

5 Considerações Finais

Na primeira parte deste trabalho, foram analisados os critérios de seleção de itens representativos de diferentes níveis de proficiência empregados por cinco programas de avaliação em larga escala: *Sistema Nacional de Avaliação da Educação Básica (SAEB)*, *National Assessment for Educational Progress (NAEP)*, *Programa Internacional de Avaliação de Alunos (PISA)*; *Trends in International Mathematics and Science Study (TIMSS)* e *Estudo Longitudinal sobre a Qualidade e Equidade no Ensino Fundamental Brasileiro - Projeto GERES - 2005*.

Todos os programas considerados têm em comum alguns aspectos básicos, como a constituição de uma sucessão de itens selecionados, determinados ao longo de um *continuum* de proficiência e que fornecem elementos fundamentais à interpretação das escalas de proficiência. Porém, se os diversos programas se assemelham quanto a esse aspecto geral, por outro lado, apresentam marcantes diferenças quanto à determinação dos níveis de proficiência e aos critérios de seleção dos itens característicos desses níveis.

Uma variação observada entre os diferentes programas associou-se à quantidade de itens selecionados por cada critério, como mostra o Quadro 8. Verificou-se que há critérios mais restritos (SAEB 1995-1997, NAEP e TIMSS), segundo os quais apenas um subconjunto do total de itens é selecionado, ao passo que outros programas (SAEB 1999-2007, PISA e GERES) permitem a inclusão de todos os itens selecionados. A opção pela inclusão de todos os itens demonstrou ser mais adequada, pois permite um maior aproveitamento do banco de itens, o que, por sua vez, favorece uma exploração mais completa das possibilidades de interpretação pedagógica da totalidade de níveis de habilidade considerados.

Quadro 8

Programa de Avaliação	Número de Itens Selecionados	%
SAEB (1999-2007); PISA; GERES	168	100
TIMSS	50	30
SAEB (1995-1997); NAEP	39	23

Outro ponto de interesse foi a amplitude dos níveis de seleção constituídos de forma contínua, ou descontínua, como mostra o Quadro 9, que permitiu verificar a existência, ou não, de intervalos de proficiência entre níveis consecutivos. Percebeu-se, então, que o NAEP e o TIMSS basearam-se em níveis de seleção consideravelmente estreitos, centrados em torno de pontos de seleção específicos. No caso específico do NAEP, esse procedimento resultou em apenas dois níveis de referência para a 4ª série do Ensino Fundamental, visto que os níveis posteriores referiam-se às séries mais avançadas consideradas por esse programa (8ª e 12ª). Essa opção de se concentrar em apenas alguns níveis de largura muito limitada no *continuum* de habilidade faz com que um grande percentual de itens e de alunos avaliados sejam excluídos da análise, o que parece corresponder a uma restrição indesejável do banco de dados disponível.

Por outro lado, os demais programas preferiram dividir todo o espectro de proficiência em níveis de amplitude predominantemente iguais entre si e dispostos de modo contínuo ao longo da escala de proficiência. Essa segunda opção parece ser mais interessante, pois permite interpretações da escala ao longo do espectro completo de proficiência, o que, também oferece possibilidades mais ricas de interpretação pedagógica das habilidades desenvolvidas.

Quadro 9

Programa de Avaliação	Distribuição dos Níveis de Referência
SAEB (1995-1997); SAEB (1999-2007); PISA; GERES	Contínua
TIMSS; NAEP	Descontínua

Um problema comumente constatado foi a concentração de itens selecionados nos níveis médios de proficiência, com um conseqüente “despovoamento” desses itens na direção das extremidades, tanto inferior, quanto superior, de proficiência¹⁴. Observou-se que tal problema foi, na verdade, inevitável, pois a informação de cada item é maior na região central de sua respectiva curva característica, e é, geralmente difícil encontrar itens cujas curvas características estejam muito deslocadas na direção desses extremos. Um modo de diminuir esse efeito foi a adoção de relaxamentos dos critérios de seleção conforme mostra o Quadro 10. Desse modo, foi possível, em sistemas como o SAEB1999 - 2007, por exemplo, conseguir um número razoável de itens nesses níveis extremos.

Quadro 10

Programa de Avaliação	Adoção de Mais de 1 Critério de Seleção
SAEB (1999-2007); PISA; GERES	Não
SAEB (1995-1997); NAEP; TIMSS	Sim

Outra questão fundamental relacionada à seleção dos itens representativos foi o problema da arbitrariedade na definição dos níveis de seleção, e também na destinação dos itens a esses níveis. Em alguns casos, itens de dificuldades relativamente próximas entre si foram

¹⁴ No NAEP, essa ausência relativa de itens foi verificada apenas no nível inferior de proficiência, devido ao fato de que só foram considerados dois níveis-âncora, com o segundo nível correspondendo a valores intermediários de proficiência para a escala como um todo (que vai até a 12ª série). Dessa forma, a escala considerada para a 4ª série foi truncada, de modo a conservar somente os dois primeiros níveis originais.

incluídos em níveis diferentes, e também houve casos de itens com curvas características bastante diferentes entre si serem incluídos em níveis iguais de seleção. Muitas vezes, essas distorções resultaram de critérios de seleção dependentes dos percentuais de acerto por nível de proficiência (como no caso do SAEB, tanto o 1995-1997, quanto o 1999-2007). Por outro lado, em sistemas como o GERES, que não se basearam nesse tipo de critério, conseguiu-se uma determinação da seleção pautada em características mais fundamentais, como a dificuldade de cada item, ou o ponto de consolidação da proficiência necessária para acertá-lo, o que, conseqüentemente, resultou numa diminuição considerável do grau de arbitrariedade.

Nesse sentido, dois sistemas se destacaram dos demais, o PISA e o GERES, pois ambos basearam-se em pontos fixos das curvas características de cada item, a saber, os pontos correspondentes à discriminação máxima (50% de probabilidade de acerto para o modelo de Rash, ou $50\% + c/2$ para o modelo de três parâmetros) e à consolidação da aprendizagem (80% de probabilidade de acerto). Além de se basear em critérios mais fundamentais e, conseqüentemente, menos arbitrários, a adoção desse procedimento possibilitou a inclusão da totalidade dos itens do banco e a cobertura de todo o nível de proficiência considerado, abrindo caminho para uma interpretação pedagógica dos itens baseada em dois pontos de importância crucial na aprendizagem, quais sejam, o auge do desenvolvimento e a consolidação das habilidades necessárias ao acerto de cada item.

Quadro 11

Programa de Avaliação	Critério para a Seleção dos Itens Característicos dos Níveis de Proficiência
SAEB (1995-1997); SAEB (1999-2007); TIMSS	Diferenças entre percentuais de acerto nos níveis considerados
PISA	Probabilidades de acerto dos itens associadas às CCI de 1 parâmetro
NAEP	Diferenças entre percentuais de acerto ou probabilidades de acerto associadas às CCI de 3 parâmetros
GERES	Parâmetros b e s das CCI de 3 parâmetros

Uma alternativa proposta para o tratamento dessas questões foi o emprego de uma análise de conglomerados (*cluster analysis*) para a escolha de um subconjunto de itens característicos de cada nível de proficiência. O agrupamento de itens assim realizado teve como referência dois pontos notáveis: o ponto **b** (correspondente ao ponto de maior inclinação da curva característica do item - CCI - e interpretado como sendo o ponto onde a habilidade encontra-se no auge de seu desenvolvimento); e o ponto **s** (correspondente à maior taxa de decrescimento da inclinação da CCI e interpretado como o ponto onde se consolida a habilidade requerida para se acertar o item). A utilização simultânea desses dois pontos notáveis permitiu a seleção dos *clusters* e a constituição de grupos homogêneos em função do posicionamento desses dois pontos e, também, da forma das CCIs dos itens considerados.

Uma das vantagens decorrentes dessa metodologia reside no fato de que a interpretação da escala assim obtida é orientada pelo agrupamento dos itens com base em suas respectivas curvas características, e não pela seleção de itens a partir de níveis de proficiência predeterminados. Tal método tinha, entre outros inconvenientes, o fato de apresentar uma grande incerteza prévia quanto ao número de itens considerados para cada nível de referência.

Outra vantagem obtida com o uso do método de conglomerados foi o fato de que o seu emprego evitou a perda de itens, cujos respectivos parâmetros **b** e **s** estão situados próximo da fronteira de proficiência entre dois níveis pré-fixados (um problema comumente verificado em alguns dos programas de avaliação aqui considerados), permitindo a agregação das informações fornecidas por esses itens aos seus respectivos *clusters*, o que favoreceu a interpretação da escala.

Também relevante foi a utilização exclusiva, por parte do método de conglomerados, de itens com uma elevada capacidade de discriminação, visto que todos eles tinham um parâmetro **a** $\geq 1,15$. Tal característica, acrescida do fato de os itens encontrarem-se próximos do centróide de seus respectivos *clusters*, conferiram-lhes um elevado grau de representatividade, relacionada as suas habilidades específicas, o

que, conseqüentemente, proporcionou uma interpretação pedagógica mais rica e acurada das habilidades e competências desenvolvidas pelo alunado.

Ainda outro aspecto vantajoso do processo de clusterização foi o fato de ter permitido a adoção de um critério único para a seleção dos itens representativos de cada nível, o que conferiu uma maior coerência e homogeneidade ao processo de seleção.

O modelo de três parâmetros para as curvas características dos itens possibilitou uma descrição mais precisa do comportamento de cada item, visto que levou em conta não apenas sua dificuldade, mas também seu poder de discriminação (variável com valores elevados para todos os itens selecionados) e o acerto casual, por se tratar de itens de múltipla escolha.

Um papel relevante cumprido pela constituição dos *clusters* que caracterizam os níveis de proficiência foi minimizar a concentração de itens nas fronteiras entre eles, o que permitiu uma interpretação mais nítida e acurada de cada nível de proficiência, propiciando conclusões fundamentais para a efetiva compreensão dos resultados alcançados e levou às seguintes constatações:

- À medida que as médias dos centróides que definem os *clusters* de proficiência vão aumentando, uma habilidade desenvolvida parece funcionar como um recurso mobilizável por habilidades mais complexas.

- Uma habilidade desenvolvida pressupõe a existência de outros recursos mobilizáveis, mas não se confunde com eles.

- O desenvolvimento das habilidades ocorre de modo gradual e progressivo, e esse processo vai sendo enriquecido, à medida que os *clusters* considerados passam a se referir a graus progressivamente maiores de dificuldade ou complexidade de cada habilidade.

- As habilidades desenvolvidas ao longo dos *clusters* de proficiência consistem notadamente em detectar, pouco a pouco, analogias que não se mostram à primeira vista.

- Um conjunto de habilidades desenvolvido demonstra a constituição de competências importantes relacionadas às séries iniciais do Ensino Fundamental.

De posse desses elementos, pode-se considerar que a construção e a interpretação da escala de proficiência constituída pelos *clusters* representam uma possibilidade concreta de oferecer aos educadores subsídios valiosos para o planejamento, acompanhamento e revisão dos processo de ensino e aprendizagem desenvolvidos pelas escolas.

Acrescente-se a isso a possibilidade de estabelecer uma melhor comunicação pedagógica com os gestores e os professores, tendo em vista o debate, nas redes de ensino e unidades escolares, sobre as metas que podem ser estabelecidas a partir da compreensão e análise das habilidades desenvolvidas pelos alunos situados em cada um dos 6 *clusters*. Assim, a interpretação qualitativa da escala de proficiência pode propiciar às escolas condições para detectar os principais problemas de aprendizagem dos alunos e, ao mesmo tempo, estabelecer planos de intervenção pedagógica que visem reduzir os índices de reprovação e de abandono da escola, elevando o nível de desempenho dos alunos e, conseqüentemente, promovendo a igualdade de oportunidades educacionais.