

# As Disciplinas de Introdução à Programação na USP: um Estudo Preliminar

Yorah Bosse, Marco Aurélio Gerosa

Instituto de Matemática e Estatística – Departamento de Ciência da Computação -  
Universidade de São Paulo (USP)

Rua do Matão, 1010 – CEP 05508-090 – São Paulo - SP - Brasil

{yorah,gerosa}@ime.usp.br

**Abstract.** *Introduction to Programming is part of computer and informatics curricula and other bachelor courses. There is little quantitative data about the results of this course. This paper quantifies the failure rate and the number of students who dropped out of the course at the University of São Paulo from 2010-1 to 2014-2. We performed analyses differentiating the computer science students from the others; first-time attendants; professors; and the programming language used. As a result, we found that students from other areas fail more frequently than computer and informatics related courses, and the C language is more used than others, but the adoption of Python is increasing in one of the USP schools.*

**Resumo.** *A disciplina de Introdução à Programação-IP faz parte da grade curricular dos cursos da área de computação e informática e de outras áreas. Há uma percepção de que é uma disciplina difícil, porém há poucos dados sobre seus resultados. Este artigo reporta um levantamento quantitativo do índice de reprovações e trancamentos nas disciplinas de IP na Universidade de São Paulo (USP) no período de 2010-1 a 2014-2. Foram realizadas análises diferenciando, por exemplo, alunos de computação dos demais, alunos que frequentam a disciplina pela primeira vez e geral, por professor e linguagem utilizada. Como resultado obtivemos que alunos que não são da área de computação reprovam mais e que a linguagem mais utilizada pelos professores de um dos institutos da USP (IME) é C, porém, perdendo espaço para a linguagem Python.*

## 1. Introdução

Uma disciplina que ensina os conceitos básicos de programação faz parte da grade curricular de diversos cursos das áreas de exatas, humanas e biológicas. Muitos são os nomes atribuídos a ela, como Introdução à Programação, Introdução à Ciência da Computação, Introdução à Programação de Computadores, Lógica de Programação, Algoritmos, entre outras. Neste artigo adotaremos o nome genérico de Introdução à Programação (IP). O foco principal da disciplina de IP é que o aluno aprenda fazer algoritmos, que, segundo Forbellone e Eberspächer [2005], são sequências de passos ordenados para atingir um objetivo específico. As disciplinas de IP adotam linguagens de programação como C, C++, Java, Python, VBA, entre outras, para treinar a aplicar os conceitos passados em sala de aula.

Para Santos e Costa [2006], os alunos sentem uma grande dificuldade em entender os conceitos de introdução à programação, levando a reprovações ou

trancamentos na disciplina e até desistência do curso. Bennedsen e Caspersen [2007] afirmam que cursos de computação são difíceis e apresentam alta taxa de insucesso. Poucas pesquisas mostram dados quantitativos do índice de reprovações e linguagens de programação utilizadas. Esta pesquisa busca levantar dados que gestores, professores e desenvolvedores de ferramentas educacionais possam usar para guiar suas ações.

Em pesquisa anterior (Bosse e Gerosa, 2015), identificamos que o índice de reprovação e trancamento em IP na USP é por volta de 29%. Neste artigo, expandimos o escopo dessa análise inicial, investigando este índice diferenciando, por exemplo, alunos que são da computação dos demais, alunos que frequentam a disciplina pela primeira vez e geral, por professor, linguagem utilizada, entre outras. A seguinte questão de pesquisa norteou esta pesquisa: quais são os resultados em termos de notas e reprovações/trancamentos das disciplinas de introdução à programação na USP?

Este artigo está organizado em 7 Seções, sendo esta a primeira. Na Seção 2, temos a revisão da literatura. Na Seção 3 é apresentada a metodologia do estudo. Na Seção 4 temos os resultados obtidos e suas análises, seguida da Seção 5, onde são apresentadas ameaças à validade e limitações do estudo. Na Seção 6 apresentamos a conclusão e agradecimentos, e por último, na Seção 7, as referências bibliográficas.

## 2. Revisão da Literatura

Entender e aplicar os conceitos estudados nas disciplinas de introdução à programação é fundamental para que o aluno consiga se desenvolver em programação de sistemas mais avançados. Helminen e Malmi [2010] afirmam que a programação é uma competência essencial nos cursos de Ciência da Computação. Porém, normalmente os estudantes apresentam uma grande dificuldade em assimilar os conteúdos abordados, ocasionando em reprovação ou desistência. Corney e colegas [2010] afirmam que em média 30% dos alunos da disciplina de IP da Universidade de Queensland Technology não conseguiram ser aprovados.

A linguagem de programação textual pode representar uma barreira inerente a compreensão do usuário (Lieberman, 2000). Em um estudo feito por Readt, Watson e Toleman (2002), foi visto que no período entre 1971 e 1997 o Pascal era a linguagem de programação mais utilizada na Austrália, chegando ao índice de 92% de universidades que a utilizavam. Este estudo realizou um Survey que identificou que, em 2002, dos 57 cursos participantes, de 37 universidades da Austrália, 23 utilizavam Java, 14 Visual Basic, 8 C++ e 4 usavam C.

A mudança da linguagem de programação pode ocorrer por decorrência das linguagens e paradigmas utilizados no comércio e indústria, de uma possível tentativa de melhorar o entendimento do aluno dos conceitos que envolvem a programação, entre outros fatores. Uma pesquisa realizada por Bennedsen e Caspersen [2007], com a participação de 63 instituições de diversos países, obtiveram um índice de 67% de aprovações, ou seja, 33% de reprovações, trancamentos e desistência caracterizando insucesso na tentativa de aprender a programar.

A dificuldade na disciplina de IP pode estar entre os motivos que fazem com que os cursos da área de computação estejam entre os maiores no índice de evasões. Uma pesquisa realizada por Silva Filho e colegas [2007] mostra que a média de evasão nos cursos superiores no Brasil, entre os anos 2000 e 2005, era de 22% e, os cursos da área de computação estavam entre os maiores, chegando à 36%.

### 3. Metodologia

Nesta pesquisa, realizamos um estudo quantitativo a partir dos dados do sistema acadêmico da universidade. Consideramos os alunos que cursaram uma das disciplinas de Introdução à Programação da USP, no período que compreende o ano 2010 ao 2014.

Na USP, são várias as disciplinas de IP existentes e que são ministradas para diversos cursos da área de computação e de outras áreas. Iniciamos esta pesquisa com a seleção das disciplinas que iriam participar da análise. As disciplinas foram selecionadas por meio de uma pesquisa no sistema acadêmico da universidade<sup>1</sup> com as seguintes palavras chaves: “programação”, “algoritmos” e “computação”. A Tabela 1 apresenta a relação da quantidade apresentada pelo sistema e a quantidade escolhida para a análise.

**Tabela 1: Totais de disciplinas apresentadas pelo sistema acadêmico para cada palavra de busca utilizada**

Palavra de Busca	Total de Disciplinas Apresentadas	Total de Disciplinas Escolhidas
Programação	66	12
Algoritmos	29	0
Computação	112	19
Total	207	31

Esta etapa foi realizada manualmente e o conteúdo programático de cada disciplina foi analisado. As disciplinas foram escolhidas ou descartadas conforme critério de inclusão e exclusão apresentado na Tabela 2.

**Tabela 2: Critérios de inclusão e exclusão das disciplinas.**

Critérios de Inclusão		Critérios de Exclusão	
CI1	Ter no seu programa pelo menos 4 dos seguintes assuntos: ⇒ Conceitos de linguagens de programação, por exemplo: tipos de variáveis ⇒ Estruturas de programação: seleção e repetição ⇒ Variáveis estruturadas ou tipos de dados compostos: vetores e matrizes ⇒ Modularização de programas: funções e/ou procedimentos ⇒ Utilização de alguma linguagem de programação	CE1	Tratar de uma comparação de paradigmas de programação como programação funcional, procedural, orientada à objetos, entre outras. Sendo feita programação para ver o funcionamento de cada uma.

Foram utilizadas 26 disciplinas das 31 selecionadas na primeira etapa. As de código SCC0222 e SCC0221 não possuíam dados por terem sido ativadas apenas em 2015, estando fora do período de análise e, as de códigos SSC0102, SSC0601 e SSC0801 por serem disciplinas de laboratório, ministradas no mesmo período da disciplina de IP que aborda todo conteúdo do critério, sendo assim, os mesmos alunos normalmente cursam as duas. Essas disciplinas de laboratório só são ministradas no

<sup>1</sup> Disponível no site: <https://uspdigital.usp.br/jupiterweb/jupDisciplinaBusca?tipo=D&codmnu=4526>

Campus da USP de São Carlos. Das matrículas realizadas nas disciplinas teóricas e as de laboratório, 96,3% das matrículas cursam as duas no mesmo ano e semestre. A média das notas das disciplinas teóricas, ou seja, a SSC0101, SSC0600 e SSC0800 é 5,7 e das de laboratório é 6,8. Quanto ao percentual de reprovações, as três disciplinas teóricas apresentaram juntas 30% e as três de laboratório, 17%. Para uma melhor compreensão da correlação entre as notas das disciplinas teóricas e de laboratório, foi feita uma análise entre as disciplinas SSC0600 e SSC0601, SSC0101 e SSC0102, SSC0800 e SSC0801. Os coeficientes obtidos foram 0,633, 0,670 e 0,694, respectivamente, o que indica uma correlação moderada entre as disciplinas teóricas e de laboratório das duplas estudadas.

Os dados para análise foram disponibilizados pela USP, após autorização da Pró-Reitoria de Graduação. Os campos que identificam os alunos foram alterados para números aleatórios, porém consistentes, para que pudéssemos identificar unicamente o aluno dentro da análise, mantendo a confidencialidade dos dados.

Foram disponibilizadas tabelas de dados com informações referentes aos alunos que cursaram as disciplinas abordadas nesta pesquisa, as médias e resultados alcançados, o ano e semestre em que ela foi ministrada, a turma e o docente que a conduziu. Além disso, foi fornecido os históricos dos alunos que se matricularam em alguma dessas disciplinas no período de 2010 e 2014. Os resultados obtidos pelos alunos são apresentados levando-se em consideração notas e frequências. Na USP, a média de aprovação é cinco, em uma faixa de zero a dez.

#### 4. Resultados

Os dados utilizados para esta análise são de 5 anos (2010-1 a 2014-2), sendo consideradas 18.784 de 19.879 matrículas existentes em todas as disciplinas juntas. As 1.095 matrículas desconsideradas são as que correspondem a situação “Não cadastrado”, não tendo como identificar se nesta matrícula o aluno foi ou não aprovado. Com esses dados, algumas análises foram feitas para responder questões que esta pesquisa se propôs a responder.

**Quantos alunos cursam IP por ano?** Em média são 3.756 matrículas anuais. O total de matrículas e média de alunos por turma pode ser visto na Tabela 3.

**Tabela 3: Total de turmas, matrículas e média de alunos por turma, por ano.**

Ano	Total de Turmas	Total Matrículas por Ano	Média Matrículas por Turma
2010	73	4073	55,8
2011	66	3787	57,4
2012	63	3679	58,4
2013	59	3643	61,7
2014	60	3602	60,0
	321	18784	58,7

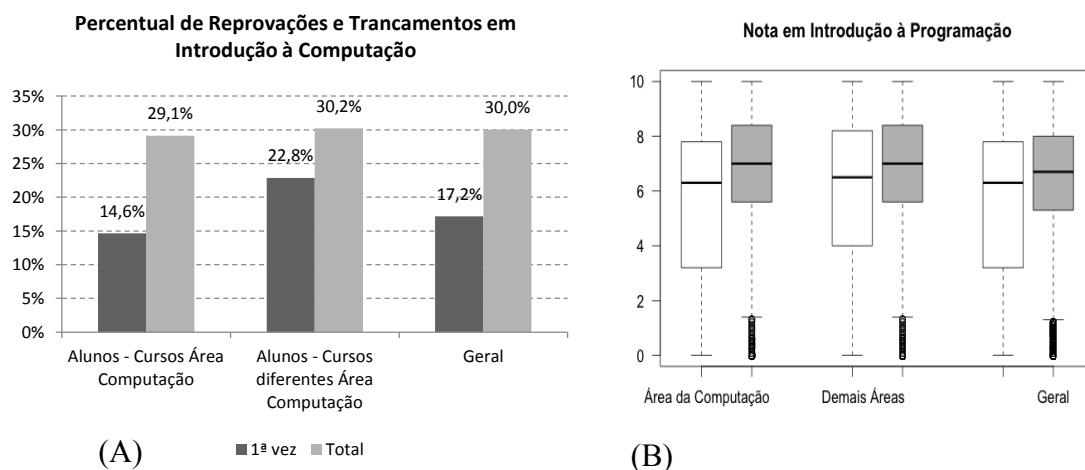
**Qual o percentual de reprovações e trancamentos geral e por ano?** Considerando cada matrícula com o mesmo peso, obtivemos que 70% resultaram em aprovação, 27% em reprovação por nota e/ou frequência e 3% em trancamentos. Quando consideramos ano a ano, temos como resultado um percentual de 31% de reprovações e trancamentos juntos em 2010, o maior índice de 32% em 2011 e 2014,

30% em 2012 e o menor índice em 2013, com 26%. Estatisticamente, não é possível afirmar que há diferença significativa entre os anos analisados quanto às médias de reprovação, segundo o teste não-paramétrico de Kruskal-Wallis.

***Qual o percentual de reprovações e trancamentos por disciplina?***

Considerando cada disciplina isoladamente obtivemos um resultado bastante variado, no qual a de código 5952010, disciplina oferecida ao curso de Bacharelado de Matemática e Negócios ficou com um percentual de reprovações e trancamentos de 62,18% (n=3). Em contrapartida, a ZAB0264, disciplina oferecida ao curso de Engenharia de Alimentos, obteve 5,91% (n=10) de reprovações/trancamentos. A mediana entre as reprovações e trancamentos é de 27,42%.

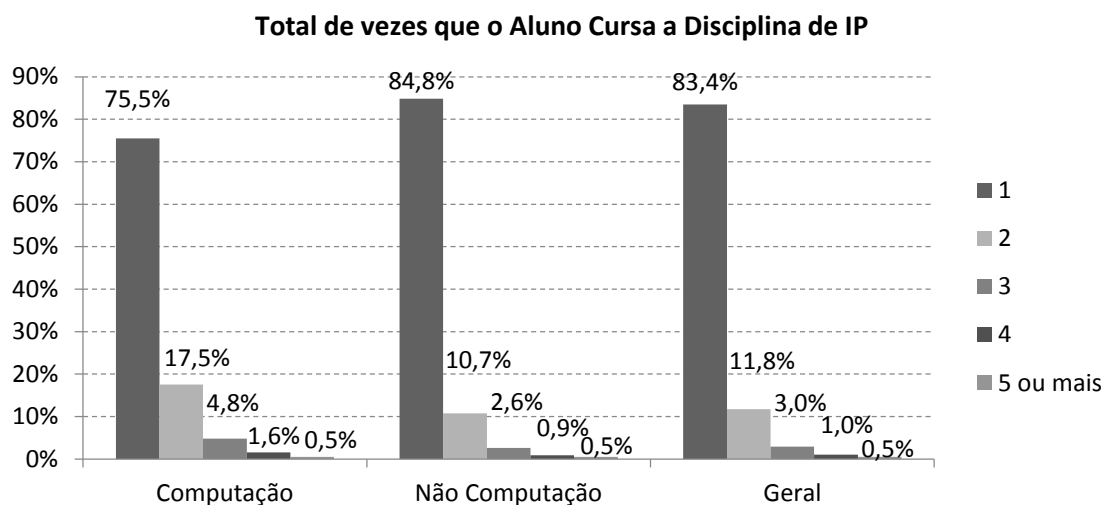
***Qual a diferença entre o percentual de reprovações de alunos dos cursos da área de computação e os demais?*** Observando a Figura 1A podemos verificar que os alunos dos cursos da área de computação reprovam menos que os demais, tendo uma diferença estatisticamente significativa de 1,1% no total, considerando todas matrículas. No caso de alunos cursando pela primeira vez, essa diferença percentual é maior (8,2%). Para uma melhor compreensão da distribuição das notas destes alunos, foram feitos boxplots (Figura 1B). As médias das notas para os alunos que cursaram a disciplina pela primeira vez é de 6,6, 6,1 e 6,2 para as Áreas de Computação, Demais Áreas e Geral, respectivamente, enquanto que as médias das notas no total são 5,8, 5,4 e 5,5, também respectivamente. A partir desse gráfico, pode-se observar que as notas dos alunos que cursaram a matéria pela primeira vez e foram reprovados distribuíram de forma semelhante para as três áreas analisadas, uma mediana igual a 7, o que indica uma maior dispersão das notas nos primeiros 50% dos dados. As notas totais também possuem uma distribuição semelhante, uma mediana de 6,3 e 50% dos dados acima de 6,3.



**Figura 1. (A) Percentual de Reprovações dos alunos que cursaram a disciplina pela primeira vez e geral, separados por cursos de Computação, dos demais cursos e visão completa da situação. (B) Média das notas dos alunos na mesma separação feita para a Figura 1A.**

***Quantas vezes os alunos cursam a disciplina para serem aprovados?*** 24,5% dos alunos dos cursos da área de computação cursam duas ou mais vezes a disciplina para obter aprovação. Já para os cursos de outras áreas, esse percentual cai para 14,8% (Figura 2). É importante observar que para essa análise foram consideradas apenas as matrículas dos alunos que já obtiveram aprovação na disciplina, totalizando 12.722

matrículas, 10.821 dos alunos de cursos de outras áreas e 1.901 de alunos dos cursos da área de computação.



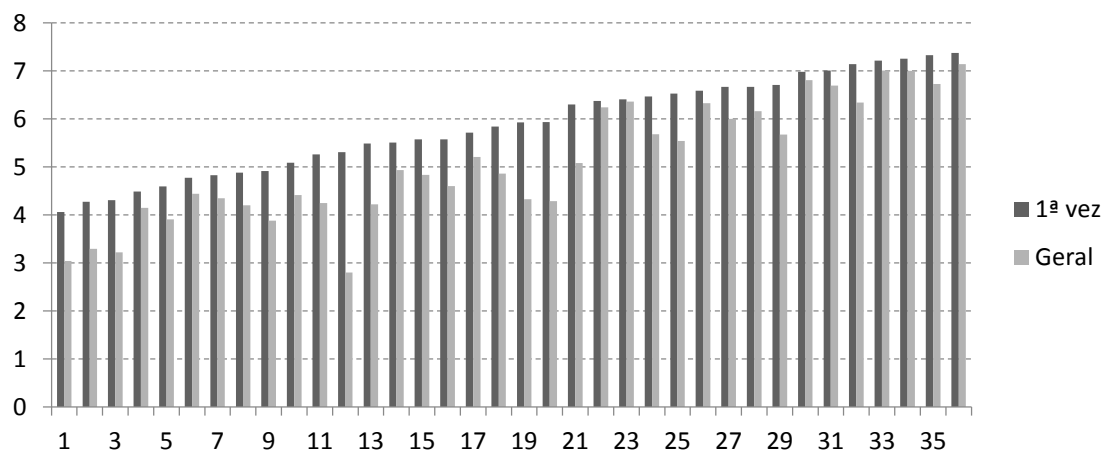
**Figura 2. Percentual de vezes que os alunos de cursos de computação e demais áreas necessitam cursar a disciplina de IP para obterem aprovação na disciplina.**

Há diversos fatores que podem influenciar no aprendizado dos alunos nas disciplinas ministradas em uma faculdade. Um deles é o conhecimento prévio do aluno em conteúdos considerados básicos para o desenvolvimento daquela habilidade. Outro ponto a se considerar são os professores que ministram a disciplina, que possuem diferentes níveis de cobrança, estratégias de ensino-aprendizagem, tecnologias adotadas, conteúdo didático, etc. Outro ponto é se a disciplina é da área afim ao curso. Mais um fator a ser considerado neste contexto é a linguagem de programação adotada.

Para analisar professor e linguagem fizemos uma pesquisa com os professores vinculados ao Instituto de Matemática e Estatística – Departamento de Ciência da Computação da USP de São Paulo, Capital. Foram feitas três perguntas por e-mail para 39 professores, dos quais obtivemos 36 retornos, ou seja, uma participação de 92%. Juntando as respostas obtidas e os dados que já possuíamos, fizemos análises complementares.

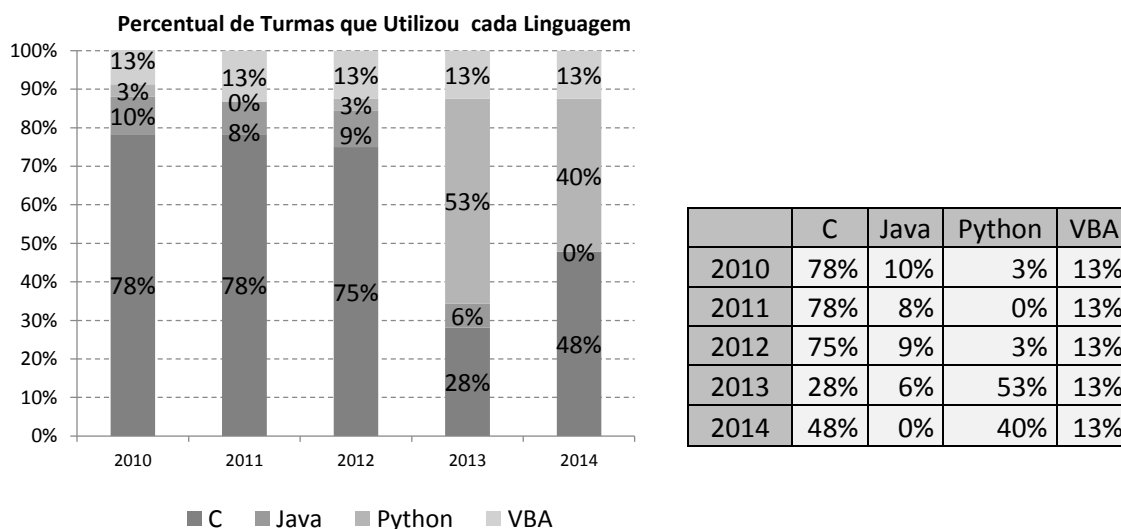
***Há diferença entre as médias das notas obtidas nas turmas dos diversos professores que participaram?*** A análise dos dados mostrou que as médias das notas por professor variaram de 3,0 a 7,1 no geral, sendo que 20 dos 36 professores tiveram nota média nas turmas abaixo de 5,0, que é a média de aprovação da USP, conforme Figura 3.

**Médias das Notas nas Disciplinas de IP, por Professor Ministrante do IME**



**Figura 3. Médias das notas obtidas pelos alunos dos professores do IME-USP-SP, nas disciplinas de IP.**

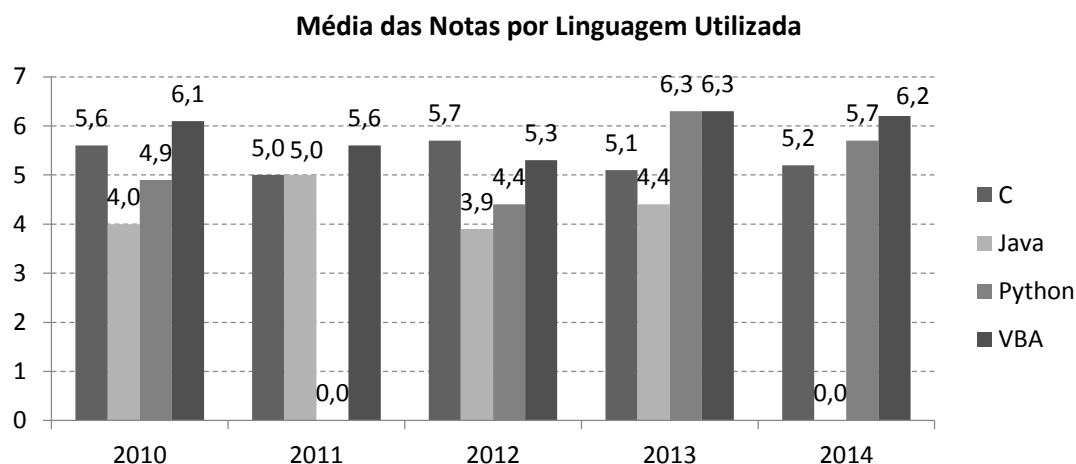
*Quais as linguagens de programação mais utilizadas?* Na Figura 4, podemos observar que ao longo dos cinco anos a linguagem de programação C foi a mais utilizada, com exceção do ano 2013, no qual Python sobressaiu. Python ficou como a segunda linguagem mais utilizada, seguida de VBA e por último, Java. Cabe ressaltar que em algumas turmas o professor tem liberdade de escolher a linguagem, enquanto em outras é padronizada para o curso (exemplo, VBA para as 4 turmas anuais na Faculdade de Economia, Administração e Contabilidade).



**Figura 4. Percentuais de utilização das quatro linguagens citadas pelos professores do IME-USP-SP, nas disciplinas de IP entre 2010 e 2014.**

*Ao longo desses anos, as médias foram muito diferentes de uma linguagem para outra?* A linguagem com maior média foi o VBA no Excel, sendo esta uma linguagem utilizada exclusivamente para ensinar programação aplicada para alunos de ciências humanas. Em segundo vem o Python e C, empatados e por último Java, conforme Figura 5. Cabe ressaltar que como algumas linguagens estão atreladas ao

curso em que a turma é ministrada e estão associadas às preferências dos professores, há muitos fatores que podem gerar as diferenças de média.



**Figura 5. Média das notas dos alunos que cursaram a disciplina de IP na USP, entre 2010 e 2014, com um dos professores do IME-USP-SP.**

*Quais são as IDEs e ferramentas utilizadas por professores do IME-USP?* Os professores do IME-USP utilizam Code::Blocks, Eclipse, OpenGL e Dev-C++ para ensinar com C; DrJava, BlueJ e Greenfoot para ensinar com Java, Excel quando utilizam a linguagem VBA e IDLE para Python.

## 5. Limitações e Ameaças à Validade

Alguns fatores podem levar a possibilidade de erros nos dados divulgados, como a possibilidade de erro na extração e compilação dos dados do sistema. Para evitar isso, foram selecionados vários dados para checagem manual, comparando com outras fontes (planilha de professor, relatório do sistema, etc.).

Um fator relevante é o das matrículas que possuem situação final “Não Cadastrado”. Essa situação refere-se ao fato da disciplina estar vinculada a uma outra e, nesta estar a situação final. Esses casos foram desconsiderados neste estudo e isto poderia gerar alguma diferença quanto ao percentual de aprovações e reprovações/trancamentos.

## 6. Conclusão

A pesquisa mostrou que nos cinco anos analisados, em 18.784 matrículas, 30% tinham como resultado reprovações ou trancamentos, mantendo-se razoavelmente constantes ao longo dos anos. Para cada disciplina, esse índice teve uma variação bastante acentuada, chegando a uma diferença de 56,27% entre a que mais e menos reprovou.

Os resultados apresentados mostram ainda que a linguagem C, com percentual médio de 62%, é a linguagem mais utilizada pelos professores do IME-USP-SP. Porém, ela vem perdendo espaço para a linguagem Python. Em 2014 a linguagem C era utilizada em apenas 48% das turmas ministradas por esse grupo de professores, enquanto que a linguagem Python subiu de 3% em 2010 para 40% em 2014.

Analisando-se as médias de nota por linguagem utilizada nestes cinco anos em que a pesquisa se encaixa, as linguagens C e Python empatam no segundo lugar em



termos de nota média, enquanto na linguagem VBA, os alunos obtiveram melhor nota, com média de 5,9 (cabe lembrar que as linguagens foram adotadas em diferentes cursos e por diferentes professores). Turmas ministradas com Java obtiveram média 4,3. Esses valores podem ter várias influências, como o nível de cobrança do professor que a leciona, visto que as médias por professor variam entre 3,0 e 7,1.

Outro dado verificado com esta pesquisa é que existe um índice maior de reprovações e trancamentos por parte dos alunos que frequentam cursos que não são da área de computação, chegando a 30,3% em comparação a 25,1% dos alunos que são da área.

Sugestão de trabalho futuro é fazer uma investigação dos motivos que levam a existir uma variação tão elevada de reprovações entre as disciplinas de IP, buscando descobrir se é a metodologia aplicada ou o nível de cobrança ou se alguma ferramenta de ensino aprendizagem é utilizada. Podem ser feitas também análises das notas obtidas nas disciplinas em comparação com a nota de corte para entrar nos cursos.

### **Agradecimentos**

Nossos sinceros agradecimentos aos professores das disciplinas de Introdução à Programação, vinculados ao Instituto de Matemática e Estatística - IME, do departamento de Ciência da Computação, que responderam ao nosso e-mail, pelo rápido retorno dado à pesquisa realizada.

### **Referências**

- Bennedsen, J. e Caspersen, M. (2007). Failure Rates in Introductory Programming. The SIGCSE Bulletin, Vol.39. N.2. June.
- Bosse, Y e Gerosa, M.A. (2015). Reprovações e Trancamentos nas Disciplinas de Introdução à Programação da Universidade de São Paulo: Um Estudo Preliminar. XXIII WEI – Workshop sobre Educação em Informática. Recife, Julho.
- Corney, M., Teague, D., Thomas, R. (2010). Engaging students in programming. Em: ACE '10 Proceedings of the Twelfth Australasian Conference on Computing Education - Volume 103, pp. 63-72. Austrália.
- Creswell, J. (2010). Projeto de Pesquisa: métodos qualitativos, quantitativos e misto. 3ª. edição. Porto Alegre : Artmed.
- Forbellone, A. e Eberspächer, H. (2005). Lógica de Programação : a construção de algoritmos e estruturas de dados. 3. ed. São Paulo : Pearson Prentice Hall.
- Helminen, J. e Malmi, L. (2010). Jype - a program visualization and programming exercise tool for python. Em: SOFTVIS '10 Proceedings of the 5th international symposium on Software visualization. pp.153-162. USA.
- Lieberman, H. (2000). Your Wish is My Command: Programming by Example. Media Laboratory. Massachusetts Institute of Technology. Publicado em Morgan Kaufmann na Elsevier. Cambridge, MA, USA. Disponível no site: <http://web.media.mit.edu/~lieber/Your-Wish/>
- Readt, M. de, Watson, R. e Toleman, M. (2002). Language Trends in Introductory Programming Courses. InSITE – “Where Parallels Intersect” – Informing Science – June 2002.
- Santos, R. e Costa, H. (2006). Análise de Metodologias e Ambientes de Ensino para Algoritmos, Estruturas de Dados e Programação aos Iniciantes em Computação e Informática. In: INFOCOMP, Volume 5, nº.1, ISSN 1807-4545.