

Academic effort patterns in contexts of vulnerability *

Claudia Navarro and Valeria Henríquez^[0000-0002-9003-2254] y Cristian Olivares-Rodríguez^[0000-0002-4991-5784]

Instituto de Informática, Universidad Austral de Chile
General Lagos 2086, Valdivia, Chile
claudia.navarro.galindo@alumnos.uach.cl
[valeria.henriquez,colivares]@inf.uach.cl
<http://www.inf.uach.cl>

Abstract. Latin American countries shown a low performance in mathematics according standardized tests such as PISA

In particular, Chile shows 51.5 % of students have low performance in mathematics. This figures increase in institutions with more vulnerable students.

The results motivate the incorporation of b-learning in formal education, using educational platforms such as Khan Academy, which allows students to develop autonomous learning by using available learning resources and leaving traces of their own knowledge building experience.

This paper presents a study of academic effort patterns, using data of the students of an educational institution in Valdivia. The analysis use the participant's temporal behavior and socioeconomic information with the purpose of understanding the risk of failure and school dropout. The final analysis with clusters shown different levels of effective monthly work throughout the academic year and different study profiles. In both analyzes, a strong relationship with academic performance was found, which makes it possible to have better understanding of the learning patterns about who are at greater risk of failing in their studies. The results represent a clear opportunity to provide effective support to risky students.

Keywords: Academic effort · Clustering · Vulnerability.

1. Introducción

La educación universal es un indicador que refleja el desarrollo de un país, tal como declara la OCDE [1], por ende la calidad de la educación y el acceso universal a ella es un desafío del cual los gobiernos, las instituciones educativas y las comunidades deben velar por brindar las mismas oportunidades a las personas [2]. Dentro de estas condiciones de partida se han detectado brechas entre los estudiantes que provienen de diferentes contextos socioeconómicos, de

* Estudio realizado con el soporte de Puentes Digitales.

sus hábitos de estudio o de otros factores que influyen y se reflejan en el bajo rendimiento y la deserción escolar. Según el segundo Informe Observatorio de Niñez y Adolescencia, Infancia [3], en Chile el 7% de los estudiantes abandonan la educación básica y el 11% la educación media. Estas cifras adquieren una mayor gravedad en establecimientos educacionales de alta vulnerabilidad social. Existen estudios en Chile que han detectado las variables que inciden en el bajo rendimiento y la deserción escolar [2], la mayoría de ellos están basados en datos históricos, lo que constituye una guía para detectar casos de riesgo. No obstante, estos estudios dificultan la creación de modelos actualizados que estén basados en indicadores estándares para tomar acciones preventivas.

La educación escolar en América Latina debe hacer frente a altas tasas de deserción o abandono escolar. Un estudio ha establecido una correlación entre el mal rendimiento académico, la baja escolaridad de la madre y la inasistencia, y su rol preponderante en la decisión de abandonar [4]. En Chile se elaboró un modelo que determina cómo medir la deserción escolar en Chile, haciendo uso de dos tipos de tasas de deserción: incidencia y prevalencia [2]. Por otra parte Portillo-Torres [5], hace una propuesta para reducir el abandono escolar basado en tres niveles de acción: 1) Prevención Universal, 2) Prevención Selectiva, y 3) Prevención indicada. No obstante, el estudio indica que no hay evidencia concreta del aporte del modelo. En consecuencia, conociendo oportunamente el rendimiento académico de los estudiantes y sus patrones de esfuerzo académico, se podría predecir su riesgo de abandono y tomar acciones preventivas.

La Fundación San Carlos del Maipo, en Chile, impulsa un sistema para prevenir la deserción escolar a través de un Sistema de Alerta Temprana (SAT) junto con un programa que intenta motivar a los estudiantes con sus colegios a través de un espacio recreativo para incentivar a los niños a reencontrarse con sus colegios. SAT considera como factor de riesgo la insistencia a clases, por lo que provee un seguimiento automático de las inasistencias de los estudiantes a clases y activa a una red de profesionales cuya tarea es intervenir preventivamente. Este sistema ha beneficiado a más de 7000 estudiantes, logrando disminuir en promedio un 37% la deserción escolar y aumentar la asistencia en un 79% de los casos que ingresan al programa [6]. Desde el punto de vista de los datos, un estudio desarrollado por Bowers [7] propone un modelo de predictor del riesgo escolar basado en el promedio de notas de los estudiantes de enseñanza secundaria, en todas sus asignaturas, y agrega datos obtenidos como exámenes de entrada y test estandarizados. Este modelo se sustenta en la construcción de modelos computacionales basados en minería de datos. Asimismo, un estudio realizado por Dekker [8] propone modelos de predicción de aprendizaje con una muestra de 648 estudiantes de Ingeniería Eléctrica en Holanda. En la muestra trabajan con más de 100 variables para cada uno de los estudiantes.

En este trabajo se propone analizar la información recopilada automáticamente durante el año 2016 a través de la plataforma Khan Academy, la cual cuenta con la herramienta BA-Khan (de Blended Analytics y Khan Academy) para recopilar una gran cantidad de datos de las actividades desarrolladas por los estudiantes. Estos datos resultan de utilidad en el análisis automático de patrones

de comportamiento. Por lo tanto, se busca descubrir patrones inherentes en los datos de comportamiento potenciados con información socio-económica de los estudiantes. La información obtenida desde este estudio de manera automática representa un soporte concreto para que los establecimientos educacionales comprendan los patrones de comportamiento académicos de sus estudiantes, en particular, los patrones de esfuerzo asociados a la deserción escolar.

Por lo tanto, las preguntas que guían este trabajo son:

- RQ1: ¿La información socio-económica determina la deserción escolar?
- RQ2: ¿Existen patrones de esfuerzo académico claramente reconocibles entre los participantes?
- RQ3: ¿Es posible determinar automáticamente patrones de esfuerzo académico entre los participantes?
- RQ4: ¿Existe un patrón de esfuerzo académico vinculado a la deserción escolar?

2. Materiales y métodos

Para este estudio se utilizaron datos recopilados en un establecimiento de educación secundaria (liceo) en Valdivia-Chile, seleccionado debido a que cuenta con dos años de utilización de la plataforma BA-Khan. El conjunto de datos construido incorpora información socioeconómica, información de rendimiento académico y de trabajo de los estudiantes en la plataforma Khan Academy sobre la asignatura de matemáticas. Los datos obtenidos corresponden a 281 estudiantes (43.77% hombres), de los cuales 173 pertenecen a primer año de educación secundaria (41.61% hombres).

Para la recolección de datos en el establecimiento se siguió el siguiente protocolo: 1) visita al establecimiento, 2) solicitud del consentimiento informado de parte de padres/tutores de los estudiantes con el propósito de obtener los datos requeridos para el análisis, 3) recuperación de datos de rendimiento académico desde los libros de clases y datos socio-económicos, 4) asegurar la anonimidad de los estudiantes en los datos y 5) vincular datos académicos, socio-económicos y de comportamiento. En este trabajo solo se consideran los datos correspondientes a los estudiantes del establecimiento en estudio.

La metodología b-learning presenta la combinación de Internet y medios digitales con clases formales que requieren la presencia física de profesor y estudiantes. Han surgido numerosas adecuaciones que combinan estos dos elementos, según Horn, Gu & Evans se pueden diferenciar 4 modelos para b-learning en la educación primaria y secundaria: Rotation Model, Flex Model, Self-Blend model y Enriched-Virtual model [9]. En particular los datos de este estudio se recopilaban a partir de la utilización del modelo de rotación de estaciones y de laboratorio debido a que no todos los estudiantes pertenecientes al establecimiento tienen acceso a Internet ni cuentan con los dispositivos adecuados para acceder a los contenidos desde sus casas.

La herramienta Khan Academy [10] fue la elegida como LMS pues es una plataforma web de acceso libre y gratuito, con contenido validado por decenas

de profesores especialistas de matemáticas y otras disciplinas. En la figura 1a se puede apreciar cómo los estudiantes pueden ejercitar en Khan Academy, ver pistas en casos de no saber resolver los ejercicios o revisar los contenidos asociados a través de un video tutorial de los conceptos asociados a la ejercitación. En la figura 1b los estudiantes pueden analizar su esfuerzo académico basado en la cantidad de ejercicios que realizaron sobre una temática, el tiempo dedicado a cada uno de ellos, si pudieron completarlo de manera correcta y si utilizaron los videos y pistas para contestar.

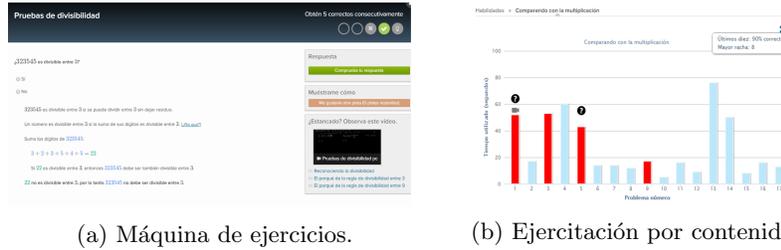


Fig. 1: Plataforma Khan Academy, ejercitación.

El equipo ha enriquecido las herramientas Khan Academy desarrollando BA-Khan Academy, la que incorpora funcionalidades de apoyo al profesor. BA-Khan Academy periódicamente consume los servicios de Khan Academy API Explorer y recopila la información del trabajo realizado por cada uno de los estudiantes. La herramienta tiene principalmente dos macro funcionalidades de apoyo a la docencia: que son: 1) Apoyo a la evaluación y calificación y 2) Gestión de aula. Desde la perspectiva del tratamiento de los datos se utilizó los nueve pasos del método KDD (del inglés Knowledge Discovery in Databases)[11].

3. Resultados

Los resultados se encuentran dirigidos por las cuatro preguntas de investigación definidas en este estudio.

3.1. RQ1: ¿La información socio-económica determina la deserción escolar?

Para responder a esta pregunta se analizan mediante la herramienta de IBM Watson Analytics los factores que influyen en la aprobación de los estudiantes debido a su relación estrecha con la deserción escolar. La Figura 2 muestra que la situación final se ve influenciada fuertemente por el Promedio Final, el Promedio de Matemáticas y la Asistencia, los tres con un 93%. Por lo que se puede comprobar que obtener buenas calificaciones y asistir regularmente a clases nos lleva a una Situación Final satisfactoria (Aprobado).



Fig. 2: Situación final.

Por otra parte, se analiza cómo afecta el nivel educacional de los padres en el rendimiento de los estudiantes. Se puede apreciar en el Cuadro 1 que existe una correlación positiva entre el promedio de matemáticas obtenido por los estudiantes y el nivel educacional de los padres. No obstante, cuando la educación del padre corresponde a Superior Instituto Profesional (SIP) se observa un resultado superior en promedios de matemáticas, mientras que el promedio más bajo se alcanza cuando la educación del padre corresponde a Básica Incompleta (BI).

Table 1: Nivel educacional de los padres.

Escolaridad	(Padre)	(Madre)
	Promedio de Matemáticas	
Sin Educación (SE)	-	4,50
Educación Básica Incompleta (BI)	47,29	52,38
Educación Básica Completa (BC)	50,41	50,41
Educación Media Incompleta (MI)	53,95	53,09
Educación Media Completa (MC)	50,60	49,71
Educación Superior Instituto Profesional Incompleta (SIPI)	-	62,00
Educación Superior Instituto Profesional Completa (SIP)	54,67	57,00
Educación Superior CFT Incompleta (SCFTI)	-	52,00

3.2. RQ2: ¿Existen patrones de esfuerzo académico claramente reconocibles entre los participantes?

Para responder a esta pregunta se analiza el comportamiento de los estudiantes en la plataforma Khan Academy asociado al aprendizaje de matemáticas, extrayendo información de los ejercicios realizados por los estudiantes e información vinculada a la visualización de videos.

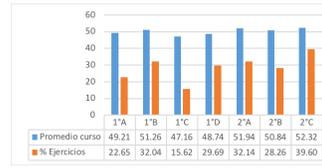
El horario del día en el cual se utiliza más la plataforma es una variable que determina el esfuerzo académico. Para ello, se determina que el pico de trabajo durante el día es alcanzado durante la estadía de los estudiantes en el establecimiento educacional, apreciando que 150 estudiantes de primer año de secundaria y 100 estudiantes de segundo año de secundaria representan el mayor

esfuerzo conjunto. El resultado de este análisis fue calculado considerando los datos de todo el año 2016 (Figura 3a).

Al analizar la cantidad de ejercicios totales realizados por los estudiantes durante el año la cifra asciende a 114.516 en primero de secundaria y 71.049 en segundo de secundaria. Mientras que los cursos con mayor porcentaje de ejercicios realizados alcanzan un mayor rendimiento académico (Figura 3b), por lo que se puede concluir que al realizar mayor cantidad de ejercicios se logran mejores promedios en matemáticas.



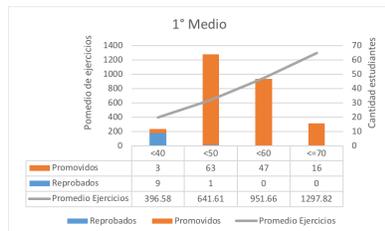
(a) Horario.



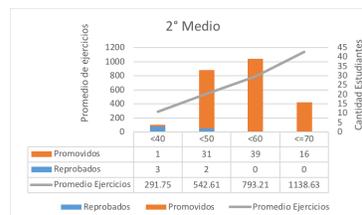
(b) Ejercicios.

Fig. 3: Esfuerzo en horario y ejercicios.

Al analizar la cantidad de ejercicios realizados por cada estudiante, se consideran rangos de calificaciones promedio en < 40 , < 50 , < 60 y ≤ 70 . Luego se calcula el promedio de ejercicios realizados y el promedio de notas para cada rango, además de la cantidad de estudiantes promovidos y reprobados. Las Figuras 4a y 4b evidencian que a mayor cantidad de ejercicios resueltos más alto será el rango alcanzado por el estudiante en las notas de matemáticas. Al momento de hacer ejercicios, los estudiantes tienen la opción de pedir pistas (count_hints) y de realizar varios intentos de respuesta (count_attempt). Después de analizar estos datos para los estudiantes de primero y segundo, no se aprecian variaciones significativas en los datos.



(a) Primero medio.



(b) Segundo medio.

Fig. 4: Ejercicios resueltos y promedio de notas.

3.3. RQ3: ¿Es posible determinar automáticamente patrones de esfuerzo académico entre los participantes?

Para responder a esta pregunta se analiza la información mensual (de marzo a noviembre) del trabajo realizado por los estudiantes de primero dentro de la plataforma BA-Khan Academy. El número de estudiantes en cada mes varía debido a que no siempre trabajaron los mismos estudiantes.

Antes de llevar a cabo el análisis de grupos, se estudian los índices de verosimilitud para el algoritmo de clustering EM con 3 (-32,19) y 6 (-31,47) grupos, los cuales resultan ser similares en cuanto a su confiabilidad. Sin embargo, con la obtención de 3 clusters presenta grupos más reconocibles debido a que se aprecian características que hacen más fácil su diferenciación.

Posteriormente, se realiza un análisis de clúster mensual con el propósito de obtener una visión de las diferencias de comportamiento de los estudiantes a través del año y, asimismo, observar el cambio de comportamiento de cada estudiante en función de los grupos automáticamente reconocidos. En consecuencia, los tres clusters son Trabajo Alto (TA), Trabajo Medio (TM) y Trabajo Bajo (TB), ya que se comprueba que los estudiantes se dividen de acuerdo al esfuerzo académico invertido al momento de hacer ejercicios. Así, los nombres de los cluster se definen en función del promedio de ejercicios existente en cada clúster.

Al analizar los patrones de los estudiantes de cada cluster se aprecian claras diferencias. Los estudiantes pertenecientes al clúster TB realizan más ejercicios a inicio de año pero a medida que va pasando el tiempo la cantidad disminuye, volviendo a subir a finales de año. Mientras que los estudiantes que pertenecen al clúster TM trabajan de una manera más constante a lo largo del año, siendo este trabajo mayor durante el primer semestre y disminuye de forma notoria en el segundo semestre. Finalmente, para el clúster TA los estudiantes trabajan de manera continua durante todo el año.

3.4. RQ4: ¿Existe un patrón de esfuerzo académico vinculado a la deserción escolar?

Para responder a esta pregunta se analiza el patrón de esfuerzo académico de los estudiantes y su relación con la reprobación del curso y, en consecuencia, con el riesgo de deserción escolar. Para ello se realiza un nuevo análisis de cluster, descartando del estudio aquellos estudiantes que fueron retirados del establecimiento ya que no es posible determinar su desempeño académico. En este estudio también se utilizó del algoritmo EM, el cual arroja como resultado un número óptimo de tres cluster. La interpretación de los clústeres nos indica que hay tres tipos de estudiantes, los cuales se pueden clasificar en los siguientes grupos: 1) Estudiante precario (EP), 2) Buen estudiante (BE) y 3) Estudiante Excelente (EE).

En la Tabla 2 se aprecian las características de los estudiantes en los diferentes grupos. Se aprecia una fuerte relación entre el patrón de trabajo a lo largo del año con los resultados que se obtienen en cuanto a promedios de calificaciones.

En el caso de un estudiante EE, el trabajo mensual en casi todos los meses fue TA y TM y, finalmente, obtiene a final de año una calificación promedio de 5,83 (sobre 7,0) en matemáticas. Un estudiante EP, que a lo largo del año realiza un trabajo TB, lo lleva a un promedio de matemáticas menor a un estudiante EE y la asistencia promedio se encuentra cercana al mínimo permitido para aprobar el curso.

Table 2: Perfiles de estudiantes.

Indicador	Cluster 1	Cluster 2	Cluster 3
Esfuerzo	(TA,TB,TB,TB, TB,TM,TM,TB)	(TA,TM,TA,TM, TB,TM,TM,TB)	(TA,TA,TM,TA, TA,TM,TM,TA)
Promedio PM	44,90	51,50	58,30
Promedio PF	51,20	55,60	60,00

En las Figuras 5a y 5b se presenta la cantidad de ejercicios realizados en el primer y segundo semestre académico. Durante el primer semestre se observa que un estudiante EE (cluster 3) trabaja más que sus pares, siendo muy superior a un estudiante EP (cluster 1) donde se encuentran los estudiantes con menores promedios. También se puede notar que, para los tres clúster el trabajo fue similar en los meses de marzo, septiembre y octubre, por lo que se puede concluir que los estudiantes EE y BE (clúster 2 y 3) con mejores calificaciones disminuyen su trabajo al finalizar el año. Esto puede vincularse a que los estudiantes tenían certeza de su aprobación. En cambio, un estudiante EP (clúster 1) aumenta su trabajo en el último mes, muy probablemente por encontrarse en dificultades de alcanzar la aprobación. Finalmente, al analizar los ejercicios incorrectos se ve que los estudiantes pertenecientes a EE (clúster 3) en general cometen más errores, pero debe considerarse que este grupo es el que más ejercicios realiza y que realizar ejercicios incorrectos también es una vía de aprendizaje.

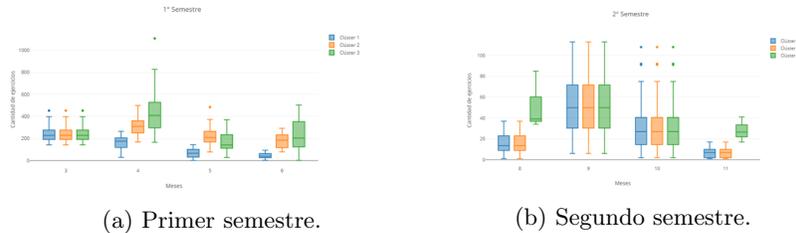


Fig. 5: Ejercicios resueltos y promedio de notas.

4. Conclusiones

Respecto de las preguntas de investigación planteadas podemos concluir lo siguiente:

1. ¿La información socio-económica determina la deserción escolar? Si bien los resultados arrojan una cierta relación entre el ingreso familiar y la deserción, esta relación es indirecta pues la influencia se ve ejercida por el nivel de asistencia a clases. Por otra parte dado el contexto de vulnerabilidad del establecimiento los ingresos familiares son muy similares y rondan el ingreso mínimo legal. Por tanto no podemos afirmar que el ingreso familiar en establecimientos educacionales de vulnerabilidad determinen la deserción escolar.
2. ¿Existen patrones de esfuerzo académico claramente reconocibles entre los participantes? En este caso, los resultados nos permiten afirmar que existen claros patrones de comportamientos en el uso de la plataforma Khan Academy y que las características que definen el comportamiento están vinculadas principalmente a esfuerzo, esto es, cantidad de ejercicios realizados, sin importar la correctitud de ellos y las pistas revisadas para realizar la ejercitación. Por tanto cobra vital importancia la labor docente que debe fomentar la ejercitación de los estudiantes sin reprender en los casos de error.
3. ¿Es posible determinar automáticamente patrones de esfuerzo académico entre los participantes? La determinación del esfuerzo académico a partir de las características relevadas del uso de la plataforma está construida a partir de datos de recopilación automática, por ende, es posible la determinación automática de patrones de esfuerzo académico en la medida que el proceso de limpieza y normalización de los datos pueda ser realizada de manera automática.
4. ¿Existe un patrón de esfuerzo académico vinculado a la deserción escolar? Al analizar el patrón de ejercitación mensual por grupos de estudiantes se concluye que el esfuerzo académico y la constancia contribuyen positivamente en el desempeño académico, disminuyendo el riesgo de deserción y reprobación. Además cobra relevancia el trabajo realizado en los primeros meses de clases, pues los resultados muestran que los estudiantes con bajo esfuerzo no logran repuntar durante el año, por ende se podría considerar que los primeros tres meses son representativos del trabajo durante el resto del año.

De esta manera y de acuerdo a lo descrito en la literatura, este trabajo contribuye a entender cómo los patrones de uso de plataformas educativas pueden servir para determinar el riesgo escolar de los estudiantes, principalmente porque al incorporar la metodología b-learning al contexto formal se puede recopilar de manera automática datos que pueden ser traducidos en asistencia, indicador clave de la deserción escolar [6]. Adicionalmente al encontrar una relación entre el esfuerzo en la ejercitación de los estudiantes en la asignatura de matemáticas y las calificaciones obtenidas en el promedio de la misma asignatura y en el promedio final, indicador del riesgo escolar [7], es posible crear indicadores tempranos del nivel de esfuerzo de los estudiantes para la intervención oportuna de

los estudiantes con esfuerzo bajo y por ende en riesgo. Finalmente este trabajo aporta evidencias de la importancia de la equivocación como parte del proceso de aprendizaje, pues los estudiantes que logran un mayor desempeño son aquellos que cometen más errores durante la ejercitación y por tanto las prácticas pedagógicas deben contribuir a que los estudiantes erren en un entorno seguro.

Referencias

1. OCDE, "Programa para la Evaluación Internacional de Alumnos Resultados principales PISA Estudiantes de bajo rendimiento Por Qué se Quedan Atrás y Cómo Ayudarles a Tener Éxito", <http://www.oecd.org/pisa/keyfindings/PISA-2012-Estudiantes-de-bajo-rendimiento.pdf>. Último acceso 2016/09/01.
2. MINEDUC. Centro de Estudio, "Serie Evidencias: Medición de la deserción escolar en Chile", <https://bit.ly/2Ig8AZB>. Último acceso 2016/08/31.
3. Observatorio Niñez y adolescencia, "Infancia Cuenta Chile 2014 - 2º Informe Observatorio Niñez y Adolescencia", <http://www.observatorioninez.cl/2014/07/29/infancia-cuenta-chile-2014-2o-informe-observatorio-ninez-y-adolescencia/>. Último acceso 2016/08/31.
4. SEP, Secretaría de Educación Pública México, "Movimiento contra el abandono escolar", <https://bit.ly/2KIlh5E>. Último acceso 2016/09/01.
5. Portillo-Torres, M. C. Propuesta de un nuevo enfoque para reducir el abandono escolar en secundaria. *Revista Electrónica Educare*, 19(2), 303-316 (2015).
6. San Carlos, "Fundación San Carlos del Maipo Impulsa Novedosos Sistemas Para Combatir Deserción Escolar", <http://docenciaeinvestigacion.blogspot.cl/2015/02/fundacion-san-carlos-del-maipo-impulsa.html>. Último acceso 2016/08/31.
7. Bowers, A. J. Analyzing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out and hierarchical cluster analysis. *Practical Assessment Research and Evaluation*, 15(7), 1-18 (2010).
8. Dekker, G., Pechenizkiy, M., and Vleeshouwers, J. Predicting students drop out: A case study. *Proceedings of the 2nd International Conference on Educational Data Mining, EDM'09*, 41-50 (2009).
9. Horn M., Gu A., and Evans M. *Knocking Down Barriers: How California Superintendents Are Implementing Blended Learning*. Clayton Christensen Institute for Disruptive Innovation (2014).
10. Khan S. *The one world schoolhouse: Education reimaged*. Twelve, New York (2012).
11. Maimon, O., Rokach, L. Introduction to knowledge discovery in databases. In *Data Mining and Knowledge Discovery Handbook* pp. 1-17. Springer US (2005).