de VA reportadas después de ajustar por una serie de criterios (promedio del año académico en evaluaciones estatales de matemáticas, porcentaje de estudiantes que califican para almuerzo gratis/precio reducido, porcentaje de alumnos con educación especial, entre otros) son de 0.56 para el modelo 1 y el MQI, de 0.57 para el mismo modelo y LBMKT. Mientras que el modelo 2 tuvo correlaciones de 0.22 para MKT, 0.30 para MQI y 0.39 para LBMKT. Finalmente, el modelo 3 tuvo una correlación de 0.22 con el MKT, 0.52 con el MQI y 0.55 con LBMKT.

El estudio evaluó dos proposiciones fundamentales para la validez de los actuales usos de las puntuaciones de valor agregado: que estos resultados convergen con otros indicadores de instrucción y se apartan de constructos teóricamente no relacionados, y que los resultados son lo suficientemente precisos para respaldar las decisiones tomadas en casos específicos. En resumen, la evidencia indica que los índices de VA convergieron con las evaluaciones realizada por expertos, pero también descubrieron que estos resultados se correlacionaron de alguna forma con aspectos de la composición de los estudiantes.

Metodología

Base de Datos para los Análisis de Valor Agregado

Un aspecto muy importante en esta investigación fue la creación de una base de datos para realizar los análisis de Valor Agregado a nivel profesor. Una base de datos con estas características es difícil de encontrar en Chile y debió ser construida vinculando mediciones de rendimiento de los alumnos en pruebas estandarizadas en dos momentos del tiempo junto con información que relaciona cada alumno con el profesor que impartió clases en cada asignatura. La cohorte de alumnos estudiada es la que cursó 8° básico en 2004, 1° medio en 2005 y 2° medio en 2006 quienes fueron evaluados en la prueba SIMCE 2004 y 2006. Junto a la información de rendimiento se consideró en los análisis otra información individual de los alumnos proveniente de las bases de datos de la misma prueba SIMCE y de otras bases de datos pertenecientes del Ministerio de Educación. Finalmente, las estimaciones de Valor Agregado de cada docente modeladas en esta base de datos fueron asociadas con los resultados de los mismos docentes en la Evaluación Docente.

La base de datos e información utilizada se describe en las siguientes sub-secciones.

Nivel Individual

SIMCE 2004 8° Básico y SIMCE 2006 2° medio: corresponden a los resultados individuales de las pruebas SIMCE de matemática y lenguaje. El SIMCE es el Sistema Nacional de Evaluación de resultados de aprendizaje del Ministerio de Educación de Chile. Este programa aplica pruebas censales a estudiante de 4° básico todos los años, y alternadamente a 8° básico y 2° medio (hasta 2011). Estas pruebas fueron rendidas dos veces por la misma cohorte en 2004 y 2006, y los resultados de ambas mediciones fueron vinculadas por un identificador del alumno disponible en las mismas bases de datos. De este modo, disponemos de dos mediciones de rendimiento de los estudiantes, requisito para realizar estimaciones de Valor Agregado. Estas bases de datos contienen además antecedentes personales y socioeconómicos de los alumnos y sus familias. Las variables utilizadas para

este estudio son: sexo (hombre), nivel educacional de la madre (años de escolaridad), alumnos con padre o madre indígena (si o no), y si la familia es beneficiada con subsidios del Estado (si o no).

Rendimiento y matricula 1° medio 2005 y 2° medio 2006: contiene información individual que incluye para cada año y grado el promedio de notas general del alumno, asistencia, si ha repetido (si o no), si es integrado (si o no) y si pertenece a un grupo diferencial (si o no). Esta base de datos entrega información que permite determinar si el alumno permaneció o no en el mismo curso entre 1° y 2° medio. Para este estudio incluimos tanto alumnos que estuvieron con el mismo profesor en 1° y 2° medio, como alumnos que estuvieron con ese profesor solo en 2° medio.

Nivel sala de clase - profesor

Registro Docente 2005-2006: identifica al docente que impartió matemática y lenguaje en los cursos de los establecimientos registrados en el Ministerio de Educación. El curso está identificado por el código de enseñanza, el grado, la letra de curso y el RBD del establecimiento. Gracias a este registro que asocia profesores a salas de clases es posible vincular información de los estudiantes con información del profesor, particularmente las estimaciones de Valor Agregado del profesor basadas en el aprendizaje de sus alumnos con sus resultados de la Evaluación Docente.

Evaluación Docente 2005-2010: contiene el resultado final de los profesores municipales participantes del Sistema Nacional de Evaluación de Desempeño Docente entre los años 2005 a 2010. Como existen docentes que han sido evaluados en más de una ocasión (debido a que se han re-evaluado al cabo de cuatro años o resultaron insatisfactorios y se han re-evaluado al año siguiente) hemos considerado solo el más reciente de los resultados disponibles. Para estos análisis fueron considerados además del resultado final por docente, los resultados de cada instrumento de evaluación: auto-evaluación (AE), informe de

referencia de terceros (supervisores: director y jefe UTP) (IRT), entrevista del evaluador par (un docente del mismo municipio que hace clases en el mismo subsector pero no en el mismo colegio) (EP), y el portafolio (PF), distinguiendo en este último el módulo 1 (que corresponde a la evidencia escrita) (M1) y módulo 2 (que corresponde a la evaluación de una clase) (M2).

Participantes

Debido a que el periodo entre las mediciones de rendimiento de los alumnos en la prueba SIMCE es de dos años, nos aseguramos de incluir en los análisis solo aquellos profesores que enseñaron en 1° medio 2005 y en 2° medio 2006 a la misma cohorte. Esto es, el análisis considera solo los cursos que tuvieron clases **con el mismo profesor en ambos años** para asegurar que los alumnos estuvieran bajo el efecto del mismo docente y que el Valor Agregado estimado estuviese asociado solo a un profesor. Para controlar que un alumno no haya cambiado de curso durante 1° y 2° medio se incorporó en los análisis una variable que indica si el alumno estuvo presente los dos años en el mismo curso.

El estudio considera solo los **profesores del sector municipal** evaluados entre 2005 y 2010 en lenguaje y matemática en Enseñanza Media por el sistema de Evaluación Docente. No todos los profesores evaluados por la Evaluación Docente cuentan con datos de progreso de sus alumnos, debido a que dicha información solo existe para la cohorte de alumnos evaluada en 2° medio en 2006 y que rindió previamente la prueba SIMCE de 8° básico 2004. El total de participantes de la Evaluación Docente entre 2005 y 2010 es N=78.044 profesores. Según nuestros datos, el total de profesores evaluados entre 2005 y 2010 que impartieron clases en 1° medio en 2005 y (además) en 2° medio en 2006 es de 973 docentes en lenguaje y 1060 en matemática.

El total de estudiantes con datos SIMCE de 2° medio 2006 es de N=248.139, pero los estudiantes con datos de progreso en el aprendizaje entre 2004 y 2006 es de N=161.118

tanto en lenguaje como matemática. De ellos, solo N=36.207 alumnos tienen un profesor con resultado en la Evaluación Docente en el subsector de lenguaje, y N=36.668 para matemática. Considerando solo a los estudiantes con información completa en las variables usadas en los análisis y a **profesores con 10 o más alumnos**, el total de casos para los análisis fue N=27.156 estudiantes en lenguaje distribuidos en 1.187 cursos con 761 profesores en 476 establecimientos, y N=27.601 alumnos en matemática distribuidos en 1.197 cursos con 725 profesores en 477 establecimientos.

Trabajamos también con una sub-muestra que contiene solo profesores que se evaluaron en 2006 y que tienen datos correspondientes al SIMCE 2004-2006 de sus alumnos. Entonces, para esa sub-muestra tenemos a profesores que fueron evaluados en 2006 con los mismos estudiantes que tomaron la prueba SIMCE a fines de 2006, y que estuvieron con ese profesor tanto en 1° medio como en 2° medio. Esa sub-muestra nos ofrece entonces un vínculo más directo y “limpio” entre datos del desempeño docente y aprendizaje de alumnos. La Tabla 1 muestra la cantidad de alumnos por subsector para la muestra completa y la sub-muestra 2006 que se utilizaron para las estimaciones de valor agregado a nivel profesor (“efecto profesor”).

Tabla 1: Cantidad de alumnos, cursos, profesores y escuelas por subsector para la muestra completa y para el año 2006.

Tamaño	Lenguaje Total	Matemática Total	Lenguaje 2006	Matemática 2006
Alumnos	27156	27601	6045	5552
Cursos	1187	1197	245	233
Profesores	761	725	163	151
Cursos / Profesor	1.56	1.65	1.48	1.51
Media alumno / profesor	36	38	37	36
Min alumnos / profesor	10	10	10	10
Max alumnos / profesor	187	327	187	130
Escuelas	467	477	90	95

Los tamaños de las muestras sobre las que se basan las correlaciones entre estimaciones del “efecto profesor” y resultado de la Evaluación Docente son levemente distintas a las que se usaron para la estimación del “efecto profesor” ya que no todos tienen información completa en los instrumentos de Evaluación Docente (missings listwise) (ver Tabla 2).

Tabla 2: Cantidad de profesores por subsector para correlaciones entre “efecto profesor” y resultados en la Evaluación Docente.

N de casos	Lenguaje Total	Matemática Total	Lenguaje 2006	Matemática 2006
N de profesores para calcular correlaciones entre modelos	761	725	163	151
N de profesores para calcular correlaciones VA-ED	735	697	162	150

Análisis de datos

Un paso importante de este estudio es el desarrollo y la estimación de modelos de Valor Agregado del profesor basados en el aprendizaje de sus alumnos, los que luego van a ser conectados con los resultados de Evaluación Docente de los mismos profesores. Para la estimación del Valor Agregado del profesor fue empleada la metodología de Modelos Lineales Jerárquicos (Hierarchical Linear Modeling, HLM) siguiendo desarrollos como los de Bryk y Raudenbush (2002) y Hox (2010). Los modelos fueron definidos considerando a los alumnos como unidad de análisis de nivel-1 y el profesor como unidad de análisis de nivel-2.² Esto implica que algunos alumnos en distinto curso pero bajo la enseñanza del mismo profesor pertenecen al mismo grupo de nivel-2. De esta manera aumentamos la precisión de las estimaciones para dichos profesores al considerar todos sus alumnos y nos aseguramos de estimar solo una medida de Valor Agregado para cada docente. Fueron estimados modelos separados para matemática y lenguaje, correspondiéndole a cada análisis una muestra de profesores (nivel-2) distinta. Siguiendo las sugerencias de McCaffrey et al. (2003), en nuestro estudio – que tiene fines solo investigativos del “efecto profesor” y su relación con otra medida de evaluación docente – se considera el profesor como **efecto aleatorio**.³

La ecuación general para estimar los puntajes en el nivel-1 tiene la siguiente expresión:

² No fue posible incluir un nivel 3 “colegio,” ya que en nuestra muestra hay solo alrededor de 1.5 profesores por colegio.

³ Test de Hausman para los modelos M1, M3, M5: Tanto para Lenguaje y Matemática el test entregó un valor $\sim p < 0.001$, lo que indica que se rechaza la hipótesis nula (la hipótesis nula indica que en el modelo de valor agregado de efectos aleatorios no existe correlación entre las variables regresoras y el efecto del profesor) y nos quedamos entonces con la hipótesis alternativa (que el modelo de efectos fijos del profesor es mejor, ya que al existir correlación entre las variables regresoras y el efecto del profesor, este efecto estaría sesgado). Sin embargo, al realizar la correlación entre los efectos fijos y aleatorios del profesor esta es del orden de 0.98 para todos los modelos tanto en lenguaje y matemática, lo que podría indicar que prácticamente no hay diferencias entre usar el modelo de efectos fijos o aleatorios. Además que el test de Hausman asume que el modelo de efectos fijos no tiene ningún tipo de sesgo, lo cual no tiene porque ser verdad.

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}(1)$$

En la que Y_{ij} representa el puntaje de rendimiento del alumno i en la clase del profesor j , β_{0j} representa el rendimiento promedio de los alumnos del profesor j ; βX representa el coeficiente de regresión asociado a X (variables atribuidas al estudiante); y e_{ij} representa el término de error, el que se asume distribuido normalmente, con promedio 0 y varianza σ_e^2 .

La ecuación general para las estimaciones de nivel-2 fue especificada de la siguiente forma:

$$\beta_{0j} = \gamma_{00} + \gamma_{0j}Z_j + \mu_{0j}(2)$$

$$\beta_{1j} = \gamma_{10}(3)$$

En el que γ_{00} representa el gran promedio, γZ representa el coeficiente de regresión asociado a Z (variables atribuidas al efecto del profesor), μ_{0j} representa la desviación del rendimiento promedio de los alumnos del profesor respecto el gran promedio de rendimiento. **Este término de error μ_{0j} representa para nosotros una estimación del valor agregado del profesor.** La distribución de μ_{0j} se asume normal, con promedio 0 y varianza $\sigma_{\mu_{0j}}^2$.

Finalmente la especificación del modelo es de la siguiente forma:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{0j}Z_j + e_{ij} + \mu_{0j}$$

Los modelos fueron ajustados usando HLM 6.08 y PROC MIXED en SAS 9.2. Las estimaciones de los μ_{0j} fueron obtenidas usando Empirical Bayes (EB).

En total, fueron estimados seis modelos, cada uno de ellos con distintas especificaciones de

covariables. Cada modelo fue estimado para Lenguaje y para Matemática. El detalle de la especificación de los seis modelos se detalla a continuación y las ecuaciones se encuentran en el Anexo 1.

M1: incluye solo variables relevantes para el aprendizaje que no son influenciadas por el profesor (variables individuales), sin controlar por rendimiento previo.

M2: incluye todas las variables individuales más variables de aprendizaje que operan a nivel curso (variables nivel 2), sin rendimiento previo ni individual ni del curso.

M3: incluye solo las variables rendimiento previo (Matemática y Lenguaje).

M4: incluye variables de rendimiento previo a nivel individual y curso.

M5: incluye solo variables individuales más rendimiento previo.

M6: incluye variables individuales y de nivel 2 más el rendimiento previo individual y del curso.

Los primeros dos modelos (M1 y M2) los llamamos “tribute to the economists” o *pseudo valor agregado*. No incluyen como variable predictora al rendimiento previo del estudiante lo cual para nosotros es un requisito para los modelos de valor agregado.

Los modelos M3 y M4 los llamamos “tribute to Tennessee” ya que son inspirados por la práctica establecida por Sanders y colegas en el sistema de responsabilidad docente del estado de Tennessee, EE.UU., de no incluir covariables más allá del rendimiento previo de los estudiantes.

Los últimos dos modelos (M5 y M6) los llamamos “tribute to complexity” ya que son modelos completos de Valor Agregado que incluyen tanto rendimiento previo como otras

covariables respecto de características individuales y del curso, primero solo a nivel estudiante (M5) y después tanto a nivel estudiante como a nivel profesor (M6).

La intención de utilizar esta metodología es poder observar cómo cambia la estimación del valor agregado del profesor entre los distintos modelos, de modo de poder llegar a conclusiones respecto de la (no) comparabilidad de los modelos. Queremos saber si importa para las estimaciones de VA y para las relaciones entre VA y ED cual modelo utilizamos.

La idea de incluir co-variables tanto a nivel individual como a nivel grupal es de controlar la asignación no aleatoria de estudiantes a profesores y de crear condiciones equitativas entre profesores respecto de las variables exógenas o fuera del control del profesor que inciden en su labor docente y en el aprendizaje de sus estudiantes.

Las variables incorporadas en los distintos modelos se especifican en la siguiente tabla.

Tabla 3: Variables del estudiante que no son influenciadas por el profesor.⁴

Variable	Item	Respuesta
Rendimiento previo Lenguaje y Matemática	Resultado de la prueba SIMCE 2004 en lenguaje y matemática	Puntaje estandarizado
Sexo	Sexo del estudiante	0. Mujer 1. Hombre
Educación de la madre	Años de escolaridad de la madre (2006)	De 0 a 20 años de estudio
Subsidio	¿Es beneficiada la familia del estudiante con subsidios del estado? (2006)	0. No 1. Si
Promedio de notas general⁵	Promedio de notas general en 1o medio y 2do medio.	Escala de 1 a 7.

⁴ Consideramos la variable “asistencia” pero análisis descriptivos sugieren que es poco confiable y por ende no se incluyó en los modelos. Las siguientes variables no fueron incluidas por tener un N debajo 1% en las muestras finales: estatus del alumno como repitente, indígena, diferenciado, o integrado.

⁵ Esta variable se incluye porque se estima como un indicador de habilidades no cognitivas del alumno. Esta variable está centrada al promedio de la escuela, debido a que las escalas de notas que se utilizan en las escuelas varían considerablemente entre ellos.

Tabla 4: Variables que operan a nivel profesor.

Variable	Descripción
Promedio de rendimiento previo	Promedio del rendimiento previo de los alumnos por profesor en Lenguaje y Matemática
Proporción de hombres	Proporción de hombres por profesor
Proporción de alumnos con subsidio	Proporción de alumnos con subsidio por profesor
Promedio de notas general	Promedio de notas general de los alumnos por profesor
Heterogeneidad Rendimiento (2004)	Variabilidad (desviación estándar) de rendimiento de los estudiantes por profesor.
Heterogeneidad Social (2006)	Variabilidad (desviación estándar) de años de escolaridad de la madre de los estudiantes por profesor.

Una vez obtenido las estimaciones de los μ_{0j} (Valor Agregado del profesor) se analizaron las correlaciones de las estimaciones entre los distintos modelos. Después correlacionamos las estimaciones de Valor Agregado producidas en los distintos modelos con los resultados de Evaluación Docente, expresados en el puntaje final y en el puntaje de cada instrumento (Autoevaluación, Evaluador Par, Informe de Referencia de Tercero, Portafolio). Las correlaciones utilizadas entre las estimaciones de valor agregado y con los resultados de la Evaluación Docente son correlaciones de Spearman, ya que queremos evaluar la asociación que existe entre los resultados de valor agregado y Evaluación Docente considerando la ubicación relativa, o ranking, que obtiene un docente en ambas medidas de desempeño docente.

Resultados

A continuación presentamos primero los resultados descriptivos, seguidos por resultados de los distintos modelos de Valor Agregado con los cuales estimamos el “efecto profesor.” También correlacionamos entre sí los resultados de los distintos modelos presentados en la sección anterior. Finalmente, correlacionamos los resultados que obtiene cada docente a través de las dos medidas de efectividad docente (Evaluación Docente y Valor Agregado).

Resultados descriptivos

Respecto de los resultados descriptivos de las variables en el estudio llama la atención que el rendimiento de la muestra general tanto en Lenguaje como en Matemática entre 2004 y 2006 disminuye en vez de aumentar. Recordemos que las pruebas SIMCE de 8° básico y 2° medio no son verticalmente escaladas y que tienen distintas desviaciones estándar.⁶ Tenemos una muestra de mayoritariamente mujeres (56%) con un promedio de notas de 5.5 y cuyas madres tienen en promedio entre 9 y 10 años de escolaridad. Un 31% de las familias recibe algún subsidio del estado por su condición de vulnerabilidad socioeconómica. Las correlaciones entre variables no superan el 0.4 salvo entre las variables de rendimiento, donde alcanzan un 0.5 entre notas y rendimiento SIMCE, y 0.7 entre las dos mediciones en el tiempo de lenguaje y matemática, respectivamente (ver Anexo 2).

Respecto del rendimiento SIMCE bruto de los estudiantes para los profesores de cada nivel de desempeño en la Evaluación Docente, cabe destacar que hay un ordenamiento de menor rendimiento para profesores insatisfactorios hacia mayor rendimiento para profesores destacados en el caso de matemática, pero no así tan lineal en lenguaje. No se trata de una relación causa-efecto entre el nivel de desempeño de los docentes y el rendimiento de sus

⁶ Eso puede constituir una limitación del estudio ya que algunos expertos consideran como requisito para MVA la disponibilidad de datos longitudinales medidas por pruebas verticalmente escaladas (Lissitz, 2006).

estudiantes, ya que no controlamos por ningún factor adicional de asignación de estudiantes a profesores, pero sí podemos constatar que los profesores de peor desempeño trabajan con niños de más bajo rendimiento SIMCE matemática, y los de mejor desempeño con niños de más alto rendimiento SIMCE matemática (ver Anexo 2).

Análisis de descomposición de la varianza

En el análisis multinivel la descomposición de la varianza del modelo nulo nos ayuda a entender que niveles constituyen fuentes de la varianza más y menos relevantes. Después, al incluir variables de control, se demuestra como se modifica la varianza en cada nivel. En nuestro caso nos interesa sobre todo como se comporta la varianza del nivel profesor al incluir co-variables que describen características de los alumnos sobre los cuales el profesor no tiene control pero que influyen en su labor de enseñanza.

Tabla 5: Descomposición de la varianza.

LENGUAJE

	MN	M1	M2	M3	M4	M5	M6
Nivel1 (estudiante)	0.66	0.53	0.53	0.37	0.37	0.35	0.35
Nivel2 (profesor)	0.35	0.19	0.1	0.04	0.02	0.04	0.02
Total	1.01	0.72	0.63	0.41	0.39	0.39	0.37
Nivel1 (estudiante) %	65.35	73.61	84.13	90.24	94.87	89.74	94.59
Nivel2 (profesor) %	34.65	26.39	15.87	9.76	5.13	10.26	5.41
Total %	100	100	100	100	100	100	100
R2 nivel1	-	0.20	0.20	0.44	0.44	0.47	0.47
R2 nivel2	-	0.46	0.71	0.89	0.94	0.89	0.94

MATEMÁTICA

	MN	M1	M2	M3	M4	M5	M6
Nivel1 (estudiante)	0.59	0.41	0.41	0.31	0.31	0.27	0.27
Nivel2 (profesor)	0.43	0.25	0.15	0.1	0.05	0.1	0.05
Total	1.02	0.66	0.56	0.41	0.36	0.37	0.32
Nivel1 (estudiante) %	57.84	62.12	73.21	75.61	86.11	72.97	84.38
Nivel2 (profesor) %	42.16	37.88	26.79	24.39	13.89	27.03	15.63
Total %	100	100	100	100	100	100	100
R2 nivel1	-	0.31	0.31	0.47	0.47	0.54	0.54
R2 nivel2	-	0.42	0.65	0.77	0.88	0.77	0.88

Los resultados demuestran que la varianza atribuible al nivel profesor es de 35% en Lenguaje y de 42% en Matemática. En ambas asignaturas el mejor modelo para dar cuenta de la varianza que existe a nivel profesor es el modelo 4 (“tribute to Tennessee”) que incluye la variable de rendimiento previo tanto a nivel estudiante como a nivel grupal (promedio de todos los alumnos del profesor). En ese modelo solo quedan 5% de la varianza a nivel profesor por explicar en Lenguaje, y 14% en Matemática. Ese modelo es incluso levemente más explicativo que el modelo 6 (“tribute to complexity”) que incluye, además de las variables de rendimiento previo individual y grupal, todas las otras co-variables que describen características relevantes de los alumnos. Claramente inferior

respecto de su capacidad de dar cuenta de la varianza a nivel profesor son los modelos 1 y 2 (“tribute to the economists”) que no incluyen el rendimiento previo de los alumnos. También es relevante notar que los modelos que incluyen variables a nivel grupal (respecto del grupo de pares que enseña un profesor) siempre dan cuenta de más varianza del nivel profesor que modelos sin esas variables grupales, lo que apunta a la relevancia del efecto del grupo de pares en el proceso de enseñanza-aprendizaje (ver también Taut & Escobar, 2012; San Martín & Carrasco, 2012).

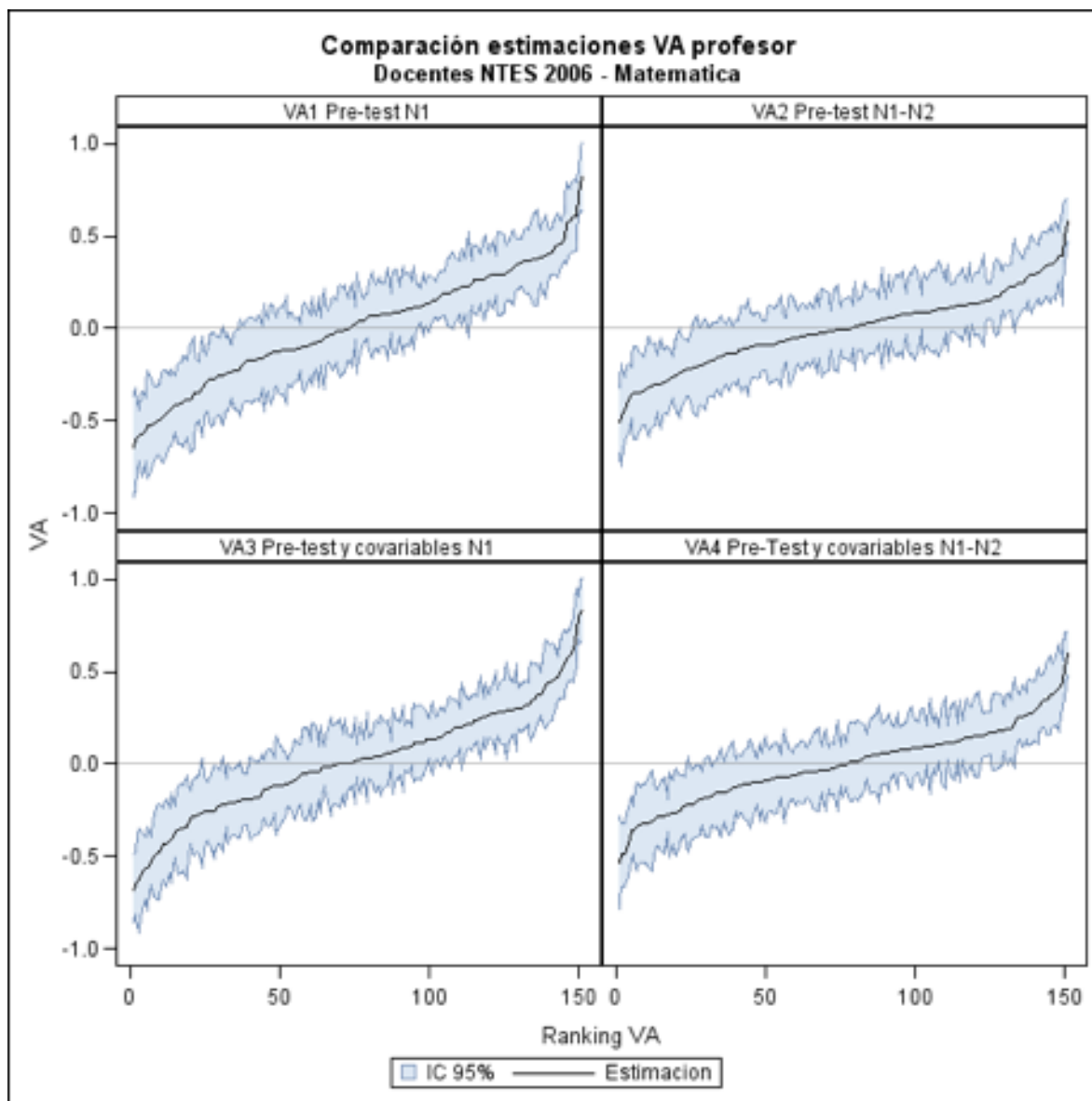
Resultados de valor agregado a nivel profesor

En esta sub-sección relatamos los resultados de las estimaciones de Valor Agregado de la sub-muestra 2006⁷. Esa sub-muestra incluye solo profesores que se han evaluado por la Evaluación Docente en 2006 con el mismo grupo de estudiantes de los cuales tenemos datos de aprendizaje y que nos sirvieron para estimar el VA de esos profesores. Nos enfocamos en los resultados de esa sub-muestra ya que ahí el vínculo entre los resultados de Valor Agregado del profesor y sus resultados en la Evaluación Docente es más directo y válido.

El siguiente gráfico (Figura 1) muestra las estimaciones de VA para los profesores de Matemática con un intervalo de confianza de 95% para los cuatro modelos que incluyen rendimiento previo. Al incluir variables a nivel grupal la curva se atenúa y hay menos profesores que se distinguen significativamente del promedio. Sin embargo, la pregunta central en el presente estudio es si los profesores que se escapan de manera significativa tanto por sobre y por debajo el promedio corresponden a profesores que tienen resultados en la Evaluación Docente también mejores o peores que el promedio.

⁷ Para ver los resultados de la muestra completa, ver Anexo 3. El Anexo 3 incluye resultados de los 6 modelos (coeficientes no estandarizados y estandarizados, y model fit).

Figura 1. Intervalos de confianza de estimaciones de VA en Matemática para los cuatro modelos que incluyen rendimiento previo.



Los siguientes gráficos (Figuras 2 y 3) demuestran la relación entre el Valor Agregado del profesor en Matemática y su resultado en el portafolio de la Evaluación Docente (ED).

Figura 2. Scatter plot de la relación entre VA (Modelo 6) y resultado en el portafolio ED.

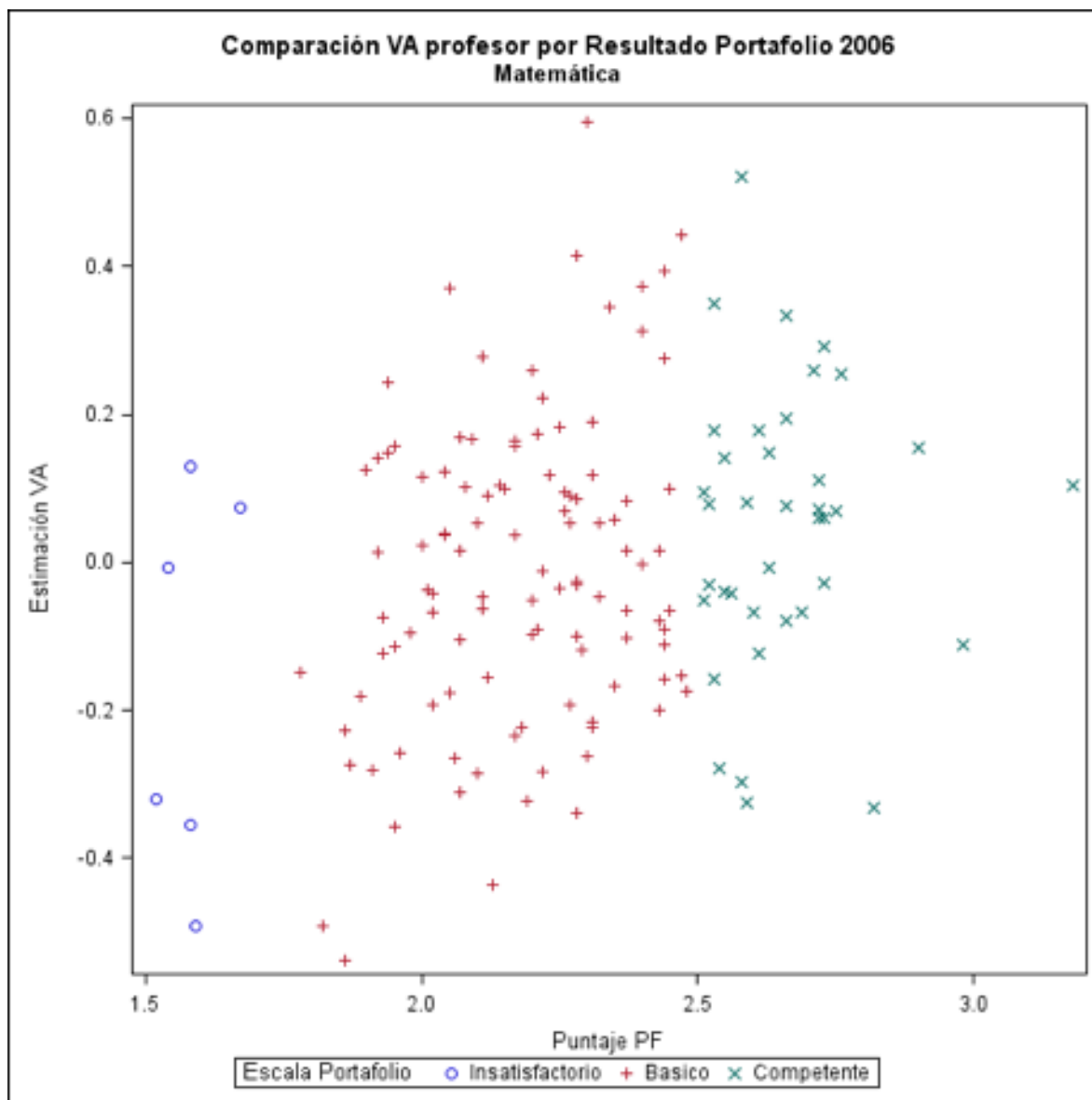
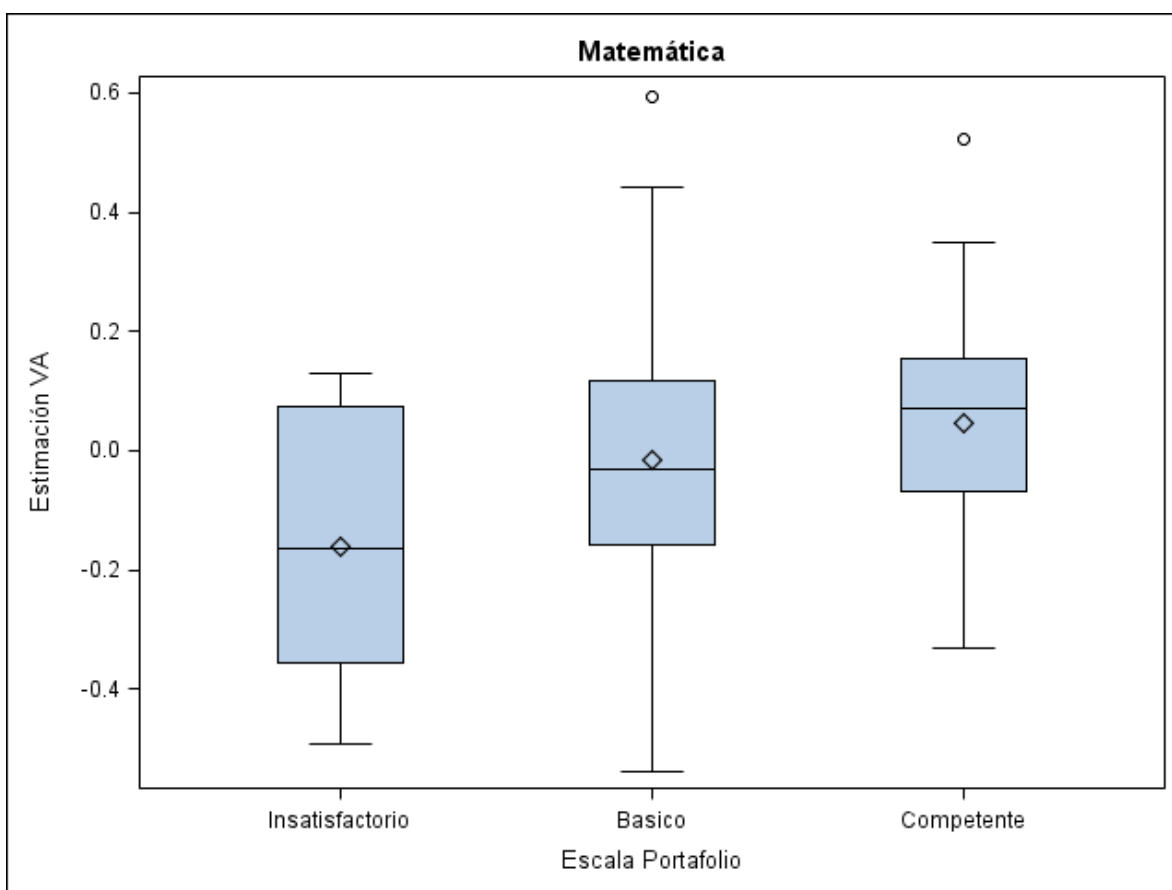


Figura 3. Boxplots de la relación entre VA (M6) y resultado en el portafolio de la ED.



Los promedios de VA de los profesores con distintos niveles de desempeño en el portafolio de la Evaluación Docente se comportan de manera esperada (a mayor VA, mejor resultado en la ED). Sin embargo, llama la atención la gran varianza en las estimaciones de VA para los docentes con un nivel insatisfactorio y básico en el portafolio. Cabe notar que la cantidad de profesores varía bastante por categoría de desempeño. La siguiente Figura 4 y la correspondiente Tabla 6 muestran los mismos resultados pero se agruparon a los docentes en terciles de acuerdo al resultado que obtuvieron en el portafolio.

Figura 4. Boxplots de la relación entre VA (M6) y resultado en el portafolio de la ED (agrupado en terciles).

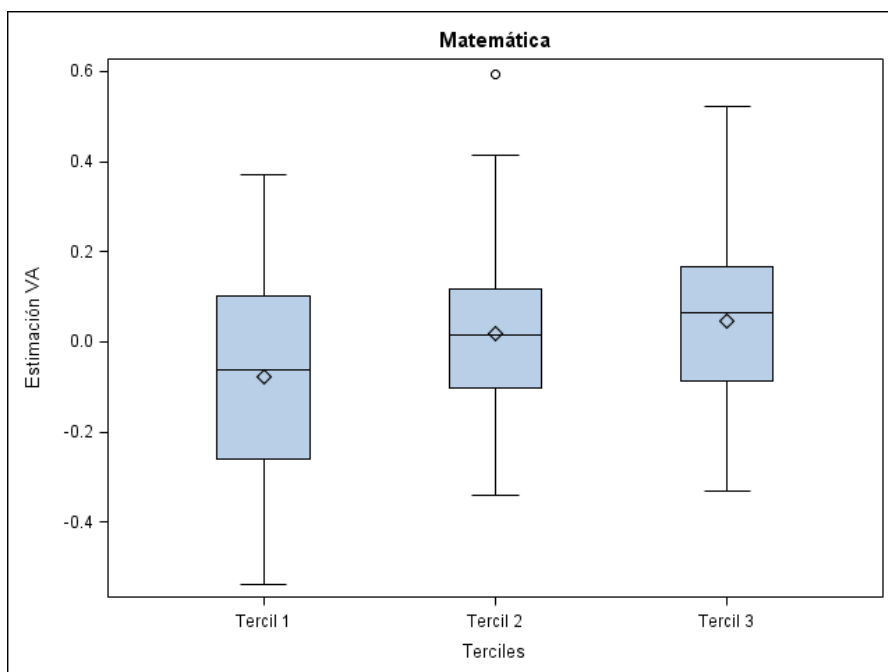


Tabla 6. Estimaciones de VA (M6) por terciles en el portafolio de la ED.

Portafolio	N	Promedio VA (M6)	D.E.	Error estándar	%VA>0
Lenguaje					
Tercil1	58	-0.03	0.12	0.02	39.65
Tercil2	52	0.02	0.11	0.02	55.77
Tercil3	53	0.01	0.11	0.02	56.60
Matemática					
Tercil1	50	-0.07	0.22	0.03	44.00
Tercil2	49	0.01	0.21	0.03	51.02
Tercil3	52	0.05	0.20	0.03	57.69

Asociación entre resultados de la Evaluación Docente y Valor Agregado del profesor

A continuación se presenta el análisis de la asociación entre las estimaciones de Valor Agregado y los puntajes en los instrumentos de la Evaluación Docente, tanto para la muestra completa como para la sub-muestra 2006. Presentamos los coeficientes de correlación de Spearman separando a los docentes por subsector (Lenguaje y Matemática). Se presentan las correlaciones entre los distintos instrumentos de la Evaluación Docente y distintas medidas de rendimiento de los alumnos: logro promedio en 2006 de los alumnos (“logro”), progreso promedio de los alumnos entre 2004 y 2006 (“dif”), y cada uno de nuestros seis modelos de VA que estiman el “efecto profesor” descritos anteriormente.

Tabla 7. Correlaciones puntajes ED y VA en Lenguaje (muestra completa, N=735).

	logro	m1	m2	dif	m3	m4	m5	m6
AE	0.01	0.02	0.04	0.02	0.02	0.02	0.02	0.02
EP	-0.02	-0.03	0.04	0.03	0.01	-0.00	0.03	0.02
IRT	0.14**	0.12**	0.10**	0.08*	0.13**	0.12**	0.08*	0.07
M1	0.04	0.05	0.06	0.09*	0.09*	0.10**	0.10**	0.10**
M2	0.06	0.07	0.09*	0.08*	0.10**	0.10**	0.08*	0.09*
PF	0.07	0.08*	0.09*	0.12**	0.13**	0.13**	0.13**	0.13**
TOTAL	0.08*	0.08*	0.13**	0.13**	0.13**	0.13**	0.13**	0.13**

Tabla 8. Correlaciones puntajes ED y VA en Matemática (muestra completa, N=697).

	logro	m1	m2	dif	m3	m4	m5	m6
AE	0.00	0.02	0.03	0.01	0.01	0.02	0.00	0.01
EP	0.08*	0.07	0.09*	0.09*	0.09*	0.08*	0.09*	0.07
IRT	0.15**	0.15**	0.12**	0.20**	0.20**	0.18**	0.18**	0.17**
M1	0.13**	0.15**	0.15**	0.18**	0.17**	0.18**	0.14**	0.15**
M2	0.14**	0.12**	0.09*	0.13**	0.13**	0.12**	0.07	0.06
PF	0.16**	0.18**	0.17**	0.20**	0.19**	0.20**	0.15**	0.16**
TOTAL	0.18**	0.18**	0.18**	0.22**	0.22**	0.21**	0.19**	0.18**

Tabla 9. Correlaciones puntajes ED y VA en Lenguaje (muestra 2006, N=162).

	logro	m1	m2	dif	m3	m4	m5	m6
AE	-0.07	-0.07	-0.09	-0.03	-0.06	-0.08	-0.05	-0.06
EP	0.05	0.01	-0.01	0.09	0.10	0.07	0.11	0.10
IRT	0.05	0.05	-0.05	0.06	0.05	0.05	0.03	0.02
M1	0.06	0.03	0.02	0.11	0.10	0.10	0.13	0.13
M2	0.10	0.15	0.22**	0.20*	0.18*	0.23**	0.22**	0.24**
PF	0.11	0.10	0.12	0.17*	0.17*	0.19*	0.20*	0.21**
TOTAL	0.09	0.08	0.07	0.16*	0.16*	0.16*	0.18*	0.19*

Tabla 10. Correlaciones puntajes ED y VA en Matemática (muestra 2006, N=150).

	logro	m1	m2	dif	m3	m4	m5	m6
AE	-0.01	0.04	0.10	0.04	0.03	0.07	0.09	0.09
EP	0.13	0.11	0.07	0.10	0.10	0.08	0.02	0.00
IRT	0.11	0.11	0.08	0.16	0.16	0.14	0.18*	0.17*
M1	0.22**	0.26**	0.14	0.28**	0.25**	0.27**	0.19**	0.21**
M2	0.21*	0.26**	0.31**	0.24**	0.23**	0.27**	0.18*	0.19*
PF	0.26**	0.32**	0.24**	0.32**	0.29**	0.33**	0.21**	0.23**
TOTAL	0.24**	0.28**	0.21**	0.28**	0.26**	0.28**	0.19*	0.20*

Para analizar los resultados de las correlaciones conviene hacer una distinción entre subsectores. En el caso de Lenguaje, en general las correlaciones son más débiles que en Matemática para todos los instrumentos de la ED y todos los modelos. Para la muestra completa, las relaciones más fuertes en Lenguaje se encuentran para el puntaje del portafolio, el puntaje total y la evaluación de los supervisores (IRT). Los modelos no se

distinguen mucho entre ellos. En el caso de la sub-muestra 2006, estas relaciones aumentan y los más fuertes se encuentran para el Módulo 2 (video), y nuevamente los modelos no arrojan resultados muy distintos. Cabe destacar que – contrario a nuestra hipótesis – hay relaciones similares a los que se comentaron para el modelo que incluye co-variables a nivel individual y grupal, pero excluyendo rendimiento previo.

En el caso de la Matemática los resultados son similares pero de magnitud más alta. Para la muestra completa, las relaciones más fuertes están entre el portafolio, el puntaje total y la evaluación de los supervisores por un lado, y el VA basado en un modelo que controla por rendimiento previo a nivel individual por otro lado (hasta 0.22).

Las relaciones en Matemática (tal como las de Lenguaje) son más fuertes para la muestra 2006. En esta última alcanzan hasta 0.33 entre portafolio y el modelo que incluye rendimiento previo a nivel individual y grupal. En la sub-muestra también hay fuertes relaciones entre el Módulo 1 y 2 con todas las medidas basadas en el rendimiento de los alumnos (incluyendo logro bruto).

En general se podría constatar que en Matemática el resultado del profesor en la ED se relaciona tanto con el logro bruto de sus estudiantes como el valor que agrega a su aprendizajes, o sea, los mejores profesores en términos de su resultado en la ED tienen niños con un alto rendimiento, y también agregan más valor en términos de su aprendizaje, y profesores con un bajo resultado en la ED tienen niños con un bajo rendimiento, y también agregan menos valor en términos de su aprendizaje. Comparativamente, en Lenguaje existe una asociación más baja entre ED y el logro bruto de los alumnos, siendo levemente más alta la relación entre ED y Valor Agregado.

También de modo general podemos decir que el instrumento de la ED que más se relaciona con el VA es el portafolio (con ambos módulos). También, y en segundo lugar, correlaciona significativamente (sobre todo en Matemática) el informe de los directivos. No hay relación con la auto-evaluación y la entrevista del profesor par.

Comparación entre modelos

En esta sub-sección se presentan las comparaciones entre los distintos modelos del presente estudio. Fueron calculados coeficientes de la correlación de Spearman para las muestras de profesores de Lenguaje y de Matemática por separado. Incluimos en estas comparaciones el nivel de logro y progreso promedio de los alumnos del profesor, las estimaciones de pseudo-Valor Agregado, y estimaciones de Valor Agregado controlando por distintas covariables. Los resultados de las correlaciones de Spearman para la muestra completa están reportados en las Tablas 11 y 12 para Lenguaje y Matemática respectivamente (para la muestra de docentes evaluados en 2006 consultar Anexo 4).

Cabe destacar que no hay *per se* una relación pre-establecida entre el nivel del rendimiento de un grupo de alumnos y el valor agregado por el profesor. Puede haber un docente que tiene un grupo de estudiantes con un alto rendimiento pero el agrega poco valor en términos del progreso en el aprendizaje de sus alumnos. También puede haber un docente con estudiantes de bajo rendimiento pero el agrega mucho valor en términos de sus aprendizajes. En situaciones de alto rendimiento también puede haber docentes que aun así agregan mucho valor, y situaciones de bajo rendimiento donde profesores agregan poco valor (Briggs, 2008).

Tanto para Lenguaje como Matemática es posible observar que todas las medidas de efectividad del profesor basada en rendimiento y progreso de los estudiantes incluidos en este estudio muestran asociaciones positivas y estadísticamente significativas entre sí. Sin embargo, los grados de asociación difieren de manera importante en cuanto a magnitud o fuerza de la asociación dependiendo del subsector o de los modelos contrastados.

En primer lugar, la correlación entre logro y progreso es positiva y de magnitud moderada en Lenguaje (0.37) pero fuerte en Matemática (0.84) e indica que en general los cursos con más alto rendimiento obtienen mayores progresos, tendencia que es mucho más acentuada en los docentes/cursos de Matemática. Esta evidencia por sí misma sugiere que existen

diferencias importantes entre una medida de logro y otra de progreso. Por otro lado, las pruebas SIMCE 8° básico 2004 y SIMCE 2° medio 2006 no tienen escalamiento vertical, por lo que estrictamente no es correcto hablar de progreso en el aprendizaje (sobre todo en lenguaje).

En general, ordenamientos de profesores basados en logro podrían llegar a ser muy diferentes de aquellos que se obtendrían al utilizar medidas de Valor Agregado. De hecho, las correlaciones más bajas (<0.5) se encuentran entre modelos basados en logro (bruto o con co-variables) y modelos de valor agregado completos (con rendimiento previo y otras co-variables). Estos resultados sugieren que considerar rendimiento previo y la incorporación de otras covariables en el modelamiento multinivel del efecto del profesor tiene un importante impacto en dichas estimaciones tanto en Lenguaje como en Matemática. Sin embargo, se relacionan fuertemente los modelos de estatus y los modelos “tribute to Tennessee” (>0.7 en matemática y >0.6 en lenguaje).

Finalmente, se observa que las correlaciones entre diferentes modelos de Valor Agregado dependen de los controles estadísticos considerados y varían entre 0.77 y 0.97 en Lenguaje y entre 0.72 y 0.99 en Matemática.

Tabla 11. Correlaciones entre estimaciones VA Lenguaje (muestra completa, N=746)

	Logro	m1	m2	dif	m3	m4	m5	m6
Logro	1.00							
m1	0.93	1.00						
m2	0.56	0.74	1.00					
Dif	0.37	0.41	0.44	1.00				
m3	0.81	0.81	0.62	0.80	1.00			
m4	0.74	0.85	0.75	0.79	0.95	1.00		
m5	0.30	0.36	0.42	0.96	0.78	0.77	1.00	
m6	0.30	0.42	0.53	0.94	0.77	0.81	0.97	1.00

Tabla 12. Correlaciones entre estimaciones VA Matemática (muestra completa, N=703)

	logro	m1	m2	dif	m3	m4	m5	m6
Logro	1.00							
m1	0.92	1.00						
m2	0.60	0.78	1.00					
Dif	0.84	0.83	0.68	1.00				
m3	0.89	0.87	0.70	0.99	1.00			
m4	0.84	0.93	0.81	0.95	0.95	1.00		
m5	0.38	0.44	0.56	0.78	0.74	0.72	1.00	
m6	0.38	0.49	0.60	0.77	0.73	0.75	0.99	1.00

Discusión y conclusiones

El presente estudio se dedicó a investigar la relación entre los resultados de profesores evaluados por la Evaluación Docente (ED) en Chile y el aprendizaje de sus alumnos en matemática y lenguaje durante 1o y 2o medio. Utilizamos la Metodología de Valor Agregado para estimar el “efecto profesor” y correlacionamos estas estimaciones con los resultados de la ED. Estos análisis correlativos se enmarcan en una agenda de validación de la Evaluación Docente (Taut, Santelices & Stecher, 2012). El supuesto es que una medida de evaluación de desempeño docente debería mostrar una relación positiva y de tamaño débil a mediana con el progreso en el aprendizaje de los alumnos de los profesores evaluados. Adicionalmente, nuestro trabajo tiene un objetivo metodológico al investigar la relación entre distintos modelos que tratan de estimar el “efecto profesor”. Específicamente, nos interesa comparar modelos que controlan por el rendimiento previo de los alumnos con modelos que solo controlan por otras co-variables.

Los análisis de descomposición de la varianza demuestran que hay más varianza entre profesores en matemática (42%) que en lenguaje (35%) – algo que ya se ha constatado en trabajos previos tanto en EE.UU. (Milanowski, 2004) como en Chile (Taut & Escobar, 2012). Es importante notar que el nivel profesor en nuestro modelo incluye la varianza entre colegios, ya que no pudimos incluir estos como un tercer nivel en los modelos. En trabajos previos el nivel de varianza entre colegios en Chile esta entre 23% y 34% para Lenguaje y entre 26% y 45% en Matemática, dependiendo del nivel de enseñanza (Manzi et al., 2008). Sin embargo, solo entre colegios municipales, utilizando los mismos datos de este estudio, Taut y Escobar (2012) encontraron una varianza atribuible al nivel colegio de 20% para Lenguaje y 26% para Matemática.

En el modelo de valor agregado más completo (que incluye todas las covariables a nivel individual y del grupo de pares) la varianza a nivel profesor se ha disminuido hasta un 5%

en lenguaje (16% en matemática), reforzando la idea que las co-variables a nivel alumno y aula que se incluyen en nuestros modelos dan cuenta de factores que influyen en el aprendizaje pero que están fuera del control del profesor. Es interesante notar que la misma reducción de varianza a nivel profesor se puede lograr con un modelo más parsimonioso que solo incluye rendimiento previo a nivel alumno y promedio de rendimiento previo a nivel grupo de pares, y que modelos que incluyen solo otras co-variables pero excluyen rendimiento previo no logran una tan alta varianza explicada.

Respecto de las correlaciones entre la ED y el VA de los profesores, la relación para la muestra completa es más débil de lo que demuestran otros estudios del contexto norteamericano, aunque estos trabajos también encuentran una gran variabilidad de las correlaciones dependiendo del grado y subsector (ver Milanowski, 2004; Grossman et al., 2010; Hill et al., 2011). En nuestro caso, las correlaciones de Spearman para el subsector de lenguaje alcanzan un 0.13 entre VA y puntaje portafolio, puntaje total y la evaluación de los supervisores (IRT), mientras que para matemática son más fuertes y alcanzan 0.22 entre VA y puntaje total. Es importante tomar en cuenta que en el caso de esa muestra completa, gran parte de los resultados de ED corresponden a un año distinto que los datos de aprendizaje de los alumnos. Por ende, se puede asumir un vínculo menos fuerte entre ambas medidas.

Es por eso que hemos analizado aparte la muestra de docentes que se evaluó en 2006, durante el mismo año de la segunda medición del aprendizaje de sus alumnos, así estableciendo un vínculo más directo entre ED y VA. Efectivamente, las correlaciones para la muestra 2006 son más fuertes y alcanzan hasta un 0.33 entre portafolio y VA para matemática, y hasta un 0.24 entre Módulo 2 (video) y VA para lenguaje. Queda claro que es mejor establecer relaciones finas entre la práctica docente y el aprendizaje de los alumnos, utilizando idealmente datos del mismo año. También sería interesante examinar la estabilidad de las correlaciones usando varios años de datos de un profesor.

Adicionalmente, al comparar las correlaciones de distintos instrumentos de la ED para la muestra 2006, podemos constatar que medidas directas del trabajo en aula se relacionan más fuertemente con el aprendizaje de los alumnos (Bill & Melinda Gates Foundation, 2012). Eso sería un argumento a favor de seguir incluyendo ese tipo de medidas en la ED. También llama la atención que la Auto-evaluación no tiene ninguna relación con el aprendizaje de los alumnos, y que la Entrevista Par tiene una relación débil. Respecto de la evaluación de los supervisores la relación con VA es similar a la del portafolio para la muestra completa, pero es más débil para la muestra 2006. Estamos en el proceso de examinar también la relación entre VA y las distintas dimensiones del portafolio.

Respecto del objetivo metodológico de contrastar distintos modelos de VA y de “pseudo-VA” nuestro estudio muestra que no llegan a estimaciones similares (ver también San Martín & Carrasco, 2012), y que hace una diferencia si se incluyen co-variables a nivel individual o también grupo de alumnos del profesor en el modelo, pero estas diferencias son mucho más fuertes en el caso de lenguaje que en matemática. Las correlaciones más bajas se encuentran entre logro bruto de los estudiantes y estimaciones de VA que incluyen todas las co-variables disponibles. Desafortunadamente, en la política educacional chilena todavía es común utilizar el logro bruto -sobre todo a nivel escuela, pero también a nivel aula- como indicador de efectividad o calidad. Ejemplos son los rankings SIMCE en los diarios, las cartas enviadas por el Ministerio de Educación a todos los hogares chilenos comunicando los resultados SIMCE de los colegios con subvención estatal utilizando semáforos, y la comunicación de resultados a nivel aula en los informes SIMCE.

Un tema relevante en la aplicación de modelos de Valor Agregado es la pregunta si estos modelos pueden controlar y dar cuenta de la asignación no-aleatoria de niños a escuelas y profesores. En Chile hay suficiente evidencia de que la asignación de niños a escuelas depende fuertemente del nivel socio-económico de los niños y sus familias, y que el NSE es

un fuerte predictor de rendimiento escolar (Torche & Mizala, 2012). Además, hay evidencia que los mejores docentes se van a escuelas de mayor nivel socio-económico y rendimiento, acentuando así la inequidad del sistema educacional (Meckes & Bascopé, 2012; Valencia & Taut, 2011). En nuestro estudio utilizamos solo datos del sector municipal, tanto de profesores como de sus estudiantes. Por ley no existe la posibilidad de selección en los colegios municipales. Al comparar docentes solo dentro del sistema municipal, la no-aleatoriedad no debe ser tan acentuada como la sería en el sistema educacional chileno completo. Adicionalmente, debido al diseño del sistema educacional chileno, muchos estudiantes del sistema municipal se cambian de colegio entre 8° básico (fin de la educación básica) y 1° medio (comienzo de la educación media). Por lo tanto, la asignación no-aleatoria dentro de los colegios en 1° medio no debe ser tan pronunciada como podría ser por ejemplo durante la enseñanza básica. También incluimos rendimiento previo y otras variables relevantes en los modelos para controlar cualquier efecto de asignación no aleatoria. Adicionalmente, corrimos análisis de correlaciones parciales, controlando por el efecto del nivel socioeconómico al correlacionar VA con ED, pero la magnitud de las correlaciones no se vio modificada.

A modo de conclusión, este trabajo mostró que existe una relación mediana entre los resultados que obtiene un profesor en la ED y el resultado de aprendizaje de los alumnos que tuvo a su cargo el mismo año de su evaluación. Es evidencia que apoya la validez de la Evaluación Docente, especialmente del portafolio con evidencia directa del trabajo en aula del profesor, pero también del conjunto de medidas respecto del desempeño docente incluidas en la ED.

Queremos destacar que nosotros apoyamos el uso de la metodología de valor agregado para estimar el “efecto profesor” solo con fines investigativos y no para basar en sus resultados decisiones de alto impacto en un contexto de una evaluación docente. Además, estamos conscientes de la multitud de condiciones para su adecuado uso que en general no se cumplen con los datos de aprendizaje disponibles en Chile hoy. De hecho, en la misma

línea argumentan varios expertos de educación de EE.UU., llamando a un uso limitado de VAM en sistemas de evaluación docente, enriqueciendo la evidencia respecto del desempeño docente con otras medidas más directas de la labor en aula de los profesores que además ofrecen información formativa a los docentes evaluados (Darling-Hammond et al., 2011; Darling-Hammond, 2012; Baker et al., 2010).

Referencias

- Alvarado, M., Cabezas, G., Falck, D., & Ortega, M. E. (2012). *La Evaluación Docente y sus instrumentos: Discriminación del desempeño docente y asociación con los resultados de los estudiantes*. Centro de Estudios MINEDUC & PNUD.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (AERA, APA & NCME) (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- Baker, E. L. et al. (2010). *Problems with the use of student test scores to evaluate teachers*. (EPI Briefing Paper #278). Washington, DC: Economic Policy Institute.
- Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment for teachers. *Journal of Educational and Behavioral Statistics*, 29 (1), 37–65.
- Bill & Melinda Gates Foundation (2012). *Gathering Feedback for Teaching*. Research Report. Disponible: www.metproject.org/.../MET_Gathering_Feedback_Research_Paper.pdf
- Braun, H. (2005). Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models. *Education Testing Service*. Disponible: <http://www.ets.org/Media/Research/pdf/PICVAM.pdf>
- Braun, H., Chudowsky, N., & Koenig, J. A. (2010). *Getting Value Out of Value-Added: Report of a Workshop*. Washington, DC: The National Academies Press. Disponible: <http://www.nap.edu/catalog/12820.html>
- Bravo, D., Falck, D., González, R., Manzi, J., & Peirano, C. (2008) *La relación entre la evaluación docente y el rendimiento de los alumnos: Evidencia para el caso de Chile*. Disponible: http://www.microdatos.cl/docto_publicaciones/Evaluacion%200docentes_rendimiento%20escolar.pdf
- Briggs, D. C. (2008). Synthesizing causal inferences. *Educational Researcher*, 37(1), 15-22.

- Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical Linear Models in Social and Behavioral Research: Applications and Data Analysis Methods* (First Edition). Newbury Park, CA: Sage Publications.
- Glazerman, S., Loeb, S., Goldhaber, D., Staiger, D., Raudenbush, S., Whitehurst, G. (2010). *Evaluating Teachers: The Important Role of Value-Added*. Brown Center on Education Policy at Brookings.
- Grossman, P., Loeb, S., Cohen, J., Hammerness, K., Wyckoff, J., Boyd, D., Lankford, H. (2010). Measure for measure: The relationships between measures of instructional practice in middle school English language arts and teachers' value-added scores. (*Working Paper N° 16015*) Extraído de National Bureau of Economic Research Working Paper website <http://www.nber.org/papers/w16015>
- Hill, H. C., Kapitula, L. & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794–831. doi:10.3102/0002831210387916
- Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications* (2nd ed.). New York: Routledge.
- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Darling-Hammond, L. (2012). Desarrollo de un enfoque sistémico para evaluar la docencia y fomentar una enseñanza eficaz. *Pensamiento Educativo. Revista de Investigación Educativa Latinoamericana*, 49 (2), 1-20. Disponible: <http://dx.doi.org/10.7764/PEL.49.2.2012.1>
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E. H., & Rothstein, J. (2011). Getting teacher evaluation right: A background paper for policy makers. *Research Briefings*. Washington, DC: American Educational Research Association, National Academy of Education.
- Kane, M. T. (2006). Validation. En R. L. Brennan (Ed.), *Educational measurement* (4th ed.), pp. 17–64. Westport, CT: Praeger Publishers.

- Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2010). *Identifying effective classroom practices using student achievement data* (NBER Working Paper No. 15803). Cambridge, MA: National Bureau of Economic Research.
- Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee value added assessment system. *Educational Evaluation and Policy Analysis*, 25(3), 287-298.
- Lissitz, R. (ed.) (2005). *Value Added Models in Education: Theory and Applications*. Maple Grove: JAM Press.
- Lissitz, R. (2006). *Longitudinal and Value-added Models of Student Performance*. Maple Grove: JAM Press.
- Manzi, J., Strasser, K., San Martin, E. & Contreras, D. (2008). *Quality of education in Chile*. Santiago, Chile: Centro de Medición de la Pontificia Universidad Católica de Chile (Mide UC).
- Meckes, L., & Bascopé, M. (2012). Uneven Distribution of Novice Teachers in the Chilean Primary School System. *Education Policy Analysis Archives*, 20(30). Disponible: <http://epaa.asu.edu/ojs/article/view/1017>
- Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education*, 79(4), 33 — 53. doi: 10.1207/s15327930pje7904_3
- McCaffrey, D. F., Lockwood, J. R., Hamilton, L. & Koretz, D. (2003). *Evaluating Value-Added Models for Teacher Accountability*. Santa Monica, CA: RAND.
- Newton, X., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-Added Modeling of Teacher Effectiveness: An Exploration of Stability across Models and Contexts. *Education Policy Analysis Archives*, 18(23). Disponible: <http://epaa.asu.edu/ojs/article/view/810>
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How Large Are Teacher Effects?. *Educational Evaluation And Policy Analysis* (26),3, 237-257. doi: 10.3102/0162373702600323

- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: SAGE.
- Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics*, 29(1), 121-129.
- Raudenbush, S. (2008). Advancing Educational Policy by Advancing Research on Instruction. *American Educational Research Journal*, 45(1), 206-230.
- Sanders, J. & Horn, S. (1998). Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research. *Journal of Personnel Evaluation in Education*, 12(3), 247-256.
- San Martín, E. & Carrasco, A. (2012). Clasificación de escuelas en la nueva institucionalidad educative: contribución de modelos de valor agregado para una responsabilización justa. *Temas de la Agenda Pública*, 7(53), 1-16. Centro de Políticas Públicas, Pontificia Universidad Católica de Chile.
- Sartain, L., Stoelinga, S. R., & Brown, E.R. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation*. (Research Report) Consortium on Chicago School Research at the University of Chicago.
- Steele, J.L., Hamilton, L.S., & Stecher, B.M. (2010). *Incorporating Student Performance Measures into Teacher Evaluation Systems*. Distribution. Santa Monica, CA: RAND Corporation. Disponible: http://www.rand.org/pubs/technical_reports/TR917/
- Taut, S., & Escobar, J. (2012). El efecto de las características de los pares en el aprendizaje de estudiantes chilenos de enseñanza media. *Informes Técnicos MIDE UC*, IT1201, Junio 2012. En: www.mideuc.cl/category/publicaciones
- Taut, S., Santelices, V. & Stecher, B. (2012). Validation of a national teacher assessment and improvement system. *Educational Assessment Journal*, 17(4), 163-199. Disponible: <http://dx.doi.org/10.1080/10627197.2012.735913>

- Valencia, E., & Taut, S. (2011). *Evidencia de inequidad en el acceso a los docentes mejor calificados en la Enseñanza Básica Chilena* (MIDE UC Technical Report IT-1108). Santiago, Chile: Pontificia Universidad Católica de Chile, Escuela de Psicología, Centro de Medición MIDE UC.
- Wenglinsky, H. (2002) How schools matter. The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives*, 10 (12).
Extraído de <http://epaa.asu.edu/epaa/v10n12/>

ANEXO 1: Especificación de los seis modelos utilizados

Modelos: “tribute to the economists”

M1:

$$Y_{ij} = \gamma_{00} + \gamma_{10}Hombre_{ij} + \gamma_{20}Educ.Madre_{ij} + \gamma_{30}Subsidio_{ij} + \gamma_{40}Promedio\ de\ notas_{ij} + e_{ij} + \mu_{0j}$$

M2:

$$Y_{ij} = \gamma_{00} + \gamma_{10}Hombre_{ij} + \gamma_{20}Educ.Madre_{ij} + \gamma_{30}Subsidio_{ij} + \gamma_{40}Promedio\ de\ notas_{ij} + \gamma_{01}Proporción_Hombres_j + \gamma_{02}Proporción_Subsidio_j + \gamma_{03}Notas_Por_Docente_j + \gamma_{04}Heterogeneidad\ lenguaje_j + \gamma_{05}Heterogeneidad\ matemática_j + \gamma_{06}Heterogeneidad\ social_j + e_{ij} + \mu_{0j}$$

Modelos: “tribute to Tennessee”

M3:

$$Y_{ij} = \gamma_{00} + \gamma_{10}Lenguaje04_{ij} + \gamma_{20}Matematica04_{ij} + e_{ij} + \mu_{0j}$$

M4:

$$Y_{ij} = \gamma_{00} + \gamma_{10}Lenguaje04_{ij} + \gamma_{20}Matematica04_{ij} + \gamma_{01}Promedio_Lenguaje04_j + \gamma_{02}Promedio_Matematica04_j + e_{ij} + \mu_{0j}$$

Modelos: “tribute to complexity”

M5:

$Y_{ij} =$

$$\gamma_{00} + \gamma_{10}Hombre_{ij} + \gamma_{20}Educ.Madre_{ij} + \gamma_{30}Subsidio_{ij} + \gamma_{40}Promedio\ de\ notas_{ij} + \gamma_{50}Lenguaje04_{ij} + \gamma_{60}Matemática04_{ij} + e_{ij} + \mu_{0j}$$

M6:

$$\begin{aligned} Y_{ij} = & \gamma_{00} + \gamma_{10}Hombre + \gamma_{20}Educ.Madre + \gamma_{30}Subsidio + \gamma_{40}Promedio\ de\ notas \\ & + \gamma_{50}Lenguaje04 + \gamma_{60}Matemática04 + \gamma_{01}Proporción_Hombres \\ & + \gamma_{02}Proporción_Subsidio + \gamma_{03}Notas_por_docente \\ & + \gamma_{04}Heterogeneidad\ Lenguaje + \gamma_{05}Heterogeneidad\ Matemática \\ & + \gamma_{06}Heterogeneidad\ Social + \gamma_{07}Promedio\ Lenguaje04 \\ & + \gamma_{08}Promedio\ Matemática04 + e_{ij} + \mu_{0j} \end{aligned}$$

ANEXO 2: Resultados Descriptivos

Lenguaje: N=27156

Variable	Mean	SD	Min	Max
Puntaje Lenguaje 2° EM 2006	254.71	50.24	122.49	398.16
Puntaje Lenguaje 8° EB 2004	257.17	48.51	98.19	391.67
Puntaje Matematica 8° EM 2004	256.95	47.13	123.61	406.09
Puntaje Z Lenguaje 2° EM 2006	0.00	1.00	-2.63	2.86
Puntaje Z Lenguaje 8° EB 2004	0.00	1.00	-3.28	2.77
Puntaje Z Matematica 8° EM 2004	0.00	1.00	-2.83	3.16
Hombre	0.44	0.50	0	1
Promedio general 2005-2006	5.52	0.55	2.20	7.00
Años de escolaridad de la madre	9.68	3.54	0	19
Recibe subsidio del estado	0.31	0.46	0	1

Matemática: N=27601

Variable	Mean	SD	Min	Max
Puntaje Matematica 2° EM 2006	254.58	61.29	94.39	426.58
Puntaje Matematica 8° EM 2004	257.51	47.05	123.89	406.09
Puntaje Lenguaje 8° EB 2004	257.96	48.24	98.19	391.67
Puntaje Z Matematica 2° EM 2006	0.00	1.00	-2.61	2.81
Puntaje Z Lenguaje 8° EB 2004	0.00	1.00	-3.31	2.77
Puntaje Z Matematica 8° EM 2004	0.00	1.00	-2.84	3.16
Hombre	0.44	0.50	0	1
Promedio general 2005-2006	5.52	0.55	2.20	7.00
Años de escolaridad de la madre	9.80	3.56	0	19
Recibe subsidio del estado	0.30	0.46	0	1

Correlaciones co-variables Lenguaje

	leng06z	leng04z	mat06z	mat04z	sex	pg	educM	sub
leng06z	1							
leng04z	0.75	1						
mat06z	0.70	0.64	1					
mat04z	0.65	0.69	0.76	1				
sex	-0.08	-0.09	0.08	0.11	1			
pg	0.51	0.48	0.55	0.48	-0.12	1		
educM	0.30	0.30	0.31	0.30	0.02	0.19	1	
sub	-0.12	-0.12	-0.12	-0.12	-0.03	-0.04	-0.24	1

Correlaciones co-variables Matemática

	leng06z	leng04z	mat06z	mat04z	sex	pg	educM	sub
leng06z	1							
leng04z	0.75	1						
mat06z	0.69	0.65	1					
mat04z	0.65	0.69	0.76	1				
sex	-0.06	-0.07	0.10	0.12	1			
pg	0.53	0.49	0.55	0.49	-0.11	1		
educM_C	0.31	0.30	0.31	0.31	0.01	0.20	1	
sub	-0.12	-0.12	-0.12	-0.12	-0.02	-0.05	-0.25	1

Cantidad de profesores por categoría de la evaluación docente por subsector para la muestra

completa y la muestra 2006

	Lenguaje		Matemática	
	Muestra 2006	Muestra completa	Muestra 2006	Muestra completa
Insatisfactorio	1	5	2	7
Básico	55	164	44	195
Competente	96	511	85	457
Destacado	11	80	20	66
Total	163	760	151	725

Resultado SIMCE Profesores Evaluación Docente Lenguaje

Escala CCE	N	SIMCE	Media	DE	Min	Max
Insatisfactorio	118	Lenguaje 2° EM	250.36	44.72	164.66	369.06
		Lenguaje 8° EB	253.01	46.66	150.44	343.18
		Progreso	-2.65	32.04	-73.77	90.74
Básico	4936	Lenguaje 2° EM	249.26	48.88	127.23	390.44
		Lenguaje 8° EB	254.45	47.98	105.63	391.67
		Progreso	-5.18	35.01	-154.49	184.94
Competente	18572	Lenguaje 2° EM	255.23	49.79	122.49	398.16
		Lenguaje 8° EB	257.31	48.15	98.19	391.67
		Progreso	-2.08	34.95	-194.00	237.55
Destacado	2981	Lenguaje 2° EM	256.70	52.52	126.29	398.16
		Lenguaje 8° EB	256.92	48.99	99.16	384.82
		Progreso	-0.22	35.80	-181.87	160.90

Resultado SIMCE Profesores Evaluación Docente Matemática

Escala CCE	N	SIMCE	Media	DE	Min	Max
Insatisfactorio	244	Lenguaje 2° EM	237.16	56.66	110.67	385.88
		Lenguaje 8° EB	246.98	42.27	131.22	343.22
		Progreso	-9.82	38.94	-115.98	107.71
Básico	6775	Lenguaje 2° EM	248.67	60.51	96.79	426.58
		Lenguaje 8° EB	255.21	47.15	124.53	406.09
		Progreso	-6.54	39.17	-256.15	162.79
Competente	17236	Lenguaje 2° EM	256.13	61.18	94.68	426.58
		Lenguaje 8° EB	258.57	47.06	123.89	406.09
		Progreso	-2.43	39.49	-255.03	168.84
Destacado	2774	Lenguaje 2° EM	263.62	61.88	94.39	426.58
		Lenguaje 8° EB	258.99	46.85	127.01	403.69
		Progreso	4.63	39.84	-173.31	188.67

Resultado SIMCE nivel estudiante por Desempeño Evaluación Docente Lenguaje 2006

Escala CCE	N	SIMCE	Media	DE	Min	Max
Insatisfactorio	47	Lenguaje 2° EM	278.34	40.59	193.41	369.06
		Lenguaje 8° EB	277.49	40.45	164.67	343.18
		Progreso	0.84	31.42	-73.77	66.92
Básico	1728	Lenguaje 2° EM	249.45	48.36	128.33	381.06
		Lenguaje 8° EB	254.40	46.89	105.63	378.68
		Progreso	-4.94	34.39	-154.49	140.83
Competente	3847	Lenguaje 2° EM	255.42	48.39	127.30	398.16
		Lenguaje 8° EB	256.66	45.34	101.99	391.67
		Progreso	-1.24	34.50	-140.46	153.95
Destacado	399	Lenguaje 2° EM	260.46	54.21	126.29	393.15
		Lenguaje 8° EB	262.79	49.08	129.43	371.01
		Progreso	-2.33	34.18	-138.19	113.95

Resultado SIMCE nivel estudiante por Evaluación Docente Matemática 2006

Escala CCE	N	SIMCE	Media	DE	Min	Max
Insatisfactorio	140	Matemática 2° EM	240.72	59.33	110.79	385.88
		Matemática 8° EB	248.29	38.28	166.86	336.56
		Progreso	-7.57	40.62	-104.05	107.71
Básico	1612	Matemática 2° EM	249.80	58.52	105.48	425.04
		Matemática 8° EB	256.39	45.78	135.06	391.04
		Progreso	-6.59	38.26	-141.23	162.79
Competente	3016	Matemática 2° EM	265.01	60.83	101.75	426.58
		Matemática 8° EB	264.45	45.96	123.89	406.09
		Progreso	0.56	38.80	-218.81	157.18
Destacado	737	Matemática 2° EM	248.75	53.10	111.26	425.04
		Matemática 8° EB	250.68	42.41	139.01	371.27
		Progreso	-1.93	39.67	-173.31	157.48

ANEXO 3: Resultados de Estimación de VA

Estimaciones de VA (M6) por nivel de desempeño en el portafolio de la ED (muestra completa)

Portafolio	N	VA (M6)	D.E.	Error estándar	%VA>0
Lenguaje					
Insatisfactorio	16	-0.07	0.10	0.03	6.25
Básico	517	-0.01	0.12	0.01	7.54
Competente	213	0.02	0.12	0.01	9.39
Matemática					
Insatisfactorio	20	-0.06	0.27	0.06	25.00
Básico	519	-0.01	0.20	0.01	17.34
Competente	164	0.04	0.21	0.02	26.22

Estimaciones de VA (M6) por nivel de desempeño en el portafolio de la ED. (muestra 2006)

Portafolio	N	VA (M6)	D.E.	Error estándar	%VA>0
Lenguaje					
Insatisfactorio	10	-0.08	0.08	0.03	0
Básico	137	0	0.11	0.01	8.03
Competente	17	0.02	0.11	0.03	5.88
Matemática					
Insatisfactorio	6	-0.16	0.26	0.11	16.67
Básico	107	-0.01	0.21	0.02	17.76
Competente	38	0.05	0.19	0.03	21.05

Estimaciones de VA (M6) por nivel de desempeño en el puntaje final de la ED (muestra completa)

Puntaje final	N	VA (M6)	D.E.	Error estándar	%VA>0
Lenguaje					
Insatisfactorio	4	0.00	0.04	0.02	0.00
Básico	156	-0.02	0.12	0.01	3.85
Competente	502	0.00	0.12	0.01	8.37
Destacado	84	0.03	0.13	0.01	14.29
Matemática					
Insatisfactorio	7	-0.09	0.23	0.09	14.29
Básico	181	-0.03	0.21	0.02	17.13
Competente	448	0.01	0.21	0.01	18.53
Destacado	67	0.07	0.22	0.03	34.33

Estimaciones de VA (M6) por nivel de desempeño en el puntaje final de la ED (muestra 2006)

Puntaje final	N	VA (M6)	D.E.	Error estándar	%VA>0
Lenguaje					
Insatisfactorio	2	0.03	0.02	0.01	0.00
Básico	56	-0.03	0.12	0.02	5.36
Competente	95	0.01	0.11	0.01	9.47
Destacado	11	-0.01	0.12	0.04	0.00
Matemática					
Insatisfactorio	4	-0.09	0.19	0.10	25.00
Básico	44	-0.04	0.23	0.03	18.18
Competente	82	0.01	0.20	0.02	17.07
Destacado	21	0.02	0.19	0.04	23.81

Estimaciones de VA (M6) por tercil en el portafolio de la ED (muestra completa)

Portafolio	N	VA (M6)	D.E.	Error estándar	%VA>0
Lenguaje					
Tercil1	252	-0.02	0.13	0.01	7.54
Tercil2	248	0.00	0.12	0.01	6.85
Tercil3	246	0.02	0.12	0.01	9.76
Matemática					
Tercil1	236	-0.04	0.21	0.01	16.53
Tercil2	238	0.00	0.18	0.01	15.55
Tercil3	229	0.04	0.22	0.01	27.07

Estimaciones de VA (M6) por tercil en el portafolio de la ED (muestra 2006)

Portafolio	N	VA (M6)	D.E.	Error estándar	%VA>0
Lenguaje					
Tercil1	54	-0.04	0.11	0.01	5.56
Tercil2	57	0.02	0.11	0.02	8.77
Tercil3	53	0.01	0.11	0.01	7.55
Matemática					
Tercil1	51	-0.08	0.21	0.03	15.69
Tercil2	52	0.02	0.20	0.03	17.31
Tercil3	48	0.05	0.20	0.03	22.92

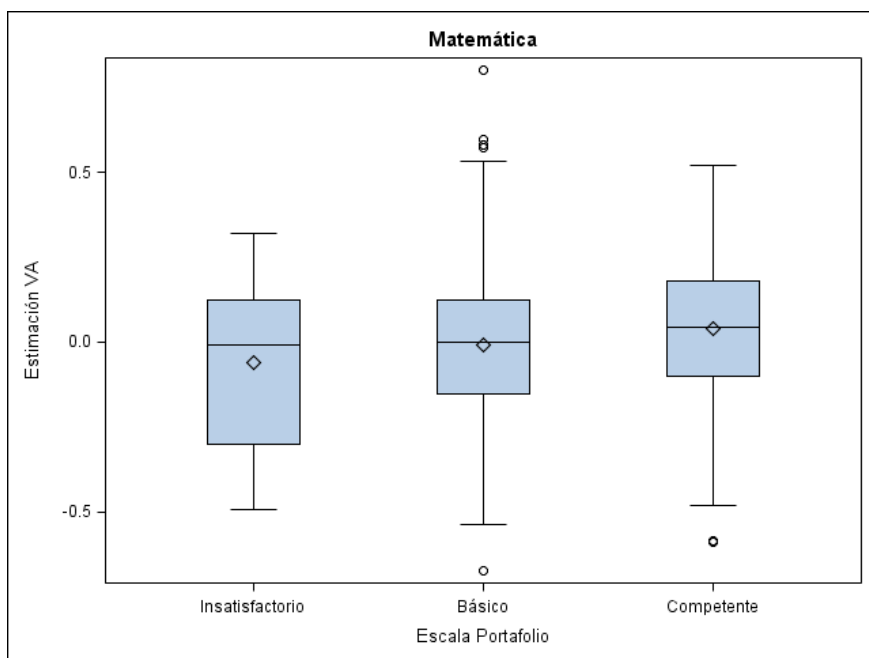
Estimaciones de VA (M6) por tercil en el puntaje final de la ED (muestra completa)

Puntaje final	N	VA (M6)	D.E.	Error estándar	%VA>0
Lenguaje					
Tercil1	261	-0.02	0.12	0.01	4.98
Tercil2	256	0.01	0.11	0.01	7.81
Tercil3	229	0.02	0.13	0.01	11.79
Matemática					
Tercil1	246	-0.03	0.21	0.01	17.07
Tercil2	242	-0.02	0.19	0.01	14.46
Tercil3	215	0.06	0.22	0.01	28.37

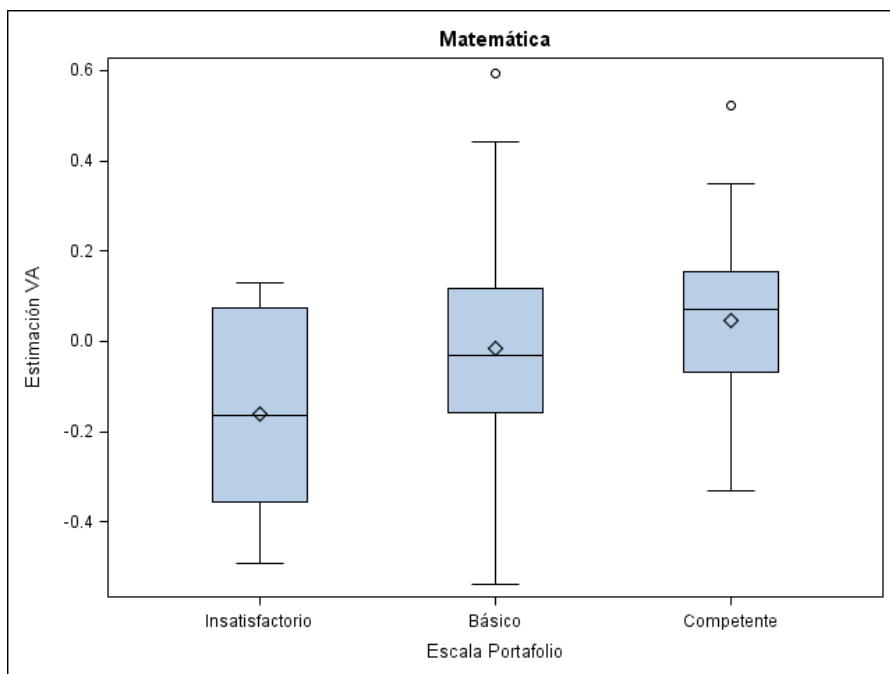
Estimaciones de VA (M6) por tercil en el puntaje final de la ED (muestra 2006)

Puntaje final	N	VA (M6)	D.E.	Error estándar	%VA>0
Lenguaje					
Tercil1	55	-0.03	0.10	0.01	1.82
Tercil2	56	0.01	0.11	0.01	8.93
Tercil3	53	0.02	0.12	0.02	11.32
Matemática					
Tercil1	51	-0.04	0.22	0.03	19.61
Tercil2	52	-0.04	0.16	0.02	7.69
Tercil3	48	0.07	0.22	0.03	29.17

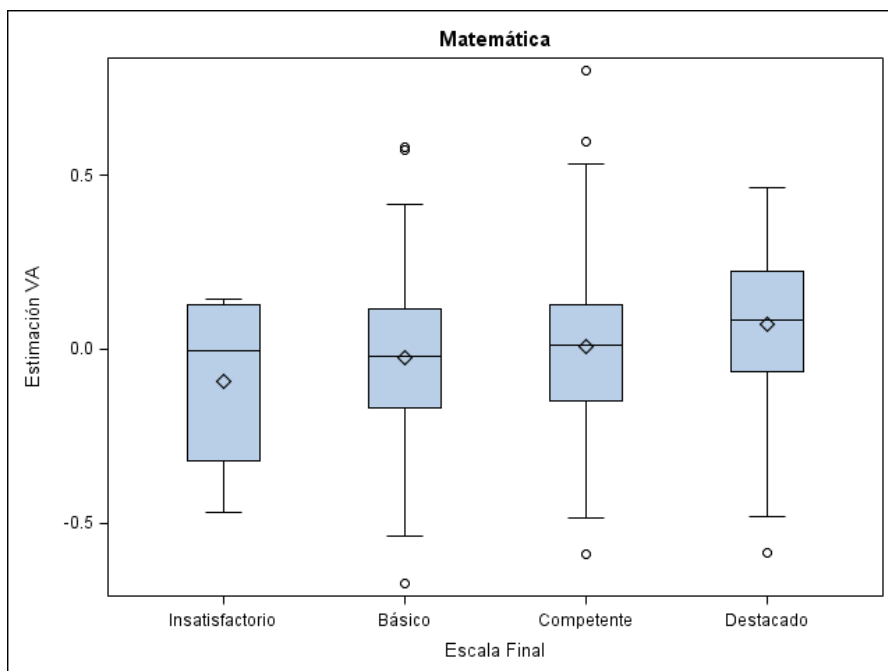
Estimaciones de VA (M6) por nivel de desempeño en el portafolio de la ED (muestra completa)



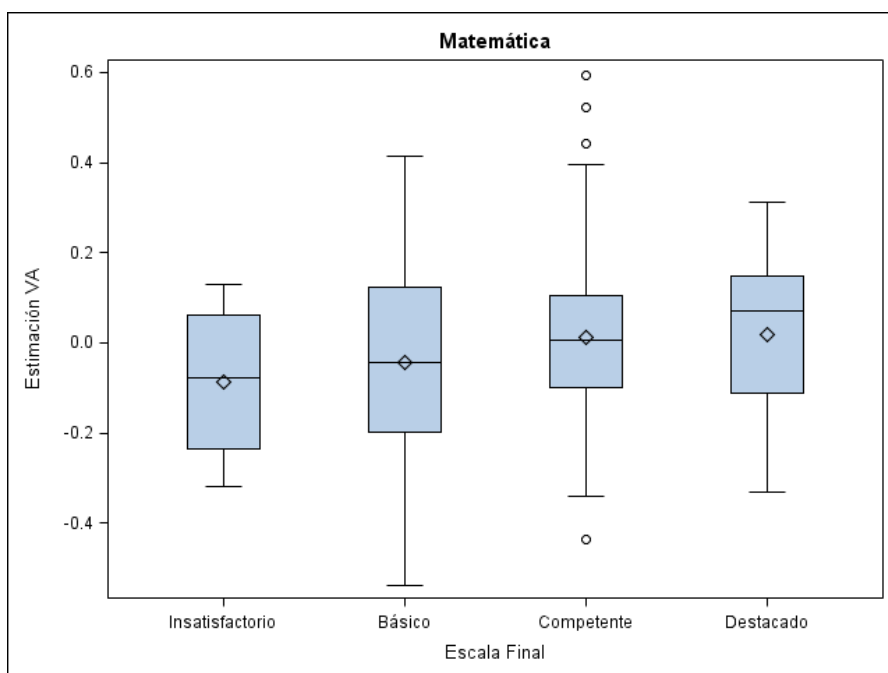
Estimaciones de VA (M6) por nivel de desempeño en el portafolio de la ED (muestra 2006)



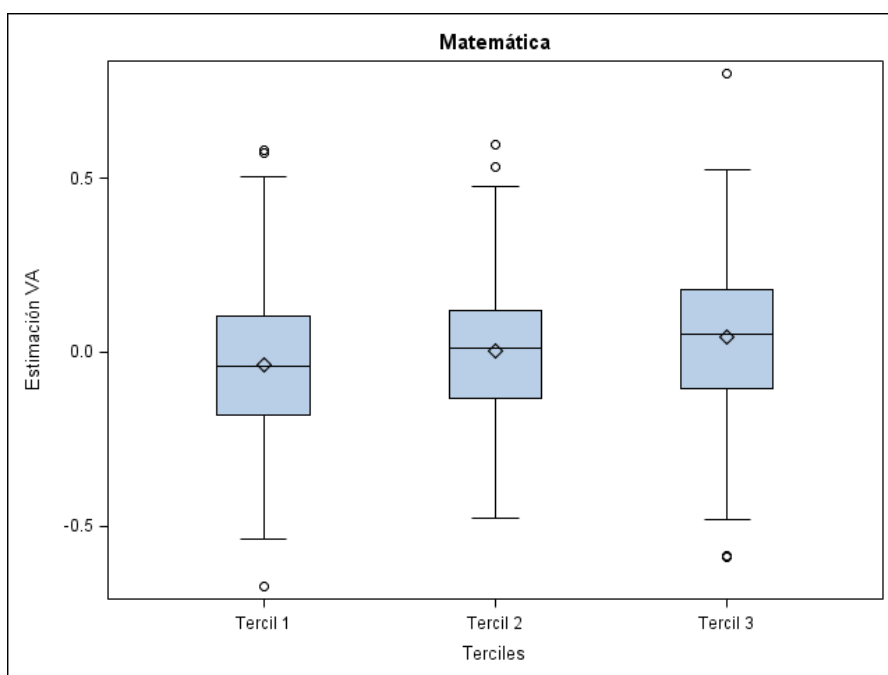
Estimaciones de VA (M6) por nivel de desempeño en el puntaje final de la ED (muestra completa)



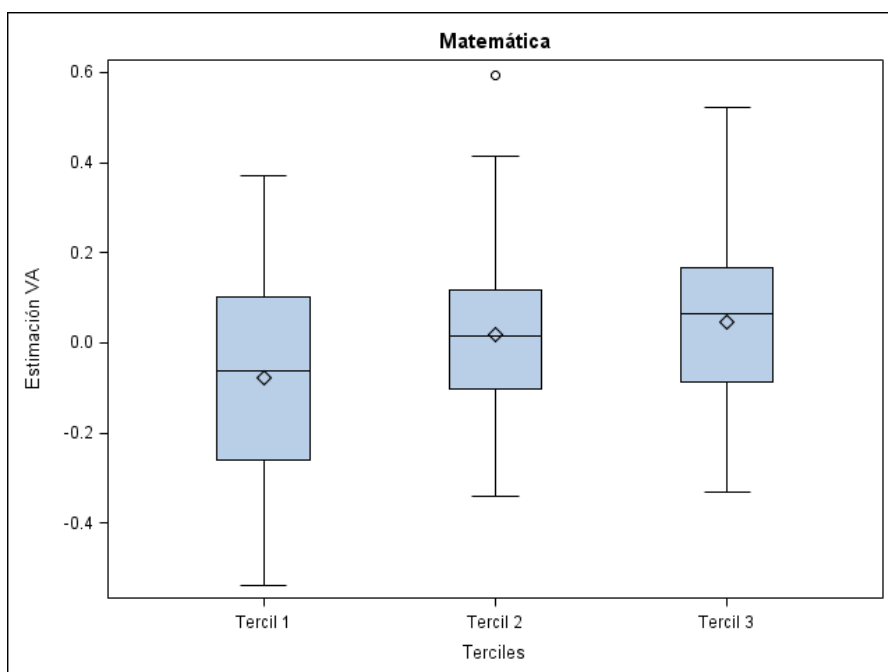
Estimaciones de VA (M6) por nivel de desempeño en el puntaje final de la ED (muestra 2006)



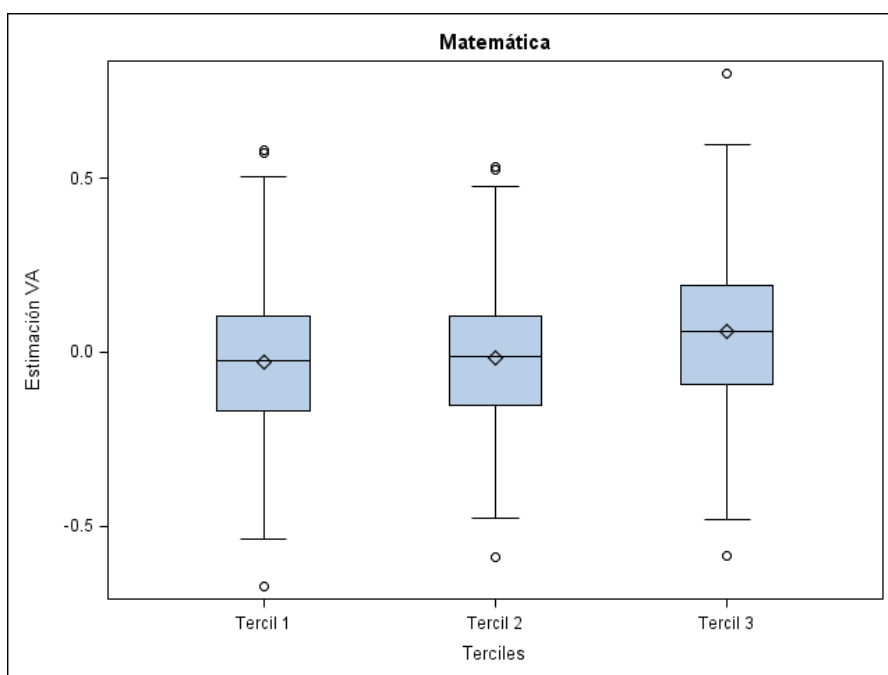
Estimaciones de VA (M6) por tercil en el portafolio de la ED (muestra completa)



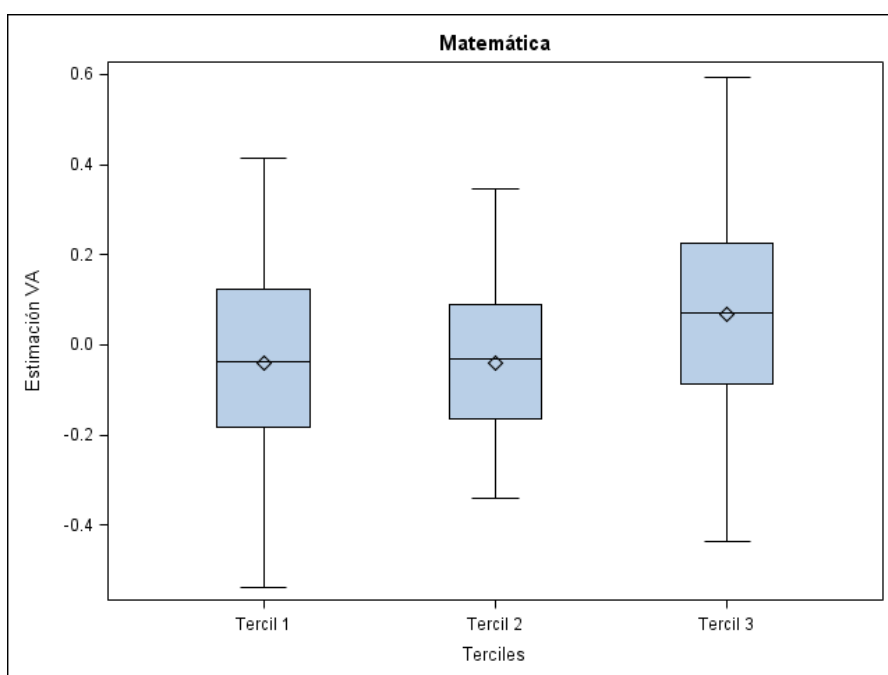
Estimaciones de VA (M6) por tercil en el portafolio de la ED (muestra 2006)



Estimaciones de VA (M6) por tercil en el puntaje final de la ED (muestra completa)



Estimaciones de VA (M6) por tercil en el puntaje final de la ED (muestra 2006)



Resultados de las estimaciones de los seis modelos de valor agregado

Lenguaje (coeficientes no estandarizados)

	M1	M2	M3	M4	M5	M6
<i>Nivel individual</i>						
Lenguaje 2004			0.535** (100.25)	0.520** (95.98)	0.486** (89.87)	0.472** (86.20)
Matemática 2004			0.215** (39.74)	0.202** (36.63)	0.176** (31.42)	0.159** (27.87)
Hombre	-0.010* (-2.06)	-0.008 (-1.65)			-0.028** (-6.86)	-0.022** (-5.38)
Educación madre	0.068** (12.92)	0.064** (12.16)			0.030** (7.23)	0.018** (4.44)
Subsidio	-0.007 (-1.62)	-0.002 (-0.62)			-0.006 (-1.57)	-0.002 (-0.64)
Promedio de ⁸ notas	0.368** (78.70)	0.364** (77.94)			0.142** (33.78)	0.149** (35.58)
(Intercepto)	-0.277** (-10.98)	-4.662** (-15.44)	-0.028** (-3.27)	-0.057** (-8.13)	-0.086** (-5.59)	-0.134 (-0.75)
<i>Nivel profesor</i>						
Lenguaje 2004 por profesor				0.153** (7.92)		0.127** (5.74)
Matemática 2004 por profesor				0.003 (0.15)		0.051* (2.26)
Proporción de hombres		-0.013 (-0.89)				-0.012 (-1.54)
Subsidio por profesor		-0.175** (-12.21)				0.000 (-0.02)
Promedio de notas por profesor		0.272** (18.78)				0.007 (0.80)
Heterogeneidad lenguaje		-0.058** (-3.86)				0.002 (0.27)
Heterogeneidad matemática		0.004 (0.30)				-0.010 (-1.47)
Heterogeneidad social		-0.015 (-1.23)				-0.003 (-0.54)

Estimación máxima verosimilitud. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.

⁸ Promedio de notas está centrado al promedio de notas por escuela

Lenguaje (coeficientes estandarizados)

	M1	M2	M3	M4	M5	M6
<i>Nivel individual</i>						
Lenguaje 2004			0.535** (100.25)	0.520** (95.98)	0.486** (89.87)	0.472** (86.20)
Matemática 2004			0.215** (39.74)	0.202** (36.63)	0.176** (31.42)	0.159** (27.87)
Hombre	-0.020* (-2.06)	-0.016 (-1.65)			-0.057** (-6.86)	-0.045** (-5.38)
Educación madre	0.019** (12.92)	0.018** (12.16)			0.008** (7.23)	0.005** (4.44)
Subsidio	-0.016 (-1.62)	-0.006 (-0.62)			-0.013 (-1.57)	-0.005 (-0.64)
Promedio de ⁹ notas	0.749** (78.70)	0.741** (77.94)			0.289** (33.78)	0.304** (35.58)
(Intercepto)	-0.277** (-10.98)	-4.662** (-15.44)	-0.028** (-3.27)	-0.057** (-8.13)	-0.086** (-5.59)	-0.134 (-0.75)
<i>Nivel profesor</i>						
Lenguaje 2004 por profesor				0.167** (7.92)		0.139** (5.74)
Matemática 2004 por profesor				0.003 (0.15)		0.054* (2.26)
Proporción de hombres		-0.057 (-0.89)				-0.055 (-1.54)
Subsidio por profesor		-1.044** (-12.21)				-0.001 (-0.02)
Promedio de notas por profesor		0.951** (18.78)				0.026 (0.80)
Heterogeneidad lenguaje		-0.009** (-3.86)				0.000 (0.27)
Heterogeneidad matemática		0.000 (0.30)				-0.001 (-1.47)
Heterogeneidad social		-0.035 (-1.23)				-0.007 (-0.54)

Estimación máxima verosimilitud. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.

⁹ Promedio de notas está centrado al promedio de notas por escuela

Matemática (coeficientes no estandarizados)

	M1	M2	M3	M4	M5	M6
<i>Nivel individual</i>						
Lenguaje 2004			0.176** (36.09)	0.169** (34.40)	0.132** (27.69)	0.125** (26.14)
Matemática 2004			0.532** (106.72)	0.523** (104.05)	0.444** (88.93)	0.434** (86.52)
Hombre	0.308** (34.28)	0.310** (34.39)			0.161** (21.53)	0.167** (22.12)
Educación madre	0.016** (11.80)	0.015** (11.16)			0.005** (4.93)	0.003** (3.07)
Subsidio	-0.008 (-0.92)	-0.002 (-0.23)			-0.003 (-0.45)	0.000 (0.06)
Promedio de notas¹⁰	0.845** (100.14)	0.839** (99.54)			0.423** (55.02)	0.431** (56.09)
(Intercepto)	-0.417** (-15.57)	-4.830** (-13.92)	-0.061** (-4.95)	-0.092** (-9.62)	-0.193** (-10.90)	-0.143 (-0.59)
<i>Nivel profesor</i>						
Lenguaje 2004 por profesor				0.081** (2.81)		0.128** (3.69)
Matemática 2004 por profesor				0.158** (5.66)		0.161** (4.90)
Proporción de hombres		-0.013 (-0.19)				-0.081 (-1.74)
Subsidio por profesor		-1.044** (-10.49)				0.058 (0.90)
Promedio de notas por profesor		1.008** (17.51)				0.013 (0.30)
Heterogeneidad lenguaje		-0.017** (-6.16)				-0.002 (-1.01)
Heterogeneidad matemática		0.004 (1.34)				0.001 (0.57)
Heterogeneidad social		-0.072* (-2.17)				-0.032 (-1.74)

Estimación máxima verosimilitud. * p<0.05, ** p<0.01. En paréntesis se muestra el valor t.

¹⁰ Promedio de notas está centrado al promedio de notas por escuela

Matemática (coeficientes estandarizados)

	M1	M2	M3	M4	M5	M6
<i>Nivel individual</i>						
Lenguaje 2004			0.176** (36.09)	0.169** (34.40)	0.132** (27.69)	0.125** (26.14)
Matemática 2004			0.532** (106.72)	0.523** (104.05)	0.444** (88.93)	0.434** (86.52)
Hombre	0.152** (34.28)	0.153** (34.39)			0.079** (21.53)	0.082** (22.12)
Educación madre	0.055** (11.80)	0.052** (11.16)			0.018** (4.93)	0.011** (3.07)
Subsidio	-0.003 (-0.92)	-0.000 (-0.23)			-0.001 (-0.45)	0.000 (0.06)
Promedio de notas ¹¹	0.414** (100.14)	0.411** (99.54)			0.207** (55.02)	0.211** (56.09)
(Intercepto)	-0.417** (-15.57)	-4.830** (-13.92)	-0.061** (-4.95)	-0.092** (-9.62)	-0.193** (-10.90)	-0.143 (-0.59)
<i>Nivel profesor</i>						
Lenguaje 2004 por profesor				0.072** (2.81)		0.114** (3.69)
Matemática 2004 por profesor				0.151** (5.66)		0.154** (4.90)
Proporción de hombres		-0.003 (-0.19)				-0.019 (-1.74)
Subsidio por profesor		-0.171** (-10.49)				0.009 (0.90)
Promedio de notas por profesor		0.294** (17.51)				0.003 (0.30)
Heterogeneidad lenguaje		-0.106** (-6.16)				-0.010 (-1.01)
Heterogeneidad matemática		0.023 (1.34)				0.005 (0.57)
Heterogeneidad social		-0.033* (-2.17)				-0.015 (-1.74)

Estimación máxima verosimilitud. * p<0.05, ** p<0.01. En paréntesis se muestra el valor

¹¹ Promedio de notas está centrado al promedio de notas por escuela

Model fit

	LEN							MAT						
	MN	M1	M2	M3	M4	M5	M6	MN	M1	M2	M3	M4	M5	M6
U0	0.35	0.28	0.14	0.04	0.02	0.05	0.02	0.43	0.37	0.19	0.10	0.05	0.13	0.05
R	0.66	0.53	0.53	0.37	0.37	0.35	0.35	0.59	0.42	0.42	0.31	0.31	0.28	0.27
StdDev U0	0.59	0.53	0.37	0.21	0.15	0.22	0.15	0.66	0.61	0.43	0.32	0.23	0.35	0.23
StdDev R	0.81	0.73	0.53	0.61	0.61	0.59	0.59	0.77	0.65	0.65	0.56	0.56	0.52	0.52
Deviance	68114	61136	61630	51087	50711	49785	49308	65977	56646	56179	47769	47359	44698	44153
Estimated Parameters	3	7	15	5	7	9	19	3	7	15	5	7	9	19
AIC	68126	62193	61742	51122	50764	49867	49472	65989	56703	56289	47804	47411	44781	44311
BIC	68151	62259	61865	51163	50822	49949	49628	66014	56769	56412	47845	47469	44863	44467
ICC (ICC parcial)	0.35	0.35	0.21	0.11	0.06	0.13	0.06	0.42	0.47	0.31	0.24	0.15	0.31	0.17
Reduccion Varianza	-	0.00	0.40	0.48	0.45	-1.16	0.54	-	-0.11	0.34	0.23	0.38	-1.07	0.45
Reduccion Varianza N2	-	0.20	0.50	0.71	0.50	-1.50	0.60	-	0.14	0.49	0.47	0.50	-1.60	0.62
Reduccion Varianza N1	-	0.20	0.00	0.30	0.00	0.05	0.00	-	0.29	0.00	0.26	0.00	0.10	0.04
N Level 1	27156	27156	27156	27156	27156	27156	27156	27601	27601	27601	27601	27601	27601	27601
N Level 2	761	761	761	761	761	761	761	725	725	725	725	725	725	725

ANEXO 4: Correlaciones entre estimaciones VA

Lenguaje (muestra 2006, N=164)

	Logro	m1	m2	dif	m3	m4	m5	m6
logro	1.00							
m1	0.93	1.00						
m2	0.59	0.76	1.00					
dif	0.44	0.42	0.35	1.00				
m3	0.84	0.78	0.55	0.82	1.00			
m4	0.79	0.86	0.72	0.79	0.94	1.00		
m5	0.40	0.38	0.34	0.97	0.82	0.78	1.00	
m6	0.38	0.42	0.46	0.94	0.79	0.82	0.97	1.00

Matemática (muestra 2006, N=151)

	logro	m1	m2	dif	m3	m4	m5	m6
logro	1.00							
m1	0.91	1.00						
m2	0.61	0.79	1.00					
dif	0.87	0.84	0.62	1.00				
m3	0.91	0.87	0.64	0.99	1.00			
m4	0.86	0.94	0.79	0.94	0.94	1.00		
m5	0.42	0.45	0.51	0.78	0.73	0.71	1.00	
m6	0.40	0.50	0.57	0.76	0.71	0.74	0.98	1.00