

3

O uso da análise de conglomerados na seleção de um conjunto de itens característicos dos níveis de proficiência

Nesta segunda investigação, desenvolve-se uma abordagem alternativa para a seleção dos itens característicos e a definição dos níveis representativos de uma escala de proficiência. Utilizam-se métodos baseados em análise de conglomerados (*cluster analysis*) para a escolha do subconjunto de itens que caracterizam cada nível, e, como conseqüência, os centros e os limites dos níveis de proficiência, que são utilizados na interpretação da escala, emergem dessa escolha.

Nesse processo de clusterização, leva-se em consideração uma característica importante da TRI, que é a representação das medidas dos parâmetros de dificuldade dos itens e de proficiência dos alunos ao longo da mesma dimensão da escala de habilidades. Para tal, utiliza-se a noção de proximidade ou semelhança dos itens expressa em termos das medidas de distância euclidianas, calculadas para esses dois pontos notáveis dos itens: (ponto b) ponto de maior inclinação da Curva Característica do Item-CCI; e o (ponto s) ponto de maior taxa de decrescimento da inclinação da CCI. A utilização simultânea desses dois pontos notáveis permite a seleção dos *clusters* dos itens e a constituição de grupos homogêneos não apenas em função do posicionamento quanto a esses dois pontos, mas também quanto à forma da CCI. Estes critérios são utilizados para a formação dos *clusters*, cujo número pode ser especificado *a priori*.

E importante salientar que esses dois pontos notáveis estão entre os três utilizados na interpretação da escala de proficiência do Projeto Geres e, como já se mencionou na primeira parte deste trabalho, eles estão associados a duas etapas importantes do processo de aprendizagem dos alunos, respectivamente habilidade no auge do desenvolvimento e habilidade consolidada. Isso permite agrupar os itens em subgrupos (*clusters*) relativamente homogêneos, a partir de suas medidas de proximidades ou semelhança, ou seja, dentro de um mesmo *cluster*,

essas medidas sejam as menores possíveis, enquanto que, entre os *clusters*, elas sejam as maiores possíveis.

Este estudo está dividido em duas partes: (i) primeira parte - uma breve apresentação da análise de agrupamento (*cluster analysis*); (ii) segunda parte - uma apresentação da construção de *clusters* para a seleção de itens característicos dos níveis da escala de proficiência.

3.1 **Análise de *clusters*: considerações gerais**

A análise de agrupamento ou *clusters* é um conjunto de técnicas multivariadas de interdependência, cuja finalidade primária é agregar objetos baseando-se nas suas características. Assim sendo, sujeitos ou variáveis são agrupados de forma a constituir grupos homogêneos ou compactos com base em uma ou mais características (HAIR JR, 2005:384).

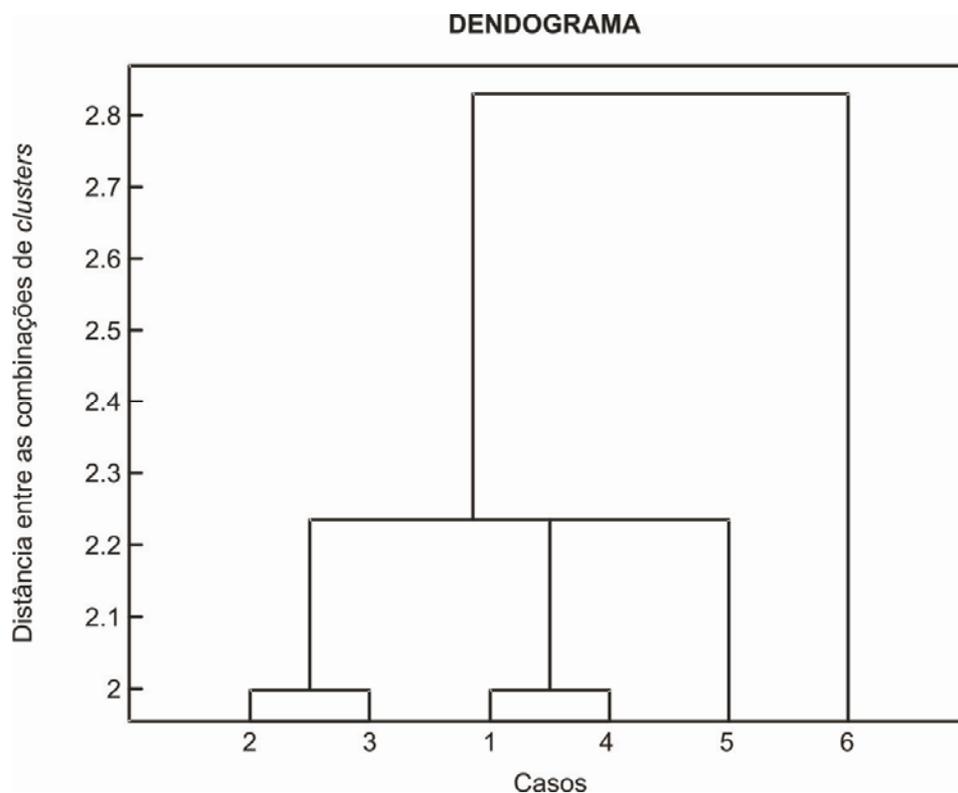
Conseqüentemente, a análise de *clusters* dispõe os elementos, simultaneamente, segundo o grau de semelhança entre si e segundo a maior dessemelhança possível em relação aos objetos pertencentes a outros grupos. A idéia inerente a esse propósito está em minimizar a variabilidade interna dos grupos, ao mesmo tempo em que maximiza a heterogeneidade entre os grupos.

3.1.1 **Formação dos agrupamentos**

Basicamente, para a formação dos agrupamentos, devem ser respondidas três questões fundamentais: Qual é a estrutura da formação dos agrupamentos nas diferentes etapas da análise? Como formar o melhor agrupamento possível com um número pré-especificado de *clusters*? Qual a melhor quantidade de grupos para se obter, simultaneamente, o mínimo de variabilidade dentro dos grupos e o máximo de variabilidade entre os grupos?

Vários métodos podem responder a essas questões, porém, dependendo da especificidade requerida, recorre-se a um deles que, normalmente, prioriza uma delas. Os primeiros tipos são os chamados métodos para a detecção de aglomerados (*clustering*) não-supervisionados ou objetivos. Fundamentalmente, eles se dividem em hierárquicos e não-hierárquicos. Os métodos hierárquicos dão preferência à interconexão entre os grupos em diferentes estágios do agrupamento e não apresentam, *a priori*, o número de agrupamentos que se deseja formar. De início, forma-se um número muito elevado de agrupamentos (geralmente, pouco menor ou igual ao número de casos), e, depois, seguindo-se critérios pré-estabelecidos de inclusão, várias etapas de agrupamento são realizadas até se chegar a um único *cluster*. Assim, os elementos vão-se juntando para formar os *clusters*, ou vão-se juntando aos *clusters* já formados, ou, ainda, *clusters* diferentes vão-se unindo, formando novos *clusters*, e assim sucessivamente. O processo de agrupamento hierárquico pode ser acompanhado com o auxílio de um dendrograma, onde é possível definir, por exemplo, o número de *clusters* que melhor expressa a natureza dos dados ou os interesses de quem analisa. Podem ser empregadas, ainda, estatísticas para medir a relação entre as variabilidades interna e externa do conjunto dos dados e decidir por um número adequado de *clusters* e sua configuração. Evidentemente, esse é um processo que deve ser utilizado com poucos casos, pois, de outro modo, a análise pode tornar-se muito trabalhosa. O dendrograma apresentado na Figura 12 ilustra a situação considerada.

Figura 11 - Dendograma



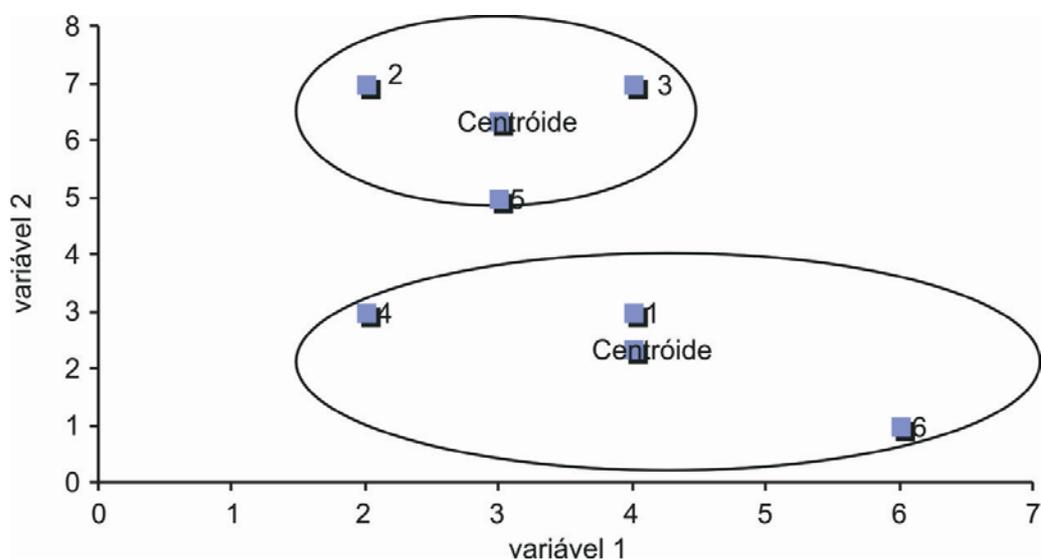
Esse dendrograma mostra que os elementos unidos em um determinado estágio do agrupamento permanecerão assim até o final do processo que se desenvolve pelo agrupamento de casos, de casos a grupos e de grupos, de maneira a consolidar-se um único e grande grupo ao final da análise.

Outros métodos muito utilizados são os não-hierárquicos, como, por exemplo, o *k-means clustering*, em que o número de grupos é pré-especificado. Comparado aos métodos hierárquicos, esse apresenta como principal vantagem a facilidade com que é aplicado a um conjunto muito grande de dados.

O método é aplicado obedecendo-se aos seguintes passos: (i) são formados os centróides iniciais (no mesmo número de *clusters* que se deseja obter), que são pontos no espaço, formados pelas variáveis; (ii) calculam-se as distâncias de cada elemento a cada um dos centróides; (iii) formam-se os primeiros *clusters* com os elementos que apresentam a menor distância em relação àquele centróide; (iv) cada elemento passa a

pertencer ao *cluster* de cujo centróide está mais próximo; (v) recalculam-se os centróides de cada *cluster* com base nas coordenadas dos elementos classificados conforme essa nova disposição; (vi) calculam-se novamente as distâncias de todos os elementos aos centróides; (vii) os elementos, novamente, passam a pertencer ao *cluster* de cujo centróide estão mais próximos. Nesse passo, alguns elementos podem trocar de *cluster*. Esse processo tende a repetir-se até que nenhum elemento mude de *cluster*, ou até que um número pré-especificado de iterações seja atingido. Nesse caso, os centróides e os elementos dos grupos chegam a uma solução ótima, à medida que se reproduz, a cada nova iteração, uma estrutura de grupos mais adequada, tendo em vista o número previsto de *clusters*. A Figura 13 ilustra a solução de um processo de *k-means clustering*, formando dois grupos com duas variáveis e seis casos.

Figura 12 - Aplicação de *k-means*



3.2

Construção de *clusters* para a seleção de itens característicos dos níveis da escala de proficiência

3.2.1

A seleção de itens

O foco deste estudo é a construção e a interpretação da escala de proficiência do Sistema de Avaliação da Educação Básica - SAEB - para o teste de Matemática da 4ª série do ensino fundamental. Para tal, além da base de itens da avaliação do SAEB 2003, foi utilizado o banco de itens das avaliações do SAEB amostral 2001 e da Prova Brasil 2005, encontrando-se todos esses itens na mesma escala do SAEB. Os testes do SAEB 2001 e 2003 foram compostos de 169 itens cada; e os da Prova Brasil, de 70 itens. Após a equalização de todos os itens na mesma escala, processo que utiliza itens comuns às diferentes avaliações, chegou-se a uma base constituída de 366 itens.

Para a escolha dos itens característicos, representativos dos diferentes níveis da escala, foram utilizados dois critérios: (i) itens que apresentaram boa discriminação, ou seja, valores do parâmetro $a \geq 1,15$ na escala original¹²; (ii) itens que apresentaram um parâmetro b de até 275 pontos na escala do SAEB, uma vez que, após esse ponto, encontra-se um percentual muito reduzido de alunos da 4ª série, como mostra a distribuição dos alunos da 4ª série do ensino fundamental, avaliados na Prova Brasil em 2005, apresentada no Quadro 6.

Quadro 6 - Distribuição dos alunos da 4ª série EF - Prova Brasil 2005

4ª série – 2005											
0-125	125-150	150-175	175-200	200-225	225-250	250-275	275-300	300-325	325-350	350-375	375 ou mais
13,55%	16,78%	20,64%	19,92%	14,95%	8,55%	4,00%	1,21%	0,37%	0,04%	0,00%	0,00%

¹² O valor desse parâmetro é expresso na escala produzida pelos programas computacionais usualmente utilizados na produção das medidas por meio da TRI (como o BILOG-MG e outros, que calculam os valores do parâmetro, de tal modo que esses, em geral, apresentam uma variação entre 0 e 3).

Com base nesses dois critérios, foram selecionados 109 itens, que constam do Anexo XI.

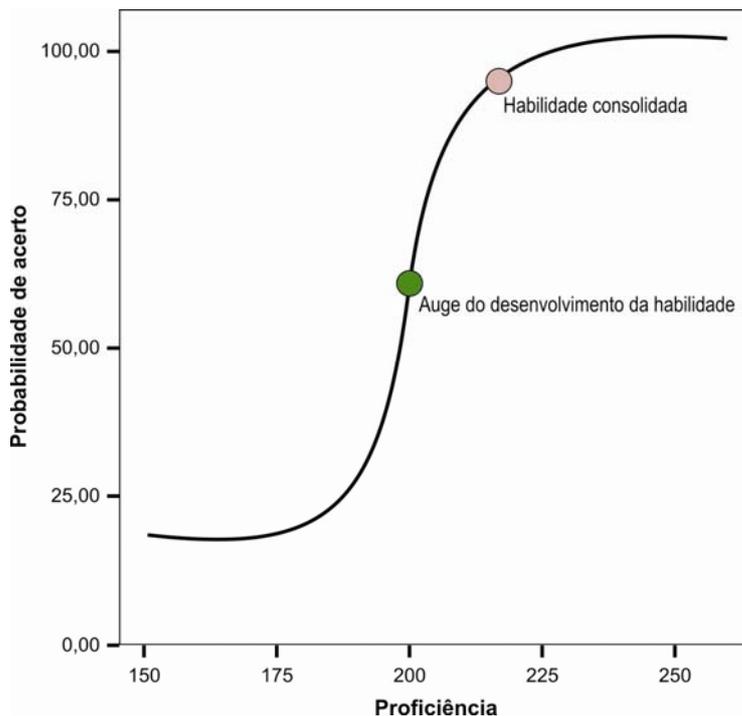
A vantagem de se utilizarem itens com maior discriminação está na sua forte representatividade das etapas do desenvolvimento cognitivo dos alunos avaliados ao longo da escala de proficiência. Evita-se, assim, a análise de itens cuja CCI apresenta baixa inclinação, o que, geralmente, ocorre quando eles estão associados a múltiplas habilidades, tornando sua utilidade pouco expressiva na interpretação da escala. Por outro lado, deve-se considerar que as curvas características pouco inclinadas podem ser decorrentes de itens representativos de habilidades que apresentam um processo alongado de consolidação na escala de proficiência. Em geral, esse tipo de habilidade não é pertinente apenas ao período de escolaridade considerado.

A partir dessa seleção, será utilizada a análise de agrupamentos (*cluster analysis*) para se escolher um subconjunto de itens que caracterizam os níveis de proficiência da parte da escala do SAEB considerada neste estudo. Desse processo, resultarão a classificação dos itens e a definição dos centros dos *clusters*, com a posterior definição dos limites dos *clusters* que emergirão dessa escolha para a interpretação da escala.

3.2.2 O agrupamento inicial dos itens selecionados

O processo de clusterização adotado utilizará dois pontos notáveis da CCI, e que estão dispostos na própria escala de habilidades: o ponto de maior inclinação da CCI (habilidade em seu auge do desenvolvimento), que será chamado de ponto **b**, e o ponto de maior taxa de decrescimento da inclinação da CCI (habilidade já consolidada), que será chamado de ponto **s**. A Figura 14 mostra o exemplo de um item com esses dois pontos notáveis.

Figura 13 - Exemplo de uma Curva Característica do Item



Ponto **b**: auge do desenvolvimento da habilidade - ocorre no ponto em que a 1ª derivada da CCI do Item é máxima (2ª derivada da CCI = 0 = parâmetro **b**);

Ponto **s**: Habilidade consolidada - a taxa de decréscimo da 1ª derivada da CCI é máxima (3ª derivada da CCI = 0 (segunda raiz)).

Os *clusters* foram formados tendo como variáveis esses dois pontos notáveis da Curva Característica do Item - CCI: o primeiro ponto (ponto **b**) é estatisticamente igual ao parâmetro **b**, que mede a dificuldade do item. Nesse ponto, a 2ª derivada da Curva Característica do Item é igual a zero, correspondendo ao ponto de maior inclinação da curva e, conseqüentemente, ao de maior discriminação do item. Pode-se dizer que, nesse ponto, a habilidade encontra-se no auge de seu desenvolvimento. Nele, a probabilidade de acerto do item corresponde a $0,5 + c/2$, sendo a probabilidade **c** de acerto ao acaso nos itens de múltipla escolha presentemente considerados.

O segundo ponto (ponto **s**) sinaliza a consolidação da habilidade requerida pelo item. Nele, a 3ª derivada da CCI é igual a zero.

Estatisticamente, nesse ponto, localiza-se a maior taxa de decréscimo da inclinação da CCI; a partir daí, observa-se pouca variação na probabilidade de acerto do item (que já é bastante elevada), indicando que a habilidade já está consolidada. Normalmente, neste ponto, a probabilidade de acertar o item está em torno de 0,8 (80%).

Esses dois pontos constituem as variáveis utilizadas para o agrupamento de itens, o que será realizado com base na metodologia *k-means*. Para a definição de padrões de “proximidade” entre os itens e os *clusters* ao longo da escala de proficiência, serão agrupados em um mesmo *cluster* os itens que possuírem os pontos notáveis mais próximos entre si, com os diferentes *clusters* apresentando as maiores distâncias possíveis entre si.

Para se determinar a proximidade entre os itens com base em tais pontos, utiliza-se a distância euclidiana simples, que é a distância entre os pontos representados pelas coordenadas dadas por **b** e **s** para os itens selecionados, de acordo com a seguinte fórmula:

$$d_{i,j} = \sqrt{(b_i - b_j)^2 + (s_i - s_j)^2}$$

onde

d_{ij} é a distância euclidiana, item_i-item_j;

b_i = ponto **b** do item *i*;

b_j = ponto **b** do item *j*;

s_i = ponto **s** do item *i*;

s_j = ponto **s** do item *j*.

Calculando-se a distância euclidiana entre os pares de itens selecionados, encontram-se, basicamente, duas situações características e distintas, que estão representadas pelas CCIs apresentadas nos Gráficos 4 e 5.

Gráfico 4

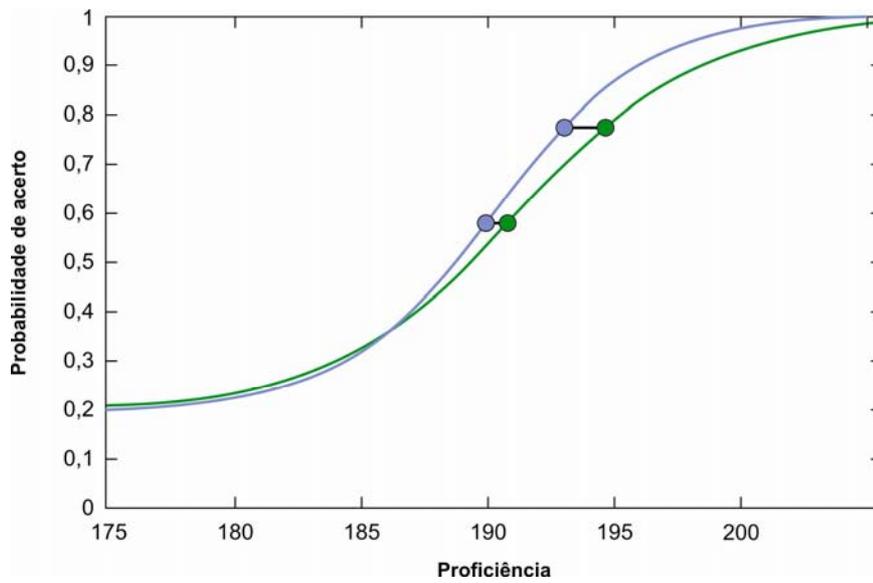
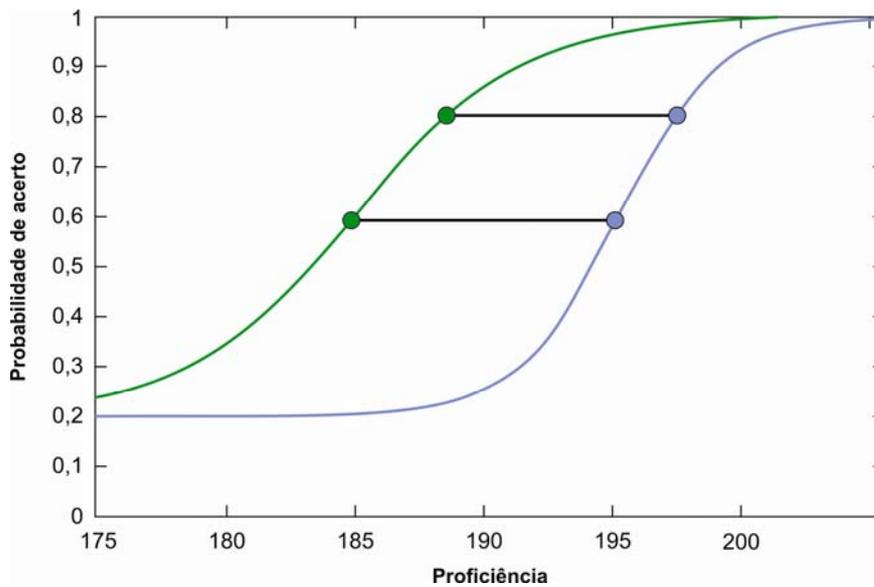


Gráfico 5



No Gráfico 4, observa-se que os dois itens possuem seus respectivos pontos notáveis próximos, o que produz uma distância euclidiana menor e torna-os similares. No Gráfico 5, por outro lado, observa-se que os itens possuem seus respectivos pontos notáveis mais distantes entre si e, conseqüentemente, menos similares. Assim, tendo-se como referência os dois pontos notáveis do item, quanto maior for a distância euclidiana entre eles, menor será o grau de similaridade.

Nesse sentido, a noção de proximidade dos itens pode ser expressa em termos das medidas de distância euclidiana, calculadas para os pontos notáveis dos itens, permitindo agrupar itens em subgrupos (*clusters*) relativamente homogêneos, com base na sua proximidade ou semelhança; ou seja, são selecionados subconjuntos de itens que estejam “próximos” entre si na escala de habilidades.

Com o emprego dessa metodologia, foram obtidos 11 *clusters* que agruparam os 109 itens que haviam sido selecionados pelos critérios referentes aos parâmetros **a** e **b** anteriormente mencionados. Em seguida, os 11 *clusters* formados pelo procedimento *k-means* foram inseridos no Módulo de Interpretação de Escalas do SisAni, desenvolvido pelo Centro de Políticas Públicas e Avaliação da Educação - CAEd/UFJF. Esse software oferece informações gerais sobre os resultados obtidos, tanto em relação aos *clusters*, quanto em relação aos itens que os compõem, além de permitir a realocação de itens em diferentes *clusters* ou a eventual eliminação de alguns deles.

Entre as informações fornecidas pelo SisAni, estão o número de itens utilizados e os tipos de descritores presentes em cada cluster; além disso, o SisAni fornece a ilustração gráfica das curvas características dos itens, dos limites dos *clusters*, das curvas de informação plotadas dentro de cada um deles, e outras funcionalidades, tais como:

(i) Cálculo do centróide dos *clusters* (no qual cada centróide é definido como um vetor especificado pela média dos pontos **b** e pela média dos pontos **s** dos itens) e da distância de cada item ao centróide de seu respectivo *cluster*, utilizando-se a distância euclidiana. A importância dessa informação está no fato de que, quanto menor for essa distância, mais representativo do *cluster* é o item selecionado, o que permite identificar os itens que deverão ser eliminados ou realocados, isto é, aqueles que estão excessivamente afastados do centróide.

(ii) Cálculo dos limites do *cluster*, que, aqui, é definido como tendo seu início correspondendo à média dos pontos **b**, e seu final correspondendo à média dos pontos **s** dos itens, o que faz esses limites

coincidirem com que esses limites coincidam com as componentes dos centróides.

(iii) Cálculo da integral da Curva de Informação do Item (CII), ou seja, da área delimitada entre a CII o eixo das abscissas, valor esse que representa a quantidade máxima de informação que o item pode oferecer. Além disso, calcula-se a quantidade de informação dos itens em cada *cluster*, com base na curva de informação do item integrada nos limites do *cluster*. Esse dado é importante para se estabelecer a comparação entre a quantidade de informação que o item oferece dentro de seu *cluster* e a informação máxima oferecida pelo item na escala.

(iv) Cálculo da quantidade de informação fornecida por cada *cluster*, que corresponde ao somatório das quantidades de informação de cada um de seus respectivos itens componentes. Essa medida permite identificar quais são os *clusters* que apresentam maior quantidade de informação ao longo da escala de proficiência.

(v) Cálculo da proporção de informação do item no *cluster*, obtido pela divisão entre a quantidade de informação do item no *cluster* e a quantidade de informação total do *cluster*. Quanto maior o valor percentual, maior será o impacto da eliminação desse item sobre a configuração dos elementos de composição do *cluster*, e maior será a atenção pedagógica que o item deverá merecer. Vale mencionar que *clusters* com muitos itens, geralmente, ocupam posições mais centrais na escala de proficiência e são menos sensíveis à eliminação desses, o que não acontece com os *clusters* que ocupam as extremidades da escala.

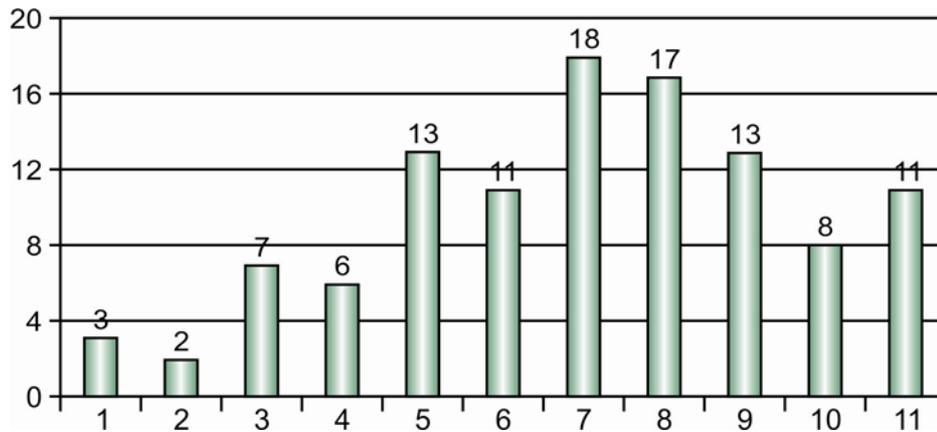
Na combinação dos resultados do procedimento do *k-means* com os indicadores produzidos pelo SisAni, foram produzidas, para cada *cluster*, as informações apresentadas na Tabela 17, compreendendo: (i) o número do *cluster*, (ii) o número de itens pertencentes a cada *cluster*, (iii) a quantidade de informação dos *clusters* dentro de seus respectivos limites, (iv) os limites de cada *cluster*, (v) os centróides dos *clusters*, (vi) o menor e o maior valor de **b** e (vii) o menor e o maior valor de **s**.

Tabela 17 - Resumo dos 11 clusters originais

N. do <i>cluster</i>	N. de itens	Quantidade de informação	Coordenada x do Centróide (média dos b) = Limite Inferior do grupo	Coordenada y do Centróide (média dos s) = Limite Superior do grupo	menor b	maior b	menor s	maior s
1	3	0,027506764	119,7	155,98	115,6	124,15	152,42	159,58
2	2	0,020289296	138,71	167,96	135,5	141,92	164,04	171,87
3	7	0,061354133	157,06	189,31	152,15	161,77	182,64	194,47
4	6	0,052239873	170,05	202,97	163,05	175,68	195,07	208,21
5	13	0,127362022	183,71	215,04	178,7	189,32	209,06	221,62
6	11	0,089681414	193,75	226,2	187,48	198,93	222,99	230,45
7	18	0,175985838	208,92	238,53	200,67	218,55	231,33	245,92
8	17	0,152157568	218,94	250,95	212,01	226,38	245,16	258,14
9	13	0,127695743	236,33	266,28	229,4	244,32	259,42	273,59
10	8	0,088990043	253,03	281,28	247,73	260,36	272,87	290,77
11	11	0,105748598	265,81	297,29	258,18	273,05	289,9	301,61

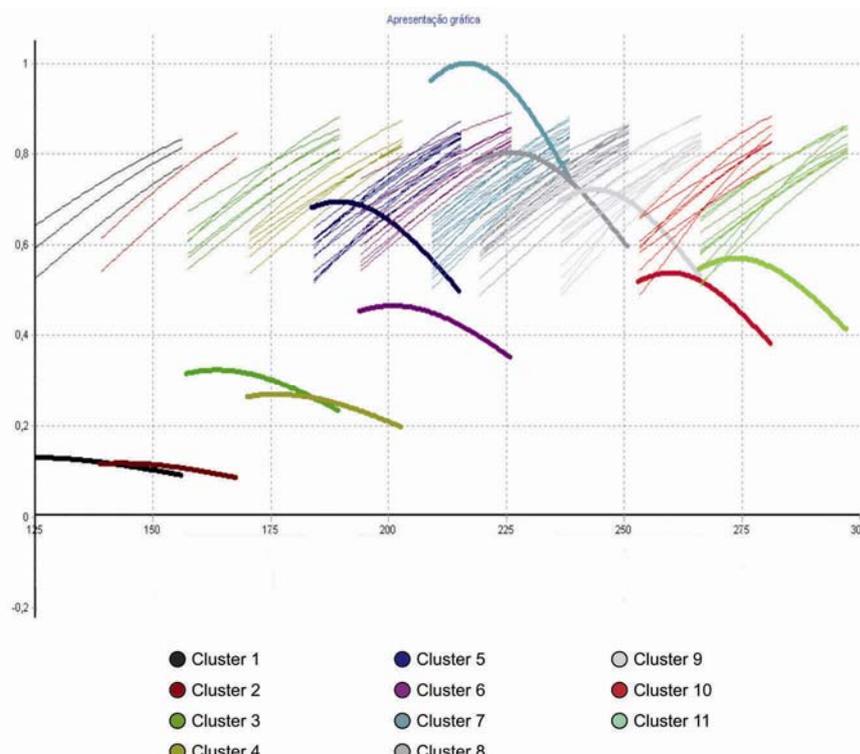
- Observa-se que a distribuição dos itens por *cluster* apresenta um padrão bastante irregular, que pode ser visualizado pela série de picos e vales indesejáveis mostrada no Gráfico 6.

Gráfico 6 - Distribuição do número de itens ao longo dos 11 *clusters* formados



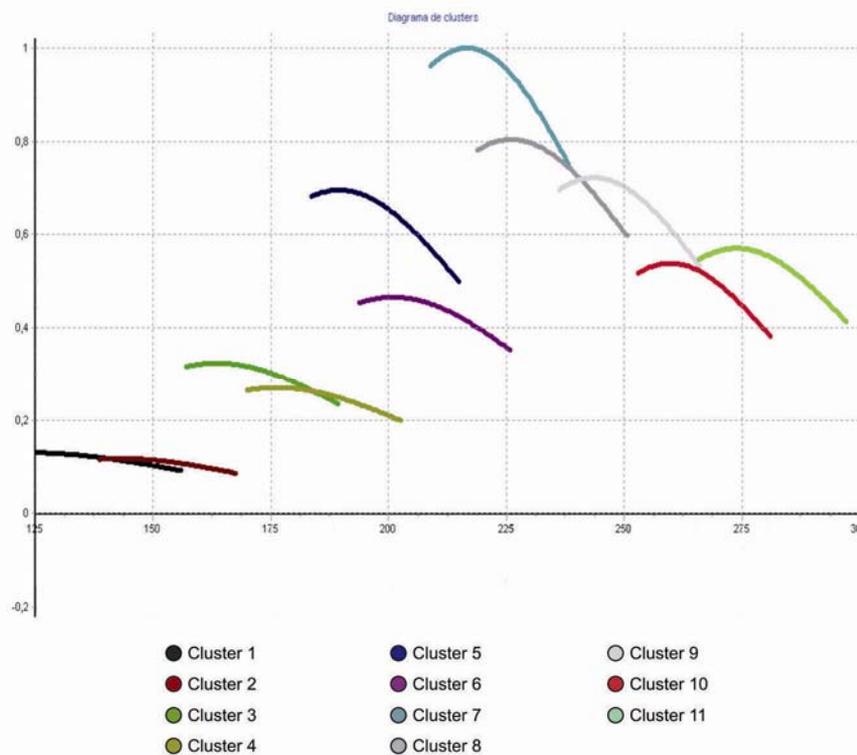
No Gráfico 7, são apresentadas na mesma cor as curvas de informação dos *clusters* e as curvas características dos itens (CCI) de cada *cluster* dentro de seus respectivos limites.

Gráfico 7 - Distribuição dos itens ao longo dos 11 *clusters* formados



Por sua vez, o Gráfico 8 mostra a distribuição apresentada pelos *clusters*:

Gráfico 8 - Distribuição dos 11 *clusters* formados



Cabe, aqui, a explicitação de que as quantidades de informação e sua representação gráfica evidenciam a importância do *cluster* para a constituição geral da escala considerada, uma vez que estão associadas ao número de itens presentes em cada *cluster*, ao poder de discriminação e ao grau de dificuldade dos itens. Em outros termos, elas demonstram o grau de representatividade daqueles conjuntos de itens em relação às habilidades avaliadas.

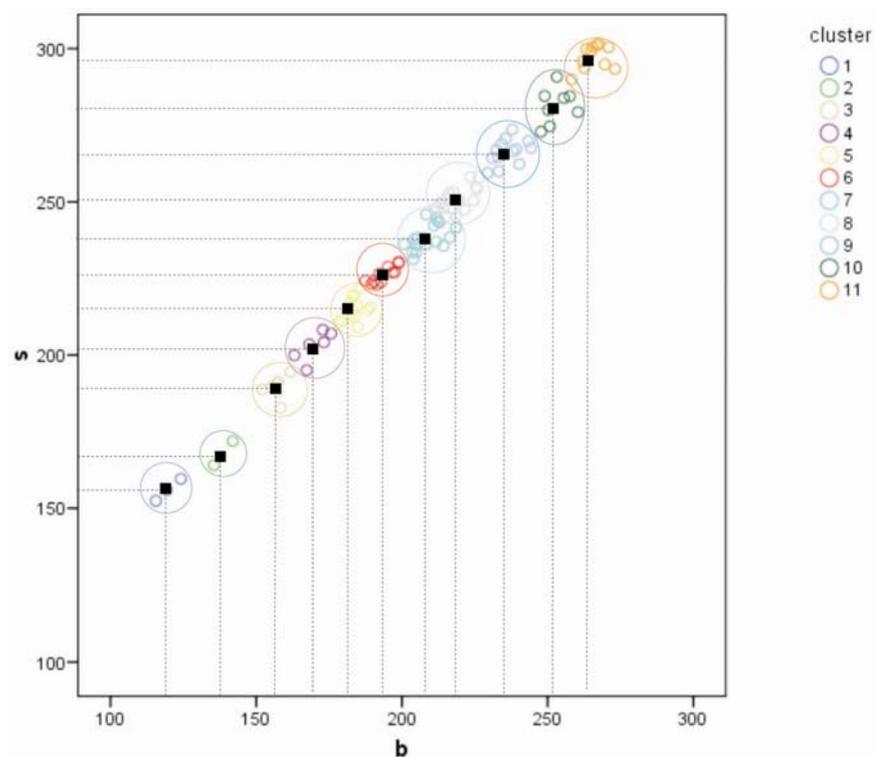
Pode-se também observar que o gráfico da Curva de Informação do *cluster*, na maioria das vezes, apresenta-se em formato de sino, atingindo seu ápice próximo ao ponto médio do *cluster*. Quanto mais alto for esse vértice, mais informações o *cluster* tem, e, conseqüentemente, maior é a sua importância para a interpretação da escala.

Na distribuição das curvas de informação dos *clusters*, apresentada no Gráfico 8, pode-se observar que há uma série de sobreposições, notadamente no intervalo entre 225 e 250 pontos.

Constata-se, ainda, a ausência de uma distribuição equilibrada das curvas de informação dos *clusters*, como se observa, por exemplo, no *cluster* 4, que é sobreposto pelos *clusters* 3, 5 e 6. Isso dificulta sua interpretação pedagógica, pois, além de ele apresentar apenas 6 itens, quase toda a sua informação está presente nos *clusters* adjacentes. Isso quer dizer que o nível de dificuldade dos itens que constituem o *cluster* 4 é semelhante aos níveis de dificuldade dos itens que compõem os *clusters* 3, 5 e 6, demandando, portanto, dos alunos, esforço semelhante de raciocínio. Quando esse fato ocorre, observa-se que o conjunto de habilidades do *cluster* sobreposto, de certo modo, estende-se para a região correspondente ao auge e à consolidação das curvas de informações dos *clusters* adjacentes.

Esse aspecto também pode ser verificado, a seguir, no Gráfico 9, que apresenta a distribuição dos itens por cada um dos seus respectivos *clusters*, ao mesmo tempo em que evidencia o caráter bidimensional do processo de conglomerados aqui adotado, ou seja, apresenta as centróides dos *clusters* e seus respectivos itens na forma de pontos definidos pelas duas variáveis notáveis aqui consideradas: **b** e **s**. Percebe-se, neste caso, a ocorrência de uma proximidade acentuada entre os valores de **b** pertencentes a *clusters* diferentes. Tal fato é, naturalmente, indesejável, visto que corresponde a uma baixa capacidade de diferenciação entre *clusters* adjacentes.

Gráfico 9 - Representação bidimensional das centróides dos 11 *clusters* formados e seus respectivos itens



Outro fato que se percebe é a ocorrência de itens não muito distantes entre si e agrupados em *clusters* distintos, como, por exemplo, os itens que compõem os *clusters* 5 e 6, o que significa que esses itens, provavelmente, deveriam pertencer a um mesmo *cluster*.

Pelas considerações aqui apontadas, constatam-se resultados indesejáveis na solução obtida com 11 *clusters*. Isso sugere a necessidade de uma nova organização dos agrupamentos, com o objetivo de se produzir uma melhor interpretação pedagógica. Pode-se, assim, reutilizar o método automático do *k-means*, variando-se a quantidade de *clusters* desejada com base nas questões apontadas. Para se verificar se uma solução adequada foi alcançada, são feitas as seguintes considerações:

i) O intervalo característico dos *clusters* não deve ser muito amplo, visto que os *clusters* devem ser, internamente, mais homogêneos e, externamente, mais disjuntos, ou heterogêneos.

ii) O número de *clusters* não pode ser muito grande, pois um conjunto de habilidades próximas pode não pertencer a *clusters* diferentes; tampouco o número de *clusters* pode ser reduzido a ponto de agrupar habilidades consolidadas em distâncias muito longas na escala de proficiência.

iii) Os *clusters* devem ser distribuídos de tal forma que permita a abrangência de todos os níveis da escala de proficiência relevantes para o período de escolaridade avaliado, desde os níveis mais baixos até os mais altos. É importante ressaltar-se que os níveis de proficiência mais altos têm importância, pois se referem às habilidades comuns ao período de escolaridade subsequente, o que garante as condições de equalização entre diferentes etapas escolares, apesar de poucos alunos se encontrarem nesse nível.

iv) Finalmente, a distribuição dos *clusters* deve levar em consideração a obtenção de mais informação em níveis de proficiência onde se espera encontrar a maior parte dos escores atribuídos aos alunos avaliados em um determinado período de escolaridade. Isso permitirá que as habilidades desses alunos sejam melhor caracterizadas e que se estabeleçam precisos padrões de discriminação entre eles.

Para alcançar-se uma solução que contemple adequadamente essas questões, produziu-se uma readequação do número de *clusters*, visando à obtenção de um grau significativo de informação sobre o nível de desempenho escolar dos alunos avaliados. Chegou-se, assim, a um número de 6 *clusters*, correspondente à solução que será descrita a seguir.

3.2.3

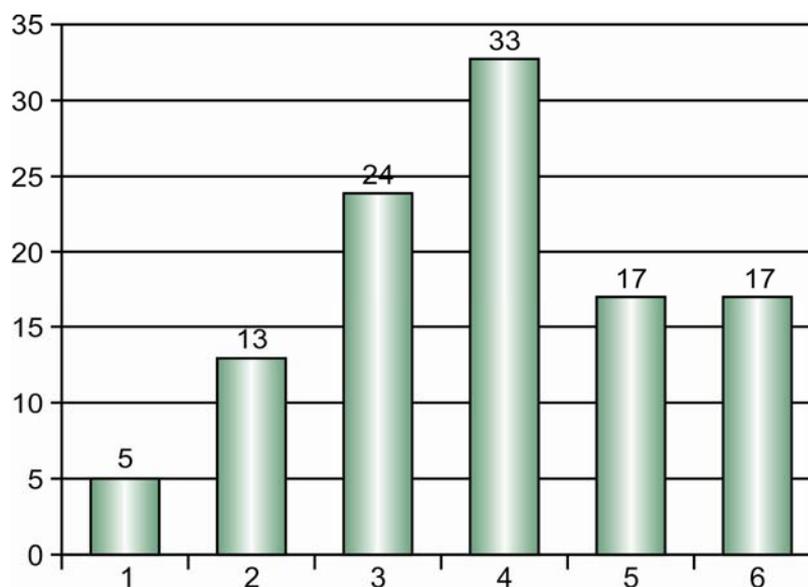
Características gerais do agrupamento de itens após a obtenção da solução com 6 *clusters*

Os resultados obtidos com a solução de 6 *clusters* estão apresentados de forma detalhada no Anexo XII e de forma sintética na Tabela 18.

Tabela 18 - Os 6 *clusters* formados

N. do cluster	N. de itens	Quantidade de informação	Coordenada x do Centróide (média dos b) = Limite Inferior do grupo	Coordenada y do Centróide (média dos s) = Limite Superior do grupo	menor b	maior b	menor s	maior s
1	5	0,47860076	127,3	160,77	115,60	141,92	152,42	171,87
2	13	0,111522463	163,05	195,62	152,15	175,68	182,64	208,21
3	24	0,213235272	188,31	220,16	178,70	198,93	209,06	230,45
4	33	0,306057297	213,11	243,75	200,67	225,82	231,33	254,61
5	17	0,170836133	236,5	266,19	223,47	250,75	257,96	274,60
6	17	0,166757204	261,75	292,52	248,96	273,05	279,25	301,61

Nesse novo agrupamento, constata-se uma boa distribuição dos itens dentro dos *clusters*, uma vez que apresenta um padrão aproximadamente simétrico, com poucos itens nas bordas e uma maior quantidade de itens nos *clusters* centrais, conforme mostra o Gráfico 10, a seguir:

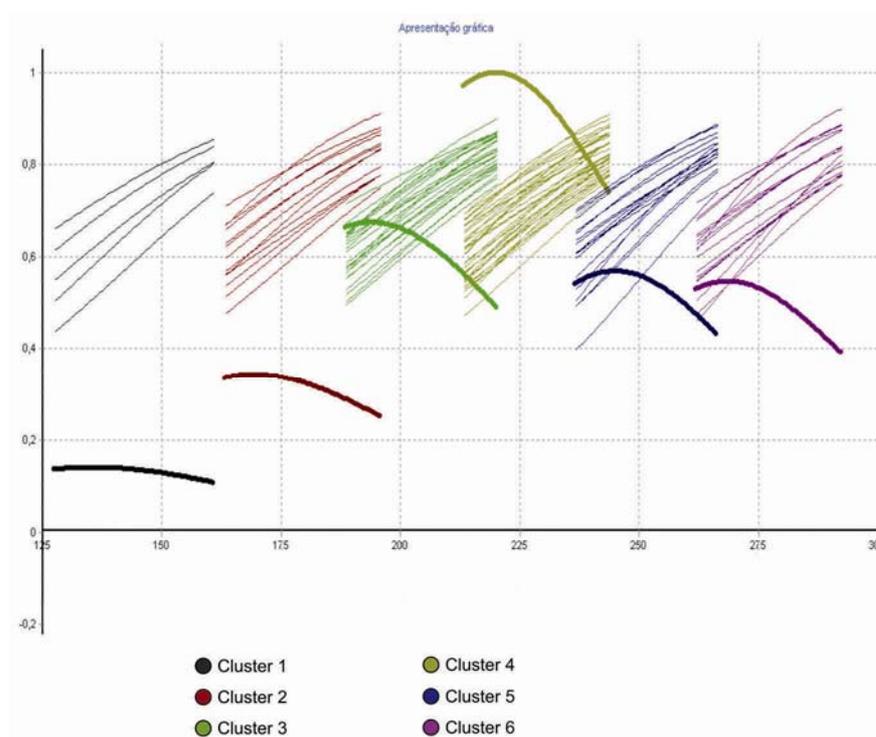
Gráfico 10 - Distribuição do número de itens ao longo dos 6 *clusters* formados

Essa disposição apresenta uma pequena assimetria, uma vez que ocorre uma distribuição mais prolongada à medida que os *clusters*

caminham para a direita. Esse fato é proveniente de dois fatores: (i) é mais difícil que sejam elaborados itens com boa discriminação em níveis baixos da escala; (ii) é necessário que se produzam informações sobre as habilidades desenvolvidas pelos alunos situados nos níveis extremos da escala, principalmente nos níveis mais altos, com vistas à equalização com o período escolar subsequente. A distribuição, no entanto, é adequada, à medida que permite uma seleção de itens bastante representativos do período de escolaridade considerado, o que resulta informações pedagógicamente mais ricas.

O Gráfico 11 apresentado a seguir mostra esse novo agrupamento com as curvas de informação dos *clusters* combinadas com as curvas características dos seus respectivos itens componentes.

Gráfico 11 - Distribuição dos itens ao longo dos 6 *clusters* formados



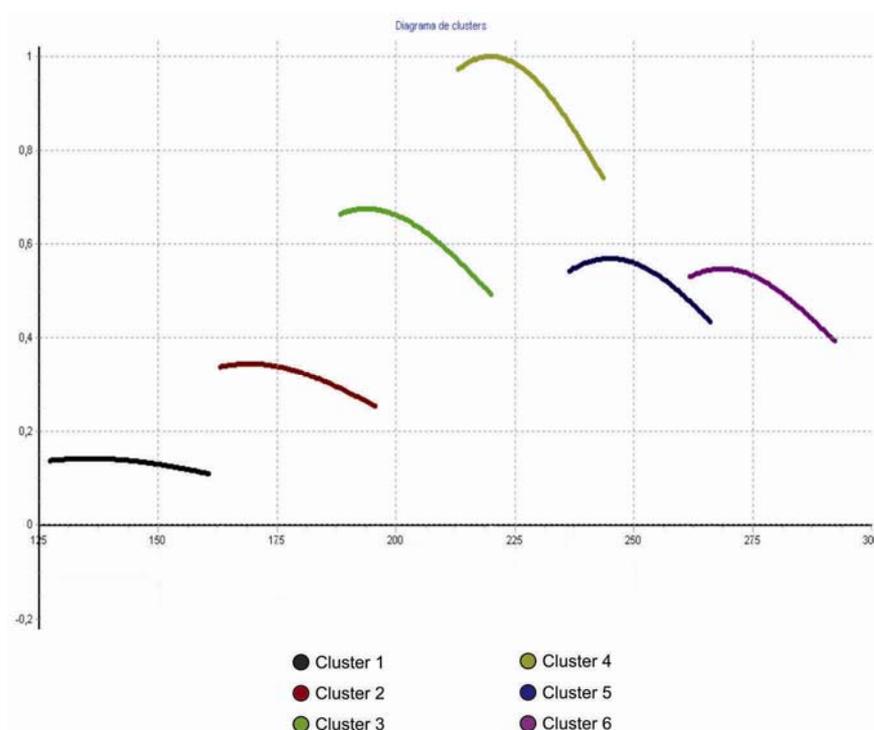
Percebe-se que os *clusters* formados apresentam larguras adequadas: nem muito estreitas, que os impeçam de agrupar itens próximos, nem muito largas, que lhes permitam agrupar itens com dificuldades muito distintas. A distância entre os centróides dos *clusters* foi expressiva. E, embora ainda ocorram intercessões entre seus limites,

essas intercessões localizam-se depois do vértice da curva de informação, onde se obtém a maior informação do *cluster*, não comprometendo, portanto, sua interpretação.

É importante lembrar, no entanto, que se trata de uma escala **contínua**, onde intercessões inevitavelmente ocorrerão. Afinal, elas representam situações que passam por diferentes etapas de consolidação das habilidades avaliadas, de forma que, enquanto um conjunto dessas habilidades começa a se desenvolver, outras habilidades já estão consolidadas, e algumas outras sequer começaram o seu desenvolvimento. Portanto, intercessões localizadas nos ramos ascendentes ou descendentes das curvas de informação adjacentes não prejudicam, substantivamente, a análise a que este estudo se propõe.

O Gráfico 12 a seguir explicita a distribuição desses 6 novos *clusters* obtidos ao longo da escala de proficiência considerada.

Gráfico 12 - Distribuição dos 6 *clusters* formados



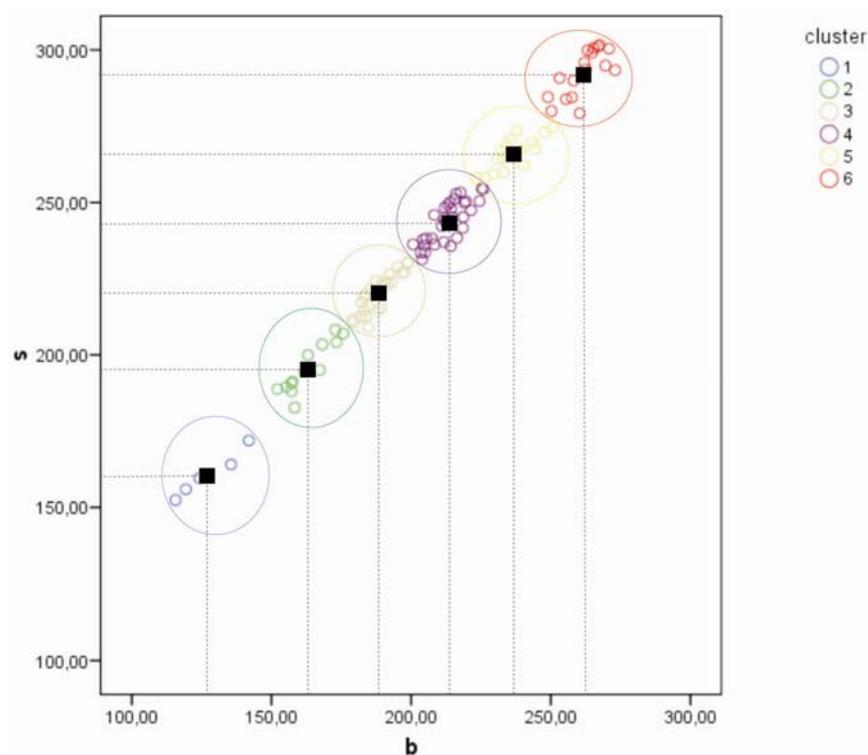
Observa-se que a largura média dos *clusters* gira em torno de 31 pontos na escala de proficiência. Apesar de as dimensões serem um pouco maiores que as usualmente adotadas pelo SAEB, quando são considerados os intervalos de 25 pontos na parte da escala que vai até 200, e menores que os 50 pontos adotados nos intervalos da escala de 200 a 350 (ela retorna aos 25 pontos no intervalo 350-375), pode-se constatar que eles não perdem homogeneidade, pois os elementos norteadores na construção desses *clusters* tiveram como finalidade buscar a maior proximidade possível entre os itens que compõem cada um deles.

Outra observação importante se refere à probabilidade de acerto do item dentro de cada *cluster*. Verifica-se que, no início dos *clusters*, a maioria dos itens apresenta um percentual de acerto esperado entre 50% e 70%, sendo raros os itens com probabilidade superior a 70%. Ao final dos *clusters*, o percentual de acerto esperado está entre 70% e 80%. Assim, o conjunto específico de habilidades de cada *cluster* já se apresenta praticamente consolidado em seu limite superior.

De maneira geral, esse resultado é consequência de duas características-chave muito importantes: os itens possuem boa discriminação e estão próximos do centróide, características que lhes conferem um elevado grau de representatividade relacionada às suas habilidades específicas, o que, conseqüentemente, proporcionará uma interpretação pedagógica mais rica e acurada das habilidades e competências desenvolvidas pelo alunado.

O Gráfico 13 evidencia novamente o caráter bidimensional do processo de conglomerados, ilustrando a distribuição dos *clusters* e de seus respectivos itens constituídos em função das variáveis **b** e **s**. Diferentemente do que se havia constatado no Gráfico 9, que foi construído com base em 11 *clusters*, percebe-se agora um maior distanciamento entre os valores de **b** de *clusters* adjacentes, o que, naturalmente, corresponde a uma maior diferenciação entre os *clusters*.

Gráfico 13 - Representação bidimensional das centróides dos 6 *clusters* formados e seus respectivos itens



Um aspecto vantajoso e imediatamente perceptível após a obtenção dos *clusters* finais é o fato de o número de *clusters* - seis - ser adequado aos propósitos de interpretação pedagógica dessa parte da escala de proficiência do SAEB, uma vez que tais *clusters* apresentam as habilidades essenciais para a 4ª série do ensino fundamental.

Assim sendo, percebe-se que os itens pertencentes a cada grupo apresentam fortes semelhanças quanto ao nível de dificuldade das habilidades, ao mesmo tempo em que as distâncias entre os *clusters* são grandes o suficiente para permitirem uma análise mais abrangente das habilidades avaliadas e dos saltos qualitativos significantes. Esses saltos possibilitam a associação dos diferentes *clusters* a diferentes níveis de proficiência, o que garante boas condições para a interpretação de uma escala que se baseia no comportamento de itens.

No modelo proposto, em particular, e no modelo do GERES, em geral, a utilização das curvas características dos itens favorece, conforme já se mencionou na primeira parte deste estudo, uma

interpretação mais fundamentada do comportamento dos itens, de modo a evitar-se a arbitrariedade observada em modelos como o do SAEB, no qual, com frequência, itens de comportamento significativamente diferente entre si eram agrupados em um mesmo nível, apenas por satisfazerem a determinados critérios de percentuais de acerto por nível. A interpretação aqui proposta evita esse procedimento.

Outra vantagem associada à utilização das curvas características dos itens é o uso do modelo de três parâmetros, que possibilita, nesse caso, uma descrição mais precisa do comportamento de cada item, pois leva em conta não apenas sua dificuldade, mas também seu poder de discriminação (variável, muito embora seja elevado para todos os itens selecionados) e o acerto casual, por se tratar de itens de múltipla escolha. A utilização desse modelo confirma ser mais conveniente que o modelo de Rasch, utilizado, por exemplo, no PISA, uma vez que os três parâmetros possibilitam um tratamento mais acurado e realista dos itens da amostra em questão.

Outro ponto favorável à solução de seis *clusters* é a abrangência de sua cobertura aos níveis de habilidade. Percebe-se que essa cobertura é praticamente completa para o espectro de habilidades referentes ao período de escolaridade considerado neste estudo, correspondente à primeira etapa do Ensino Fundamental. Ou seja, os seis *clusters* formados distribuem-se ao longo da escala em um intervalo compreendido entre 125 e 275 pontos de proficiência, com larguras semelhantes entre si, cerca de 31 pontos, e com a ocorrência de intervalos muito estreitos separando os diversos *clusters*. Dessa forma, a solução proposta evita a indesejável seleção de itens característicos somente para faixas muito restritas de habilidade, limitação apontada nas escalas do NAEP e do TIMSS, conforme apresentado na primeira parte deste trabalho.

Portanto, o que aqui se propõe é a constituição de seis níveis de proficiência univocamente associados aos seis *clusters* formados, conforme mostra o Quadro 7, a seguir.

Quadro 7 - Níveis de habilidade e proficiências características¹³

Nível	Auge do desenvolvimento da habilidade	Consolidação da habilidade
1	127	161
2	163	196
3	188	220
4	213	244
5	237	266
6	261	292

Nesse quadro, percebe-se que os seis níveis de proficiência estão associados aos pontos **b** e **s** de seus *clusters* correspondentes. Portanto, cada nível fica caracterizado pelos dois pontos críticos de cada cluster: a média dos pontos referentes ao auge da aprendizagem e a média dos pontos onde ocorre a consolidação da aprendizagem.

Os alunos do banco de dados considerado foram então agrupados ao longo dos intervalos que compõem cada um desses seis níveis de proficiência. O objetivo deste procedimento foi distribuir os percentuais de alunos ao longo de quatro intervalos que se consideraram como relevantes para fins de interpretação de cada nível: o intervalo onde a proficiência dos estudantes se encontra aquém do ponto que lhes permitiria acertar os itens do *cluster* em questão; o intervalo onde se observa o máximo crescimento da aprendizagem necessária ao acerto dos itens; o intervalo no qual se inicia a consolidação dessa aprendizagem e, finalmente, o intervalo correspondente à região do *continuum* onde a aprendizagem já se encontra consolidada. Isto é o que se observa na Tabela 19, a seguir:

¹³ A descrição e a interpretação dos níveis de habilidade aqui considerados serão tratadas na Parte 3 deste estudo.

Tabela 19 - Distribuição dos alunos pelos níveis de habilidades

Nível	Aquém das habilidades do nível		Auge do desenvolvimento das habilidades do nível		Em vias de Consolidação das habilidades do nível		Habilidades do nível consolidadas	
	Intervalo	% alunos	Intervalo	% alunos	Intervalo	% alunos	Intervalo	% alunos
1	< 116	7,3	116 a 142	15,9	142 a 161	16	>= 161	60,8
2	< 152	31,6	152 a 176	21,1	176 a 196	15,6	>= 196	31,8
3	< 179	55	179 a 199	15,3	199 a 220	12,3	>=220	17,4
4	< 201	71,7	201 a 226	13,6	226 a 244	6,7	>= 244	8
5	< 223	84	223 a 251	9,8	251 a 266	2,8	>= 266	3,4
6	< 249	93,3	249 a 273	4,2	273 a 292	1,4	>= 292	1,1

Definidos, portanto, os seis níveis de proficiência com base no processo de agrupamento dos itens, passa-se à interpretação pedagógica da escala de proficiência, que será abordada na terceira parte, apresentada a seguir.