



ESTADO DE LA NACIÓN



ESTADO DE LA EDUCACIÓN

QUINTO INFORME DEL ESTADO DE LA EDUCACIÓN

Estado de la base de datos del PIAD para el distrito 07 de Desamparados, 2013

Catherine Mata Hidalgo



01 de marzo de 2015

Nota: Las cifras de las ponencias pueden no coincidir con las consignadas por el Quinto Informe Estado de la Educación (2015) en el tema respectivo, debido a revisiones posteriores. En caso de encontrarse diferencia entre ambas fuentes, prevalecen las publicadas en el Informe.

Introducción

En el 2014 el Programa Estado de la Nación y la Fundación CRUSA acordaron la realización de un conjunto de estudios para el período 2014-2015 entre las cuales se propuso la investigación denominada: “Estudio sobre el potencial del PIAD como herramienta para evaluar el rendimiento escolar”. El mismo tenía los siguientes objetivos principales:

- Valorar la calidad de la información generada por el PIAD para analizar el rendimiento de los estudiantes y mejorar la gestión de los centros educativos.
- Aplicar técnicas estadísticas de análisis multivariado a la información generada por el PIAD en el distrito 07 de Desamparados que permitan identificar factores asociados al rendimiento de los estudiantes en las materias básicas (matemáticas, español, ciencias) considerando características socio-demográficas de los estudiantes, características de los docentes y características de los centro educativo y los servicios que estos ofrecen.
- Discutir con actores claves los hallazgos y resultados.

Este informe presenta los principales resultados obtenidos respecto al primer objetivo planteado que implicaba la revisión exhaustiva del estado de la base de datos del Programa de Informática para el Alto Desempeño (PIAD) correspondiente al año 2013 de los centros educativos del distrito 07 de Desamparados. Se analizó la calidad de la información contenida en esta base de datos, para determinar las potencialidades que ofrece para elaborar indicadores y análisis sobre el desempeño de los estudiantes.

Este documento está organizado de la siguiente manera: después de esta introducción, la primera sección presenta una descripción de los antecedentes del PIAD; la segunda describe la metodología utilizada para la revisión de la información contenida en la base de datos. En la tercera sección, se describen los resultados de esa revisión. En la cuarta parte se concluye sobre el estado de la base de datos y finalmente, en la quinta sección se plantean usos y aplicaciones factibles con la base de datos que se cuenta actualmente.

1. Antecedentes

El PIAD es un sistema de información para el seguimiento y atención individual de los estudiantes, que captura datos sobre la situación y rendimiento educativo de cada uno de ellos. Esta herramienta se ha desarrollado a partir de la premisa de que las causas principales de los rezagos en la educación pública son la falta de organización del centro educativo y la disponibilidad de información oportuna sobre los estudiantes para promover mejores estrategias de atención por parte de los docentes (PIAD, s.f.)

En el año 2008 se generó una alianza entre el Ministerio de Educación Pública (MEP), la Asociación Nacional de Educadores (ANDE), la Asociación Empresarial para el Desarrollo (AED) y la Asociación para la Innovación Social (ASIS), con el objetivo de mejorar la gestión educativa y convertir a los centros con bajo desempeño organizacional en centros de alto desempeño. Se promueve el uso de herramientas que *"hacen posible mejorar la calidad de vida y el desempeño de los docentes, directores y personal administrativo en función de atender la población estudiantil en riesgo de abandonar el centro educativo"* con un sistema –PIAD- El sistema hace posible identificar a quienes están en riesgo y atenderlos antes de que abandonen el centro educativo y sus estudios. (PIAD, s.f.)

Desde el 2011, mediante el decreto ejecutivo 36451-MEP se oficializó el uso del PIAD y se asignó a la dirección de Gestión de Calidad y Evaluación del MEP la responsabilidad de su implementación. CRUSA-Costa Rica ha apoyado este proceso mediante aportes financieros importantes para el desarrollo de este programa, como muestra del interés en su implementación.

El programa se implementó inicialmente en centros de educación primaria y actualmente abarca desde centros de educación preescolar hasta secundaria. En el 2013, el MEP y otros socios del PIAD acordaron aplicar el programa a todos los centros educativos del distrito 07 de Desamparados. Este plan piloto ha generado información de rendimiento de cada estudiante del distrito (aproximadamente 13.000 estudiantes en total) y constituye una base de información inédita en Costa Rica sobre el rendimiento individual de los estudiantes. En los próximos cinco años la gestión del PIAD enfrentará varios retos: ampliar la cobertura, la conectividad de banda ancha y el número de computadoras en los centros educativos, fortalecer la capacitación para docentes, administrativos y directores, así como demostrar la utilidad de la información generada para mejorar la gestión educativa de los centros y el rendimiento de los estudiantes.

2. Metodología

En esta sección se describe la base de datos recibida y se determina la metodología utilizada para valorar el estado de la información disponible. Para esto la sección se subdivide en dos apartados: La descripción de la información recibida y el estado de la información disponible

2.1 Descripción de la información recibida

La base de datos proviene de archivos en formato *SQL Server* y de estos se generó un archivo en formato *Excel*, separando en diferentes hojas la información de acuerdo a los siguientes temas: beneficios sociales recibidos por el estudiante, datos socioeconómicos del hogar, datos de asistencia del estudiante, calificaciones en las materias básicas recibidas por el alumno, características del centro educativo al que asiste y datos de los padres o encargados del alumno.

Cada uno de estos archivos contenía el mismo número de observaciones o individuos (13,198), así como dos posibles variables que permitirían identificar entre archivos cada individuo: ID e IDENTIFICACION. Ambas variables constituían una identificación numérica por individuo que permitieran conservar el anonimato de los estudiantes. La variable IDENTIFICACION no estaba disponible en el archivo de características del centro educativo al que asiste el estudiante; sin embargo, la variable ID de este archivo coincidía exactamente con la numeración de la variable IDENTIFICACION de las demás bases, por lo que se procede a darle el mismo nombre. Con esta variable para todos los individuos, en todos los archivos, se procedió a fusionar la información en una sola base, logrando un empate del 100% entre todos ellos; es decir, en la nueva base todos los individuos tenían la información contenida en las variables disponibles en las 6 hojas del archivo de *Excel*.

Los manuales entregados con la base de datos explican las generalidades del programa, las posibilidades de información que se pueden obtener de él y sobre todo, una descripción detallada de cómo funciona el programa a nivel de sistemas informáticos. Sin embargo, no se encontró en estos manuales información detallada referente al significado, interpretación o detalle de las variables contenidas en la base. Así por ejemplo, consultas como la unidad de medida del ingreso, nomenclatura para los datos no disponibles o que no aplican, entre otras, debieron hacerse a los encargados del PIAD posteriormente, ante la ausencia de un directorio de variables (ver anexo 1)

La base de datos del PIAD para el distrito 07 de Desamparados, del año 2013, comprende información para 15 centros educativos, de los cuales 3 son centros de educación preescolar, 5 de educación primaria (3 de las cuales incluyen información de preescolar y primaria) y 7 de secundaria (2 de los cuales son nocturnos). En términos de distribución de los 13,198 estudiantes que abarca la base, el 10% corresponden a niños de preescolar, 38% estudiantes de primaria, 52% jóvenes de secundaria y 0,5% se encuentra en alguna modalidad de educación especial -aula edad y aula integrada-. El cuadro 1 presenta la distribución de los estudiantes por centro educativo.

Cuadro 1
Número de estudiantes por centro educativo y niveles
Distrito 07 de Desamparados

Código	Centro Educativo	Niveles Impartidos	Número de estudiantes
480	Escuela San Rafael	preescolar y primaria	1.040
488	JN María Jiménez	preescolar	387
492	Escuela José T.Mora	preescolar y primaria	1.049
504	Escuela García Monge	primaria	1.176
566	JN María Retana	preescolar	239
567	Escuela Elías Jiménez	primaria	1.329
594	Escuela San Jerónimo	preescolar y primaria	698
604	JN La Valencia	preescolar	223
3982	Liceo Monseñor Rubén Odio	secundaria	1.977
3983	Liceo Calle Fallas	secundaria	1.078
3987	Liceo Roberto Gamboa Valverde	secundaria	1.129
4159	CTP Monseñor Sanabria	Educación Diversificada	1.135
4837	Colegio Nocturno Desamparados	secundaria	937
5735	UP La Valencia	secundaria	522
5827	Secc.Noct. T.V. Monseñor Sanabria	CINDEA -I a III nivel-	279
Total			13.198

Fuente: Elaboración propia con datos del PIAD

La base de datos está principalmente compuesta por variables dicotómicas o categóricas. Las variables dicotómicas son aquellas cuya respuesta toma el valor de 0 o 1, siendo 0 si la respuesta es negativa (no) y 1 si la respuesta es afirmativa (sí). Las variables categóricas toman valores discretos (como 1,2,3...) para indicar una condición o respuesta particular, definida de antemano por quién plantea las preguntas; como por ejemplo, cuando se consulta por el estado de la casa del estudiante, donde la respuesta se restringe a (0) excelente, (1) muy buena, (2) regular o (3) mala.

En menor proporción se encuentran variables discretas y continuas. Son consideradas variables discretas aquellas como el número de miembros del hogar, la edad, el número de ausencias, así como las notas de los estudiantes, pues toman valores enteros (no decimales). Por su parte, el ingreso familiar y el ingreso per cápita son las dos variables consideradas continuas, pues pueden tomar valores fraccionados o decimales. Un mayor detalle de cada variable, según el tipo de variable y los valores que toma dentro de la base de datos, se encuentra en el anexo 1 de este documento.

2.2 Información disponible

Tanto la base de datos, como sus manuales, no son claros en cuanto a la nomenclatura utilizada para el faltante de información. No se distingue claramente entre un faltante de información debido a que no se cuenta con la información

(conocido como “no disponible”-n.d.-) y entre el faltante de información de cuando esa variable no aplica –n.a.- para el individuo.

Por esta razón, para determinar lo robusto de la base, en términos de la cantidad de información contenida, se tuvo que proceder de la siguiente forma:

1. Contar el número de espacios en blanco por cada variable (“blancos”)
2. Contar el número de espacios con el valor de -1 dentro de la base, además del 0 en las variables discretas y continuas (“missing”)
3. Estimar el número de individuos que se espera no debían tener información por cada variable, basados en el nivel educativo que cursaban. Por ejemplo, el número de estudiantes en preescolar constituye la cantidad de observaciones que se espera que no tengan información sobre las notas de las asignaturas¹. Es decir, estos constituyen el número de “no aplica” esperados para esta variable. En el anexo 3 se detalla el número de estudiantes que “no aplican” por nivel, es decir, que no deben contar con información de las notas por materia.
4. Una vez definido el número de “no aplican”, el número de observaciones sin información, en cada variable, se calculó de la siguiente forma:
Observaciones sin información = (blancos + missing) – no aplican
5. Finalmente, el porcentaje de observaciones sin información o el porcentaje de faltante de información se define como,

$$\frac{\# \text{ observaciones sin información}}{13,198} * 100$$

El detalle de los resultados numéricos de estos cálculos se encuentra en el anexo 2 de este documento. Allí se realiza el cálculo del faltante de información para cada una de las variables contenidas en la base.

Debe tomarse en cuenta que, según fue indicado posteriormente por los encargados del PIAD², en una serie de n cantidad de elementos, el n+1 se usa cuando no hay datos. Esto no es digitado por el profesor, se selecciona a través de un botón de radio que solo les permite seleccionar sí o no. Sin embargo, en la base de datos se encuentran casos de falta de información con otra nomenclatura, como el espacio en blanco o el -1, por lo que se procedió a consultar cuál era la diferencia entre estas nomenclaturas y el n+1, a la fecha de cierre de este informe³ no se había obtenido respuesta, por lo que no se tomó en cuenta el n+1 como valores no disponibles y, por tanto, podría estar subestimándose el faltante de información al no contabilizar estos casos.

¹ Por ejemplo, ciencias se imparte de primero a noveno; mientras que física, química y biología de décimo a duodécimo.

² De acuerdo a correo de fecha 17/12/2014 de Laura Monge, ver detalle en anexo 1.

³ Al 01/03/2015 no se había recibido respuesta.

3. Descripción de resultados: integridad y completitud de la Base de Datos

Cuatro variables de la información del hogar y cuatro referentes a los encargados del alumno, son las variables que presentan más de 50% de observaciones sin datos. Específicamente 59% de las observaciones no tienen el dato del tipo de vivienda y estado de la casa en que habitan. Además, 79% de los individuos no reportó el ingreso familiar y per cápita del hogar. Así mismo, las notas de los alumnos para cada materia, tanto trimestralmente como el promedio final, no están disponibles en al menos el 50% de los casos.

En los datos de los encargados destaca la disparidad entre la información disponible de la madre y del padre. Mientras que 19% de las observaciones no tienen información de la madre, el 63% no tiene datos sobre el estado civil del padre sobre su nacionalidad, escolaridad y si vive o no con el estudiante.

Las variables de código presupuestario, año cursado, sección, país, adecuación, tipo de adecuación, repitencia, deserción, caso de ausentismo, prioridad en el comedor y si trabaja el estudiante, no tienen valores desconocidos. Es decir, los valores de estas variables están disponibles para el total de observaciones.

3.1. Características del estudiante y de su hogar

En la variable *adecuación* todos los estudiantes son clasificados de acuerdo a si tienen adecuación (1) o si no tienen adecuación (0). Posteriormente, en tipo de adecuación se asigna 0 a aquellos que no tienen adecuación y los que sí la tienen se clasifican en tres tipos de adecuación (1: no significativa, 2: significativa, 3: de acceso).

En *repitencia* todos los individuos son clasificados de acuerdo a si están repitiendo el nivel (1) o si lo llevan por primera vez (0). Así como en *deserción* todos los individuos son clasificados de acuerdo a si desertaron (1) o no (0)⁴.

En la variable de *caso_ausentismo* todos los individuos se clasifican de acuerdo a si son (1) o no (0) un caso especial de ausentismo. Un estudiante se convierte en un caso de ausentismo cuando falta al 10% de sus lecciones en primaria, al 15% en secundaria o bien tiene 9 ausencias en preescolar. Por otra parte, en *trabaja_estudiante* todos los individuos se clasifican de acuerdo a si trabajan (1) o no (0). Mientras que, en la variable *prioridad_comedor* todos los individuos se clasifican de acuerdo a si son (1) casos prioritarios en el comedor o si no lo son (0).

La variable extranjero, si bien no tiene individuos sin la información, solo toma el valor de cero, que en una dicotómica implicaría que ninguno de los estudiantes es

⁴ Los encargados del PIAD afirmaron que la base de datos guarda información histórica del estudiante. De esta forma, podría identificarse si fueron casos de deserción o ausentismo en el pasado. Sin embargo, estas variables se refieren al año en curso o año de la base, en este caso 2013.

extranjero. Sin embargo, al cruzarlo con el país de procedencia se encuentra que hay observaciones cuyo país de nacimiento no es Costa Rica (ver cuadro 2).

Cuadro 2

Número de observaciones por lugar de procedencia Distrito 07 de Desamparados

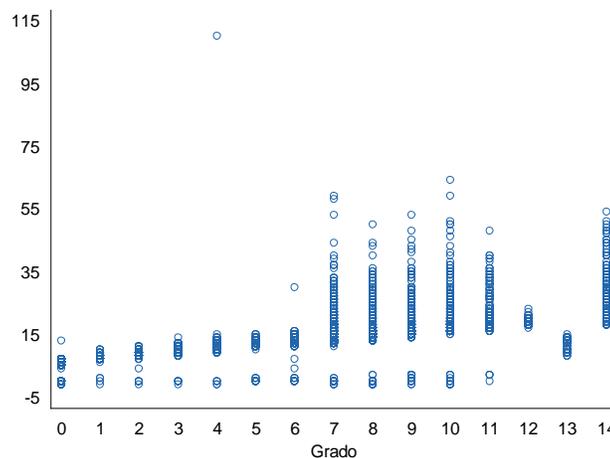
Asia	1
Canadá	2
Colombia	17
Costa Rica	12.968
Cuba	2
El Salvador	1
Estados Unidos	7
Europa	3
Guatemala	3
Honduras	3
México	1
Nicaragua	187
Panamá	1
Perú	2
Total	13198

Fuente: Elaboración propia con datos del PIAD

La variable de edad tiene un 1% de valores desconocidos, dado que se encuentran 29 individuos sin su correspondiente edad (valor de -1 en la base). Además, se encuentran algunos otros casos con un valor de edad atípico: 97 con edad 0, 15 con 1 año, 14 con 2 años y 1 persona con 110 años. Estos últimos casos deben cruzarse con otras variables que confirmen si se trata de un dato erróneo, en caso que se utilicen en futuros estudios; no obstante es un porcentaje pequeño de individuos. La distribución de la edad de las observaciones por nivel académico cursado se encuentra en el gráfico 2.

Gráfico 2

Edad de los individuos de acuerdo al grado en que se encuentran Distrito 07 de Desamparados



0= preescolar, 1 a 12= primero a duodécimo (se incluyen centros educativos nocturnos), 13= aula integrada y aula edad, 14= CINDEA I-II-III nivel.

Fuente: Elaboración propia con datos del PIAD

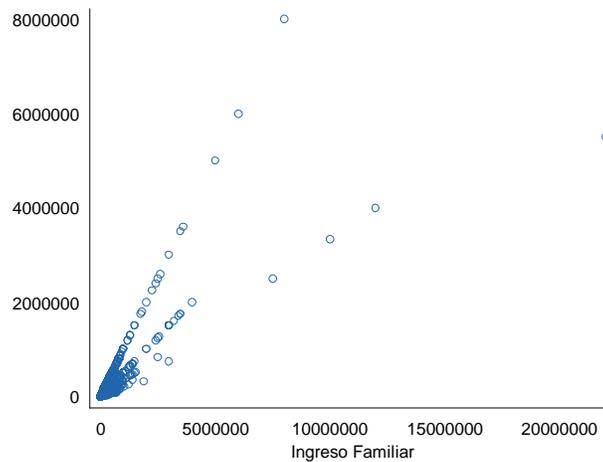
Para el 12% de los estudiantes no se sabe si tienen o no acceso a internet. Por su parte, el número de miembros del hogar no está disponible para el 14% de los individuos y en 45% de los casos se reporta un solo miembro del hogar. Fue indicado por los encargados del PIAD que en esta variable debía indicarse el total de miembros del hogar, incluyendo al estudiante. De aquellos casos que reportan un solo miembro, 5% asiste a preescolar y 54% a primaria, por lo que no parece razonable que a lo interno del hogar el único miembro sea el estudiante. Como veremos más adelante, esta situación también repercute en la variable que calcula el ingreso per cápita.

Además, como se indicó antes, las variables de ingreso del hogar son de las que se tiene menos información (79% de valores desconocidos). Pero también presentan algunos valores que podrían considerarse atípicos. En el gráfico 3 se muestra la distribución de estos valores. En particular algunos valores que se pueden considerar muy altos, en contraposición a otros muy bajos. Esto plantea la duda de si todos los centros están reportando el ingreso en la misma unidad de medida, aunque según los encargados del PAID, en esta variable debe expresarse el monto mensual en colones.

En la misma línea, el valor del ingreso per cápita parece estar generándose automáticamente por el sistema, a partir de la variable del número de miembros del hogar y el ingreso familiar. Aunque en la variable de miembros del hogar deben contabilizarse tanto el estudiante como el resto de miembros, existen casos en que parece estar reportando solamente el número de miembros diferentes al estudiante (pues aparecen individuos con un solo miembro en el hogar. En estos casos el ingreso familiar y el ingreso per cápita coinciden. En otros casos, donde el número de miembros del hogar es desconocido (-1) pero sí

se cuenta con el dato del ingreso familiar, la base está generando un ingreso per cápita negativo.

Gráfico 3
Distribución del ingreso familiar y el ingreso per cápita
Distrito 07 de Desamparados



Fuente: Elaboración propia con datos del PIAD

3.2. Asistencia del estudiante

Las variables de asistencia (periodo1-periodo3) se refieren al número de ausencias de cada estudiante en cada uno de los tres trimestres del año lectivo. Sin embargo, en período 1 aparecen 9 observaciones con números negativos que van desde -19 hasta -3. Los datos de los períodos 2 y 3 presentan los números discretos positivos esperados en este tipo de variable. En ninguna de las 3 variables hay valores desconocidos, por lo que el dato está para todos los estudiantes en los tres trimestres.

3.3. Beneficios sociales recibidos por el estudiante

Las variables de beneficios sociales tienen bajos porcentajes de valores desconocidos. En la variable de *beca* el 4% de las observaciones no tiene un valor para esta variable, mientras que en *bono escolar*, *transporte escolar* y *comedor* este porcentaje es de 12%. Si bien la mayoría de observaciones no tiene valor alguno para la variable *otro* (que se refiere a otro beneficio social), no se considera un problema pues en esta variable solo deben tener datos para aquellos individuos con alguna otra ayuda social.

Por su parte, la variable *beca* se refiere a si tienen (1-5)⁵ o si no tienen (0) beca para estudios. *Bono escolar* indica si el estudiante recibió (1) o no recibió (0) bono escolar. En la variable de *transporte*, que indica si el estudiante cuenta con ayuda de transporte por parte del Estado, se encuentran solamente 28 individuos que cuentan con esta ayuda (1). Es entonces una variable dicotómica, donde se le asigna el valor de 0 a los estudiantes que no tienen este beneficio y 1 a los que sí.

⁵ 1 Avancemos, 2 Fonabe, 3 Municipalidad, 4 Otros y 5 Comedor

Por su parte, la variable *comedor* indica si el estudiante asiste (1) o no (0) al comedor. Cruzando la variable de *comedor* con la variable de *prioridad_comedor* se encuentran solamente 2 casos en que afirman tener prioridad en el comedor pero que aparecen como que no reciben el servicio de comedor (ver cuadro 3).

Cuadro 3

Número de observaciones por asistencia y prioridad en el comedor Distrito 07 de Desamparados

	No prioridad	Prioridad	Sin datos	Total
No asiste	9.535	2	16	9.553
Asiste	223	1.810	5	2.038
Sin datos	223	235	1.149	1.607
Total	9.981	2.047	1.170	13.198

Fuente: Elaboración propia con datos del PIAD

En la variable de otros beneficios sociales *-otro-*, que se refiere a otras ayudas sociales que puede estar recibiendo el estudiante, las opciones más mencionadas en la base son ayudas del IMAS, pensiones y becas de Fonabe. En esta variable, 13,156 observaciones (99,68% del total) tienen el espacio en blanco, sin embargo, este espacio en blanco se interpreta como que estas observaciones no tienen otras ayudas, más que como valores desconocidos de la variable. En los casos que reportan becas, estas debieron imputarse en la variable *beca* y no en la categoría de otros beneficios.

Al cruzar la variable *becas* con la variable de otros beneficios sociales, se encuentra que de aquellos individuos que dijeron no tener beca anteriormente ("*beca*"=0), 1 afirma tenerla ("*otro*"=Beca), 3 tienen beca de Fonabe, 2 tienen beca del IMAS y se encuentran además 15 casos en que incorporan alguna categoría como "No, Ninguno, Ninguna" dentro de la variable "*otro*". Todos estos casos debieron haberse reportarse adecuadamente en la variable *becas*, sin llenar espacios como otros beneficios sociales. Sin embargo, son pocas observaciones comparadas con el total de individuos de la base de datos.

3.4. Registro de calificaciones del estudiante

Para el análisis de los datos correspondientes a las notas de los estudiantes en las materias básicas, se utilizó el dato del año cursado para estimar la cantidad de valores desconocidos (que no aplican) que se esperaban por cada materia. Así por ejemplo, los estudiantes de preescolar y primaria serían valores desconocidos esperados o que no aplican para las materias de física, química y biología, sumado a los estudiantes de secundaria de séptimo a noveno. Para cada materia se siguió el procedimiento descrito en la sección 2.2, con base en el número de estudiantes por cada nivel académico (ver en anexos el número de estudiantes a los que no aplica la variable de notas de acuerdo a su nivel cursado).

En general, se encuentra que alrededor del 50% de los estudiantes -que debieran tener sus notas en la base de datos- no tienen información sobre las notas de Español, Matemática y Estudios Sociales. Este porcentaje es alrededor de 40% en Ciencias y disminuye a 19% cuando se trata de Química y 16% en Física y Biología. El faltante de información es similar en el caso de los datos de promedio final en cada una de las materias, pero mejora cuando se trata de indicar si el estudiante aprobó o no la materia.

En la variable de aprobación de las materias de química, física y biología el porcentaje de valores desconocidos sobre el total de observaciones es negativo (ver anexo 2). Esto se debe a que en la base de datos le fue asignado un valor de cero (en lugar de un -1 de no aplica) a individuos cursando niveles donde estas materias no son impartidas (preescolar, primaria y secundaria de séptimo a noveno). Esto ocasiona que el número calculado de estudiantes para los que no aplica la variable sea mayor al número real de valores desconocidos (-1 o espacio en blanco) encontrados en la base, con lo que la diferencia entre ambos es negativa. Recordemos que el 0 en variables dicotómicas, como la de aprobación, no se considera un valor no disponible o que no aplica, si no el valor que indica respuesta negativa, en este caso indicaría que el estudiante no aprobó la materia. Sin embargo, se encuentra que a 95% de los estudiantes de preescolar se les asigna el valor de cero (no aprobado) en las materias, cuando debería indicarse como un valor que no aplica (-1, que sí se presenta en 5% de las observaciones)

Los estudiantes que se encuentran con un promedio de cero en las materias, son producto de valores de -1 o 0 en las notas trimestrales. De acuerdo a los encargados del PIAD, el promedio final es calculado por el sistema, con base en la regla: $(\text{nota primer trimestre} * 30\%) + (\text{nota segundo trimestre} * 30\%) + (\text{nota tercer trimestre} * 40\%)$. Esto trae a colación nuevamente la importancia de que en la base de datos se distinga entre los valores no disponibles, las variables que no aplican a un estudiante y el cero para determinar una condición, pues en una variable continua como el promedio, el cero debería ser parte de los valores posibles y no el indicativo de “no aplica”.

En los casos de las notas de convocatorias, se espera tener poca información o muchos valores desconocidos, dado que no habrá información para todos aquellos estudiantes que no reciben la materia –como en el caso de las notas de los trimestres- y de aquellos estudiantes que aprobaron el curso en el período regular de los tres trimestres. Al analizar los datos para las observaciones que cuentan con esta información, se encuentra que en ningún caso hay información para la tercera convocatoria. Además, las notas consignadas en la primera convocatoria son exactamente iguales a las consignadas en la segunda convocatoria para cada individuo. Al consultar a los encargados del PIAD sobre el motivo al qué se debe tal simetría, ya que sucede en todas las materias de la base, no recibimos respuesta (ver anexo 1)

3.5. Información de los padres o encargados del estudiante

Como se mencionó anteriormente, en los datos de los encargados de los estudiantes hay una gran disparidad entre la información disponible sobre la madre y la disponible sobre el padre. Mientras que en el 19% de las observaciones no se tienen datos sobre la madre, el 63% no presenta datos sobre el padre. En cada caso el porcentaje de valores desconocidos se mantiene a lo largo de las cuatro variables contempladas por cada encargado: estado civil, nacionalidad, escolaridad y si vive con el estudiante.

En el caso de los datos de la localización geográfica del encargado (aunque no se define quién es el encargado) se encuentra que no se tiene la información para el 12% de los estudiantes y que, además en el 1% de los casos el encargado no habita en la provincia de San José (71 en Cartago, 2 en Guanacaste y 6 en Heredia), aun cuando la base de datos se refiere a estudiantes de centros educativos del cantón de Desamparados. Además, no es claro qué sucede con los estudiantes mayores de edad, respecto a la definición de quién es el encargado del estudiante, él mismo u otra persona.

Además, se encuentra que un mismo distrito aparece varias veces dentro de la base, con ligeras diferencias debido a errores de ortografía o bien por asignarle un nombre ligeramente diferente, a pesar de que, según los encargados del PIAD⁶, el docente selecciona el distrito a partir de listas enlazadas (provincias con sus cantones, cantones con sus distritos) en el sistema.

4. Conclusiones sobre el estado de la base de datos

Una vez revisado el total de la base, variable por variable, fue posible identificar y clasificar las variables de acuerdo a su completitud –porcentaje de observaciones sin información- y el porcentaje de observaciones con datos pero con valores atípicos o dudosos. Todas las variables fueron clasificadas y en el cuadro 4 se resume su estado, de acuerdo al rango porcentual de cada limitación en que se encuentran.

⁶ Ver en anexo 1 el detalle de las consultas hechas a los encargados del PIAD y sus respuestas.

Cuadro 4

Clasificación de las variables según información no disponible y valores atípicos

Información no disponible

	0	1-15%	16-40%	41-70%	71%-más
ID	Tipo Hogar Uniparental	Años Experiencia_ESP	Tipo Vivienda	Ingreso Familiar	
Identificación	Edad	Grado Académico_ESP	Estado Casa		
Código presupuestario	Beca	Sexo_ESP	Nota Periodo1_ESP		
Año cursado	Acceso internet	Grupo Profesional_ESP	Nota Periodo2_ESP		
Sección	Bono Escolar	Edad_ESP	Nota Periodo3_ESP		
País	Transporte	Nota Periodo1_CIE	Aprobado_ESP		
Adecuación	Comedor	Nota Periodo2_CIE	Promedio_ESP		
Tipo adecuación	Años Experiencia_QUI	Nota Periodo3_CIE	Nota Periodo1_SOC		
Repitencia	Grado Académico_QUI	Aprobado_CIE	Nota Periodo2_SOC		
Deserción	Sexo_QUI	Promedio_CIE	Nota Periodo3_SOC		
Caso Ausentismo	Grupo Profesional_QUI	Años Experiencia_CIE	Aprobado_SOC		
Prioridad Comedor	Edad_QUI	Grado Académico_CIE	Promedio_SOC		
Trabaja Estudiante	Años Experiencia_FIS	Sexo_CIE	Nota Periodo1_MAT		
Período1-Período3	Grado Académico_FIS	Grupo Profesional_CIE	Nota Periodo2_MAT		
	Sexo_FIS	Edad_CIE	Nota Periodo3_MAT		
	Grupo Profesional_FIS	Años Experiencia_SOC	Aprobado_MAT		
	Edad_FIS	Grado Académico_SOC	Promedio_MAT		
	Años Experiencia_BIO	Sexo_SOC	Estado Civil Padre		
	Grado Académico_BIO	Grupo Profesional_SOC	Nacionalidad Padre		
	Sexo_BIO	Edad_SOC	Escolaridad Padre		
	Grupo Profesional_BIO	Años Experiencia_MAT	Vive con Estudiante Padre		
	Edad_BIO	Grado Académico_MAT			
	Provincia	Sexo_MAT			
	Cantón	Grupo Profesional_MAT			
	Distrito	Edad_MAT			
		Nota Periodo1_QUI			
		Nota Periodo2_QUI			
		Nota Periodo3_QUI			
		Promedio_QUI			
		Nota Periodo1_FIS			
		Nota Periodo2_FIS			
		Nota Periodo3_FIS			
		Promedio_FIS			
		Nota Periodo1_BIO			
		Nota Periodo2_BIO			
		Nota Periodo3_BIO			
		Promedio_BIO			
		Estado Civil Madre			
		Nacionalidad Madre			
		Escolaridad Madre			
		Vive con Estudiante Madre			
16-40%					
41-70%		Miembros Hogar			Ingreso Per Cápita
70%-más	Extranjero				

Información con valores atípicos o dudosos

15% o menos

Fuente: Elaboración propia con datos del PIAD

Destacan las variables del estado de la casa (59%), de ingresos económicos (79%), de datos del padre (63%) y de las notas de los estudiantes en materias básicas (Español, Matemática y Estudios Sociales, 58%), como las variables con importantes limitaciones en términos de la falta de información. Además, dentro de los estudiantes que sí poseen los datos de ingresos, se presenta el problema sobre la calidad de esta información. Específicamente, alrededor de 15% de los valores consignados en el ingreso familiar, se consideran atípicos. Así mismo, el ingreso per cápita presenta alto porcentaje de valores atípicos, por estar generándose a partir de los valores del ingreso familiar y del número de miembros del hogar, ambas variables con limitaciones de calidad de su información.

La base de datos trae a la luz un problema de gestión del PIAD, pues los altos porcentajes de información no disponible denotan que no todos los docentes están haciendo uso del sistema, así como los altos porcentajes de información atípica o dudosa denotan fallos en el ingreso de los datos. Se considera que algunas variables con datos cuestionables plantean la disyuntiva de si todos los docentes están ingresando la información en el mismo entendido, con la misma comprensión de qué deben ingresar como datos. Esto plantea también la necesidad de una revisión de valores atípicos a nivel de centro educativo, que permita corregir estos datos antes de pasar a ser estadísticas más generales.

La lista de variables contenidas en el PIAD es importante, no solo para el control docente, sino para una mejor comprensión del rendimiento académico de los estudiantes y su entorno más cercano. Sin embargo, la riqueza de esta información está en su disponibilidad y confiabilidad. No se puede hablar de controlar y comprender el rendimiento académico de cada estudiante cuando no se tiene información sobre sus notas, siendo esta una de las variables con menos información disponible dentro de la base.

Mediante un ejercicio se identificaron los centros y las secciones dentro de esos centros, que poseen los porcentajes más altos de información disponible, para las variables con menor completitud. Se encuentra que los centros con más información son centros de preescolar (JN María Jiménez y JN La Valencia). Ante la necesidad de tener información para primaria y secundaria, se buscó el centro con más información en cada uno de estos niveles, los cuales corresponden a la Escuela San Rafael (cód.480) y el UP La Valencia (cód.5735). No obstante, la información sobre ingresos en primaria y secundaria sigue siendo limitada.

A nivel de base de datos y sobre la claridad de la información contenida, el principal problema que se encuentra -además de la falta de información- es que no hay claridad ni estandarización para distinguir cuando una variable no aplica a un estudiante o cuando aplica pero no se tiene la información disponible. Esto torna complejo el uso de los datos de la base, desde una perspectiva sistemática o agrupada del total de centros educativos y estudiantes.

Por lo tanto, con el diseño y las variables que contempla el sistema del PIAD, los análisis que podrían generarse serían importantes. Sin embargo, con el estado y la calidad de la información contenida, al menos en esta muestra, el análisis se torna limitado.

Finalmente, es importante que el PIAD no sea sólo una herramienta de uso para el docente, sino también para las autoridades en términos de políticas educativas. Esto último subraya la necesidad de conservar datos históricos, a la vez que sus actualizaciones; es decir, se plantea la importancia de contar con datos de panel.

5. Aplicaciones posibles con el estado actual de la base

Una de las principales limitaciones de la base de datos es que los estudios que se generen a partir de ella no pueden realizarse para el distrito completo, al no estar preparada para ser sujeto de algún tipo de análisis sistemático o que abarque conclusiones para todo el Distrito 07 de Desamparados. Teniendo que discernir entre cuáles centros educativos utilizar la información, con base en cuáles de ellos se presenta la información más completa y menos sesgada, y con ello hacer estudios exploratorios con base en muestras pequeñas no sesgadas.

Por ejemplo, en la muestra de individuos con información, se encuentra una relación directa entre el ingreso familiar y la educación del padre. Sin embargo, la muestra que reporta el ingreso familiar está sesgada hacia individuos con niveles bajos de educación del padre y con ello, se presume, también está sesgada hacia hogares con bajos niveles de ingreso. Esto denota una muestra no aleatoria y por ende, no representativa de la población (distrito 07 de Desamparados). Resultados similares se encuentran cuando se analiza la distribución de las variables del estado de la vivienda.

La base de datos actual presenta datos de corte transversal, por lo que un análisis de la evolución de los resultados de los mismos estudiantes en un período de tiempo mayor a un año no es posible, ante la ausencia de datos de panel. No se tienen tampoco series de tiempo.⁷

Teniendo en cuenta estas limitaciones de la base de datos, los análisis que se pueden tener para un año se conformarían de herramientas de estadística descriptiva y modelos econométricos⁸ o no paramétricos que establecen relaciones de correlación entre variables, para aquellas con la información para toda la población o en las cuales la muestra de respuestas no está sesgada. Específicamente se pueden tener datos de aprobación, ausentismo, deserción y beneficios sociales; correlacionando los último tres con la aprobación o no del estudiante por materia.

⁷ Según los encargados del PIAD, en la base de datos se guarda toda la información histórica del estudiante. Sin embargo, dicha información no nos fue entregada para evaluar su existencia y calidad.

⁸ Ver en anexo 3 un ejemplo de un modelo econométrico que determina la probabilidad de deserción de los estudiantes.

La estadística descriptiva permitiría identificar individuos, grupos y centros educativos que estén presentando problemas en alguna de las áreas e intervenirlas antes de que acabe el año lectivo para evitar su reprobación o deserción. Sin embargo, su uso se circunscribirá a una herramienta docente para el control de algunas pocas variables como el número de ausencias, cuando no se cuente con el récord de notas del estudiante.

Por ejemplo, los datos de deserción (ver cuadro 5) muestran que la mayor deserción se da a nivel de secundaria, lo cual coincide con la tendencia de las estadísticas nacionales. Trae a colación también, que en preescolar se está teniendo deserción dentro del distrito 07 de Desamparados. Por otra parte, filtrando los datos de deserción por centro educativo (ver cuadro 6) se encuentra que el Colegio Nocturno de Desamparados es el de mayor deserción (17%) en el distrito, por encima de la media total (4%) y de la media en secundaria (7%) del distrito. Le sigue el Liceo de Calle Fallas (13%) y el Liceo Monseñor Rubén Odio (7%), mostrando nuevamente el problema de deserción en secundaria, tanto diurna como nocturna. Por su parte, llama la atención la deserción a nivel de preescolar y esta se concentra en el JN María Jiménez (7% de deserción). En la mayoría de escuelas la deserción es baja y prácticamente cero, sin embargo, destaca el 5% de deserción en la escuela nocturna UP La Valencia y la Escuela San Jerónimo -escuela diurna- con 2% de deserción.

Cuadro 5
Deserción por nivel académico
Distrito 07 de Desamparados

	Preescolar	Primaria	Secundaria	Especiales	Total
Total					
Estudiantes	1.320	4.973	6.844	61	13.198
Desertores	34	13	462	0	509
% Deserción	3%	0%	7%	0%	4%

Fuente: Elaboración propia con datos del PIAD

Cuadro 6
Deserción por centro educativo
Distrito 07 de Desamparados

Centro educativo	Total estudiantes	Total Desertores	Porcentaje
CTP Monseñor Sanabria	1.135	1	0,1%
Col. Nocturno Desamparados	937	157	16,8%
Esc. Elías Jiménez	1.329	1	0,1%
Esc. García Monge	1.176	0	0,0%
Esc. José T. Mora	1.049	2	0,2%
Esc. San Jerónimo	698	11	1,6%

Esc. San Rafael	1.040	3	0,3%
JN La Valencia	223	1	0,4%
JN María Jiménez	387	26	6,7%
JN María Retana	239	0	0,0%
Liceo Calle Fallas	1.078	140	13,0%
Liceo Monseñor Rubén Odio	1.977	136	6,9%
Liceo Roberto Gamboa	1.129	1	0,1%
Secc. Nocturna TV Monseñor Sanabria	279	3	1,1%
UP La Valencia	522	27	5,2%

Fuente: Elaboración propia con datos del PIAD

Por lo tanto, la primera utilidad de la base de datos es de índole descriptivo como un sistema de alertas, que permita al personal del Ministerio de Educación y padres de familia actuar en la inmediatez del período lectivo, para evitar reprobaciones y deserciones entre los estudiantes. Identificando los centros y grupos con mayores problemas, y las relaciones de correlación con las características particulares de sus estudiantes.

Por otra parte, se considera que el éxito escolar depende de características de los estudiantes, características de sus hogares y características de los centros educativos, que incluyen las características de sus docentes (Krueger, 1997; Hanushek, 2005; Salas, 2008). Para cada uno de estos grupos de características el PIAD ofrecería información; sin embargo, la base de datos muestra que en la práctica no se tiene toda la información que el diseño del PIAD plantea.

En un análisis de la función de producción educativa, se definen los inputs o insumos para lograr un resultado u output. De acuerdo a Salas (2008) los outputs o resultados del proceso productivo educativo pueden ser de dos tipos: conocimientos y elementos no cognitivos. Los conocimientos corresponden a la adquisición de destrezas matemáticas o verbales y pueden ser medidas a través de los resultados de exámenes o pruebas. Los elementos no cognitivos por su parte incluyen actitudes y valores, y son más difíciles de medir cuantitativamente.

Sobre los inputs o insumos, Salas (2008) considera que pueden diferenciarse entre aquellos insumos escolares y los no escolares. Los insumos escolares incluyen los recursos físicos (características de los edificios y demás espacios escolares, equipamiento del colegio -cantidad y calidad-), los recursos humanos (maestros -educación y experiencia-, orientadores o psicólogos escolares, personal de apoyo disponible) y el número de alumnos por aula (tamaño de la clase). Mientras que, los inputs no escolares contemplan la inteligencia, esfuerzo personal, estatus socioeconómico (reflejado por ejemplo en la educación de los padres), número de libros en casa, disponibilidad del ordenador en casa, influencia del grupo de iguales o pares (*peer group effect*), y adicionalmente, los conocimientos que acumularon los alumnos en el pasado.

Nuevamente, estos inputs y outputs pudiesen haber sido medidos, en su mayoría, a través de las variables contenidas en el sistema del PIAD; sin embargo, no se cuenta con toda esta información en la base de datos entregada, especialmente ante la falta de información en las variables de ingreso o sus aproximaciones al nivel socioeconómico del hogar (estado de la vivienda, tipo de la vivienda, número de miembros del hogar, acceso a internet). Además, el output de rendimiento académico podría haberse medido a través de la variable "aprobado" de cada materia, las variables de notas de cada período o bien la variable de promedio anual por materia.

Por otro lado, sí se cuenta con características de sus padres (especialmente en el caso de la madre: estado civil, escolaridad y si vive con el hijo o no), características del docente en cuanto a sexo, edad, grado académico y años de experiencia, características del estudiante como el sexo, la edad, nacionalidad, condición de adecuación de algún tipo, si es un caso de repitencia o deserción, si recibe alguna ayuda social por parte del gobierno y el número de ausencias⁹. Adicionalmente, para obtener algunas otras características del centro educativo, especialmente relacionadas con la calidad de la infraestructura se puede obtener la información de la Megabase de datos georreferenciados para la educación primaria y secundaria de los centros educativos de Costa Rica (2000-2011) (Estado de la Nación). Algunas otras variables como el número de estudiantes por grupo o resultados de aprobación de los pares (para los grupos en que se tiene la información), podrían construirse a partir de la información disponible en el PIAD.

Por lo tanto, aunque el análisis de los factores que influyen o están correlacionados con el rendimiento académico podría ser factible con el diseño de la información del PIAD, no lo es con el nivel de información contenida actualmente, pues al menos la mitad de sus observaciones no cuenta con el dato de las notas de los estudiantes ni de las características de sus hogares, y la muestra que posee esta información parece no ser aleatoria e insesgada.

V. Referencias

- Castro, C. (2013). *Desempeño de la educación general básica y el ciclo diversificado en Costa Rica. Cuarto informe del Estado de la Educación*. Obtenido de Programa Estado de la Nación:
http://www.estadonacion.or.cr/files/biblioteca_virtual/educacion/004/castro_desempeno-ed-basica-y-diversificado.pdf
- De Witte, K., & Csillag, M. (2014). Does anybody notice? On the impact of improved truancy reporting on school dropout. *Education Economics*, 22(6), 549-568.

⁹ El ausentismo está altamente correlacionado con el bajo rendimiento académico (Lavecchia, Liu, & Oreopoulos, 2014)

Estado de la Nación. (s.f.). *Megabase de datos georreferenciados para la educación primaria y secundaria de los centros educativos de Costa Rica (2000-2011)*. Obtenido de Programa Estado de la Nación:
<http://www.estadonacion.or.cr/estadisticas/estadisticas-bases-de-datos/bases-de-datos-costa-rica/185-estadisticas/estadisticas-de-costa-rica/121-bases-de-datos-disponibles-en-linea>

Hanushek, E. (2005). The economics of school quality. *German Economic Review*, 269-286.

Krueger, A. (1997). *Experimental estimates of education production functions*. Massachusetts: National Bureau of Economic Research.

Lavecchia, A., Liu, H., & Oreopoulos, P. (2014). *Behavioral Economics of Education: Progress and Possibilities*. Obtenido de National Bureau of Economic Research: <http://www.nber.org/papers/w20609>

PIAD. (s.f.). *ASIS: Impacto e Innovación Social*. Obtenido de Programa de Informatización para el Alto Desempeño: <http://www.piad.or.cr>

Salas, M. (2008). *Economía de la Educación*. Madrid, España: Pearson.

Anexos

Anexo 1

Bitácora 1 proceso para obtener las bases de datos

Fecha	Actividad
Mayo, 2014	Firma de acuerdo con el Ministerio de Educación para tener acceso a las bases de datos del PIAD en el Distrito 07
14 de agosto de 2014	Recepción de 16 archivos .bak en el software MS SQL 2008 R2 con la información de los centros educativos del Distrito 07 de Desamparados.
18 de agosto 2014.	Se solicita apoyo al Centro Informática del CONARE para leer y extraer la información de los archivos (. bak).no obstante, surgen dificultades técnicas para generar un solo archivo con la información de todos los centros educativos.
20 de agosto de 2014	Se realiza una nueva solicitud al MEP para revisar los formatos de la información y ordenar los datos en una sola base según formato sugerido por el PEN al PIAD en febrero de 2014.
29 de agosto, 2014	Se recibe una segunda base por parte del PIAD con la estructura solicitada por el PEN, en formato SQL Server
1 setiembre, 2014	Inicia el proceso de revisión de contenido de la base datos, con el objetivo de verificar cuan completa y depurada se encuentra la misma

Anexo 2

Descripción de las variables contenidas en la base de datos, de acuerdo a los valores que toma y la falta de información. Distrito 07 de Desamparados

Variable	Valores	Descripción	Missing ()	Missing (-1)	Missing (0)	n.a. estimados	Total Missing	Total Observaciones	Porcentaje Missing
ID	0-14507		0				0	13198	0%
IDENTIFICACION	0-13197	llave de empate	0				0	13198	0%
Centro_estudiante									
TIPO_VIVIENDA	[0,3]	categoría	0	7774		0	7774	13198	59%
ESTADO_CASA	[0,3]	categoría	0	7774		0	7774	13198	59%
MIEMBROS_HOGAR	[1,9]	discreta	0	1791			1791	13198	14%
INGRESO_FAMILIAR	continuo, positivo		0	7774	2679		10453	13198	79%
INGRESO_PER_CAPITA	continuo, negativo y positivo		0	7774	2679		10453	13198	79%
ACCESO_INTERNET	[0,1]	dicotómica	0	1609			1609	13198	12%
TIPO_HOGAR_UNIPARENTAL	[0,1]	dicotómica	0	472			472	13198	4%
CODIGO_PRESUPUESTARIO	varios	código/nombre del centro educa	0	0	0	0	0	13198	0%
ANNO_CURSADO	niveles	categoría	0	0	0	0	0	13198	0%
SECCION	grado y número de sec	categoría	0	0	0	0	0	13198	0%
EDAD	discreta	discreta	0	29	97		126	13198	1%
EXTRANJERO	0		0	0	0	0	0	13198	0%
PAIS	14 países	categoría	0	0	0	0	0	13198	0%
ADECUACION	[0,1]	dicotómica	0	0			0	13198	0%
TIPO_ADECUACION	[0,3]	categoría	0	0			0	13198	0%
REPITENCIA	[0,1]	dicotómica	0	0			0	13198	0%
DESERCIÓN	[0,1]	dicotómica	0	0			0	13198	0%
CASO_AUSENTISMO	[0,1]	dicotómica	0	0			0	13198	0%
PRIORIDAD_COMEDOR	[0,2]	categoría	0	0			0	13198	0%
TRABAJA_ESTUDIANTE	[0,2]	categoría	0	0			0	13198	0%
Asistencia_estudiante									
PERIODO1	min -19, max 291	discreta, negativa y positiva	0	0			0	13198	0%
PERIODO2	[0,235]	discreta positiva	0	0			0	13198	0%
PERIODO3	[0,216]	discreta positiva	0	0			0	13198	0%
Beneficios_Sociales									
BECA	[0,5]	categoría	0	472			472	13198	4%
BONO_ESCOLAR	[0,1]	dicotómica	0	1607			1607	13198	12%
TRANSPORTE	[0,1]	dicotómica	0	1607			1607	13198	12%
COMEDOR	[0,1]	dicotómica	0	1607			1607	13198	12%
OTRO	pensiones, IMAS, FONABE, etc	categoría							
Calificaciones_estudiante									
NOTA_PERIODO1_ESP	min 0, máx 100	discreta desde 0	6,882	1949	30	1366	7495	13198	57%
NOTA_PERIODO2_ESP	min 0, máx 100	discreta desde 0	6,882	2017	93	1366	7626	13198	58%
NOTA_PERIODO3_ESP	min 0, máx 100	discreta desde 0	6,882	1971	180	1366	7667	13198	58%
NOTA_CONVOCATORIA1_ESP	[11,65]	discreta	6,882	6264	0				
NOTA_CONVOCATORIA2_ESP	[11,65]	discreta	6,882	6264	0				
NOTA_CONVOCATORIA3_ESP	no hay observaciones	no hay observaciones	6,882	6316	0				
APROBADO_ESP	[0,1]	dicotómica	6,882	472		1366	5988	13198	45%
PROMEDIO_ESP	min 0, máx 100	discreta desde 0	6,882	472	1465	1366	7453	13198	56%
ANIOS_EXPERIENCIA_ESP	min 0, máx 42	discreta desde 0	0	4967	321	1366	3922	13198	30%
GRADO_ACADEMICO_ESP	[0,3]	categoría	0	4967		1366	3601	13198	27%
SEXO_ESP	[0,1]	dicotómica	0	4967		1366	3601	13198	27%
GRUPO_PROFESIONAL_ESP	ASP, MT, VT, PT	categoría	472	4495		1366	3601	13198	27%
EDAD_ESP	[-46,65]	discreta	0	4967	0	1366	3601	13198	27%
NOTA_PERIODO1_CIE	[0,100]	discreta desde 0	6,882	1949	25	3950	4906	13198	37%
NOTA_PERIODO2_CIE	[0,100]	discreta desde 0	6,882	2014	95	3950	5041	13198	38%
NOTA_PERIODO3_CIE	[0,100]	discreta desde 0	6,882	1995	175	3950	5102	13198	39%
NOTA_CONVOCATORIA1_CIE	[27,65]	discreta	6,882	6281	0				
NOTA_CONVOCATORIA2_CIE	[27,65]	discreta	6,882	6281	0				
NOTA_CONVOCATORIA3_CIE	no hay observaciones	discreta desde 0	6,882	6316					
APROBADO_CIE	[0,1]	dicotómica	6,882	472		3950	3404	13198	26%
PROMEDIO_CIE	[0,100]	discreta desde 0	6,882	472	1462	3950	4866	13198	37%
ANIOS_EXPERIENCIA_CIE	[0,35]	discreta desde 0	0	7411	705	3950	4166	13198	32%
GRADO_ACADEMICO_CIE	[0,3]	categoría	0	7411		3950	3461	13198	26%
SEXO_CIE	[0,1]	dicotómica	0	7411		3950	3461	13198	26%
GRUPO_PROFESIONAL_CIE	ASP, MT, PT	categoría	6939	472		3950	3461	13198	26%
EDAD_CIE	[-46,60]	discreta	0	7411		3950	3461	13198	26%
NOTA_PERIODO1_SOC	[0,100]	discreta desde 0	6,882	1948	30	1366	7494	13198	57%
NOTA_PERIODO2_SOC	[0,100]	discreta desde 0	6,882	2014	96	1366	7626	13198	58%
NOTA_PERIODO3_SOC	[0,100]	discreta desde 0	6,882	1995	178	1366	7689	13198	58%
NOTA_CONVOCATORIA1_SOC	[25,65]	discreta desde 0	6,882	6245					
NOTA_CONVOCATORIA2_SOC	[25,65]	discreta desde 0	6,882	6245					
NOTA_CONVOCATORIA3_SOC	no hay observaciones	discreta desde 0	6,882	6316					
APROBADO_SOC	[0,1]	dicotómica	6,882	472		1366	5988	13198	45%
PROMEDIO_SOC	[0,100]	discreta desde 0	6,882	472	1466	1366	7454	13198	56%
ANIOS_EXPERIENCIA_SOC	[0,34]	discreta desde 0	0	4967	1185	1366	4786	13198	36%
GRADO_ACADEMICO_SOC	[0,6]	categoría	0	4967		1366	3601	13198	27%
SEXO_SOC	[0,1]	dicotómica	0	4967		1366	3601	13198	27%
GRUPO_PROFESIONAL_SOC	ASP, MT, PT	categoría	4495	472		1366	3601	13198	27%
EDAD_SOC	[-46,64]	discreta	0	4967	0	1366	3601	13198	27%

Continúa...

...continuación

Variable	Valores	Descripción	Missing ()	Missing (-1)	Missing (0)	n.a. estimados	Total Missing	Total Observaciones	Porcentaje Missing
NOTA_PERIODO1_MAT	[0,101]	discreta desde 0	6,882	1950	28	1366	7494	13198	57%
NOTA_PERIODO2_MAT	[0,100]	discreta desde 0	6,882	2017	94	1366	7627	13198	58%
NOTA_PERIODO3_MAT	[0,100]	discreta desde 0	6,882	1994	177	1366	7687	13198	58%
NOTA_CONVOCATORIA1_MAT	[23,65]	discreta	6,882	6226					
NOTA_CONVOCATORIA2_MAT	[23,65]	discreta	6,882	6226					
NOTA_CONVOCATORIA3_MAT	no hay observaciones		6,882	6316					
APROBADO_MAT	[0,1]	dicotómica	6,882	472		1366	5988	13198	45%
PROMEDIO_MAT	[0,100]	discreta	6,882	472	1465	1366	7453	13198	56%
ANIOS_EXPERIENCIA_MAT	[0,37]	discreta	0	4967	343	1366	3944	13198	30%
GRADO_ACADEMICO_MAT	[0,3]	categoría	0	4967	0	1366	3601	13198	27%
SEXO_MAT	[0,1]	dicotómica	0	4967	0	1366	3601	13198	27%
GRUPO_PROFESIONAL_MAT	ASP, MT, PT	categoría	4495	472		1366	3601	13198	27%
EDAD_MAT	[-46,59]	discreta	0	4967	75	1366	3676	13198	28%
NOTA_PERIODO1 QUI	[21,83]	discreta	6,102	7053	0	10599	2556	13198	19%
NOTA_PERIODO2 QUI	[25,84]	discreta	6,102	7054	0	10599	2557	13198	19%
NOTA_PERIODO3 QUI	[2,92]	discreta	6,102	7058	0	10599	2561	13198	19%
NOTA_CONVOCATORIA1 QUI	[25,70]	discreta	6,102	7084	0				
NOTA_CONVOCATORIA2 QUI	[25,70]	discreta	6,102	7084	0				
NOTA_CONVOCATORIA3 QUI	no hay observaciones	discreta	6,102	7096	0				
APROBADO QUI	[0,1]	dicotómica	6,102	472	0	10599	-4025	13198	-30%
PROMEDIO QUI	[0,84]	discreta	6,102	472	6581	10599	2556	13198	19%
ANIOS_EXPERIENCIA QUI	[0,21]	discreta	0	11438	261	10599	1100	13198	8%
GRADO_ACADEMICO QUI	[0,3]	categoría	0	11438		10599	839	13198	6%
SEXO QUI	[0,1]	dicotómica	0	11438		10599	839	13198	6%
GRUPO_PROFESIONAL QUI	ASP, MT	categoría	10966	472		10599	839	13198	6%
EDAD QUI	[29,61]	discreta	0	11438	0	10599	839	13198	6%
NOTA_PERIODO1 FIS	[0,100]	discreta	4,666	8094	1	10599	2162	13198	16%
NOTA_PERIODO2 FIS	[28,100]	discreta	4,666	8101	0	10599	2168	13198	16%
NOTA_PERIODO3 FIS	[20,100]	discreta	4,666	8107	0	10599	2174	13198	16%
NOTA_CONVOCATORIA1 FIS	[0,70]	discreta	4,666	8501	2				
NOTA_CONVOCATORIA2 FIS	[0,70]	discreta	4,666	8501	2				
NOTA_CONVOCATORIA3 FIS	no hay observaciones	discreta	4,666	8532	0				
APROBADO FIS	[0,1]	dicotómica	4,666	472	0	10599	-5461	13198	-41%
PROMEDIO FIS	[0,99]	discreta	4,666	472	7624	10599	2163	13198	16%
ANIOS_EXPERIENCIA FIS	[0,26]	discreta	0	11603	629	10599	1633	13198	12%
GRADO_ACADEMICO FIS	[0,2]	categoría	0	11603	0	10599	1004	13198	8%
SEXO FIS	[0,1]	dicotómica	0	11603	0	10599	1004	13198	8%
GRUPO_PROFESIONAL FIS	ASP, MT	categoría	11131	472		10599	1004	13198	8%
EDAD FIS	[29,61]	discreta	0	11603	0	10599	1004	13198	8%
NOTA_PERIODO1 BIO	[0,100]	discreta	4,666	8094	1	10599	2162	13198	16%
NOTA_PERIODO2 BIO	[28,100]	discreta	4,666	8101	0	10599	2168	13198	16%
NOTA_PERIODO3 BIO	[20,100]	discreta	4,666	8107	0	10599	2174	13198	16%
NOTA_CONVOCATORIA1 BIO	[0,70]	discreta	4,666	8501	2				
NOTA_CONVOCATORIA2 BIO	[0,70]	discreta	4,666	8501	2				
NOTA_CONVOCATORIA3 BIO	no hay observaciones	discreta	4,666	8532	0				
APROBADO BIO	[0,1]	dicotómica	4,666	472	0	10599	-5461	13198	-41%
PROMEDIO BIO	[0,99]	discreta	4,666	472	7624	10599	2163	13198	16%
ANIOS_EXPERIENCIA BIO	[0,26]	discreta	0	11603	629	10599	1633	13198	12%
GRADO_ACADEMICO BIO	[0,2]	categoría	0	11603	0	10599	1004	13198	8%
SEXO BIO	[0,1]	dicotómica	0	11603	0	10599	1004	13198	8%
GRUPO_PROFESIONAL BIO	ASP, MT	categoría	11131	472		10599	1004	13198	8%
EDAD BIO	[29,61]	discreta	0	11603	0	10599	1004	13198	8%
Datos_encargado									
PROVINCIA	Cartago, Guanacaste, Heredia, San José	categoría	1544	0		0	1544	13198	12%
CANTON		categoría	1544	0		0	1544	13198	12%
DISTRITO		categoría	1544	0		0	1544	13198	12%
ESTADO_CIVIL_MADRE	[0,5]	categoría	0	2527			2527	13198	19%
NACIONALIDAD_MADRE	Asia, Países	categoría	2055	472			2527	13198	19%
ESCOLARIDAD_MADRE	[0,6]	categoría	0	2540			2540	13198	19%
VIVE_CON_ESTUDIANTE_MADRE	[0,1]	dicotómica	0	2527			2527	13198	19%
ESTADO_CIVIL_PADRE	[0,5]	categoría	0	8280			8280	13198	63%
NACIONALIDAD_PADRE	Asia, Países	categoría	7808	472			8280	13198	63%
ESCOLARIDAD_PADRE	[0,6]	categoría	0	8293			8293	13198	63%
VIVE_CON_ESTUDIANTE	[0,1]	dicotómica	0	8280			8280	13198	63%

Notas:

1/ El cero es un missing sólo para aquellas variables continuas o discretas, no para dicotómicas o categóricas.

2/ No es posible estimar el total de missing en las notas de convocatorias, porque no se cuenta con información de cuántos estudiantes pasaron la materia en el período regular.

3/ El porcentaje de missing en física, química y biología es negativo por cuanto los n.a. estimados son mayores a los reales encontrados. Es decir, en la base se les asignó un valor a estudiantes que no debían tener información de estas materias.

Fuente: Elaboración propia con datos del PIAD

Anexo 3

Cuadro A.3

Número de valores que no aplican esperados (n.a. estimados) por materia y nivel académico

Distrito 07 de Desamparados

Nivel	Año cursado	Español	Ciencias	Estudios Sociales	Matemática	Química	Física	Biología
Preescolar	Materno	498	498	498	498	498	498	498
	Transición	822	822	822	822	822	822	822
Primaria	Primero					801	801	801
	Segundo					817	817	817
	Tercero					745	745	745
	Cuarto					836	836	836
	Quinto					791	791	791
	Sexto					770	770	770
Secundaria	Sétimo					1.816	1.816	1.816
	Octavo					1.275	1.275	1.275
	Noveno					1.169	1.169	1.169
	Décimo		1.217					
	Undécimo		922					
CINDEA	Duodécimo		379					
	I NIVEL					143	143	143
	II NIVEL					70	70	70
Especiales	III NIVEL		66					
	Aula Edad					15	15	15
	Aula Integrada	46	46	46	46	46	46	46
Total		1.366	3.950	1.366	1.366	10.599	10.599	10.599

Fuente: Elaboración propia con datos del PIAD

Anexo 4

Modelos como los propuestos por De Witte y Csillag (2014) determinan la probabilidad de deserción con base en el número de ausencias o caso de ausentismo (dummy), considerando el ausentismo de los pares y otras variables de control. A través de regresiones estimadas mediante mínimos cuadrados ordinarios (MCO) y con efectos fijos por estudiante y centro educativo, encuentran una relación positiva y significativa entre el ausentismo y la deserción. La función estimada por los autores es la siguiente

$$deserción_i = \alpha + \beta \text{ausentismo}_{ij} + \gamma \text{edad}_i + \delta T_{-i} + v_i + n_i + \varepsilon_i$$

Donde i representa a un estudiante en un centro educativo dado j ; α es una constante; β , γ y δ son los coeficientes estimados; T_{-i} es el ausentismo de los pares excepto el propio; v y n son los efectos fijos y ε es el error estándar. No es posible estimar los efectos fijos por estudiante y centro educativo con la base actual del PIAD, ya que sólo se cuenta con un periodo (un año). Alternativamente se puede considerar determinar la probabilidad de deserción, incluyendo factores económicos del hogar (de la base de datos del PIAD), otras características del estudiante (PIAD) y características del centro educativo (de la megabase de datos georreferenciados del Estado de la Nación). Otra opción, para poder estimar los efectos fijos es considerar los trimestres para tener un panel de 3 periodos, para esta opción se necesitaría sin embargo, tener el dato de deserción por trimestre y no solamente anual.

Una vez que el PIAD logre buenos niveles de información y supere las limitaciones actuales, se puede considerar en el futuro generar otras líneas de análisis. Por ejemplo, análisis del rendimiento o la deserción con base en el cumplimiento de las tareas, la presencia de trabajadores sociales o psicólogos en el centro educativo, o bien evaluaciones del impacto de la implementación del PIAD sobre los niveles de deserción de los centros educativos.

La deserción en secundaria es antecedida por la falta o nula entrega de las tareas para el hogar por parte de los estudiantes. Además, aquellos estudiantes que exponen mayores muestras de auto control tienen mejores hábitos de estudio, son más propensos a hacer sus tareas regularmente, obtener mejores notas, mantenerse en el sistema educativo y culminarlo. (Lavecchia, Liu, & Oreopoulos, 2014). Según Castro (2013) el PIAD es una herramienta informática con un componente de registro digital, que sirve de herramienta para uso docente, utilizada principalmente para registrar los resultados obtenidos en la evaluación, en los trabajos extra-clase, observaciones del trabajo cotidiano, pruebas, conducta y ausencias de los estudiantes, que una vez introducidos se traducen en puntos, notas y porcentajes. Por lo tanto, es posible aspirar a tener dentro de la base de datos del PIAD algún reporte de cumplimiento en las tareas, para incluirlo dentro de los análisis de factores determinantes de la deserción.

Por otra parte, la presencia de trabajadores sociales o psicólogos en los centros educativos aumenta la probabilidad de éxito de los estudiantes y evita la deserción cuando se les conduce a través de políticas de acompañamiento -reduciendo así la procrastinación¹⁰-, además de acompañarlos dándoles mayor información acerca de sus planes de estudio y las opciones que poseen, reduciendo también los efectos de posibles presiones sociales o de sus pares (Lavecchia, Liu, & Oreopoulos, 2014). Esta información puede ser obtenida de fuentes de datos alternativas, del MEP o el Estado de la Nación, para valorar la correlación que tiene la presencia de estos trabajadores sociales sobre la deserción de los estudiantes.

Evaluaciones de impacto se pueden llevar a cabo, por ejemplo siguiendo a De Witte y Csillag (2014) se puede considerar la implementación del PIAD como la política de mejora en reportes del ausentismo y evaluar si tiene algún impacto sobre la deserción. Para ello deben tenerse datos de deserción por estudiante y centro educativo por parte del MEP antes de la implementación del PIAD y posterior a ello, en los centros que aún no se implemente el nuevo sistema. Se tomaría como grupo tratado aquellos centros donde se va a implementar el PIAD y como grupo de control aquellos que siguen sin PIAD.

Por último, es importante que el PIAD conserve datos históricos de los estudiantes y actualice los datos socioeconómicos y de los padres que sólo se recogen en la primera matrícula de cada estudiante en un centro educativo determinado. Al no haber un panel de los estudiantes, esto impide muchas evaluaciones de impacto y el poder verificar el avance de los estudiantes. Por ejemplo, podría evaluarse el impacto de políticas particulares de ciertos centros educativos para encontrar casos exitosos y reproducirlos posteriormente o podría evaluarse el impacto de los que estuvieron en preescolar versus aquellos que no.

¹⁰ Procrastination en inglés, hace referencia a aplazar alguna actividad o tarea.