



# Um modelo preditivo da evasão de estudantes no ensino superior

*A predictive model of the evasion of students in higher education*

**Diogo Martins Gonçalves de Morais** ([pro7113@cefesa.edu.br](mailto:pro7113@cefesa.edu.br))  
*Doutor em Administração pela Universidade Municipal de São Caetano do Sul (USCS) e professor da Faculdade de Tecnologia Termomecânica (FTT).*

**Arthur Augusto Micheletti de Souza** ([lanceart.a7@gmail.com](mailto:lanceart.a7@gmail.com))  
*Graduando em Engenharia de Computação pela Faculdade de Tecnologia Termomecânica (FTT).*

**Vittoria Cassoni** ([VittoriaCassoni@outlook.com](mailto:VittoriaCassoni@outlook.com))  
*Graduanda em Engenharia de Computação Faculdade de Tecnologia Termomecânica (FTT).*

Engenharia de Computação

FTT Journal of Engineering and Business. • SÃO BERNARDO DO CAMPO, SP

DEZ. 2019 • ISSN 2525-8729

**Submissão:** 9 mai. 2019. **Aceitação:**

18 Out. 2019

**Sistema de avaliação:** às cegas dupla (*double blind review*).

FACULDADE TECNOLOGIA TERMOMECANICA, p. 50-61

## *Resumo*

O presente estudo propõe o desenvolvimento de um sistema computacional (modelo preditivo) capaz de obter a probabilidade de evasão dos estudantes matriculados na Faculdade de Tecnologia Termomecânica (FTT), a partir da análise dos dados do perfil socioeconômico e desempenho, obtidos dos alunos ingressantes na ocasião do vestibular. Para tanto, pretende-se utilizar uma abordagem quantitativa em nível explicativo, pois o estudo investigará o grau de associação existente entre o perfil dos alunos ingressantes dos cursos de graduação da FTT e a evasão medida na instituição constatadas no período compreendido entre agosto de 2017 e dezembro de 2018. Além de oferecer uma ferramenta estratégica para a gestão educacional considerando as características específicas e perfil da FTT, o resultado desta pesquisa pretende estimular nesta instituição o desenvolvimento de novas aplicações de Data Science e Big Data. Concluiu-se nesse estudo que o modelo preditivo é viável e conseguiu prever com sucesso a evasão dos estudantes, a partir de 11 variáveis explicativas, com taxa de acerto de 76%. Além disso, constatou-se que o algoritmo genético se mostrou adequado para este tipo de aplicação.

**Palavras-chave:** Algoritmo genético. Evasão de estudantes. Modelo preditivo.

## *Abstract*

This research project proposes the development of a computational system (predictive model) capable of obtaining the probability of evasion of students enrolled in the Faculdade de Tecnologia Termomecânica (FTT), based on data from the socioeconomic profile and performance, obtained from freshmen students at the time of the entrance exam. To do so, we intend to use a quantitative approach at an explanatory level, since the study will investigate the degree of association between the profile of incoming students of FTT undergraduate courses and the measured dropout in the institution observed in the period between August 2017 and December 2018. From this analysis, we intend to use this data for the creation and learning of a computer intelligence system capable of forecasting the evasion of new entrants. In addition to offering a strategic tool for educational management considering the specific characteristics and profile of FTT, the result of this research intends to stimulate in this institution the development of new Data Science and Big Data applications. It was concluded in this study that the predictive model is feasible and was able to predict successfully student avoidance, from 11 explanatory variables, with a success rate of 76%. In addition, it was found that the genetic algorithm was adequate for this type of application.

**Keywords:** Genetic Algorithm. Student evasion. Predictive model.

## *Introdução*

Nos últimos anos, o Brasil tem experimentado mudanças significativas na oferta e concepção da Educação Superior. Na leitura dos dados do Instituto Nacional de Pesquisas Educacionais (INEP) sobre a expansão do ensino superior no país, observa-se que em 2003 havia 1.859 Instituições de Ensino Superior (IES) com 16.505 cursos de graduação e 3.936.933 estudantes matriculados (BRASIL, 2003). Em 2013, dez anos depois, esses números aumentaram para 2.391 IES que passaram a oferecer 32.049 cursos, com 7.305.977 estudantes matriculados (BRASIL, 2013).

Em 2016, quando foram realizados os estudos do Ministério da Educação, até o desenvolvimento deste estudo, o número de IES subiu para 2.407 IES, com um total de 8.052.254 estudantes matriculados, sendo que 82,3% matriculados nas IES privadas (BRASIL, 2016).

De um lado, esse contexto, marcado pelo crescimento do ensino superior no país, sugere que os mantenedores e gestores das IES privadas passaram a enfrentar maiores desafios para a manutenção e desenvolvimento de seus negócios.

De outro, os estudos sobre gestão organizacional têm constatado a contribuição de alguns componentes desse processo para a construção dos resultados da organização, dentre os quais, a compreensão sobre a evasão dos estudantes se constitui como uma ferramenta estratégica para a gestão de instituições de ensino superior, em especial as IES privadas.

Dessa forma, o presente estudo se justifica por propor o desenvolvimento de um sistema com inteligência computacional capaz de prever a probabilidade de evasão dos seus ingressantes, a partir dos dados de seus perfis obtidos na ocasião do vestibular, para uso imediato dos gestores da Faculdade de Tecnologia Termomecânica.

## *Referencial teórico*

Na revisão da literatura, constatou-se que as teorias sobre evasão de estudantes de graduação estão bem fundamentadas, merecendo destaque os estudos de Tinto (1975), que desenvolveu um modelo teórico denominado Teoria de Integração dos Estudantes, e se tornou base para praticamente todos os modelos teóricos sobre o assunto.

Os estudos de Tinto (1975, 1987, 1993, 1996, 1997, 2010, 2012) apresentam um panorama abrangente sobre os motivos da evasão, avançam na construção de modelos preditivos, e se tornaram a maior referência para diversos outros estudos realizados nos Estados Unidos.

No Brasil, a Teoria da Integração dos Estudantes também se tornou referência para estudos realizados em um contexto brasileiro, e se transformando em temas para dissertações de mestrado, teses de doutorado e estudos dos mais diversos, como Tibola et al. (2012); Palácio (2012); Pereira Júnior (2012); Amaral (2013); Tontini, Walter (2014), e por fim, Sales Junior et al (2016).

Segundo Tinto (1975), alguns resultados que associam a permanência do aluno são:

- Status socioeconômico se revela inversamente relacionado à evasão;
- A renda tomada de forma isolada é menos determinante para a permanência que a qualidade das relações familiares e suas expectativas com relação à educação dos filhos;
- Nível de expectativa dos pais influencia a própria expectativa dos filhos, como também a sua permanência na faculdade;
- A expectativa educacional é tanto mais alta quanto maior é o status social do estudante;
- Desempenho é o melhor preditor para a permanência desde que as habilidades do estudante possibilitem as realizações exigidas naquele meio;
- A evasão é um comportamento mais impulsivo que persistente;
- Mais alto nível de planos (educacionais e de carreira) constitui o maior preditor para a permanência;
- Desenvolvimento intelectual como parte integral do desenvolvimento da personalidade da pessoa e como reflexo de sua integração dentro do sistema acadêmico relaciona-se à persistência;
- Grau de congruência entre o desenvolvimento intelectual do indivíduo e o clima intelectual da instituição é o elo que garante a permanência;
- Outros fatores sendo indiferentes à integração social aumentam a probabilidade de a pessoa permanecer na faculdade, etc.

Segundo Tibola et al. (2012), a evasão é considerada como a desistência do curso sem o ter concluído, ou a transferência para outra instituição de ensino superior. Fatores que levam o aluno a essa decisão são: cursos nos quais não se veem bem-sucedidos, ausência de apoio da instituição ao aluno (apoio social e pessoal), ausência de interação dentro da instituição e cursos que não promovam a aprendizagem.

De acordo com Palácio (2012), evasão é a saída definitiva do aluno do curso, sem concluí-lo tendo em vista tais motivos: falta de vocação para o curso; trabalho em horário incompatível com as aulas; falta de tempo para estudar e trabalhar; distância grande entre a residência e a faculdade;

transporte deficitário; dificuldades de acompanhamento do conteúdo; estrutura do curso, dentre muitas outras.

Amaral (2013) considera evasão como a interrupção do ciclo de estudo, temporariamente ou não, por meio de abandono, trancamento ou cancelamento da matrícula, ou transferência externa.

Conforme Pereira Júnior (2012), a evasão consiste na perda de parte do corpo discente durante o ciclo de estudos, onde o índice dela é maior em IES particulares do que em IES públicas. As variáveis que explicam a causa desse fenômeno são: integração acadêmica, integração social, compromisso com a instituição, compromisso com o graduar-se, compromisso com o curso, aprovação familiar e encorajamento de amigos, condições financeiras e de moradia, oportunidade de transferência, disponibilidade de tempo para estudo e intenção de permanência no curso. O autor utiliza em seu estudo regressão logística.

Tontini, Walter (2014) não chegou a um consenso acerca do termo evasão, analisa, então, as causas que levam a ela, onde as que mais se destacaram foram: colocação profissional e vocação do aluno, disponibilidade de tempo para estudo e fatores da vida pessoal. Em sua pesquisa, utilizou Redes Neurais Artificiais do tipo RBF para identificação do risco de evasão dos alunos e a análise de cluster.

Sales Junior et al (2016) considera evasão como o estudante que abandonou o curso antes de sua diplomação. Para determinar os fatores que levam o estudante a evadir-se, utilizou o Teste Qui-Quadrado, ajustando a um modelo de regressão logística obtendo como resposta formado ou evadido, onde os fatores são: contexto familiar (escolaridade e ocupação dos pais, renda familiar), atributos individuais (faixa etária, sexo, cor, cotista), escolaridade anterior (ENEM, fundamental, médio, cursinho), intenções em relação a metas e compromissos, experiências Institucionais (pesquisa, estágio, área do conhecimento) e compromissos externos e desempenho Acadêmico.

No presente estudo, a evasão será considerada a saída voluntária ou não do estudante de uma instituição de ensino superior, sem nenhum tipo de retorno por tempo indefinido, exceto a diplomação, seguindo as mesmas propostas de Tibola et al. (2012) e Palácio (2012).

A revisão da literatura possibilitou o embasamento teórico para a conceituação de evasão, utilizada no modelo proposto, assim como explorar os possíveis fatores de influência da evasão nos cursos de graduação de Administração, Engenharia de Controle e Automação, Engenharia de Alimentos e Engenharia de Computação de uma IES privada.

# Procedimentos metodológicos

O estudo utilizou-se de uma abordagem quantitativa em nível explicativo, pois investigou o grau de associação existente entre o perfil acadêmico e socioeconômico dos alunos ingressantes e a sua evasão dos cursos de graduação da Faculdade de Tecnologia Termomecânica, constatadas no período compreendido entre agosto de 2017 e dezembro de 2018. Esta abordagem caracteriza, segundo Gil (2010), a natureza quantitativa pesquisa, além do seu nível explicativo.

A amostra utilizada no estudo é não probabilística e intencional, formada por 448 alunos que ingressaram nos cursos de Administração (3 turmas – desde 2º semestre/2017), Engenharia de Alimentos (3 turmas – desde 2º semestre/2017), Engenharia de Controle e Automação (3 turmas – desde 2º semestre/2017) e Engenharia de Computação (3 turmas – desde 2º semestre/2017).

Em cada uma das turmas ingressaram 40 alunos, que serão classificados como matriculados ou evadidos, em um corte transversal do tempo, que será no 1º semestre de 2019, precisamente no dia 26 de fevereiro de 2019.

Para a elaboração do modelo, inicialmente foram consideradas 13 variáveis coletadas por meio do questionário socioeconômico e pelo desempenho acadêmico, agrupadas em 3 blocos de acordo com a sua natureza e com o modelo proposto por Tinto (1997).

Para determinar os fatores que levam o estudante a evadir-se, uma análise estatística bivariada dos dados foi realizada utilizando o teste Qui-Quadrado.

O teste Qui-quadrado é um teste de hipótese que compara a distribuição observada dos dados a uma distribuição esperada dos dados. Nesse estudo, verificou-se se haviam diferenças significativas entre a distribuição constatada e esperada de cada uma das variáveis para os estudantes matriculados e evadidos.

Para utilizar os dados no modelo, é necessário converter cada resposta em 0 ou 1. Por exemplo, a variável situação tem como resposta ou vaga geral ou vaga social, se uma pessoa for vaga geral ela recebe 1 em vaga geral e 0, em vaga social.

Na elaboração do modelo, utilizou-se a técnica denominada algoritmo genético, escrita com a linguagem de programação denominada Python. O algoritmo genético foi inspirado pela biologia, e está representado na Figura 1.

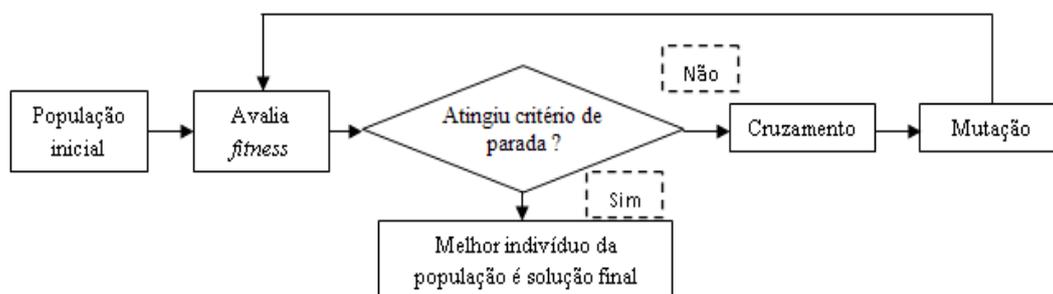


Figura 1: Fluxograma do Algoritmo genético  
Fonte: Elaborado pelos autores (2019)

Para o cálculo do Fitness de cada cromossomo de acordo com a seguinte relação matemática:

$$Fitness = \left( \frac{n^{\circ} \text{ de matriculados descobertos}}{n^{\circ} \text{ de matriculados do banco de dados}} \right) \times \left( \frac{n^{\circ} \text{ de evadidos descobertos}}{n^{\circ} \text{ de evadidos do banco de dados}} \right)$$

Em 1858 Charles Darwin apresentou sua teoria da evolução através da seleção natural dos indivíduos da natureza e após um tempo, em 1900, o trabalho de Gregor Mendel sobre herança genética ficou em alta. E então graças a estes estudos fundamentou-se a técnica de Algoritmo genético, sobre o qual John Holland começou a desenvolver as primeiras pesquisas e em 1975 lançou o livro *Adaptation in natural and artificial systems*, considerado a bíblia do algoritmo genético, que tem como finalidade solucionar problemas de forma otimizada (ICM USP).

Os algoritmos genéticos são algoritmos de busca e otimização global, baseado nos procedimentos de seleção natural e da genética.

A princípio é gerada uma população inicial, formada por diversos indivíduos (cromossomos, ou cadeia de bits onde cada bit é um gene) aleatórios, que podem ou não já ser a solução do problema.

Durante a evolução, cada indivíduo (cromossomo) é avaliado recebendo um valor de aptidão (fitness), que diz sobre sua habilidade de adaptação ao meio.

Após a avaliação uma quantia de indivíduos (cromossomos) com aptidões melhores é mantida enquanto os outros são descartados.

Essa quantia armazenada está sujeita a modificações em seus genes, por meio de cruzamento (crossover) e mutações, assim gerando descendentes para a próxima geração.

Esse processo anterior é chamado de reprodução será repetido até que apareça uma solução que seja satisfatória.

Para o desenvolvimento do modelo, utilizou-se os seguintes equipamentos utilizados: HP Compaq Pro 4300 SFF Brazil PC (Windows 7 Enterprise – 64 bits, memória RAM 8GB, Intel Core™ i7 - 3ª geração); Dell OptiPlex 7020 (Windows 10 Pro – 64 bits, memória RAM 8GB, Intel Core™ i7 - 4ª geração)

A estratégia de busca ocorre de forma paralela e estruturada, de modo aleatório, onde se busca aptidão (solução) satisfatória (ROSA; LUZ, 2009).

## *Apresentação e discussão dos resultados*

Para decidir quais variáveis serão utilizadas no estudo, realizou-se o teste Qui-Quadrado, com as variáveis organizadas em três categorias, a saber: contexto familiar, atributos individuais e escolaridade, constatando quais as variáveis se diferenciaram do que seria esperado para uma distribuição de alunos matriculados e evadidos, apresentado no Quadro 1.

Quadro 1: Resultados do Teste Qui-Quadrado para as variáveis do estudo

<b>Natureza da variável</b>	<b>Variável</b>	<b>Respostas</b>	<b>Valor-p</b>
Atributo Individual	Idade	17 ou menos, 18 anos, 19 anos, 20 anos, 21 a 24 anos, 25 anos ou mais	0,002
	Curso	Administração, Engenharia de Alimentos, Engenharia de Computação e Engenharia de Controle e Automação	0,023
	Situação	Vaga Geral e Vaga Social	0,007
Escolaridade	Instituição de Ensino Médio	Todo em escola pública, Todo em escola particular, Maior parte em escola pública, Maior parte em escola particular e CTM – Colégio Termomecanica	0,004
	Tipo de curso de ensino médio	Ensino médio tradicional Profissionalizante técnico, Profissionalizante magistério, Educação de Jovens e Adultos (EJA) e/ou supletivo ou Outra modalidade	0,035
	Ano de conclusão do ensino médio	Antes de 2002, Entre 2002 e 2011, 2012, 2013, 2014, 2015, 2016, 2017	0,011
	Nota da redação no vestibular	[0, 10]	0,168
	Nota final no vestibular	[0, 10]	0,003
Contexto Familiar	Quem incentivou a cursar graduação	Pais, Colegas/Amigos, Outros ou Ninguém	0,005
	Familiar concluiu superior	Sim ou Não	0,04
	Ocupação do Pai	Desempregado, Aposentado, Operário com pouca qualificação, Proprietário ou administrador de pequeno negócio, Técnico de nível médio, Profissional liberal, professor ou técnico de nível superior Proprietário ou administrador de grande ou média empresa	0,011
	Ocupação da Mãe	Desempregada, Aposentada, Operária com pouca qualificação, Proprietária ou administradora de pequeno negócio, Técnica de nível médio, Profissional liberal, professora ou técnica de nível superior, Proprietária ou administradora de grande ou média empresa, ou Do lar	0,003
	Quantidade de pessoas que vivem da renda	Número de pessoas	0,086

Fonte: Elaborado pelos autores (2019)

Com a realização do teste do Qui-quadrado, mantiveram-se apenas as variáveis cujas diferenças entre a distribuição constatada e esperada para os alunos matriculados e evadidos se mostraram significativas ( $p < 0,05$ ).

Desta forma, foram mantidas 11 das 13 variáveis, sendo elas: idade, instituição de ensino médio, quem incentivou a cursar graduação, familiar concluiu ensino superior, ano de conclusão do ensino médio, ocupação do pai, ocupação da mãe, nota final no vestibular, tipo de curso de ensino médio, situação e curso.

Definidas as variáveis e a situação atual de cada um dos estudantes, como matriculado ou evadido, elaborou-se o modelo, por meio da criação de uma população inicial de 6 cromossomos.

A quantidade de cromossomos utilizada no estudo possui esse valor por conveniência e a quantidade de variáveis foi definida posteriormente ao teste Qui-quadrado, relacionadas ao perfil socioeconômico e acadêmico dos estudantes.

Registra-se que o modelo preditivo foi elaborado em uma versão preliminar, considerando as 13 variáveis indicadas no Quadro 1, e conseguiu acertar 70% dos evadidos e matriculados, e quando utilizado apenas as 11 variáveis mantidas após o teste do Qui-Quadrado, o modelo passou a ter uma taxa de acertos de 76% dos evadidos e matriculados em um intervalo de tempo de 4h.

## *Considerações finais*

O modelo preditivo elaborado se mostrou eficaz na identificação de estudantes ingressantes que possuem potencial para evadir-se dos cursos de graduação, por meio de suas características socioeconômicas e desempenho no vestibular, com a taxa de acerto esperada.

O algoritmo genético se mostrou adequado para tratar desse tipo problema, uma vez que o tempo utilizado para a execução do programa escrito foi compatível com outros algoritmos utilizados na literatura.

A linguagem Python se mostrou apropriada para a elaboração do modelo preditivo, visto que é uma linguagem de programação orientada a objetos, interpretada, interativa e de código aberto. Possui clareza quanto a sua sintaxe e pode ser utilizada como linguagem de scripting associada a outras linguagens de programações.

Dada a importância da identificação prematura da evasão dos alunos ingressantes, o modelo preditivo elaborado se tornou uma ferramenta estratégica para a gestão educacional da instituição ofertante dos dados utilizados, considerando as suas características específicas, além de estimular na instituição o desenvolvimento de novas aplicações de Data Science e Big Data.

O modelo preditivo se mostrou facilmente replicável em outras instituições de ensino, independente de modalidade e nível de escolaridade, desde que possuam dados organizados acerca dos seus estudantes.

Para futuros estudos, recomenda-se a implementação de outras variáveis que não estavam disponíveis para a elaboração deste modelo, mas poderiam ter influência sobre a evasão, apontadas por autores na revisão da literatura.

## Referências

AMARAL, João Batista do. *Evasão discente no ensino superior: estudo de caso no Instituto Federal de Educação, Ciência e Tecnologia do Ceará (Campus Sobral)*. 2013. 48 f. Dissertação (Mestrado Profissional em Políticas Públicas e Gestão da Educação Superior) – Programa de Pós-Graduação em Superintendência de Recursos Humanos, Universidade Federal do Ceará, Fortaleza, 2013.

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Censo da educação superior*. Brasília, DF: Inep, 2003. Disponível em: <<http://portal.inep.gov.br>>. Acesso em 15/08/2018.

\_\_\_\_\_. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Censo da educação superior*. Brasília, DF: Inep, 2013. Disponível em: <<http://portal.inep.gov.br>>. Acesso em 15/08/2018.

\_\_\_\_\_. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Censo da educação superior*. Brasília, DF: Inep, 2016. Disponível em: <<http://portal.inep.gov.br>>. Acesso em 15/08/2018.

GIL, Antônio Carlos. *Como elaborar projetos de pesquisa*. 5 ed. São Paulo: Atlas, 2010.

PALÁCIO, Paula da Paz. *Políticas de acesso e permanência do estudante da Universidade Federal do Ceará (UFC)*. 2012. 123 f. Dissertação (Mestrado em Políticas Públicas e Gestão da Educação Superior) – Universidade Federal do Ceará, Fortaleza, 2012.

PEREIRA JÚNIOR, Edgar. *Compromisso com o graduar-se, com a instituição e com o curso: estrutura fatorial e relação com a evasão*. 2012. 414 f. Dissertação (Mestrado em Educação) – Universidade Estadual de Campinas, Campinas, 2012.

ROSA, Thatiane de Oliveira, LUZ, Hellen Souza. Conceitos Básicos de Algoritmos Genéticos: Teoria e Prática. In: XI ENCONTRO DE ESTUDANTES DE INFORMÁTICA DO TOCANTINS, 2009, Palmas. *Anais do XI Encontro de Estudantes de Informática do Tocantins*. Palmas: Centro Universitário Luterano de Palmas, 2009. p. 27-37.

SALES JUNIOR, Jaime Souza; BRASIL, Gutemberg Hespanha; CARNEIRO, Teresa Cristina Janes; CORASSA, Maria Auxiliadora de Carvalho. Fatores Associados à Evasão e Conclusão de Cursos de Graduação Presenciais na UFES. *Meta: avaliação*, Rio de Janeiro, v. 8, n. 24, p. 488-514, set. /dez. 2016.

TIBOLA, Jucélia Appio et al. Factors of students' continued attendance at a higher education institution: a confirmatory model. *Revista Alcance*, Itajaí, v. 19, n. 1, p. 83-100, jan./mar. 2012.

TINTO, Vincent. Dropout from higher education. A theoretical synthesis of recent research. *Review of Educational Research*, Washington, v. 45, p. 89-125, 1975.

TINTO, Vincent. *Leaving college: rethinking the causes and cures of student attrition*. Chicago: University of Chicago, 1987.

TINTO, Vincent. *Leaving college: rethinking the causes and cures of student attrition*. 2. ed. Chicago: University of Chicago, 1993.

TINTO, Vincent. Reconstructing the first year of college. *Planning for Higher Education, Michigan*, v. 25, n. 1, p. 1-6, 1996.

TINTO, Vincent. Classrooms as communities: exploring the educational character of student persistence. *Journal of Higher Education, New Jersey*, v. 68, n. 6, p. 599-623, 1997.

TINTO, Vincent. From theory to action: exploring the institutional conditions for student retention. In: SMART, Jonh C. (Ed.). *Higher Education: handbook of theory and research*, University of Chicago, p. 51-89. v. 25, 2010.

TINTO, Vincent. *Completing college: rethinking institutional action*. Chicago: The University of Chicago Press, 2012.

TONTINI, Géron; WALTER Silvana Anita. Pode-se identificar a propensão e reduzir a evasão de alunos? Ações estratégicas e resultados táticos para instituições de ensino superior. *Avaliação, Campinas*, v. 19, n. 1, p. 89-110, mar. 2014.