
EL RETO DE MEDIR EL TALENTO ACADÉMICO A PARTIR DE LAS NOTAS DE LA SECUNDARIA

Documento de Discusión N°8

Alfonso Rodríguez Saldarriaga,
Fredy Ruiz Condori, Víctor Salazar Córdor.

El objetivo central del presente trabajo es presentar dos metodologías de corrección del promedio de notas a nivel secundaria de postulantes a la Beca Excelencia Académica del Programa Nacional de Becas y Crédito Educativo, incorporando el componente tendencial del desempeño académico: PAS y PCTC. Para ello, se revisó una serie de trabajos anteriores que reflejan la importancia de realizar ajustes al promedio simple de notas.

Palabras Clave: Desempeño académico, Beca Excelencia Académica, ajustes.

Citar como: Salazar-Córdor, V., Rodríguez-Saldarriaga, A. y Ruiz-Condori, F. (2015). El reto de medir el talento académico a partir de las notas de secundaria. En: Evidencia para una política de inversión en el talento; Vol. 2. Hans Contreras Pulache Editor. Programa Nacional de Becas y Crédito Educativo. Ministerio de Educación.

1. Introducción

El rendimiento académico ha sido estudiado desde diversas áreas y bajo diferentes enfoques, constituyéndose así en un campo de investigación sobre el que giran muchas políticas públicas basadas en evidencia. En concreto, en nuestro país el acceso a un subsidio o crédito para la educación superior pasa invariablemente por tomar en consideración esta variable. De acuerdo con la evidencia disponible, el rendimiento académico constituye una garantía de inversión, y aquellos países que tienen mejores niveles de inversión en educación (y que consideran al rendimiento académico como eje en la estructuración de sus políticas) presentan una *performance* económica y social mejores que aquellos países en los que la educación se ve como un gasto o no se le toma como prioridad (Krueger y Lindahl, 2000¹; Aghion, Boustan, Hoxby y Vandenbussche, 2009²; Stevens y Weale, 2003³).

Sin embargo, ¿qué es el rendimiento académico?, y más precisamente, ¿cómo se operacionaliza esta variable? A manera de resumen podría decirse que el rendimiento académico es una variable que mide, cuantitativamente, los logros académicos obtenidos tras el paso de un estudiante por un sistema educativo (Lei, Bassiri y Mathew, 2001). Es decir: el rendimiento académico existe en el contexto estudiante-sistema educativo, y estaría haciendo referencia al logro de aprendizajes en un contexto educacional. En la práctica, se toman en cuenta dos etapas marcadamente distintas en la vida del estudiante para la evaluación y análisis de los determinantes y predictores del rendimiento académico: el periodo de la Educación Básica Regular y el periodo de la Educación Superior. Ambos constituyen, sin lugar a duda, una continuidad; y en tanto continuidad es que se puede dar sustento a los estudios que buscan definir los determinantes o predictores. El estudio del rendimiento académico en la educación superior ha recibido un enfoque un poco más intenso que el del periodo escolar debido a que, durante esta etapa, el estudiante hace más patente su carácter de sujeto de inversión debido a que es aquí (en la juventud) cuando entran a tallar

¹ Los autores hacen referencia a Corea del Sur y su crecimiento económico acelerado correlacionándolo con la inversión en educación básica (primaria y secundaria).

² Por ejemplo, Europa creció más rápido que EEUU después de la 2da Guerra Mundial invirtiendo en educación secundaria.

³ Se consideran aquí formas de inversión ligadas a educación superior.

contextos que trascienden el entorno educativo como tal, a saber: el mundo laboral, y con el mismo, todo el abanico de oportunidades y condicionantes del desarrollo que envuelven la vida de una persona y su entorno social cercano.

Operativamente, el rendimiento académico se ha trabajado con los resultados de una evaluación única o estandarizada. En entornos donde esto no es posible, como en nuestro país por ejemplo, en reemplazo de la operativización anterior, se consideran las notas que los estudiantes logran en la educación secundaria. Esto, dada la amplia variedad de la oferta educativa básica, entre otras cosas, constituye, sin más decirlo, en una fuente potencial de sesgos metodológicos y de inequidades al momento de la elaboración de políticas y la toma de decisión. Por ello, múltiples autores han elaborado modos de “ajustar” o “corregir” estas notas para que sirvan mejor en tanto “predictores”. Existe un fenómeno del que poco se ha estudiado en este marco de propuestas de “ajustes” metodológicos: la condición del rendimiento en tanto desarrollo progresivo, es decir: el desempeño académico. Ver las notas de todos los años de la educación y considerar su comportamiento a lo largo de dichos años (justamente lo que se denomina “desempeño académico”) y al mismo tiempo proponer una metodología que permita su operativización, es el eje central sobre el que se erige el presente trabajo.

En la primera parte, se realizará un breve recuento acerca de las medidas del rendimiento académico durante la educación escolar y su relación con el rendimiento en el periodo de educación superior. En esta parte se ponen en relieve los esfuerzos realizados por diversos autores al momento de elaborar metodologías que permitan optimizar el uso de las notas de la secundaria como predictores del rendimiento académico (los antes mencionados: “ajustes”). En la segunda parte, se caracteriza el desempeño académico y se proponen dos metodologías para operativizarlo. En la tercera parte, se elabora un modelo predictivo para evaluar la pertinencia de las metodologías propuestas. Finalmente, se muestran las conclusiones y recomendaciones de política.

2. Antecedentes

Luego de concluir los estudios de la educación secundaria, todo egresado cuenta con una serie de notas que reflejan su talento académico (estas notas se dan por cada curso y por cada año de estudio). Aun cuando hay autores que han criticado la consistencia de estas notas⁴, en materia de generación de evidencia son una herramienta de uso obligatorio para el diseño y ejecución de políticas educativas, sobre todo en un país como el Perú que no cuenta con una medida estandarizada que permita determinar el nivel de logros académicos al final de la educación básica. Así, en este contexto, el uso más común que se le da a estas notas es la del cálculo del “promedio de la secundaria”, que no es sino una media aritmética de todas las notas obtenidas en todos los cursos por cada año de educación secundaria (Noble, 2003; Hoffman y Lowitzki, 2005; Geiser y Santelices, 2007). Aun cuando este promedio pueda reflejar alguna medida de esfuerzo del estudiante (Belfied y Crosta, 2012), por si sola presenta serios problemas de sesgo y omisión de aspectos determinantes. Diversos autores (Young, 1990; Stricker, Rock, Burton, Muraki y Jirele, 1994; Caulkins y cols., 1996; Linn 1996; Lei y cols., 2001; Hoffman y Lowitzki, 2005) afirman la metodología basada en la media aritmética tiende a asumir que la dificultad de todos los cursos es la misma, que la habilidad de todos los estudiantes es bastante similar, que los estándares de calificación y evaluación son parecidos entre profesores y entre escuelas y, sobre todo: desconoce la naturaleza temporal del rendimiento académico que constituye justamente el comportamiento del estudiante a lo largo de todos los años de la educación secundaria. Estos sesgos, implícitos al momento de calcular el promedio simple de las notas del colegio, terminan operativizando una variable que, de este modo, se encuentra muy alejada, o por lo menos desconectada de la realidad.

Para evitar esto se han propuesto medidas de ajuste o corrección. La primera alternativa de corrección es la de Aiken (1963) que asigna una ponderación prefijada a los cursos, de manera que se realiza una especie de “curva”

⁴ Se ha sugerido que un porcentaje de los profesores corrigen con estándares bajos debido a que se tiene un interés por las evaluaciones de opinión de los alumnos (Lei y cols., 2001); esto puede resultar en un fenómeno que se conoce como “inflación de notas” (Bejar y Blew, 1981; Caulkins, Larkey y Wei, 1996), en el cual la nota del estudiante se ve falsamente incrementada por factores ajenos a su habilidad y/o esfuerzo (Young, 1990).

para tratar de homogenizar los estándares de evaluación. No obstante, esta propuesta tiene problemas de sesgo considerables contra aquellos alumnos que tienen una habilidad o performance de mayor nivel. La evidencia señala que los estudiantes habilidosos y no habilidosos no se encuentran distribuidos de manera homogénea⁵ (Caulkins y cols, 1996). Una segunda propuesta es la de Elliot y Strenta (1988) en la que se elabora un índice de estándares de calificación comparando las notas promedio de aquellos estudiantes que hayan llevado cursos en dos instituciones distintas, dando como resultado una diferencia de pares. El promedio de estas diferencias es el valor del “índice” que proponen Elliot y Strenta⁶. Estos autores controlan por diferencias para mejorar aún más la predictibilidad del promedio de notas ajustado con respecto a medidas de pre-admisión a la universidad de Duke⁷.

Otra alternativa es el modelo de Young (1990) en el que se propone usar un concepto llamado *Item Response Theory* (ITR o teoría de respuesta al ítem). En esta metodología, Young (1990) considera una división de los cursos en dos grandes clases: “ciencias naturales” y “humanidades y ciencias sociales”. A su vez, la segunda clase es dividida en dos subgrupos (humanidades y ciencias sociales). De esta forma, se elabora una matriz por grupo de curso con el fin de estimar la habilidad del estudiante. Se utiliza el método del *Grade Response Theory*, en el cual se propone una ecuación de probabilidad con tres parámetros tal como se muestra a continuación:

$$\pi_{nik} = \frac{\exp(\alpha_i (\Theta_n - \tau_k))}{1 + \exp(\alpha_i (\Theta_n - \tau_k))}$$

En donde: π_{nik} es la probabilidad de que un alumno n tenga un score mayor o igual a k en el ítem i . El escalar α_i es un parámetro de discriminación, τ_k es el nivel de habilidad requerida tal que la probabilidad de que la respuesta se encuentre por encima del umbral k sea de 0,5. La estimación de la habilidad

⁵ Cursos distintos necesitan curvas distintas ya que aquellos estudiantes con menor habilidad tienden a ir hacia los cursos más laxos para evaluar.

⁶ Basado en el hecho de que aquellas personas que tienen afinidad por algunas materias (tales como Física y Matemáticas) obtienen mejores notas en cursos relacionados con esos campos que en cursos que tienen que ver con otras disciplinas (ciencias sociales o humanidades).

⁷Promoción de estudiantes de 1983.

del estudiante (θ_n) se basa en un procedimiento iterativo para cada uno de los estudiantes y para cada uno de los parámetros de la función de probabilidad conjunta y se elabora el promedio-ITR ponderando el promedio normal con el parámetro θ_n .

En los resultados de esta metodología se puede ver que el promedio ajustado por ITR tiene una *performance* mejor dentro de un modelo lineal que predice (añadiendo otras variables) el rendimiento académico del estudiante en la educación superior. Aun así, como el autor menciona, que el modelo presente mejores parámetros de predicción no significa que el promedio ponderado por ITR tenga una validez mayor que el promedio normal. Se requiere evidencia que valide la construcción de ambas medidas para validar esta hipótesis (Young, 1990; Lei y cols., 2001). Este ajuste presenta un avance en intentar mejorar la metodología del cálculo del promedio de notas para poder controlarla por las heterogeneidades inherentes a los mismos estudiantes, a las organizaciones educativas y a los estándares de calificación de los profesores⁸. Las dificultades radican en la necesidad de una gran cantidad de muestra para poder estimar los parámetros de la manera más eficiente posible y en la dificultad del método de iteración.

Por último, se encuentra la propuesta por Bassiri y Schulz (2003). Aquí, el ajuste se hace mediante la utilización del modelo de Rasch (*Partial Credit Model* y *Rating Scale Model*) y la utilización de un examen estandarizado en los Estados Unidos para volver comparable al promedio de notas de colegio (High School GPA). La solución al problema de diferencias en dificultad de cursos (inter escolar e intra escolar) se realiza mediante la estimación de dos parámetros y un *ratio* de probabilidades para poder obtener una medida corregida del GPA escolar. La ecuación de cálculo es como sigue:

$$\ln \left(\frac{P_{nij}}{P_{ni(j-1)}} \right) = \beta_n - \delta_i - \tau_j$$

En donde: P_{nij} representa la probabilidad de que un alumno n tenga una nota j en el curso i y $P_{ni(j-1)}$ es la probabilidad de que un estudiante n tenga una nota

⁸El método es una adaptación del método de *Grade Response Theory* (Samejima, 1969).

$j-1$ en el curso i . Los parámetros β_n , δ_j y τ_j representan el Promedio Corregido, la dificultad del curso y el grado de exigencia requerida para obtener la nota j o más, dado que el estudiante tiene al menos $j-1$.

La estimación de los parámetros (δ_j ; τ_j) se realiza mediante la utilización del modelo de Rasch⁹ (a nivel de colegio y a nivel general respectivamente), para luego estimar el parámetro β_n del promedio corregido utilizando la nota obtenida en el examen estandarizado como variable común entre todos los individuos evaluados y corregir por dificultad (estándar de evaluación). Así como los autores antes mencionados, Bassiri y Schulls (2003) encuentran que el promedio corregido por criterios de evaluación, dificultad y habilidad del estudiante presenta un ajuste mejor en los modelos de regresión lineal (criterio de R cuadrado en un modelo multivariado con otras variables de control). Aun cuando el artículo de estos autores es el único que hace referencia a correcciones en el promedio de notas de secundaria, las alternativas propuestas por los demás autores pueden tener validez para el mismo periodo educativo. Esto debido a que los cursos pueden dividirse en grandes grupos como en Young (1990) y este promedio puede sufrir de los mismos problemas que los mencionados en la literatura revisada.

Por otro lado, dadas las condiciones en otros contextos, los investigadores han usado como predictor del éxito académico en la educación superior los resultados en un examen único y estandarizado. Exámenes tales como el *Scholastic Aptitud Test* (SAT), el *American College Test* (ACT) o la "Prueba de Selección Universitaria" (PSU) son algunos ejemplos de exámenes estandarizados que buscan medir la capacidad y aptitudes del egresado de la educación básica de cara a iniciar estudios de educación superior¹⁰. Del mismo modo, la validez de los exámenes estandarizados recae sobre la interpretación de los resultados y del corte que se elija para la aceptación (o no aceptación) del postulante. Elliot y Strenta (1988) usan este tipo de evaluación, entre otras, como medidas de predicción de éxito estudiantil en la universidad. Empero, así como con el promedio de notas, este tipo de medida también se encuentra sujeta a críticas.

⁹ El modelo de Rasch, en términos sencillos, representa la estructura que deberían tener los datos para poder realizar algún análisis estadístico de ellos. Permite estimar la dificultad de cada curso y por ende corregir por estándares de evaluación entre colegios y dentro de los colegios (Bassiri y Schulz 2003).

¹⁰ Este tipo de pruebas son de carácter binario: ¿El estudiante está listo para educación de nivel superior? (Belfield & Crosta 2012).

Geiser y Santelices (2007) encuentran que los resultados de los exámenes ACT o SAT tienen una correlación mucho más fuerte con el *background* del estudiante que con el desempeño universitario *per se*. Sus resultados exponen que las notas finales de los exámenes tipo ACT o SAT tienen una correlación negativa con aquellos estudiantes que pertenecen a una minoría. Además, el modelo multivariado que elaboran muestra que el promedio de notas tiene un coeficiente de predicción mayor¹¹.

3. Operatividad del desempeño académico

La necesidad de corregir el Promedio Simple de notas de secundaria se vuelve evidente cuando se realiza un análisis preliminar de la información que se tiene con respecto a los potenciales becarios de la Convocatoria 2014 para la modalidad de Beca Excelencia Académica (BEA)¹². El primer resultado del análisis de los datos mostrado en la Tabla 1 muestra los inconvenientes de la utilización del Promedio Simple de notas como potencial instrumento de política educativa.

Tabla 1: Comparación de puestos de acuerdo al Promedio Simple y a la nota del Evaluación Estandarizada

Puesto Nota	Nota	Promedio S	Puesto P,
637	71,75	19,28	1
241	78,53	16,47	700
1	99,77	16,15	961
700	70,71	16,11	992

Fuente: Área de Evaluación y Generación de Evidencia, 2015. Elaboración propia.

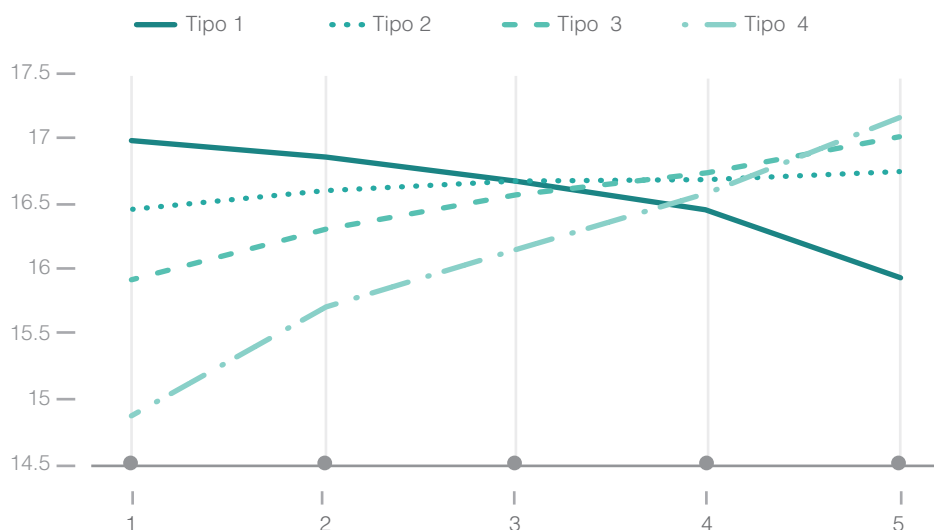
¹¹ El modelo que incluye sólo al promedio ajusta mejor que el que sólo tiene el resultado del ACT (o SAT), y cuando se combinan los dos, el modelo ajusta aún mejor (Geiser y Santelices 2007).

¹² BEA es una modalidad de Beca del Programa Nacional de Becas y Crédito Educativo (PRONABEC), destinada a los jóvenes con un promedio de notas de secundaria mayor a 16 (o que hayan ocupado el 1er o 2do puesto en su promoción).

Al realizar un análisis visual, la heterogeneidad es evidente. Se esperaría que el postulante con el mayor promedio se encuentre mucho mejor posicionado en el orden de mérito con respecto a la nota obtenida en la Evaluación Estandarizada. Sin embargo, esta expectativa se encuentra equivocada: el estudiante que obtuvo el primer puesto en el examen se encuentra por debajo del cincuenta por ciento superior en el *ranking* del promedio de notas del colegio. Análogamente, el postulante que presenta el mejor promedio de notas de la secundaria se encuentra alrededor del tercio superior de los postulantes que rindieron el examen. Esto permite dilucidar las diferencias en todos los aspectos mencionados en los antecedentes y la potencial presencia de inflación en las notas de secundaria.

Si bien es cierto que aquellas medidas ajustadas del promedio tienen un grado de predictibilidad mayor que las medidas sin ajuste y que el promedio de notas ofrece resultados superiores que las notas de un examen estandarizado, existe una suerte de falla conceptual acerca del rendimiento académico, y que consiste en la falta de incorporación del factor tiempo. Se debe tener en cuenta que las características personales tienden a variar con el tiempo, reafirmandose o cambiando hacia otras tendencias. De acuerdo con los trabajos de Heckman y Kautz (2013), existen factores de personalidad que son determinantes para el rendimiento académico, lo que comúnmente se llaman habilidades blandas, entre ellas la perseverancia (o constancia). Para incluir esta condición temporal en la definición conceptual del rendimiento académico es que se plantea la variable “desempeño académico”. Para entender en qué consiste el “desempeño académico” de manera sencilla en un contexto de propuestas de/para política(s) se procederá a mostrar el componente tendencial del comportamiento del estudiante a lo largo de los años de estudios secundarios. Para fines prácticos, se explicará la idea detrás del concepto de temporalidad mediante la descripción del Gráfico 1, elaborado a partir de los datos de una cohorte de estudiantes de “excelencia académica” (entiéndase como aquellos que muestran un carácter de excelencia en su rendimiento académico), postulantes a la BEA. Se muestran 4 grupos o “tipos” de desempeño académico. La definición por cuartiles (“tipos”) se hizo mediante el ordenamiento de las tasas de crecimiento promedio de las notas anuales para luego promediar las notas anuales de los grupos pertenecientes a cada cuartil.

Gráfico 1: Tipos de desempeño académico en la educación secundaria



Fuente: Área de Evaluación y Generación de Evidencia, 2015. Elaboración propia.

Tal y como se muestra, las tendencias o desempeños son marcadamente diferentes. Aquellos estudiantes que pertenecen al Tipo 4 (el más alto) presentan una tendencia ascendente bastante clara y muy diferente a los del Tipo 1. Los estudiantes pertenecientes al primer grupo (Tipo 1) tienen, en promedio, un descenso de poco más de un punto en el promedio final de notas (de 17 a 15,6) y los que pertenecen al último grupo (Tipo 4) muestran un incremento de poco más de dos puntos (de 15 a 17,1). Esto expone una serie de aptitudes (habilidades cognitivas) y actitudes (habilidades blandas) que determinan, o “están detrás”, del llamado rendimiento académico.

Si esta caracterización tiene algún sentido más allá del visual, se le debe someter a pruebas de hipótesis. La primera hipótesis a sostener es que el desempeño académico de los estudiantes se correlaciona con los resultados de los mismos en una evaluación única y estandarizada¹³. La segunda hipótesis que se someterá a prueba es que una metodología que busque operativizar el desempeño académico debería mostrar un compartamiento de mejor predicción (con respecto a los resultados en una evaluación única y estandarizada)

¹³ En el marco de la base de datos con la que se trabaja se cuenta con los resultados de una evaluación única y estandarizada que realizó la Pontificia Universidad Católica del Perú, y que nos servirá en el trabajo para el testeo de la hipótesis propuesta.

frente a otras propuestas del rendimiento académico (el promedio simple, por ejemplo). Para probar esta segunda hipótesis se tomará en consideración dos (2) metodologías que cuantifiquen el desempeño académico, las cuales se explicitan en el Anexo 1 y son: Promedio Ajustado de Secundaria (PAS) y Promedio Corregido por Tasa de Crecimiento (PCTC).

Del mismo modo, para probar la segunda hipótesis se realizará un modelo de regresión lineal que considerará otras variables importantes (ver más adelante para conocer más detalle sobre esto). Siguiendo la metodología de Beltrán y La Serna (2009) se postulan tres variables como fuertemente correlacionadas con el rendimiento académico: i) en primer lugar y dadas las diferencias en infraestructura, capacitación de profesores, satisfacción del personal y metodologías de enseñanza, es necesario identificar aquellas organizaciones educativas que pertenecen al sector público y privado; ii) en segundo lugar, la edad también se constituye como una variable teórica determinante sobre el rendimiento académico debido a las características de madurez mental y emocional inherentes.; iii) en tercer lugar, la variable sexo también se encuentra como determinante y se espera, como en el trabajo de Beltrán y la Serna (2009), que las mujeres tengan un rendimiento más alto que los hombres.

Para dar una idea más concreta sobre lo que hace cada ajuste del promedio de notas se muestran los resultados de la Tabla 2 donde se analizan dos casos específicos de estudiantes y se muestran los cálculos de las variables hasta ahora explicadas.

Tabla 2: Notas de la secundaria y rendimiento académico

Postulante	Educación Secundaria					Promedio Simple	PTCT	PAS
	1ro	2do	3ero	4to	5to			
A	12,82	12,64	12,45	14,55	16,18	13,73	13,79	14,28
B	13,18	13,27	13,36	14,27	14,55	13,73	13,75	13,81

Fuente: Área de Evaluación y Generación de Evidencia, 2015. Elaboración propia.

Según lo que se observa en la Tabla 2, tanto el estudiante A como el B presentan un promedio simple de 13,73. Este sería su “rendimiento académico”. Sin embargo, basta un simple análisis de las notas de la secundaria para ver que si bien ambos estudiantes presentan el mismo rendimiento académico, sus trayectorias (desempeños) muestran un comportamiento singularmente distinto. Justamente con el objetivo de “capturar” estos comportamientos es que se proponen las metodologías PAS y PTCT que, tal como se muestra, arrojan resultados diferentes. Para resumir, se podría preguntar: ¿es esto relevante? (Hipótesis 1), o: ¿cuál de las dos metodologías es mejor? (Hipótesis 2).

4. Metodología

Para el presente artículo se analizó, como se mencionó anteriormente, una base de datos conteniendo información de una cohorte de postulantes a la BEA, administrada por la Oficina de Becas Pregrado del PRONABEC. Se contó con información general, información relativa a la educación secundaria y con los resultados de una evaluación única y estandarizada. La base de datos consta de 1390 estudiantes de distintos colegios del sector público y privado. En la Tabla 3 se muestra la distribución de hombres y mujeres por tipo de gestión de colegios.

Tabla 3: Distribución de la muestra por género y tipo de gestión (%)

Género/Gestión	Pública	Privada	Total
Mujer	43,38	17,12	60,50
Hombre	29,21	10,29	39,50
Total	72,59	27,41	100,0

Fuente: Área de Evaluación y Generación de Evidencia, 2015. Elaboración propia.

La mayoría de los estudiantes analizados son mujeres y provienen de colegios públicos. La edad promedio es 17 años, siendo los hombres ligeramente (en promedio) mayores que las mujeres.

Para analizar la predictibilidad de cada una de las variables que miden el desempeño académico escolar con respecto a la nota de la evaluación única y estandarizada, se elaboró un modelo de regresión multivariada simple, tal y como se muestra en la siguiente ecuación:

$$\text{Nota Evaluación} = \beta_0 + \beta_1 X_{ik} + \beta_2 \text{Edad}_i + \beta_3 \text{Género}_i + \mu_i$$

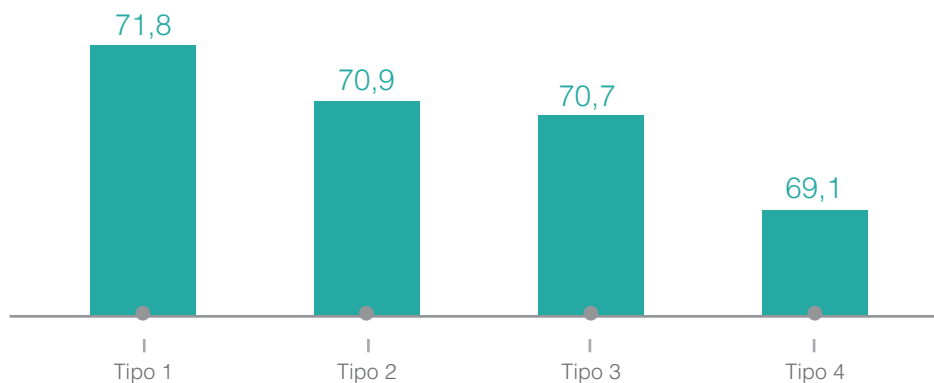
Donde X_{ik} representa cada uno de los promedios, el Promedio Simple, el PAS y el PCTC. De esta forma, se compararán los estadísticos correspondientes a cada una de las estimaciones: el R^2 ajustado, la significancia de cada uno de los componentes, y el módulo del coeficiente (el tamaño de cada Beta). Es importante resaltar que, dada la restricción de la base de datos, estas fueron las únicas variables que se pudieron identificar para poder ser incluidas en el modelo. Desde ya esto marca un escenario fértil para otros estudios de interés.

La implementación del modelo econométrico se elaboró por fases, partiendo de un modelo simple univariado para luego agregar las demás variables y observar si es que existe algún cambio en la significancia de los estimadores con una especificación similar a la que se presenta en la metodología de Beltrán y La Serna (2009).

5. Resultados

Para el contraste de la primera hipótesis se realizará un análisis visual mediante la explicación de los Gráficos 2 y 3. De esta forma, el Gráfico 2 muestra un resultado como sigue:

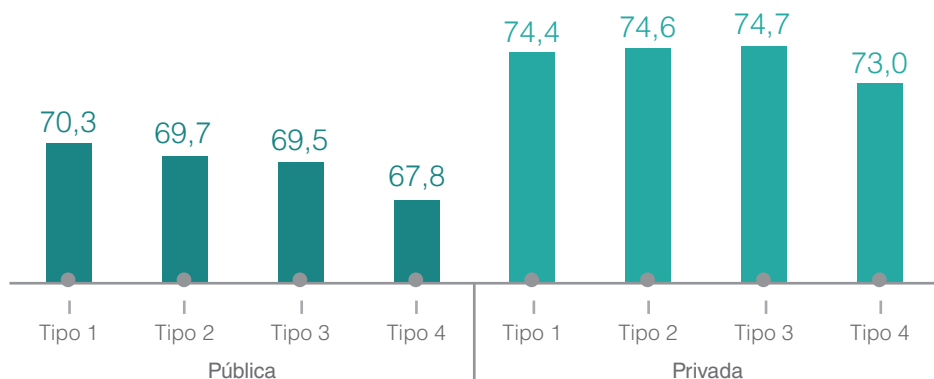
Gráfico 2: Promedios del resultado en la evaluación estandarizada por tipo de desempeño



Fuente: Área de Evaluación y Generación de Evidencia, 2015. Elaboración propia.

El análisis visual muestra un comportamiento de los datos distinto al que se esperaba en la primera hipótesis del presente trabajo. Se muestra que el resultado obtenido en la evaluación única y estandarizada es menor, en promedio, en aquellos estudiantes que pertenecen al Tipo 4¹⁴. Lo cual puede deberse a que, sin considerar ningún otro tipo de variable, el punto de partida (la nota en el primer año de secundaria) tendría una importancia mayor que la misma tasa de crecimiento en cada promedio anual. Para volver más completo el análisis, se presenta el Gráfico 3 en el que se identifica el tipo de escuela para cada grupo de estudiantes.

Gráfico 3: Promedios del resultado en la evaluación única y estandarizada por tipo de desempeño y tipo de escuela



Fuente: Área de Evaluación y Generación de Evidencia, 2015. Elaboración propia.

¹⁴ Aquellos cuyo promedio anual fue ascendiendo.

Tres incidentes son claros de acuerdo con la información expuesta en el Gráfico 3. En primer lugar, aquellos estudiantes que vinieron de escuelas privadas obtuvieron, en promedio, un resultado mucho mayor (la mínima diferencia es de 3,9 puntos y la máxima de 5,2) que los procedentes de escuelas públicas. En segundo lugar, el resultado promedio en la evaluación única y estandarizada es mucho más uniforme para los cuatro tipos de estudiante identificados que provienen de escuelas privadas. Por último, aquellos estudiantes pertenecientes al Tipo 2 y 3, en las escuelas privadas, (que han mantenido un promedio alto sin una variación muy fuerte) obtuvieron un resultado mayor. Del mismo modo, el análisis comparativo de las medias de Bonferroni (los análisis de Scheffe y Sidak dan resultados similares)¹⁵ nos indica que cuando no se controla por el tipo de escuela, todas las diferencias de medias con respecto al Tipo 4 son significativas (al 10% como mínimo) y que al introducir la variable de pertenencia a escuela privada o pública, la diferencia de medias se vuelve no significativa para los postulantes que provienen de escuelas privadas.

Para el contraste de la segunda hipótesis, se analizará el grado de ajuste, la significancia y el tamaño del coeficiente de cada una de las medidas de operativización del desempeño. Así, el resultado mostrado en la Tabla 4 presenta un comportamiento jerárquico. Las medidas que presenta un mejor grado de ajuste bajo el criterio de R cuadrado son el promedio simple y el PCTC, este último tiene el mayor coeficiente, dejando al PAS en tercer lugar.

Tabla 4: Coeficientes de Estimación (modelo univariado)

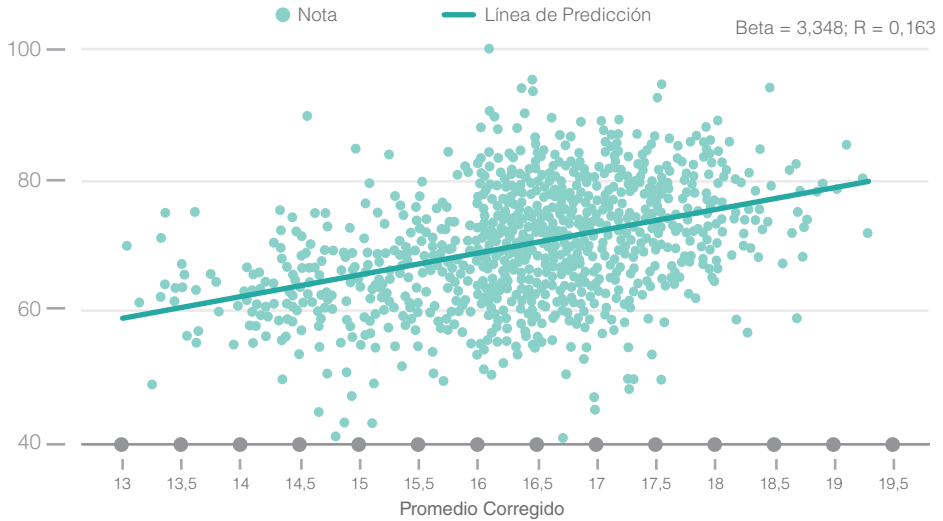
VARIABLES	EVALUACIÓN ESTANDARIZADA		
PAS	3,333*** (0,194)		
Promedio Simple		3,342*** (0,192)	
PCTC			3,348*** (0,193)
Constante	15,59*** (3,203)	15,60*** (3,167)	15,46*** (3,181)
Observaciones	1,390	1,390	1,390
R-cuadrado	0,157	0,163	0,163

***: Muestra parámetros significativos al 1%, los errores estándar robustos se encuentran entre paréntesis. Fuente: Área de Evaluación y Generación de Evidencia, 2015. Elaboración propia.

¹⁵ Ajustan los p-valores para reducir la posibilidad de Error Tipo 1 cuando se comparan más de dos grupos, ver tablas 5 y 6 de los anexos.

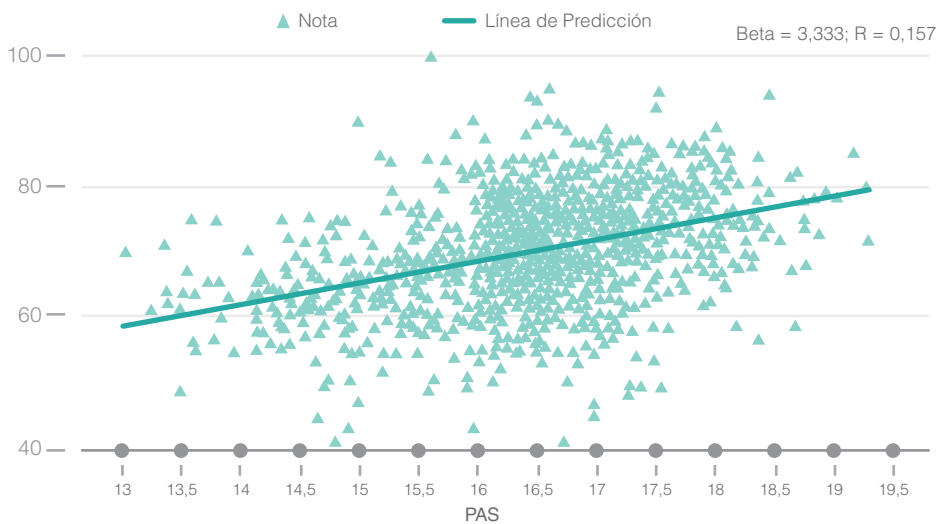
Tal y como se muestra en la Tabla 4 y los Gráficos 4 y 5, el coeficiente asociado a la estimación que utiliza al PCTC (Beta) tiene un valor de 3,348 superando al promedio simple (3,342) y PAS (3,333). Del mismo modo el factor de ajuste (R cuadrado) tiene un valor de 0,163, que es igual al ajuste del promedio simple.

Gráfico 4: Línea de ajuste MCO entre el PCTC y la nota de la evaluación única y estandarizada



Fuente: Área de Evaluación y Generación de Evidencia, 2015. Elaboración propia.

Gráfico 5: Línea de ajuste MCO entre el PAS y la nota de la evaluación única y estandarizada



Fuente: Área de Evaluación y Generación de Evidencia, 2015. Elaboración propia.

Para completar el análisis se presenta la Tabla 5 de regresión del modelo multivariado en donde la variable “género” toma el valor de 1 si el postulante es hombre (0 si es mujer), la variable “edad” representa la edad que tenía el postulante al momento de rendir la evaluación única y estandarizada y la variable “gestión” toma el valor de 1 si la organización educativa es privada (0 si es pública).

Tabla 5: Coeficientes de estimación (modelo con variables de control)

VARIABLES	EVALUACIÓN ESTANDARIZADA		
Género	3,686*** (0,402)	3,682*** (0,400)	3,689*** (0,400)
Edad	-0,788*** (0,229)	-0,779*** (0,227)	-0,779*** (0,227)
Gestión	3,675*** (0,417)	3,638*** (0,416)	3,647*** (0,416)
PAS	3,141*** (0,189)		
Promedio Simple		3,150*** (0,186)	
PCTC			3,159*** (0,187)
Constante	30,01*** (5,347)	29,87*** (5,250)	29,69*** (5,267)
Observaciones	1,390	1,390	1,390
R-cuadrado	0,248	0,253	0,253

***: Muestra parámetros significativos al 1%, los errores estándar robustos se encuentran entre paréntesis. Fuente: Área de Evaluación y Generación de Evidencia, 2015. Elaboración propia.

Al incluir las variables de género, edad y tipo de gestión los coeficientes de todos los indicadores de rendimiento escolar aumentan y el R cuadrado pasa de 0,163 a 0,253. En este caso, el ajuste del modelo que considera el promedio simple es igual al que considera el PCTC (el PAS vuelve a quedar relegado), nuevamente el PCTC cuenta con el mayor coeficiente. La variable de edad presenta un coeficiente negativo y significativo debido a que aquellos postulantes de mayor edad son los que (en promedio) han tenido un periodo de egreso anterior a los postulantes menores, y los conocimientos y nivel de dominio de las materias evaluadas puede no encontrarse en el óptimo. Los hombres tuvieron un mejor rendimiento, en promedio, que las mujeres y aquellos estudiantes que procedentes de colegios privados tienen un rendimiento mayor que aquellos de colegios públicos. Aunque la mejora de los estadísticos sea pequeña, el

resultado puede indicar que al introducir variables de control adicionales (como características del colegio y de autopercepción) la diferencia entre los modelos que contemplan al PCTC y al promedio simple puede volverse más evidente.

6. Conclusiones y Recomendaciones de Política

Debido a la inexistencia de una evaluación única y estandarizada para los jóvenes que terminan la secundaria en el Perú, es necesario (lamentablemente) utilizar la información que brinda promedio de notas de colegio como herramienta de política. Esta medida, más que una elección, representa una consecuencia inmediata e irremediable frente a la ausencia de dicha evaluación, que sería justamente la recomendación principal a la que apunta este documento. Existe pues una necesidad importante e impostergable de contar con una medición justa y equitativa de los logros de la educación secundaria. Como instrumento de política, este examen único y estandarizado permitirá optimizar el otorgamiento de becas o créditos educativos y al mismo tiempo direccionar el ejercicio de la política pública según los logros de aprendizaje de los estudiantes. Ahora bien, conscientes de las implicancias de plantear una recomendación como ésta, ya sea en términos financieros como logísticos, se sugiere empezar con una suerte de prueba piloto. Por ejemplo, un examen único y estandarizado para Beca 18 (que es la principal beca de pregrado en el Perú). Esto permitirá ir ajustando materiales o marcos conceptuales para luego aplicar una prueba con carácter censal.

Aun cuando los autores mencionados concuerdan en que es necesario un ajuste del cálculo de promedio de notas de la secundaria, ninguno de ellos investigadores ha contemplado la incorporación del factor temporal, como hemos hecho en estas páginas bajo el concepto de “desempeño académico”. Así, al separar los tipos de desempeño académico (Tipo 1 al 4) por gestión educativa (pública o privada) el primer resultado en primera instancia contraintuitivo se volvió más acorde con lo postulado en la Hipótesis 1. Aquellos estudiantes que mantuvieron un rendimiento alto durante los 5 años de secundaria obtuvieron un mejor resultado en promedio. Por otro lado, las mejoras en el ajuste de las metodologías propuestas (PAS y PCTC) no fueron tan evidentes como se esperaba que fueran. El R cuadrado permaneció casi igual y el coeficiente mejoró en un porcentaje muy pequeño. Esto podría deberse a un tema de tamaño muestral o algo más profundo y que pasa por el hecho de que al medir

el desempeño académico sin ajustar por tipo de curso, por ejemplo, se cae en los mismos errores potenciales y metodológicos del promedio simple. En este sentido, se recomienda continuar con el promedio simple para fines operativos, en tanto no se cuente con una metodología que brinde, de modo evidente, una operativización óptima del desempeño académico. Esta operativización óptima pasa por hacer ajustes a nivel de cursos recibidos. Este es un tema para próximas investigaciones que busquen trabajar con una base de datos que contenga mayor información acerca de los estudiantes con respecto a los cursos, características socioeconómicas, características de la entidad educativa de la que provienen y características de personalidad que permita corregir el promedio volviéndolo comparable entre escuelas y dentro de las mismas escuelas para luego realizar el ajuste por tendencia temporal.

7. Referencias

- **Aghion, P., Boustan, L., Hoxby, C. y Vandenbussche, J. (2009).** The Causal Impact of Education on Economic Growth: Evidence from U.S. Harvard University and CEPR, University of California, NBER, Stanford University, IMF.
- **Aiken, L. (1963).** The Grading Behavior of A College Faculty. *Educational and Psychological Measurement* 13(2).
- **Bassiri, D. y Schulz, M. (2003).** Constructing a Universal Scale of High School Course Difficulty. ACT Research Report Series.
- **Bejar, I. y Blew, E. (1981).** Grade Inflation and the Validity of the Scholastic Aptitude Test College. Entrance Examination Board.
- **Belfield, C. y Crosta, P. (2012).** Predicting Success in College: The Importance of Placement Tests and High School Transcripts. CCRC Working Paper No. 42.
- **Beltran, A. y La Serna, K. (2009).** ¿Qué explica el rendimiento académico en el primer año de estudios universitarios? Centro de Investigación de la Universidad del Pacífico.
- **Caulkins, J., Larkey, P. y Wei, J. (1996).** Adjusting GPA to Reflect Course Difficulty. The Heinz School of Public Policy and Management Carnegie Mellon University.
- **Elliot, R. y Strenta, C. (1988).** Effects of Improving the Reliability of the GPA on Prediction Generally and on Comparative Predictions for Gender and Race Particularly. *Journal of Educational Measurement*

25(4): 333-347.

- **Geiser, S. y Santelices, M. (2007).** Validity of high-School grades in predicting student success beyond freshmen year: High-School Record vs. Standardized Tests as Indicators of Four-Year College. Research & Occasional Paper Series. University of California.
- **Hiss, W. y Franks, V. (2014).** Defining Promise: Optional Standardized Testing Policies In American College And University Admissions.
- **Hoffman, J. y Lowitzki, K. (2005).** Predicting College Success with High School Grades and Test Scores: Limitations for Minority Students. The Review of Higher Education Summer 28(4): 455–474.
- **Johnson, V. (1997).** An Alternative to Traditional GPA for Evaluating Student Performance. Statistical Science 12(4): 251-269.
- **Krueger, B. y Lindahl, M. (2000).** Education and Growth: Why and for Whom? Journal of Economic Literature 39: 1101–1136.
- **Lei, P., Bassiri, D. y Mathew, E. (2001).** Alternatives to the Grade Point Average as a Measure of Academic Achievement in College. ACT Research Report Series.
- **Linn, R. (1966).** Grade Adjustments for Prediction of Academic Performance: A Review. Journal of Educational Measurement 3(4): 313-329.
- **Niu, S. y Tienda, M. (2009).** Testing, Ranking and College Performance: Does High School Matter? The Ford, Mellon and Hewlett Foundations and NSF.
- **Noble, J. (2003).** The Effects of Using ACT Composite Score and High School Average on College Admission Decisions for Racial/Ethnic Groups. American Coll Testing Program.
- **Sanchez, E. (2013).** Differential Effects of Using ACT College Readiness Assessment Scores and High School GPA to Predict First-Year College GPA among Racial/Ethnic, Gender, and Income Group. ACT Research Report Series.
- **Stevens, P., Weale, M. (2003).** Education and Economic Growth. National Institute of Economic and Social Research.
- **Stricker, L., Rock, D., Burton, N., Muraki, E. y Jirele, T. (1994).** Adjusting College Grade Point Average Criteria for Variations in Grading Standards: A Comparison of Mehots. Journal of Applied Psychology 79(2): 178-183.
- **Young, J. (1990).** Adjusting the Cumulative GPA Using Item Response Theory. Journal of Educational Measurement 27(2): 175-186.

ANEXO 1: Tablas

Tabla 6: Diferencia de medias en resultado de evaluación estandarizada

Fila-Columna	Tipo 1	Tipo 2	Tipo 3
Tipo 2	-0,909059 0,864		
Tipo 3	-1,15109 0,385	-0,242033 1,000	
Tipo 4	-2,72439 0,001	-1,81533 0,022	-1,5733 0,069

Nota: los números de las filas inferiores representan $P>|t|$ si es menor que 0.1 la diferencia es significativa al 10%. **Fuente:** Área de Evaluación y Generación de Evidencia, 2015. **Elaboración propia.**

Tabla 7: Diferencia de medias en resultado de evaluación estandarizada para escuelas públicas y privadas

Fila-Columna	PRIVADA			PÚBLICA		
	Tipo 1	Tipo 2	Tipo 3	Tipo 1	Tipo 2	Tipo 3
Tipo 2	0,176927 1,000			-0,536072 1,000		
Tipo 3	0,260637 1,000	0,08371 1,0000		-0,779016 1,000	-0,242944 1,000	
Tipo 4	-1,40047 0,907	-1,57739 0,861	-1,6611 0,777	-2,50734 0,006	-1,97127 0,038	-1,72833 0,096

Nota: Los números de las filas inferiores representan $P>|t|$ si es menor que 0.1, la diferencia es significativa al 10%. **Fuente:** Área de Evaluación y Generación de Evidencia, 2015. **Elaboración propia.**

ANEXO 2: Dos metodologías para operativizar el desempeño académico

En el artículo se hace mención a dos metodologías que operativizan el desempeño académico. A continuación se exponen de modo resumido las características de las mismas.

Promedio Ajustado de Secundaria

Se define el Promedio Ajustado de Secundaria (PAS), tal como se muestra a continuación. Sea \bar{x} el promedio simple de notas de la secundaria, este se procederá a ajustar de la siguiente manera:

$$PAS = \bar{x} + 2(\alpha - 0.5)$$

Cabe destacar que α puede tomar valores de 0 a 1, por lo que el PAS se encontrará acotado entre [0,20]. La siguiente fórmula muestra el cálculo de la variable α , que proviene de una función logística.

$$\alpha = \frac{1}{1 + e^{-\theta}}$$

Como se observa, la variable θ muestra el rendimiento académico del postulante, pues en función de dicha variable se ajustará el promedio simple de notas en la secundaria. Para el cálculo de θ se considera el siguiente procedimiento:

- i. Se calcula la diferencia entre la nota del 2do al 5to año de secundaria con respecto al primer año. Estas diferencias se elevan al cuadrado y se calcula el promedio (puesto que se tienen 4 valores, se divide entre 4 para calcular el promedio). Se debe mencionar, aunque se elevan al cuadrado las diferencias, el signo no cambiará.
- ii. Para el segundo año, se repite el procedimiento anterior, pero naturalmente el valor de referencia es la nota del segundo año (se excluye la nota del primer año, por lo que ahora la sumatoria se divide entre 3).
- iii. El procedimiento anterior se repite para el cuarto y quinto año. Cabe destacar que el cálculo es similar a la fórmula de la varianza (con la única diferencia que no cambian los signos después de elevar al cuadrado). El cálculo de θ se puede describir mediante la siguiente fórmula:

$$\theta = \frac{6}{100} \left(\sum_{i>1} \frac{\alpha (X_i - X_1)^2}{4} + \sum_{i>2} \frac{\alpha (X_i - X_2)^2}{3} + \sum_{i>3} \frac{\alpha (X_i - X_3)^2}{2} + \sum_{i>4} \frac{\alpha (X_i - X_4)^2}{1} \right)$$

Donde $\alpha = \begin{cases} 1 & \text{si } X_i - X_k > 0 \\ -1 & \text{si } X_i - X_k < 0 \end{cases}$ para todo $i > k$

Estos cuatro elementos se suman y, puesto que el rango de cada uno es de ± 100 , se re-escala multiplicando por $6/100$. El valor resultante es θ , que considera el rendimiento durante la secundaria, como se mencionó.

El PAS permitirá diferenciar entre estudiantes que, teniendo igual promedio durante la secundaria, mostraron diferentes trayectorias de rendimiento, por lo que aquel que mostró un rendimiento ascendente tendrá mayor promedio ajustado que aquel que fue decayendo. Por otra parte, aquellos estudiantes que mostraron un rendimiento constante durante la secundaria no mostrarán cambios significativos en el promedio ajustado.

Promedio de notas Corregido por Tasa de crecimiento

Ante la medida propuesta anteriormente, surgió la necesidad de una alternativa un poco más sencilla en términos metodológicos pero que también sea útil. Así como el PAS, esta medida presenta el componente tendencial. El algoritmo de corrección del promedio de notas simple es como sigue.

$$\text{Sea } P_{corr} = \bar{X} + \overline{g_{1-5}}$$

Donde \bar{X} es el promedio simple de notas anuales desde primer año de secundaria hasta el quinto año del mismo nivel $\bar{X} = \sum_{i=1}^5 \frac{X_i}{5}$ y $\overline{g_{1-5}}$ el promedio de las tasas de variación de las notas en el mismo periodo. Este Promedio Corregido castiga (o recompensa) la tendencia de las notas del estudiante.