



Análise das Interações Curriculares em um Curso de Ciência da Computação: buscando subsídios para aprimoramento curricular

Title: Analysis of the Curricular Interactions in a Computer Science Course: seeking grants for curricular improvement

Fábio Gerab

Centro Universitário da FEI – FEI
prifgerab@fei.edu.br

Irani Ferreira da Silva Gerab

Universidade Federal de São Paulo - UNIFESP
irani.silva@unifesp.br

Ivander Augusto Morais Bueno

Centro Universitário da FEI – FEI
ivanderaugusto@hotmail.com

Resumo *O processo de mudança curricular deve estar alicerçado em evidências institucionais de mudanças. Neste sentido, o presente trabalho objetiva caracterizar e discutir as interações curriculares em um curso de Ciência da Computação a partir do desempenho dos alunos egressos. Os dados de desempenho foram submetidos à Análise de Componentes Principais e à Análise por Agrupamento Hierárquico para identificar eventuais correlações entre: o rendimento nas distintas disciplinas, a influência da classificação no exame de ingresso e o tempo necessário para a conclusão do curso. Observou-se a associação de disciplinas em nove grupos, com destaque para o agrupamento rotulado de “Básicas”, que interfere na duração do curso. Notou-se que o desempenho no vestibular perde sua importância no decorrer do curso. Compreender estas relações pode contribuir para a melhoria do curso de Ciência da Computação e favorecer o bom aproveitamento do aluno.*

Palavras-Chave: *Educação Superior; Currículo; Ciência da Computação; Estatística Multivariada*

Abstract *The curriculum change process must be based on evidence of institutional changes. In this sense, the present study aims to characterize and discuss curricular interactions in a Computer Science course based on the performance of graduating students. The performance data were subjected to Principal Components Analysis and Hierarchical Cluster Analysis to identify possible correlations between: performances in different disciplines, the influence of classification in the admission test and the time required to conclude the course. It was detected the association of disciplines in nine groups, with emphasis on the group labeled "basic", which interferes with the duration of the course. It was noted that performance in the admission test loses importance during of the course. Understanding these relationships can contribute to the improvement of the course of Computer Science and promote the good student formation.*

Keywords: *Higher Education; Curriculum; Computer science; Multivariate Statistics*

1. Objetivo

Entender as interações curriculares ao longo de curso de Ciência da Computação através da análise do desempenho dos seus alunos egressos aplicando uma abordagem estatística multivariada.

2. Introdução

O curso de Bacharelado em Ciência da Computação em questão foi aprovado em janeiro de 1999 e teve seu início em março desse mesmo ano. A nota obtida nas avaliações do Ministério da Educação (MEC) CPC foi 4 para os anos de 2005 e 2008. Após a conclusão deste curso de Ciência da Computação espera-se que o profissional por ele formado tenha alta capacidade de projeto e desenvolvimento de sistemas computacionais, seja para cargos como Analistas de Sistemas, gerentes de projetos de sistemas ou como gestores em tecnologia da informação. O curso é ministrado no período noturno, com a duração de quatro anos, distribuídos em oito ciclos, tendo carga semanal variando entre 20 e 24 aulas de 50 minutos.

É sabido que o desenvolvimento científico e tecnológico tem provocado mudanças nas necessidades de formação profissional. A crescente complexidade tecnológica exige o desenvolvimento de competências como capacidade de análise e síntese, habilidade de resolução de problemas, rapidez na atuação, comunicação, habilidade de trabalhar em grupo, versatilidade no uso da linguagem, habilidade para enfrentar desafios, flexibilidade frente a mudanças e capacidade de auto aprendizagem [1]. A proposta curricular no ensino superior deve mudar de uma formação especializada para uma formação generalista.

É importante considerar que o modelo de currículo a ser implementado em uma instituição reflete, de certa forma, o profissional que se pretende formar [2]

Como muitos cursos de graduação, este apresenta uma organização curricular de base disciplinar. Esta estrutura parece, em muitos casos, mostrar-se cômoda e útil, ao conceder uma ordem lógica e linear aos conteúdos curriculares. Entretanto tem sido descrita como facilitadora da fragmentação do conhecimento e estimuladora da especialização de funções [3].

A superação deste modelo que resulte em um processo de mudança deve estar alicerçada por evidências sistemáticas identificadas no cenário institucional. Neste sentido, identificam-se as seguintes questões quanto ao curso de Ciência da Computação:

- O desempenho do aluno em uma disciplina interfere no seu resultado em outras?
- Existe correlação entre a nota obtida em determinadas disciplinas e o tempo necessário para a conclusão do curso?
- A classificação do estudante no exame vestibular interfere com o seu desempenho no curso?

Na busca de repostas, o presente trabalho tem por objetivo caracterizar e discutir as interações curriculares em um curso de Ciência da Computação a partir do desempenho dos estudantes egressos pois, embora exista a influência das competências específicas de cada aluno e do real conhecimento por ele adquirido em uma dada disciplina no seu resultado nas etapas posteriores do curso, ainda é possível buscar padrões de dependência no desempenho em grupos de disciplinas através de uma análise estatística apropriada.

Assim, o escopo deste trabalho é o ensino de computação e, por consequência o seu impacto nas mais diversas áreas e, especificamente, na educação. Para tanto, desenvolveu-se uma metodologia de análise capaz de aplicar ferramentas oriundas das análises multivariadas aos dados de desenho acadêmico dos alunos.

2. Metodologia

Analisou-se o banco de dados referente ao desempenho nas diferentes disciplinas para os 370 alunos concluintes, egressos das dez primeiras turmas do curso. Foram excluídos da análise os estudantes que cursaram apenas parte do curso na instituição ou que pediram equivalência em disciplinas, totalizando dados de 335 alunos que cursaram todas as 45 disciplinas do curso de Ciência da Computação. Estas se encontram codificadas como mostrado na tabela 1, onde são mostradas as distintas disciplinas do curso ministradas em cada um de seus oito ciclos, associadas à sigla doravante utilizada para representá-las nas próximas etapas deste trabalho.

Na análise foram utilizadas as notas efetivas de aprovação, isto é, a nota de aprovação obtida na disciplina dividida pelo número de vezes em que ela foi cursada. Este procedimento permite considerar a real dificuldade do aluno em obter a sua aprovação em cada uma das disciplinas cursadas, e não somente a sua nota. A análise incluiu também a colocação obtida pelo estudante no exame de ingresso e a duração total, em semestres, até a conclusão do curso pelo aluno. O fato de estarmos trabalhando com graduados dispensou o tratamento dos valores faltantes, uma vez que existem notas de aprovação para todas as disciplinas consideradas.

Como ferramenta de análise, foi utilizada a Análise de Componentes Principais (ACP) conjuntamente à Análise por Agrupamento Hierárquico (AAH) [4]. Estes métodos têm se mostrado eficazes na identificação de relações entre grupos de variáveis com alto grau de complexidade e vem sendo aplicados em distintas áreas do conhecimento.

Quando duas ou mais variáveis envolvidas na análise estatística não são completamente independentes, elas muitas vezes podem ser relacionadas através da criação

de uma única nova variável a partir das variáveis originais. Esse procedimento reduz a complexidade do problema em estudo, sem acarretar em perda relevante de informação e evidencia as relações entre essas variáveis. O propósito da ACP é reduzir a dimensão do espaço de variáveis que caracteriza a matriz de dados em estudo a partir da matriz de correlação linear entre as variáveis, para um número menor de componentes principais, geradas a partir das variáveis iniciais, com perda mínima de informações.

Disciplinas do curso de Ciência da Computação					
	Sigla	Nome		Sigla	Nome
1º Ciclo	Calc1	Cálculo Diferencial e Integral I	2º Ciclo	Calc2	Cálculo Diferencial e Integral II
	GA1	Geometria Analítica I		GA2	Geometria Analítica II
	Fis1	Física I		Fis2	Física II
	Alg1	Computação e Desenvolvimento de Algoritmos I		Alg2	Computação e Desenvolvimento de Algoritmos II
	CExp1	Comunicação e Expressão I		CExp2	Comunicação e Expressão II
3º Ciclo	Calc3	Cálculo Diferencial e Integral III	4º Ciclo	Calc4	Cálculo Diferencial e Integral IV
	AICn1	Álgebra Linear e Cálculo Numérico I		AICn2	Álgebra Linear e Cálculo Numérico II
	Tdig1	Técnicas Digitais I		Tdig2	Técnicas Digitais II
	Tprg1	Linguagens e Técnicas de Programação I		Tprg2	Linguagens e Técnicas de Programação II
	SoAd1	Sociologia Aplicada à Administração I		SoAd2	Sociologia Aplicada à Administração II
5º Ciclo	ProEst	Probabilidade e Estatística	6º Ciclo	Compil	Compiladores
	AlgCom	Algoritmos Computacionais		PesqOp	Pesquisa Operacional
	SisOp1	Sistemas Operacionais I		SisOp2	Sistemas Operacionais II
	EngSoft1	Engenharia de Software I		EngSoft2	Engenharia de Software II
	Redes1	Redes de Computadores I		Redes2	Redes de Computadores II
	ArqCom1	Arquitetura de Computadores I		ArqCom2	Arquitetura de Computadores II
7º Ciclo	SimSi1	Simulação de Sistemas I	8º Ciclo	SimSi2	Simulação de Sistemas II
	Bdados1	Banco de Dados I		Bdados2	Banco de Dados II
	Cgraf	Computação Gráfica		HiMid	Hipermídia
	IntArt	Inteligência Artificial		AdInf	Administração da Tecnologia da Informação
	AnaSis	Análise de Sistemas		AdSist	Administração para Analistas de Sistemas
	Pform1	Projeto de Formatura I		Pform2	Projeto de Formatura II

Tabela 1: Disciplinas do curso de Ciência da Computação e suas abreviações

A ACP baseia-se em conceitos matemáticos simples e é um caso particular da Análise de Fatores (AF) sendo uma técnica multivariada de interdependência em que todas as variáveis são simultaneamente consideradas, cada uma relacionada com as demais, a fim de estudar as inter-relações existentes entre elas.

Corrar e colaboradores [5] argumentam que o objetivo da AF é encontrar uma maneira de condensar a informação contida nas variáveis originais em um conjunto menor de variáveis estatísticas (fatores) com uma perda mínima de informação, isto é, sumarizar os dados por meio da combinação entre as variáveis explicando a relação entre elas.

A principal diferença entre a ACP e a AF é que a Análise de Fatores calcula, além dos fatores comuns, os fatores específicos, ou seja, fatores comuns explicam a variabilidade comum a toda a base de dados, e os específicos explicam a parte da variabilidade de cada variável medida que não é explicável pela sua inter-correlação com as outras variáveis medidas. Uma variabilidade específica é devida provavelmente a uma origem independente de todas as outras variáveis. A ACP, por sua vez, tenta através de fatores comuns (no caso chamados componentes) explicar o máximo da variabilidade da base de dados. Na ACP a existência de uma variável independente das demais variáveis define naturalmente uma nova componente principal.

A correlação entre notas dos alunos nas distintas disciplinas cursadas pode ser resultado de distintos fatores, tais como: tempo de dedicação ao estudo, comprometimento com o curso, conhecimentos previamente adquiridos, habilidade de raciocínio abstrato, capacidade analítica, etc. A análise de componentes principais permite identificar os fatores não diretamente observáveis a partir das variáveis conhecidas (notas). Assim, o estudo das notas dos alunos egressos no curso de Ciências da Computação, é um caso típico onde a aplicação da ACP é útil.

Na realização da Análise de Componentes Principais é usual normalizar as variáveis para média zero e variância um, isso porque o interesse se concentra na variabilidade. Além disso, essa normalização permite a realização da análise com variáveis de dimensões diferentes.

Como fora salientado, a ACP substitui um conjunto de variáveis correlacionadas por um número menor de variáveis independentes. Hair e colaboradores [4] sugerem um tamanho mínimo na base de dados para que os resultados sejam estatisticamente significativos recomendando que o número de amostras seja, no mínimo, cinco vezes maior que o número de variáveis e, além disso, nunca inferior a 50.

A ACP pode modelar as notas obtidas pelos alunos egressos do curso de Ciência da Computação partindo de

uma base de dados definida pela matriz Z , composta por n estudantes (amostras) que tiveram atribuição de notas (normalizadas) em m distintas disciplinas (variáveis), tal que:

$$Z_{ij} = \sum_{k=1}^p W_{ik} \cdot P_{kj} \quad (1)$$

em notação matricial a expressão (1) pode ser escrita da seguinte forma:

$$Z = W \cdot P \quad (2)$$

onde: Z é a matriz ($m \times n$) dos dados originais normalizados. Essa matriz é composta de n colunas (correspondentes às amostras) e m linhas (correspondentes às variáveis medidas).

P corresponde ao valor das novas variáveis (as componentes principais) sendo denominada matriz dos escores das componentes. A matriz P tem p linhas (as componentes principais) por n colunas (as amostras). Ou seja, o termo P_{kj} representa o valor da k -ésima componente na j -ésima amostra. P tem p linhas, com $p < m$, pois um dos objetivos da ACP é reduzir o número de variáveis. O procedimento usado para a definição do valor de p (número de componentes principais a serem retidas na análise) será tratado mais adiante.

W é a matriz ($m \times p$) correspondente ao peso das antigas variáveis (as notas obtidas, normalizadas para média zero e variância 1) nas novas (as componentes principais). A matriz W é chamada de matriz das cargas fatoriais.

A decomposição da matriz de dados no produto de duas matrizes de cofatores é feita através da diagonalização da matriz de correlação R [6,7] dado que o objetivo da análise é obter novas variáveis ortogonais. A diagonalização é obtida pela solução de um problema de autovalores e autovetores.

$$R = ZZ' \quad (3)$$

A solução exata para esse problema é obtida quando se tem o número de componentes principais obtidas iguais ao número de variáveis iniciais. Na solução exata são explicados 100% da variabilidade original dos dados.

Todavia, supõe-se que as variáveis iniciais tragam informação redundante pois várias tem origens comuns, isto é, suas variabilidades são similares. A variabilidade associada às últimas componentes extraídas representa o “ruído” presente na base de dados. Assim, é possível truncar o modelo em p componentes principais, com $p < m$, explicando uma fração significativa da variabilidade dos dados.

A escolha do número de componentes é um ponto crucial na elaboração da análise. Isso porque as componentes serão responsáveis pela substituição do conjunto de variáveis, ou seja, sumarização dos dados, dessa forma é válido expor que uma vez que se optou trabalhar com as componentes retidas, não se tem mais a explicação de 100% da variância observada.

A escolha do número de componentes determinará a capacidade de extrapolação das inferências que serão realizadas pela análise destas componentes. Corrar e colaboradores [5] afirmam que ao limitar demais o número de componentes retidas, pode estar analisando um conjunto de componentes que explicam uma pequena parcela da variância total dos dados, o que de fato prejudicaria na análise do pesquisador. De maneira análoga, ao optar-se trabalhar com um número demasiado de componentes (grande quantidade de informação), pode prejudicar nas inferências do pesquisador. Existem diversas técnicas para a definição do número de componentes, ou seja, para o truncamento da matriz P .

O método apresentado por White e colaboradores [8] consiste em determinar o número de componentes retidas, calculando-se o desvio padrão associado a cada autovalor. Quando o desvio padrão associado aos autovalores é da mesma ordem do espaçamento entre os autovalores de duas componentes subsequentes, considera-se que não mais existe precisão estatística para a distinção entre as componentes, delimitando-se, assim, o número de componentes retidas [9]. Entretanto, existem divergências quanto a melhor forma de se quantificar o desvio padrão associado a cada autovalor, pois a aplicação do método de White pode muitas vezes levar a retenção de um número excessivo de fatores [10].

Neste trabalho utilizou-se o critério da Raiz Latente, retendo somente componentes principais associadas a autovalores maiores que um. Componentes com autovalores menores que um não devem ser retidas, uma vez componentes associadas a autovalores menores que a unidade carregam, em princípio, menos informação sobre a totalidade da base de dados que uma única variável. Echalar [11] argumenta também, que este critério, embora bastante utilizado, deve servir como um limite inferior para o número de componentes extraídas.

Tendo em vista estas divergências apresentadas aci-

ma, optou-se, neste trabalho, por reter as componentes com autovalores iguais ou superiores a um, uma vez que a avaliação correta dos desvios padrão dos autovalores é tema bastante controverso.

As componentes retidas, muitas vezes não são de fácil compreensão. Gerab [12] argumenta que uma amostra individual pode ser vista como um ponto num espaço m -dimensional, onde m é o número de variáveis iniciais. A ACP corresponde à obtenção de uma nova base ortogonal neste espaço. O truncamento da matriz P representa a redução dimensional deste espaço, resultando em uma base ortogonal com dimensão menor que a da base original. Esta redução, realizada eliminando-se dimensões da base ortogonal completa, associadas às componentes com baixa capacidade de explicação do comportamento dos dados, pode dificultar a interpretação dos resultados.

Entretanto, a interpretação do significado das componentes obtidas é facilitada quando efetua-se uma rotação apropriada dos vetores desta nova base, compostos pelas componentes retidas [13]. Existem várias técnicas de rotações possíveis, a citar, Varimax, Quartimax, Equimax, Direct Oblimin e Promax. Nesse trabalho, utilizou-se o critério de “estrutura simples”, ou método Varimax, que é um tipo de rotação ortogonal, e tem como característica o fato de minimizar a ocorrência de uma variável possuir altas cargas fatoriais para diferentes fatores, permitindo, assim, que uma variável seja facilmente identificada com um único fator [5].

Dessa maneira, uma rotação ortogonal das componentes retidas é efetuada de modo a facilitar no entendimento do significado das componentes obtidas, preservando a independência estatística das componentes retidas [13].

A comunalidade h_i^2 definida abaixo, associada à variável i , corresponde à percentagem da variabilidade explicada para esta variável.

$$h_i^2 = \sum W_{ik}^2 \quad (4)$$

Onde W_{ik} é a carga fatorial (CF) dessa variável na componente k . O método Varimax tem também por característica preservar o valor das comunalidades encontradas durante a rotação.

A escolha das CF estatisticamente significativas a serem utilizadas na interpretação das componentes principais não tem um critério de consenso na literatura. Além disso a significância das CF depende fortemente do tamanho da amostra. Hair [4] aponta que, para amostras de tamanho 50, somente CF superiores a 0,75 são significa-

tivas (ao nível de 5%), enquanto que CF superiores a 0,3 já seriam significativas para amostras de tamanho 350. Entretanto, existem discrepâncias sobre este critério, sendo aceitos valores de CF compreendidos entre 0,1 e 0,5 [12].

Um outro critério considera significativas apenas as CF numericamente maiores que três vezes o desvio padrão associado à CF, Wi , para a variável Zi . Este desvio padrão foi estimado por Heidam [14] como:

$$\sigma_{w_i} \approx \left[(1 - h_i^2) / (n - p) \right]^{1/2} \quad (5)$$

Este valor é o mesmo antes e depois da rotação Varimax, isso porque n (número de amostras) e p (número de componentes retidas) são constantes e esta rotação preserva as comunalidades. Salientadas a diversidade de critérios e suas implicações, neste trabalho optou-se trabalhar com CF maiores que 0,3, desde que atendendo ao critério de Heidam [14].

A ACP pode ser complementada pela Análise por Agrupamento Hierárquico (AAH), que agrupa um conjunto de variáveis, mensuradas em um grande número de pontos amostrais, de acordo com suas similaridades, utilizando conceitos geométricos de distância em um espaço multidimensional. Variáveis próximas neste espaço apresentam, portanto, comportamento semelhante. Neste trabalho utilizou-se na AAH a distância euclidiana quadrática e o método da centroide para o agrupamento [4,5].

As variáveis (notas efetivas de aprovação nas disciplinas) foram previamente normalizadas para média zero e variância um. Este procedimento evita efeitos de métrica e torna a análise imune à variação natural proveniente das distintas dificuldades inerentes a cada disciplina, analisando apenas a variabilidade amostral das notas de aprovação.

A análise de agrupamentos é uma técnica analítica que identifica subgrupos semelhantes de indivíduos ou variáveis. Especificamente, o objetivo é classificar uma amostra de entidades (indivíduos ou variáveis) em um pequeno número de grupos mutuamente excludentes, com base nas similaridades entre as entidades.

Na AAH utilizam-se de conceitos essencialmente geométricos para se medir (des)semelhanças entre as variáveis ou entre as amostras, agrupando-as considerando-se as suas distâncias quando colocadas em um espaço multidimensional, tendo como coordenadas os valores medidos das variáveis para um dado ponto amostral. Duas amostras são consideradas semelhantes quando a distância

geométrica entre elas é pequena. Argumenta-se que quando o agrupamento se dá através das distâncias entre as amostras, medidas no espaço das variáveis, este é chamado de “Agrupamentos por Casos”. Todavia, na maioria das vezes é mais elucidativo proceder ao “Agrupamento por Variáveis”, onde as m variáveis são representadas no espaço n dimensional das amostras. O agrupamento das variáveis, através da medida de suas distâncias neste espaço, fornece qualitativamente os grupos das variáveis mais fortemente correlacionadas [12].

Existem diversas maneiras de se calcular as distâncias nos espaços das variáveis, ou das amostras. Neste trabalho utilizou-se a distância euclidiana quadrática:

$$d_{ij}^2 = \sum_{k=1}^x (C_{ki} - C_{kj})^2 \quad (6)$$

Uma importante característica dos procedimentos hierárquicos é que os resultados de um estágio anterior são sempre incluídos dentro dos resultados dos estágios seguintes, de forma similar a uma árvore [5]. O resultado da AAH é uma representação gráfica hierarquizada das distâncias relativas entre as distintas variáveis, denominada de dendograma.

A AAH foi utilizada de forma complementar à ACP, pois, como estas análises partem de princípios matemáticos distintos, a AAH constitui-se em um excelente teste de consistência para os resultados obtidos na ACP.

Embora a metodologia descrita neste trabalho tenha sido empregada para a análise do curso de Ciência da Computação em estudo, ela tem potencial de aplicação na análise curricular de outros cursos com estruturas disciplinares e sistemas de avaliação semelhantes.

As análises estatísticas utilizaram o programa computacional SPSS 16.0.

4 Resultados

4.1 . Identificação de Agrupamento de Disciplinas por Perfil de Desempenho.

As análises foram executadas de maneira cumulativa, ciclo a ciclo, ou seja, primeiramente estudou-se a interação entre todas as disciplinas do primeiro ciclo, em seguida foram adicionadas às disciplinas do segundo ciclo na análise e novamente verificaram-se as correlações entre essas disciplinas e assim sucessivamente até que todas as disciplinas dos oito ciclos do curso estivessem

incluídas. Tal procedimento permite visualizar a interação entre as disciplinas do curso na mesma ordem que a cursada pelo aluno.

Dando início ao estudo, realizou-se a primeira análise somente com as disciplinas do primeiro ciclo do curso. Somente uma componente principal, capaz de explicar, com um autovalor associado de 2,8, 57% da variabilidade das notas, foi retida. Esta componente apresentou cargas fatoriais bastante elevadas (0,83; 0,84; 0,85 e 0,71) para as disciplinas Calc1, GA1, Fis1 e Alg1 respectivamente. A disciplina CExp1 embora, integre esta única componente retida, apresentou uma carga fatorial mais baixa, 0,50, mostrando uma correlação mais fraca com as demais disciplinas. A retenção de uma única componente, quando da análise das disciplinas do primeiro semestre, evidencia certa consistência no desempenho do aluno no início do curso, isto é, os alunos apresentaram melhor (ou pior) desempenho no conjunto das disciplinas do ciclo.

Dando continuidade à análise acrescentaram-se as disciplinas do segundo ciclo. Dessa maneira foi possível averiguar as eventuais alterações na correlação do de-

sempenho nas disciplinas do primeiro ciclo concomitantemente ao desempenho nas disciplinas do segundo ciclo. Nesta nova condição a ACP reteve duas componentes que, conjuntamente, explicaram 63% da variabilidade do conjunto das notas. A primeira componente retida, capaz de explicar sozinha 47% da variabilidade das notas, com um auto valor associado de 5,1 manteve cargas fatoriais elevadas para as disciplinas ligadas à matemática, física e programação (Calc1, GA1, Fis1, Alg1, Cal2, GA2, Fis2 e Alg2). Estes resultados apontam para uma forte dependência no desempenho destas disciplinas iniciais do curso com denso conteúdo lógico e matemático. Já a segunda componente retida, capaz de explicar 16% da variabilidade das notas, com um autovalor associado de 1,2, associou-se às disciplinas Comunicação e Expressão I e II (CExp 1 e CExp2), disciplinas estas que atuam tanto na estruturação da língua portuguesa como na elaboração e interpretação de textos. Percebe-se que já no segundo ciclo inicia-se um processo de diferenciação entre desempenhos em conjuntos distintos de disciplinas, cada um definido por suas afinidades de conteúdos e abordagens.

Total Variance Explained

Componente	Solução Exata			Solução com nove componentes principais		
	Auto valores	% da Variância Explicada	% da Variância Acumulada Explicada	Auto valores	% da Variância Explicada	% da Variância Acumulada Explicada
1	14,859	33,019	33,019	14,859	33,019	33,019
2	2,722	6,049	39,068	2,722	6,049	39,068
3	2,579	5,731	44,799	2,579	5,731	44,799
4	1,823	4,051	48,850	1,823	4,051	48,850
5	1,501	3,335	52,185	1,501	3,335	52,185
6	1,422	3,159	55,345	1,422	3,159	55,345
7	1,308	2,907	58,252	1,308	2,907	58,252
8	1,197	2,659	60,911	1,197	2,659	60,911
9	1,147	2,548	63,459	1,147	2,548	63,459
10	,983	2,184	65,643			
11	,887	1,971	67,614			
<hr/>						
43	,197	,438	99,251			
44	,178	,395	99,646			
45	,159	,354	100,000			

Tabela 2: Variância total explicada pelo modelo com retenção de 9 componentes principais

Analogamente, foram realizadas ACP agregando novas disciplinas a cada ciclo. Os resultados obtidos na ACP considerando-se todas as disciplinas do curso, estão apresentados na tabela 2.

Percebe-se, observando-se a referida tabela, que a solução exata do modelo, onde o número de componentes consideradas é igual ao número de variáveis incluídas na análise, é capaz de explicar 100% da variabilidade dos dados. Entretanto, somente as nove primeiras componentes apresentam autovalores iguais ou superiores a um. Assim, pelo critério da Raiz Latente, o modelo reteve somente as primeiras nove componentes.

Para este modelo, com nove componentes retidas, a

medida de adequação da amostra KMO (Kaiser-Meyer-Olkin) [4] de 0.937 indicou uma excelente adequação dos dados à ACP. Este modelo mostrou-se adequado, explicando mais de 63% da variabilidade do conjunto total dos dados originais [15]. Quando avalia-se a capacidade de explicação do modelo para a variabilidade das notas de cada uma das 45 disciplinas separadamente, através da análise das comunalidades, o modelo composto por nove componentes principais retidas foi capaz de explicar entre 47 e 76% das variabilidades das notas efetivas obtidas pelos alunos egressos.

A tabela 3 apresenta a fração explicada (comunalidade) da variância das notas associadas a cada uma das disciplinas consideradas na análise.

Disciplina	Comunalidade	Disciplina	Comunalidade	Disciplina	Comunalidade
Calc1	,723	Calc4	,705	EngSof2	,574
GA1	,731	AlCn2	,664	Redes2	,560
Fis1	,725	Tdig2	,600	ArqCom2	,657
Alg1	,765	Tprg2	,710	SimSi1	,535
CExp1	,552	SoAd2	,590	Bdados1	,720
Calc2	,734	Filos	,544	Cgraf	,598
GA2	,672	ProEst	,589	IntArt	,561
Fis2	,717	AlgCom	,669	AnaSis	,750
Alg2	,682	SisOp1	,575	PForm1	,573
CExp2	,661	EngSof1	,731	SimSi2	,554
Calc3	,666	Redes1	,678	Bdados2	,657
AlCn1	,674	ArqCom1	,641	Himid	,595
Tdig1	,542	Compil	,650	AdInf	,709
Tprg1	,613	PesqOp	,504	AdSis	,559
SoAd1	,543	SisOp2	,636	PForm2	,470

Tabela 3: Comunalidades associadas à cada disciplina

A análise das cargas fatoriais das variáveis sobre as

componentes retidas, após a rotação Varimax está mostrada na tabela 4.

	Component								
	1	2	3	4	5	6	7	8	9
Calc1	,725	,236	-,016	,066	,115	,139	-,121	,143	,264
GA1	,578	,258	,082	,108	,016	,168	-,187	-,025	,498
Fis1	,595	,370	,107	,262	-,160	,069	-,130	,075	,319
Alg1	,217	,778	,043	,139	,126	,094	-,128	,055	,218
CExp1	,181	,163	,179	-,088	,144	-,036	,237	,155	,592
Calc2	,780	,238	,082	,112	,037	,126	-,045	,029	,169
GA2	,696	,104	-,001	-,082	,275	,014	,001	,043	,302
Fis2	,741	,204	,087	,144	-,140	,151	,035	,178	,148
Alg2	,509	,556	,160	,018	,119	,090	,102	,121	,200
CExp2	,097	,046	-,038	-,066	,476	-,141	,535	,016	,334
Calc3	,730	,080	,155	,027	,124	,105	,234	,142	,013
AlCn1	,688	,099	,137	,158	,093	,092	,321	,105	,131
Tdig1	,550	,336	,235	-,032	,186	,048	,124	-,051	-,121
Tprg1	,299	,661	,201	,054	,008	,150	,107	,068	,073
SoAd1	,228	-,034	,211	,044	,019	,212	,400	-,030	,487
Calc4	,758	,055	,229	,151	,084	,070	,179	,070	-,058
AlCn2	,704	,061	,244	,219	,116	,021	,207	-,003	,025
Tdig2	,553	,260	,017	,071	,162	,190	,309	,120	-,223
Tprg2	,326	,710	,078	,030	,113	,181	,174	,120	-,051
SoAd2	,110	,024	,209	-,026	,158	-,003	,708	,069	,053
Filos	,134	,088	,051	,183	-,041	,292	,623	,053	,057
ProEst	,614	,080	,356	,240	,096	,092	,020	,013	-,062
AlgCom	,338	,317	,132	-,100	,487	,432	,050	-,013	-,034
SisOp1	,162	,164	,493	-,009	-,042	,475	,083	,171	,123
EngSof1	,212	,236	,292	,444	,176	,414	-,047	-,379	-,005
Redes1	,275	,157	,270	-,002	,163	,683	,022	,099	-,030
ArqCom1	,381	,337	,406	,325	,004	,301	,084	-,002	-,117
Compil	,310	,165	,334	-,060	,595	,197	,105	,077	,037
PesqOp	,419	,173	,351	,132	,377	-,065	,038	,099	-,012
SisOp2	,156	,064	,268	,034	,455	,229	,116	,460	,224
EngSof2	,147	,087	,624	,041	,003	,069	,215	,047	,316
Redes2	,265	,107	,472	,102	,224	,323	,094	,120	,259
ArqCom2	,185	,319	,400	,099	-,056	,164	,177	,537	,039
SimSi1	,171	-,260	,409	,119	,174	,097	,028	,431	,176
Bdados1	,140	,507	,379	,478	-,108	-,006	,035	,234	-,056
Cgraf	,293	,308	,565	,171	,209	-,015	,082	,137	-,005
IntArt	,384	,188	,387	,262	,302	,163	-,077	,190	-,008
AnaSis	,195	,367	,476	,561	-,143	,035	,111	,028	,015
PForm1	,117	-,033	,066	,729	,088	-,020	-,106	,000	,058
SimSi2	,210	,076	,377	,431	,198	,046	,138	,340	-,015
Bdados2	,161	,228	,071	,185	,125	,136	,015	,711	,010
Himid	,015	-,067	-,131	,409	,605	,075	,149	,102	,040
AdInf	,142	,144	-,041	,737	-,132	,074	,198	,195	-,153
AdSis	,045	,110	-,126	,236	,080	,589	,230	,228	,123
PForm2	,139	,023	,128	,526	,347	,143	,033	,064	,102

Tabela 4: Cargas fatoriais após a rotação Varimax

Dado que a existência de elevada carga fatorial de uma disciplina em uma dada componente principal evidencia uma forte associação entre ambas, percebe-se, observando-se os valores constantes na tabela 4, que, por exemplo, as disciplinas relacionadas à matemática e a física possuem cargas altas na primeira componente. Da mesma maneira, as disciplinas Computação e Desenvolvimento de Algoritmos I e II, Linguagens e Técnicas de Programação I e II e Banco de Dados I apresentam cargas fatoriais elevadas na segunda componente. Outras relações podem ser observadas para as demais componentes.

A tabela 5 apresenta estes resultados mais claramente, sistematizando-os. Nesta tabela as disciplinas foram agrupadas abaixo da componente detentora da sua maior carga fatorial, seguindo uma ordem decrescente. Nota-se que a primeira componente, com autovalor calculado, $\lambda=14.86$, explica sozinha de 33% da variabilidade do conjunto completo dos dados e, pelo fato de relacionar-se com as nove disciplinas de matemática, duas de física e três disciplinas de computação com forte raciocínio lógico, foi denominada de Básicas. Seguindo o mesmo raciocínio as demais componentes foram denominadas por: Lógica de Programação; Ciências Computacionais; Sis-

temas e Informação; Hipermídia e Compiladores; Sistemas e Redes; Humanidades; Banco de Dados; Comunicação.

Percebe-se o agrupamento das disciplinas segundo uma boa aproximação das suas semelhanças de conteúdo, mostrando a estrutura do curso segundo a percepção do aluno, a partir do seu real desempenho.

Quando da aplicação da AAH ao conjunto de notas da 45 distintas disciplinas, esta análise forneceu o dendograma mostrado na figura 1. Para o cálculo da AAH o método da centróide foi aplicado às distâncias euclidianas quadráticas entre as notas normalizadas das distintas disciplinas do curso.

Observa-se, através da análise deste dendograma que, no geral, os mesmos grupos de disciplinas foram agrupados tanto quando da aplicação da ACP como quando do uso da AAH. Tal resultado corrobora para a consistência, tanto da metodologia de análise adotada neste trabalho, como das hipóteses nela contida, pois estes dois métodos de análise se baseiam em princípios matemáticos distintos.

Componente Principal								
Básicas	Lógica de Programação	Ciências Computacionais	Sistemas e Informação	Hipermídia e Compiladores	Sistemas e Redes	Humanidades	Banco de Dados	Comunicação
$\lambda = 14.86$	$\lambda = 2.72$	$\lambda = 2.58$	$\lambda = 1.82$	$\lambda = 1.50$	$\lambda = 1.42$	$\lambda = 1.31$	$\lambda = 1.20$	$\lambda = 1.15$
Calc2	Alg1	EngSoft2	AdInf	Himid	Redes1	SoAd2	Bdados2	CExp1
Calc4	Tprg2	Cgraf	Pform1	Compil	AdSis	Filos	ArqCom2	SoAd1
Fis2	Tprg1	SisOp1	AnaSis	AlgCom		CExp2	SisOp2	
Calc3	Alg2	Redes2	Pform2				SimSi1	
Calc1	Bdados1	ArqCom1	SimSi2					
AlCn2		IntArt	EngSoft1					
GA2								
AlCn1								
ProEst								
Fis1								
GA1								
Tdig2								
Tdig1								
PesqOp								

Tabela 5: Estrutura do curso de Ciência da computação, segundo a ACP

4.2 Relação entre rendimento dos estudantes, tempo de conclusão e classificação no exame de ingresso

no exame de ingresso) e duração (tempo em semestres necessários ao aluno até a conclusão do curso) foram introduzidas na análise o novo modelo de ACP reteve 10 componentes, explicando cerca de 65% da variabilidade dos dados.

Quando as variáveis colocação (classificação do aluno

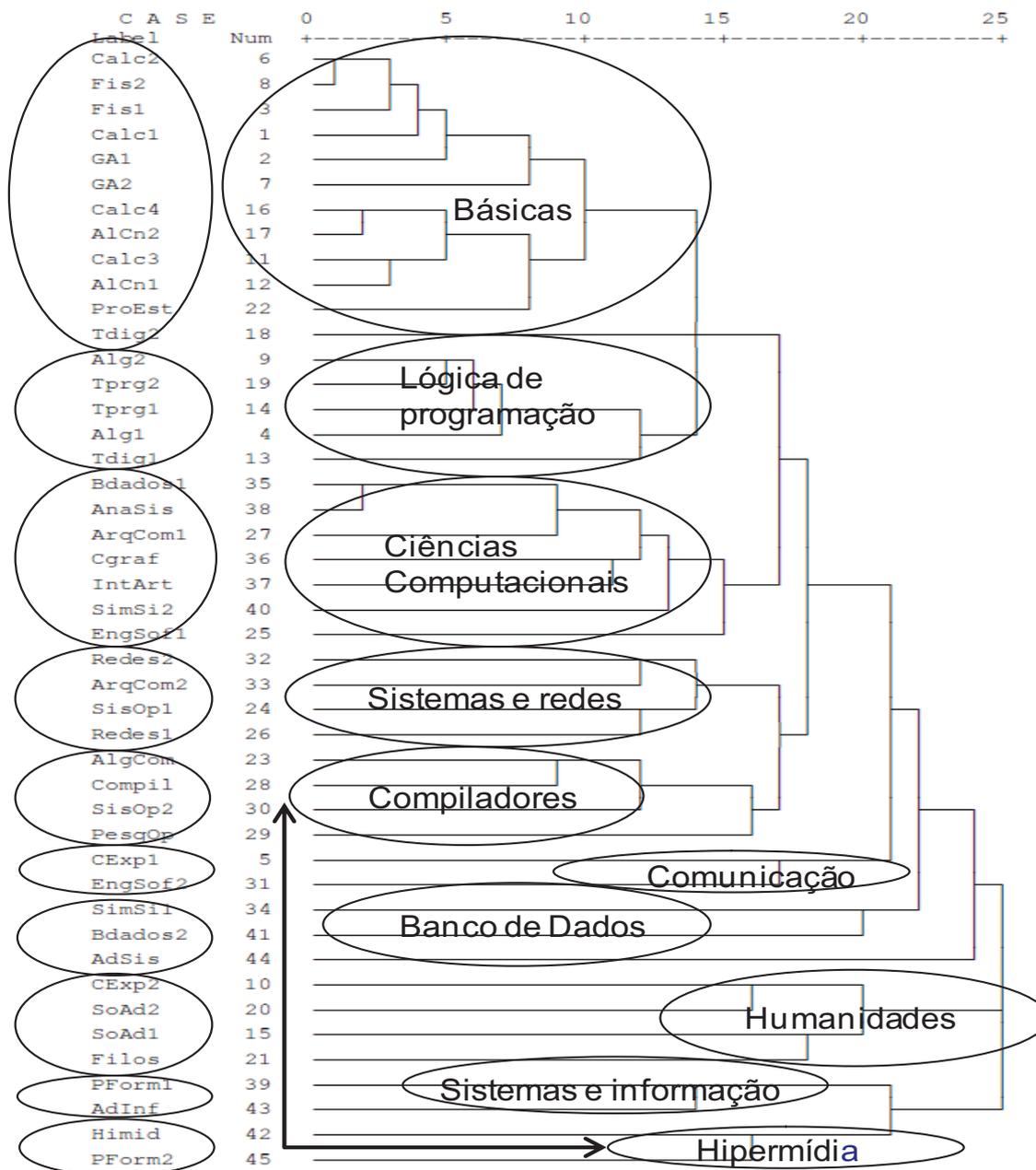


Figura 1: Dendrograma resultante da AAH

As communalidades associadas às variáveis duração e colocação foram, respectivamente, 70 e 53%.

Para evitar o surgimento de correlações negativas, as variáveis duração e colocação sofreram uma inversão de sinal pois, quanto maior duração do curso e maior a colocação exame de ingresso espera-se um pior o desempenho do aluno no conjunto das disciplinas.

A figura 2 apresenta as cargas fatoriais, após a rotação Varimax, do conjunto das 45 disciplinas acrescidas das variáveis duração e colocação.

Como mostra a figura 2 (em destaque), a variável duração associou-se fortemente à primeira componente, formada pelas disciplinas básicas.

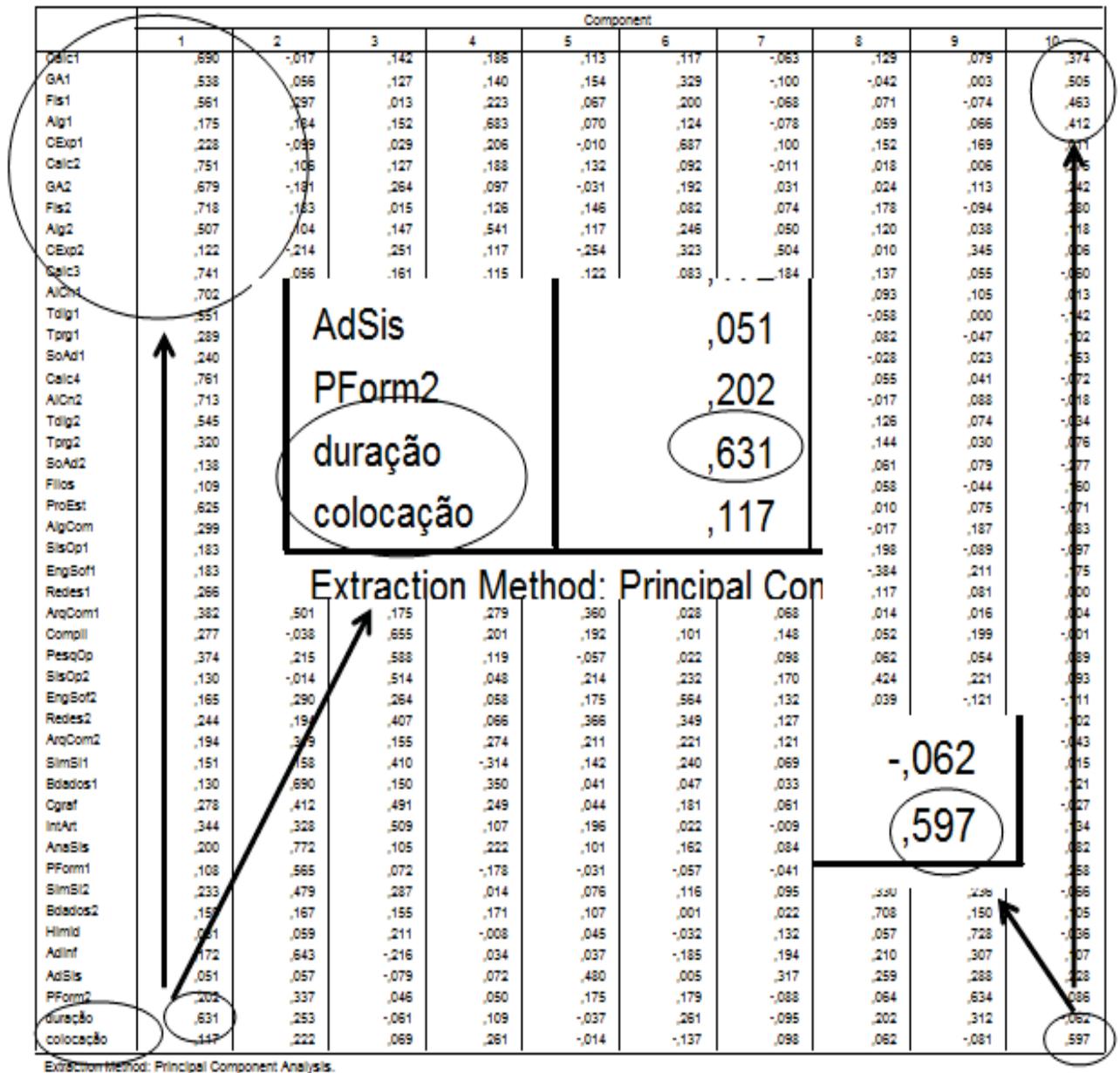


Figura 2: Cargas fatoriais após a rotação VARIMAX, incluindo-se duração e colocação.

Este resultado indica que são estas as disciplinas básicas que impõe ao aluno as maiores dificuldades, muitas vezes resultando e reprovações sucessivas. Após obter a aprovação nestas disciplinas associadas à componente principal nominada de Básicas, o aluno passa a enfrentar dificuldades menores para a conclusão bem sucedida de seu curso.

Já a variável colocação, também destacada na figura 2, que no ao início do curso estava correlacionada ao desempenho nas disciplinas básicas, vai gradativamente se diferenciando do desempenho nas disciplinas. Assim, como apresentado na figura 2, ao final do curso, quando todas as disciplinas foram cursadas, a variável colocação, embora ainda fracamente correlacionada com as disciplinas básicas do primeiro ciclo, apresenta-se com um comportamento bastante independente. Este comportamento acaba por definir a décima componente principal, indicando que a colocação obtida no exame de ingresso tem sua importância para o desempenho acadêmico diminuída ao longo do curso. Os resultados obtidos corroboram a ideia de que, ao final do curso, os alunos egressos realmente possuem um grau de conhecimento muito mais homogêneo quando comparado ao seu momento de ingresso no curso.

Os resultados ora apresentados forneceram subsídios para uma adequação Plano Pedagógico do curso estudado. Percebeu-se que as disciplinas Técnicas Digitais I e II apareceram fortemente relacionadas às disciplinas básicas, enquanto que se esperava uma associação com as disciplinas de Arquitetura de Computadores. Além disso, as disciplinas de Arquitetura de Computadores I e II posicionaram-se em componentes distintas (componentes Ciências Computacionais e Banco de dados, respectivamente, conforme tabela 5). Diante disto, em uma recente reformulação do plano pedagógicos, os conteúdos programáticos das disciplinas Técnicas Digitais I e II foram reorganizados em duas novas disciplinas: A primeira manteve o nome de Técnicas Digitais, porém foi alocada no primeiro ciclo do curso. Já a segunda, chamada de Sistemas Digitais e Projetos de Computadores, foi alocada no quarto ciclo, de maneira a fazer uma articulação importante com outra nova disciplina, agora única, de Arquitetura de Computadores.

Outra modificação importante envolveu as disciplinas de Banco de Dados I e II. Esperava-se observar estas disciplinas articuladas com disciplinas de ciclos mais adiantados do curso, entretanto, conforme observamos na tabela 5, estas disciplinas não só apareceram associadas à componentes distintas como também Banco de Dados I apareceu associada às disciplinas iniciais de algoritmos. Esta discrepância pode ser atribuída ao fato de as disciplinas Banco de Dados I e II estarem alocadas nos ciclos finais do curso (sétimo e oitavo ciclos respectivamente,

conforme tabela 1). Assim, outras disciplinas que poderiam explorar, de forma articulada, os conteúdos pertinentes à bancos de dados, não o faziam pois estavam alocadas em ciclos anteriores. Desta forma, na nova ordenação do plano pedagógico, ambas as disciplinas Banco de Dados I e II foram realocadas em ciclos intermediários (quinto e sexto ciclo respectivamente).

Ainda, avaliando a tabela 5, nota-se as disciplinas Sistemas Operacionais I e Sistemas Operacionais II, também apareceram correlacionadas com componentes principais distintas (Ciências Computacionais e Banco de Dados, respectivamente). O mesmo ocorreu com as disciplinas Redes de Computadores I e Redes de Computadores II, associadas à duas componentes principais distintas, a saber: Sistemas e Redes e Ciências Computacionais. Assim, visando uma maior aproximação entre as disciplinas de cada um destes pares, quando da elaboração do novo plano pedagógico, estas passaram por alterações de seus conteúdos buscando-se uma maior articulação entre as suas componentes curriculares I e II.

As modificações descritas acima já foram implementadas e seus impactos estão sendo monitorados. Seus efeitos serão avaliados em um futuro trabalho, entretanto, observações preliminares indicam ganhos tanto para o curso como para o aprendizado dos alunos, associados à implantação do novo plano pedagógico.

5 Conclusão

Foi possível, baseando-se nos reais desempenhos acadêmicos dos alunos egressos, identificar nove grupos coesos de disciplinas, evidenciando estruturação do curso sob a percepção destes alunos. Destes grupos destaca-se o grupo denominado Básicas, contendo as disciplinas de matemática e física, além de disciplinas de computação com forte raciocínio lógico. Esta componente representa cerca de um terço da carga horária do curso, tendo suas disciplinas concentradas no seu início. Seis outras componentes agregam, segundo suas semelhanças, as demais vinte e seis disciplinas específicas da área de ciência da computação. Já os dois fatores restantes associam-se às disciplinas relacionadas às ciências humanas. Estas associações apresentam-se compatíveis com as temáticas abordadas nas disciplinas.

A forte associação entre as disciplinas de matemática e de física indica seu alto grau de articulação interdisciplinar. O fato de o tempo necessário para a conclusão do curso se associar a estas disciplinas demonstra como são estas as disciplinas que, em geral, impõem ao aluno as maiores dificuldades. Depois de obtidas as aprovações nas disciplinas básicas, o curso tende a seguir seu ritmo natural, sem reprovações sistemáticas nas disciplinas

específicas da área de computação. Tal fato contribui para o entendimento de que estas disciplinas básicas fornecem subsídios importantes para a compreensão e a intervenção na prática do profissional de computação. Este embasamento teórico confere ao profissional a fundamentação científica para as suas ações, instrumentalizando-o para compreender os crescentes avanços do campo de conhecimento onde atua.

Quanto à classificação no vestibular observou-se uma correlação com o desempenho das disciplinas do curso apenas no seu início. Esta correlação perde importância ao longo do tempo e, ao final deste, a classificação define uma componente a parte. Tal fato corrobora a percepção de que, mesmo um aluno não tão bem classificado no exame de ingresso, quando é capaz de superar as dificuldades impostas pelo curso, muitas vezes a custa de algumas reprovações, ao final do curso tem um perfil de desempenho semelhante ao dos demais alunos egressos. Segundo Silva [16] em um estudo com alunos de ciências biológicas observou-se que o exame vestibular em um primeiro momento classifica quais são os alunos preparados para o ingresso no ensino superior, no entanto durante o curso outros fatores como família, emprego, formação básica e identificação com o curso determinam o desempenho destes alunos.

Por outro lado, é importante considerar que políticas institucionais que apoiem a permanência estudantil, principalmente relativas ao apoio pedagógico frente às dificuldades iniciais de formação são extremamente necessárias.

Em consonância ao que foi exposto, é possível apontar que o estudo aqui desenvolvido pode, se usado adequadamente, contribuir para a melhora do curso de Ciência da Computação. Adequações no plano pedagógico do curso, tendo em consideração os resultados apresentados neste trabalho, foram implementadas e estão sendo monitoradas.

Além disso, percebe-se que um melhor direcionamento dos alunos nas disciplinas do curso consideradas básicas pode trazer um melhor rendimento do aluno e uma redução no tempo médio necessário para a sua conclusão. Possibilitar equidade de progresso e resultado no ensino superior significa empenho em se propiciar superação de dificuldades iniciais de modo a que o desempenho seja resultado do esforço de cada estudante [17,18].

Agradecimentos:

Os autores agradecem ao Prof. Dr. Flávio Tonidandel por suas importantes contribuições e discussões e a Fundação Educacional Inaciana pelo apoio a este trabalho.

Referências

- [1] A. Z. Kuenzer, O que muda no cotidiano da sala universitária com as mudanças no mundo do trabalho? In Castanho, S.; Castanho, M.E. *Temas e textos em metodologia do ensino superior*. Papirus, Campinas. páginas15-28, 2001.
- [2] H. Lück, Planejamento em orientação educacional, 14º ed., Petrópolis, Ed. Vozes, 2002.
- [3] J. C. Morgado. Processo de Bolonha e ensino superior num mundo globalizado. *Educ. Soc.*, Campinas, v. 30, n. 106, abr. 2009.
- [4] J. F. Jr. Hair, R. E. Anderson, R. L. Tatham, W. C. Black, *Análise multivariada de dados*, 5º ed. Porto Alegre, Bookman, 2005.
- [5] J. L. Corrar , E. Paulo, J. M. D. Filho, *Análise Multivariada para os cursos de Administração, Ciências Contábeis e Economia*, São Paulo, Atlas, 2007.
- [6] R. C. Henry, C. W. Lewis, , P. K. Hopke, H.J. Willianson. Review of receptor model fundamentals, *Atmosph. Env.*, 18, 1507 – 1515, 1984.
- [7] H. Harman, 1976, *Modern factor analysis*, Chicago University of Chicago Press, 1976.
- [8] D. White, M. Richiman, B. Yarnal, Climate regionalization and rotation of principal components, *Intern. J. of Climatology*, 11, 1-25, 1991.
- [9] M. F. Andrade, Identificação de fontes de matéria particulada do aerossol atmosférico de São Paulo, tese de doutorado, Instituto de Física, Universidade de São Paulo, 1993.
- [10] M. A. Yamasoe, Estudo da composição elementar e iônica de aerossóis emitidos em queimadas na Amazônia, dissertação de mestrado, Instituto de Física, Universidade de São Paulo, 1994.
- [11] F. A. M. Echalar, Estudo da estrutura de fontes de aerossóis em Cubatão com o uso de Pixe e Modelos Receptores, dissertação de mestrado apresentada ao Instituto de Física da Universidade de São Paulo. São Paulo, 1991.
- [12] F. Gerab, 1996, *Técnicas Analíticas Nucleares Aplicadas à Medida em Larga Escala de Aerossóis Atmosféricos na Região Amazônica*, tese de doutorado, Instituto de Física, Universidade de São Paulo, 1996.

- [13] P. K. Hopke, Receptor modeling for air quality management, Amsterdam, Elsevier Science Publishers, 1991.
- [14] N. Z. Heidam, Atmospheric aerosol factor models, mass and missing data, *Atmosph. Env.*, 16, 1923 – 1931, 1982.
- [15] J. C. R. Pereira, Análise de Dados Qualitativos: Estratégias Metodológicas para as Ciências da Saúde, Humanas e Sociais, São Paulo, EDUSP, 2004.
- [16] M. da Silva, M. J. Padoin. Relação entre o desempenho no vestibular e o desempenho durante o curso de graduação. Ensaio: aval.pol.públ.Educ., Rio de Janeiro, v. 16, n. 58, mar. 2008.
- [17] V. L. Felicetti, M. C. Morosini. Equidade e iniquidade no ensino superior: uma reflexão. Ensaio: aval.pol.públ.Educ , vol.17, n.62, 2009.
- [18] J. E. Roemer. Equalizing opportunities (in one country) for human development. Davis, CA: University of California, 1998.