Construção e Validação de uma Prova de Matemática para Alunos do 1º ao 4º Ano de Escolaridade

Development and Validation of a Math Test for 1st to 4th Grade Students

João Lopes[*, a] & Maurício Bueno[b]
[a]*Universidade do Minho, Braga, Distrito de Braga, Portugal*
& [b]Universidade Federal de Pernambuco, Recife, Pernambuco, Brasil

Resumo
Este estudo tem como objetivo apresentar um instrumento de 46 itens para avaliação de conhecimentos matemáticos para os primeiros quatro anos de escolaridade. Participaram no estudo 505 crianças portuguesas do 1º ao 4º ano do primeiro ciclo, com idades compreendidas entre os seis e os 12 anos. Uma análise fatorial de primeira e de segunda ordem revelou a existência de cinco fatores primários que se agrupam num único fator de segunda ordem relacionado com o conhecimento geral em matemática. Uma análise psicométrica com base no modelo de Rasch da Teoria de Resposta ao Item (TRI) permitiu depurar as propriedades psicométricas dos itens. A validade desenvolvimental do instrumento para os quatro anos de escolaridade abrangidos foi igualmente investigada e discutida.
Palavras-chave: Conhecimentos matemáticos, prova de conhecimentos, teoria de resposta ao item.

Abstract
This study presents a forty-six item instrument for the assessment of mathematical knowledge for first to fourth graders. Participants are 505 primary school Portuguese students from 1st to 4th grade, between six and 12 years old. A first and second order factor analysis of the items shows that there are five primary factors that are grouped in a single factor of second order related to general mathematical knowledge. A psychometric analysis based on Rasch's model of the Item Response Theory (IRT) was conducted in order to clear up the psychometric properties of the items. The developmental validity of the instrument for primary grades is also reported and discussed.
Keywords: Mathematical knowledge, knowledge test, item response theory.

É hoje generalizadamente aceito que a construção do conhecimento matemático segue um caminho lento e gradual que vai do concreto e específico ao abstrato e geral (Citoler, 1996). É também usualmente aceito que as competências matemáticas elementares são decomponíveis em subcompetências que incluem a noção de número e sistema numérico (numeracia), o cálculo (nas suas versões de papel e lápis e de cálculo mental, entre outras), a estimativa, a resolução de problemas e os conceitos de medida e noções geométricas.

Nas últimas décadas, disciplinas como a Psicologia Cognitiva têm exercido uma influência significativa (mas de forma nenhuma única) na conceituação do conhecimento matemático, nomeadamente através do desenvolvimento de modelos progressivamente mais sofisticados de processamento da informação matemática (e. g. Brannon, 2002; Dixon, Deets, & Bangaert, 2001), os quais têm evidenciado a importância da automatização de processos básicos como forma de liberar recursos para a realização de operações mais complexas (Jensen & Whang, 1994). De fato as capacidades limitadas de processamento da informação inibem a alocação de recursos a operações complexas, quando estes recursos estão a ser utilizados na realização de operações básicas (Singer-Dudek & Greer, 2005). Os processos elementares têm pois que sofrer um processo de sobreaprendizagem até ser alcançada a automatização e inibida sua interferência nos processos mais complexos (Zentall, Smith, Lee, & Wieczoreck, 1994).

Seguindo essa tendência, em muitos países a estrutura curricular para a matemática no ensino elementar tende a privilegiar as áreas que a literatura refere como estruturantes do conhecimento matemático posterior: o conhecimento do sistema numérico, o cálculo (escrito e mental, resolução de algoritmos, etc.) e a resolução de problemas (para além da geometria, que de certa forma constitui uma área à parte; Ashlock, 2005).

O conhecimento do sistema numérico refere-se à compreensão de como opera e como se organiza um sistema numérico de base 10, de como os números se podem combinar para a realização de operações, do valor

[*] Endereço para correspondência: Departamento de Psicologia, Universidade do Minho, Campus de Gualtar, Braga, Portugal 4710-057. E-mail: joaols@psi.uminho.pt e talktomau@gmail.com

posicional dos algarismos nos números (unidades, dezenas, centenas), etc. (Dehaene, 2001; Gurganus, 2004; Jordan, 2007; Jordan, Glutting, & Ramineni, 2010). O cálculo é representado pelas operações aritméticas elementares de soma, subtração, multiplicação e divisão e respectivas representações algorítmicas e formatos (e. g. papel e lápis, cálculo mental) possíveis (Correa & Moura, 1997; Haydu, Costa, & Pullin, 2006). Implica conhecimentos procedimentais e conceituais apurados, ou seja, saber *como se faz* e compreender *o que se está a fazer* (Moore, Dixon, & Haine, 1991). Dixon et al. (2001) sugerem que para cada uma das quatro operações aritméticas básicas as pessoas utilizam quatro princípios: monotonicidade, relação com os operandos, direção, gradiente.

O princípio da monotonicidade afirma que se os operandos variam numa determinada direção, o mesmo deverá acontecer com o resultado. O princípio da relação com os operandos, especifica a relação do resultado com os operandos de acordo com o tipo de operação. Para a adição e para a multiplicação o resultado deve ser maior do que qualquer dos operandos. Para a subtração e para a divisão o resultado deve ser menor que o minuendo e o dividendo respectivamente. O princípio da direção do efeito especifica se as mudanças num operando têm como efeito a diminuição ou aumento dos resultados. Na adição e na multiplicação, o aumento em qualquer dos operandos resulta no aumento do resultado, enquanto na subtração e na divisão, o aumento do minuendo e do dividendo resulta no aumento da resposta e o aumento do subtraendo ou do divisor, resulta na diminuição do resultado. Por fim, o princípio do gradiente especifica se o efeito do aumento de um dos operandos é o mesmo independentemente do valor do outro operando. Enquanto na adição e na subtração o efeito de um operando não depende do valor do outro, na subtração e na divisão depende.

A resolução de problemas de matemática com enunciado (PMEs) é uma tarefa matemática em que o problema é colocado através de um enunciado escrito que o aluno tem de interpretar, transformar na linguagem matemática, escolher uma estratégia ou esquema, resolver e, voltando ao início, confirmar a adequação do resultado aos dados do enunciado. Segundo Newman (1983) os procedimentos dos alunos num PME são: ler o enunciado, efetuar uma operação mental com as palavras para selecionar uma estratégia matemática apropriada, aplicar o processo das competências exigidas pela estratégia escolhida e codificar a resposta numa forma escrita adequada. Trata-se pois de uma tarefa complexa, ficando patente a importância relativa das competências de cálculo e de leitura (geral) e ainda de competências específicas de leitura relacionadas com a matemática, que envolvem reconhecimento e decodificação flexível, e adequação das competências de leitura ao problema matemático escrito. Ou seja, a resolução de problemas de matemática implica, para além do conhecimento do sistema numérico, o cálculo, capacidade de resolução de algoritmos, leitura fluente e capacidade de raciocínio sobre enunciados (Vieira, 2001).

Apesar da importância da aprendizagem matemática para o desenvolvimento cognitivo dos estudantes, do impacto que o sucesso ou o fracasso nessa aprendizagem tem na vida dos estudantes e do fato de que grande parte dos encaminhamentos de crianças ao atendimento psicológico advir de problemas relacionados com essa temática, em Portugal não existem instrumentos padronizados que possam servir de base comum para interpretação dos resultados e para tomadas de decisões relativamente ao desempenho escolar na área da matemática ou à identificação de eventuais dificuldades dos alunos, ou que forneçam pistas e/ou indicadores para a sua correção. De resto a área da matemática é tradicionalmente alvo de um interesse muito menor dos investigadores do que a área da leitura (Citoler, 1996).

Por isso, o objetivo fundamental deste trabalho foi construir e avaliar as propriedades psicométricas de um instrumento de avaliação de competências matemáticas de alunos portugueses dos quatro primeiros anos de escolaridade, nas áreas do conhecimento numérico, operações aritméticas elementares e resolução de problemas (e não propriamente avaliar a pertinência do instrumento como meio de avaliar o conhecimento dos alunos). Especificamente, neste estudo investigou-se a validade fatorial e desenvolvimental do instrumento, bem como seus índices de fidedignidade. Além disso, por meio do Modelo de Rasch da Teoria de Resposta ao Item (TRI), foram analisados os índices de ajuste dos dados ao modelo (*infit* e *outfit*) e a relação entre a extensão da dificuldade dos itens e a extensão da habilidade dos sujeitos (mapa de itens).

Método

Participantes

No total, 505 crianças portuguesas do 1º ao 4º ano do primeiro ciclo, com idades compreendidas entre os 6 e os 12 anos, participaram do estudo (M = 8,46; DP = 1,27) tendo realizado a prova nos primeiros dias do ano escolar. Do total de alunos, 30,3% haviam concluído o 1º ano e apresentavam média de idade de 7,1 anos (DP = 0,45); 18,8% haviam concluído o 2º ano e tinham em média 7,9 anos de idade (DP = 0,37); 17,2% tinham concluído o 3º ano e apresentavam uma média de 8,6 anos idade (DP = 0,64); e 33,7% tinham concluído o 4º ano, e apresentavam uma média de idade de 9,9 (DP = 0,52). Apenas 485 participantes forneceram informação sobre o sexo, sendo que, no 1º ano 56,4% eram do sexo masculino, no 2º ano 47,4%, no 3º ano 55,2% e no 4º ano 54,1%, perfazendo um total de 53,6% de participantes do sexo masculino. Todos os participantes frequentavam escolas públicas da cidade do Porto (Portugal) e arredores.

Instrumento

O currículo do ensino básico em Portugal, como o de muitos outros países, estrutura-se em torno do conhecimento do sistema numérico (contagem, composição e decomposição numérica, etc.), da realização de operações aritméticas e da resolução de problemas, áreas que são consideradas por muitos autores como o núcleo central do conhecimento matemático elementar (Ashlock, 2005; Capovilla, Raad, Berberian, Dias, & Trevisan, 2007; Citoler, 1996). O primeiro bloco desse programa faz referência ao conhecimento de números e operações, a partir dos quais se desenvolvem e exercitam as operações aritméticas elementares. As operações aritméticas surgem no programa como uma forma específica de exercitação das relações entre os números. A resolução de problemas está incluída numa rubrica denominada "capacidades transversais", sendo apresentada como um tipo de desafios que exige a mobilização de múltiplas competências e conhecimentos anteriormente adquiridos.

Desta forma, na construção da prova que aqui se apresenta, procurou-se contemplar tanto o conteúdo educacional referente aos quatro primeiros anos do ensino básico em Portugal quanto as habilidades cognitivas subjacentes a esse conteúdo. Assim, a Prova de Conhecimentos de Matemática[1] apresenta um total de 46 questões, divididas em três partes, com ordenação dos itens por grau de dificuldade. A primeira parte da prova inclui 18 questões referentes ao conhecimento do sistema numérico e decomposição dos números. Seis itens dizem respeito à escrita de números por extenso, quatro à identificação do menor e do maior número dentro de uma série (como se pede a identificação de maior e menor, estes quatro itens representam na verdade oito questões) e quatro à decomposição de números em unidades, dezenas e centenas.

A segunda parte conta com 20 itens referentes à realização de operações básicas: adição (5), subtração (5), multiplicação (5) e divisão (5). Os itens de adição e subtração apresentam a mesma estrutura, com aumento progressivo da dificuldade do primeiro para o quinto item. Os três primeiros itens constituem operações simples sem transferência (transporte) de valores, aumentando apenas o número de algarismos dos operandos. O quarto item requer transferência de valores e o último implica uma operação com números decimais. Os quatro primeiros itens da multiplicação aumentam gradualmente de dificuldade em função do número de algarismos dos operandos (entre 1 e 3), enquanto o último item implica uma operação com números decimais. De forma semelhante, os três primeiros itens de divisão não apresentam resto, aumentando de dificuldade em função do número de algarismos do dividendo e do divisor; o quarto item representa uma operação simples com resto e o quinto item implica uma operação com divisor decimal.

Finalmente, a terceira parte da prova é constituída por oito problemas, cuja resolução exige a compreensão do problema (mais linguística e cristalizada – Gc e Grw) e a sua operacionalização e execução em termos quantitativos (mais relacionada com o conhecimento quantitativo - Gq). A dificuldade dos itens aumenta de acordo com o tipo de operação realizada e com a necessidade de realização de apenas uma operação ou de uma combinação de operações para resolução do problema.

Procedimento

As escolas envolvidas foram contatadas por um dos autores, tendo sido obtida autorização por parte da direção da escola bem como participação voluntária por parte dos professores aos quais foi apresentada a prova, os objetivos do estudo e informações precisas sobre os procedimentos de aplicação (uma vez que a prova foi aplicada pelos próprios professores). Para auxiliá-los, foi elaborado um roteiro de aplicação cujas instruções deveriam ser seguidas durante a aplicação da prova. Todas as dúvidas e questões foram devidamente esclarecidas e todos os aplicadores fizeram um ensaio de aplicação com quatro a cinco alunos. A prova viria a ser aplicada como parte das atividades da escola, no início do ano seguinte às séries consideradas neste estudo, tendo sido apresentada aos alunos como uma avaliação de aprendizagem relacionada com os conteúdos de matemática do ano anterior. A aplicação foi realizada pelos professores titulares de turma. As respostas dos participantes foram cotadas por estudantes de graduação em Psicologia, sob a supervisão dos investigadores responsáveis pelo projeto. Os pais dos alunos forneceram autorização escrita para participação dos filhos no estudo.

Resultados

O primeiro passo da análise de dados foi a investigação da validade fatorial. Como no instrumento construído existem itens com diferentes conteúdos (conhecimento numérico, operação básicas, resolução de problemas) dentro de um mesmo tema (Matemática), esperava-se a obtenção de fatores primários tão inter-relacionados entre si que seria também possível obter um fator geral de segunda ordem. Para realização dessa análise foram seguidos os seguintes passos: (a) análise das condições para realização da análise fatorial; (b) definição do número de fatores a serem retidos e (c) realização das análises fatoriais de primeira e segunda ordem.

A análise das condições de realização da análise fatorial foi feita por meio da inspeção da matriz de correlações tetracóricas. Optou-se pela utilização das correlações tetracóricas por serem mais adequadas a dados dicotômicos (Primi, 1998), como é o caso deste trabalho. A média dessas correlações foi de 0,418 e o desvio padrão de 0,242, indicando claramente haver correlações suficientes para a realização da análise fatorial (Clark & Watson, 1995).

O passo seguinte na investigação da estrutura fatorial foi a definição do número de fatores a serem retidos, que foi determinado por uma análise paralela. Por meio desse procedimento, os *eigenvalues* obtidos com os dados experimentais são comparados com os obtidos a partir de dados gerados aleatoriamente, retendo-se apenas os fatores experimentais cujos *eigenvalues* forem superiores aos obtidos com dados aleatórios (Horn, 1965; Reise, Waller, & Comrey, 2000). Com esta finalidade, utilizou-se

[1] Basta contatar um dos autores para obtenção de uma cópia integral do instrumento.

a sintaxe para SPSS disponibilizada por O'Connor (2000), tendo sido gerada uma matriz de correlações a partir de dados aleatórios com base no mesmo número de sujeitos e variáveis da matriz de dados deste estudo (experimental). A comparação mostrou que sete fatores experimentais apresentavam *eigenvalues* superiores aos correspondentes aleatórios, sendo, portanto, sete o número de fatores a reter na análise fatorial.

Assim, procedeu-se a análise fatorial com auxílio do *software* Testfact (Wilson, Wood, & Gibbons, 1991), com base numa matriz de correlações tetracóricas, extração dos fatores por componentes principais e rotação promax. Com vistas à obtenção da estrutura mais simples possível, procedeu-se a uma inspeção das cargas fatoriais dos 46 itens que compunham a escala inicial, eliminando-se do fator os itens que apresentassem carga fatorial inferior a 0,30, ou acima de 0,30, mas inferior à carga do mesmo item noutro fator. No último caso, o item era eliminado apenas se a diferença entre as cargas fatoriais fosse superior a 0,1 e somente do fator onde apresentou carga inferior. Caso a diferença fosse inferior a 0,1 as cargas eram mantidas em ambos os fatores. Desta forma, obteve-se a solução fatorial rotada apresentada na Tabela 1.

Tabela 1
Cargas Fatoriais da Solução com Rotação Promax

Ordem	Item	1	2	3	4	5	6	7
1	I1A	0,256	0,033	0,204	0,342	0,074	0,136	0,057
2	I1B	-0,101	0,238	0,153	0,174	0,212	0,308	-0,148
3	I1C	0,347	0,122	-0,124	0,072	-0,012	0,635	0,098
4	I1D	0,602	-0,115	0,094	0,117	-0,061	0,214	0,141
5	I1E	0,675	-0,027	0,179	-0,033	0,120	-0,026	0,169
6	I1F	0,701	-0,104	0,060	-0,027	0,085	-0,004	0,180
7	I2A_MAIO	0,147	0,038	0,875	-0,103	0,171	-0,257	-0,067
8	I2A_MENO	-0,074	0,073	0,921	-0,088	0,131	-0,093	0,054
9	I2B_MAIO	0,058	0,156	0,465	0,013	0,095	0,423	0,014
10	I2B_MENO	0,057	0,032	0,541	0,002	0,096	0,339	0,081
11	I2C_MAIO	0,525	0,002	0,479	-0,002	-0,153	0,097	0,122
12	I2C_MENO	0,583	-0,049	0,501	-0,055	-0,172	0,102	0,030
13	I2D_MAIO	0,245	-0,034	0,512	0,182	-0,096	0,242	0,039
14	I2D_MENO	0,215	-0,149	0,474	0,125	-0,054	0,236	0,120
15	I3A	-0,176	0,918	0,032	-0,062	0,003	0,033	0,019
16	I3B	-0,036	0,958	0,102	-0,060	-0,140	0,173	0,024
17	I3C	0,142	0,721	-0,063	0,090	-0,004	0,012	0,015
18	I3D	0,118	0,381	-0,073	0,248	-0,042	-0,236	0,036
19	I4A	-0,134	-0,237	0,001	0,931	0,040	-0,054	-0,107
20	I4B	0,194	0,140	-0,110	0,711	-0,101	-0,018	0,033
21	I4C	0,034	0,094	-0,060	0,559	0,117	0,310	0,068
22	I4D	0,318	0,065	0,030	-0,031	-0,061	0,738	-0,085
23	I4E	0,780	-0,148	-0,104	0,090	-0,082	0,158	0,015
24	I4F	-0,138	0,097	-0,062	0,755	0,122	-0,004	-0,024
25	I4G	0,110	0,071	-0,016	0,622	0,239	0,001	-0,072
26	I4H	0,385	-0,028	-0,202	0,141	-0,095	0,627	0,163
27	I4I	0,693	-0,154	0,043	0,241	0,008	0,122	-0,110
28	I4J	0,681	0,066	0,035	0,035	-0,080	0,017	0,122
29	I4K	0,491	-0,028	-0,072	-0,087	0,152	0,550	-0,069
30	I4L	0,573	0,044	-0,016	-0,202	0,244	0,545	-0,155

31	I4M	0,686	-0,023	0,117	-0,043	0,158	0,113	-0,027
32	I4N	0,759	-0,006	0,063	0,049	0,064	0,051	-0,058
33	I4O	0,770	0,013	0,120	0,034	-0,022	-0,064	-0,022
34	I4P	0,883	0,004	-0,155	0,044	-0,059	0,169	0,003
35	I4Q	0,997	0,111	-0,036	-0,072	0,048	-0,096	0,091
36	I4R	0,831	0,128	-0,015	-0,196	0,006	0,024	0,276
37	I4S	0,914	0,000	-0,060	-0,072	0,193	0,028	-0,025
38	I4T	-0,006	-0,032	0,070	0,144	-0,079	-0,176	0,690
39	I5A	-0,145	-0,127	0,111	0,129	0,786	0,137	0,299
40	I5B	0,222	-0,014	0,173	0,057	0,732	-0,088	0,052
41	I5C	0,331	-0,009	0,092	0,187	0,573	-0,209	0,111
42	I5D	0,182	-0,013	0,189	0,066	0,508	-0,050	0,157
43	I5E	0,244	-0,074	-0,198	-0,033	0,421	0,103	0,244
44	I5F	-0,236	-0,001	-0,177	-0,070	0,491	0,148	0,216
45	I5G	0,009	0,051	0,079	-0,128	0,019	0,022	0,308
46	I5H	0,077	0,049	0,046	-0,042	0,120	0,039	0,291
Coeficiente Alfa		0,93	0,70	0,69	0,73	0,84	0,85	0,47

Pode-se observar que o primeiro fator (F1) reúne 17 itens que envolvem conhecimento de números mais complexos, com algarismos na casa dos milhares ou com decimais e operações de multiplicação e divisão. O Fator 2 (F2) reúne os quatro itens destinados à avaliação da decomposição numérica em centenas, dezenas e unidades. Os oito itens relacionados com o conhecimento de grandeza numérica foram agrupados no terceiro fator (F3). O quarto fator (F4), agrupa seis itens relacionados principalmente com operações matemáticas simples, de adição e subtração, sem transferência de valores. O quinto fator (F5) reúne seis problemas de matemática, em que os enunciados têm que ser compreendidos e transformados em algoritmos adequados à sua resolução. O Fator 6 (F6) reúne sete itens relacionados com operações matemáticas de adição e subtração com transferências de valores e operações simples de multiplicação (com multiplicador de apenas um algarismo). Finalmente, o Fator 7 (F7) reúne três itens, um relativo a uma operação de divisão com números decimais e dois relativos a problemas com texto, todos com elevado grau de dificuldade, elaborados como desafio para os participantes. A consistência interna de cada fator foi calculada por meio da Fórmula 20 do Coeficiente de Kuder-Richardson e os resultados são todos satisfatórios, com exceção de F7, eventualmente devido ao pequeno número de itens envolvidos e à sua elevada dificuldade, resultando em pouca variância comum entre os itens.

Na sequência, os Coeficientes *Theta* (q), correspondentes às habilidades de cada sujeito, foram calculados apenas para os seis primeiros fatores. Dada a expectativa de que os fatores primários estivessem altamente correlacionados entre si, foram calculados os coeficientes de correlação de Pearson entre estes fatores. Os resultados são apresentados na Tabela 2.

Tabela 2
Coeficientes de Correlação de Pearson entre os Fatores Primários

	Fator1	Fator2	Fator3	Fator4	Fator5
Fator1	1				
Fator2	0,067	1			
Fator3	0,725*	0,080	1		
Fator4	0,309*	0,226*	0,293*	1	
Fator5	0,510*	0,142*	0,483*	0,384*	1
Fator6	0,741*	0,177*	0,708*	0,435*	0,493*

Nota. *Correlações estatisticamente significativas ao nível de 0,01.

De uma forma geral, nota-se que as correlações são positivas e estatisticamente significativas. No entanto, o F2 apresenta valores bastante inferiores aos demais, incluindo dois coeficientes, com F1 e F3, que não são estatisticamente significativos. Para verificar a possibilidade de obtenção de um fator de segunda ordem relacionado com o desempenho geral em aritmética, foi realizada uma análise fatorial de segunda ordem, com auxílio do SPSS, com extração dos fatores por análise dos componentes principais e rotação oblíqua, cujos resultados são apresentados na Tabela 3.

Tabela 3
Análise Fatorial de Segunda Ordem

	Primeira Análise		Segunda Análise	
	FSO1	FSO2		$FSO_2 1$
Fator1	0,918		Fator6	0,883
Fator3	0,897		Fator1	0,867
Fator6	0,863		Fator3	0,847
Fator5	0,662		Fator5	0,720
Fator2		0,915	Fator4	0,564
Fator4	0,363	0,542		

Como se pode observar, inicialmente formaram-se dois fatores de segunda ordem, com o FSO1 fundamentalmente relacionado com o conhecimento geral em matemática e o FSO2 mais relacionado com a decomposição numérica (F2). Verifica-se também que o F4 (operações de adição e subtração sem transferência de valores) apresenta carga em ambos os fatores de segunda ordem. Tanto a análise das correlações entre os fatores primários (Tabela 2) como a primeira análise fatorial de segunda ordem (Tabela 3), sugerem que o F2 avalia um tipo de conhecimento algo diferente dos demais fatores obtidos. Por isso, os itens que compõem o F2 foram também excluídos e uma nova análise resultou numa solução unifatorial, como pode ser observado na segunda análise fatorial de segunda ordem (Tabela 3). Deste modo a prova final ficou com 42 itens distribuídos em cinco fatores primários, que, por sua vez, se agrupam num único fator de segunda ordem, relacionado com o conhecimento geral em matemática. O coeficiente de Kuder-Richardson (KR_{20}) do fator geral é de 0,94, indicando ótima consistência interna da prova como um todo.

Em seguida, passou-se à análise psicométrica com base no Modelo de Rasch da Teoria de Resposta ao Item (TRI), tendo em vista uma depuração das propriedades psicométricas da prova geral que resultou das análises fatoriais. Nessa etapa, o primeiro passo foi a análise dos índices de ajustamento dos itens ao modelo da TRI. Para isso, foram utilizados os índices de *infit* e *outfit*. Segundo Linacre (2002), valores entre 0,5 e 1,5 indicam um ajustamento produtivo para a mensuração; valores entre 1,5 e 2,0 são considerados improdutivos, embora não degradantes para a mensuração e valores superiores a 2,0 são considerados degradantes para a mensuração.

No que se refere ao *infit*, todos os itens obtiveram índices entre 0,70 e 1,5, com exceção de um item (I5F) que apresentou 1,79. Os resultados de *outfit* podem ser divididos em três grupos: os subajustados, os ajustados e os sobreajustados. Cinco itens (I1A, I1E, I4N, I4O e I4R) mostram-se subajustados, isto é, mais previsíveis do que o esperado pelo modelo, apresentando índices inferiores a 0,5 (de 0,19 a 0,47). Vinte e dois itens (I1C, I1D, I1F, I2A_MAIO, I2A_MENO, I2B_MAIO, I2B_MENO, I2C_MAIO, I2C_MENO, I2D_MAIO, I2D_MENO, I4E, I4H, I4I, I4J, I4K, I4L, I4M, I4P, I4Q, I4S, I5C) mostram-se bem ajustados ao modelo, com índices entre 0,57 e 1,34. Quatro itens (I4D, I5B, I5D e I5E) apresentam índices de 1,51 a 1,77, estando na faixa dos sobreajustados, isto é, mais imprevisíveis do que o esperado pelo modelo, mas ainda não degradantes para a mensuração. Finalmente, oito itens (I4C, I4G, I4B, I5A, I5F, I4F, I4A, I1B) apresentaram índices superiores a 2,0 (2,11 a 7,61), sendo considerados, em princípio, sobreajustados e impróprios para a mensuração.

Os índices de *infit* e *outfit* indicam desajustamento dentro e fora da faixa de habilidade do sujeito, respectivamente. Por isso, problemas com o *infit* são muito mais graves (dificuldade de detectar o porquê e, consequentemente, de resolver o problema) do que com o *outfit* (Linacre, 2002). Portanto, considerando que todos os itens estão ajustados quanto ao *infit* (nem mesmo I5F causa degradação à medida) e, majoritariamente, também quanto ao *outfit*, optou-se pela manutenção de todos os itens que foram submetidos a esta análise.

Outro dado que apoia a manutenção dos itens na prova é o fato de que, em todos os itens, a média de habilidade dos sujeitos que responderam incorretamente ter sido inferior à média dos sujeitos que acertaram. No caso do item I1B, por exemplo, que apresenta o maior índice de *outfit* (7,61), a média de *theta* dos que responderam incorretamente foi de -2,04 e a dos que responderam corretamente foi de 0,58. Isso mostra que embora tenha havido casos de acertos ou erros inesperados, essa ocorrência não comprometeu a ordem lógica dos fatos na média da população, isto é, que os sujeitos que erraram apresentam uma média de habilidade inferior aos que acertaram.

O mapa de itens (Figura 1) mostra que houve uma boa adequação do nível de dificuldade dos itens ao nível de habilidade dos sujeitos, com os itens cobrindo toda a extensão de habilidades dos sujeitos. De fato, a média de habilidade dos sujeitos foi de 0,48 (*DP* = 2,14), muito próxima da média de dificuldade dos itens, que, por *default*, é centrada em zero (*DP* = 2,60).

```
              P - MAPA - I
              <mais>|<raro>
    6             .  +
                  .  |
                     |
                     |T
    5                +
                ##   |  I4O
                    T|
             .####   |
    4         .###   +  I4R
                     |
             .####   |  I4J       I5F
           .######   |
    3        #####   +  I4N
             #####   |  I4Q
             .#### S|S  I4S
            .#####   |  I1E       I1F
    2        #####   +  I4E       I4I       I4M
             .####   |  I5E
             .####   |
      .###########   |  I2C_MAIO  I2D_MAIO  I4P
    1      .######   +  I2C_MENO  I2D_MENO
            ######   |
              .### M|   I1D
            .#####   |  I4K       I4L
    0       .#####   +M I5D
            ######   |  I4D
     .############   |  I4H
         .########   |  I2B_MENO
   -1      .######   +  I5C
         .########   |  I2B_MAIO
         .######## S|
            .#####   |
   -2          ###   +  I1C       I5B
                     |  I4C       I4G
              ####   |S I2A_MAIO
               ###   |  I5A
   -3           ##   +  I2A_MENO
                .#   |  I4F
                    T|  I4B
               ###   |
   -4                +  I4A
                     |  I1B
                  .  |
                     |
   -5                +
                     |T
                     |
                     |
   -6                +
                     |
                     |
                     |  I1A
   -7             .  +
              <menos>|<frequente>
 CADA "#" CORRESPONDE A 3 SUJEITOS.
 CADA "." CORRESPONDE A 1 OU 2 SUJEITOS.
```

Figura 1. Mapa de itens.

O último objetivo deste estudo era investigar a validade desenvolvimental do instrumento, uma vez que era esperado que o nível de habilidade fosse aumentando em função do progresso escolar da criança. Assim, do ponto de vista estatístico era esperado que as pontuações médias das crianças aumentassem com os anos de escolaridade/séries. Neste estudo, optou-se pela utilização do ano de escolaridade /série e não da idade, conforme recomendação de Almeida, Lemos, Guisande e Primi (2008). Além disso, como a evolução e o nível de desempenho podem sofrer efeito do sexo das crianças, essa variável também foi tida em consideração.

Para conduzir essa investigação foi realizada uma Análise de Variância Univariada, com o desempenho geral na prova de aritmética como variável dependente e série e sexo como variáveis independentes. Foram significativos apenas os efeitos do ano de escolaridade/série (F=295,4; gl=3; p=0,001), que explicou 65% da variância total, e da interação entre série e sexo (F=2,8; gl=3; p=0,037), que explicou apenas 1,8% da variância total. Esses efeitos podem ser visualizados na Figura 2.

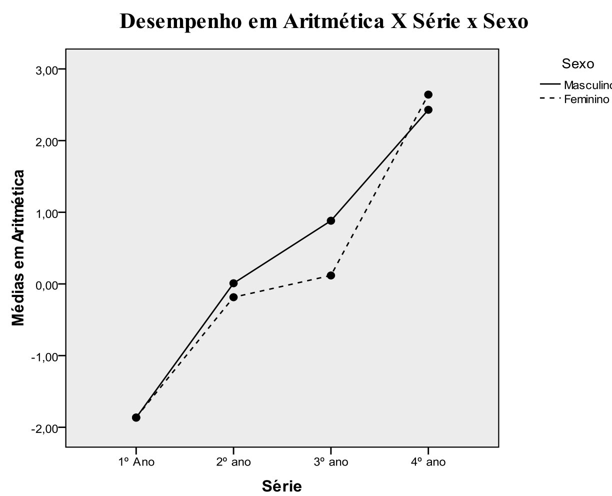

Figura 2. Representação gráfica da análise de variância univariada.

O efeito do ano de escolaridade /série é visível no perfil ascendente das médias de desempenho. Por outro lado, embora os perfis de realização nos diferentes anos/séries sejam semelhantes, verifica-se que a trajetória masculina é mais homogeneamente ascendente ao longo das séries, enquanto a feminina sofre como que uma estabilização ao longo do terceiro ano, antes de nova ascensão e equiparação com o desempenho masculino no final do quarto ano. Esses dados, ao lado dos obtidos com a análise fatorial apoiam a validade de construto do instrumento.

Discussão

A Prova de Conhecimentos de Matemática aqui apresentada parece revelar-se um instrumento com características psicométricas satisfatórias para avaliação de algumas das mais importantes competências e constructos representados na estrutura curricular do programa de matemática do 1º Ciclo do Ensino Básico de Portugal e muitos outros países (Ashlock, 2005; Capovilla et al., 2007; Delazer & Benke, 1997; Fayol, 2008; Nys & Content, 2010). A estrutura que emerge da análise fatorial da escala indicia que existe uma base sólida para considerar os constructos como relacionados ao *conhecimento do sistema numérico* (Brannon, 2002; Condry & Spelke, 2008), das *operações aritméticas* (Miller, Perlmutter, & Keating, 1984; Nys & Content, 2010) e da *resolução de problemas* (Brissiaud & Sander, 2010; Hardiman, Dufresne, & Mestre, 1989) como processos independentes com correlações significativas. É interessante salientar que na fase inicial da análise fatorial de segunda ordem, o F2 surge como fator isolado, ligado ao conhecimento matemático geral, o que poderá indicar que a decomposição numérica constitui uma importante competência matemática que tem subjacente um apurado conhecimento matemático geral nestas idades. Dehaene (2001), por exemplo, defende que ao longo do desenvolvimento e do trajeto escolar a representação do número nas suas múltiplas componentes torna-se cada vez mais integrada com outros sistemas cognitivos incluindo a compreensão de palavras e os dispositivos de rastreamento (visual) de objetos.

Uma das questões mais importantes que nos parecem suscitar os resultados se refere à dificuldade dos itens, cuja amplitude foi adequada e, por isso mesmo, capaz de captar as diferenças individuais existentes na amostra. Além disso, a análise da ordem sequencial da dificuldade dos itens identifica o caminho desenvolvimental por que passa a aprendizagem desses conceitos matemáticos. Só no que diz respeito ao conhecimento numérico, por exemplo, a investigação evidencia que a noção de número evolui gradualmente, estando muito dependente da exploração (exercitação) dos números, da sua visualização em múltiplos contextos, e do seu relacionamento em contextos que não se limitem aos algoritmos tradicionais (Burton, 1993; Howden, 1989; Reys, 1991). O mesmo sucede, aliás, com as operações aritméticas (Nys & Content, 2010; Prather & Alibali, 2008; Semenza, 2002) e com os problemas com texto (Moreau & Coquin-Viennot, 2003; Nathan, Kintsch, & Young, 1992). Por fim será de salientar que a prova se revela discriminativa em termos de anos de escolaridade/séries pelo que tem potencialidades para aferir de forma confiável os conhecimentos dos alunos nos quatro primeiros anos de escolaridade, o que permitirá, por exemplo, a sua inscrição num modelo RTI (*Response to Intervention*) de avaliação (Brown-Chidsey, Bronaugh, & McGraw, 2009; Wagner & Compton, 2011).

A evolução dos resultados por ano de escolaridade sugere, como seria de esperar, uma evolução de desempenho relativamente linear, apenas com um ligeiro abrandamento no 3º ano. Este abrandamento pode dever-se ao fato de o 1º e o 3º anos serem aqueles em tipicamente há mais matérias novas introduzidas, constituindo o 2º e 4º anos, momentos de consolidação (Lopes, 2010). Por outro lado, não há uma explicação evidente para as diferenças de desempenho para gênero no 3º ano, até porque os desempenhos voltam a equalizar-se no 4º ano. Porventura a variação será aleatória, sendo neste caso arriscada qualquer tentativa de explicação. Os resultados em qualquer destas variáveis (ano de escolaridade e gênero) deverão ser reconsiderados em futuros estudos.

As provas construídas no quadro da TRI são particularmente relevantes para a implementação de modelos RTI uma vez que asseguram o equilíbrio da dificuldade das tarefas nos diversos anos de escolaridade/séries controlando assim o enviesamento dos resultados (Berkeley, Bender, Gregg Peaster, & Saunders, 2009). Permitem, assim, aos utilizadores (i.e. escolas) não só determinar o seu nível médio de desempenho ao nível do ensino/aprendizagem, como identificar alunos que possam necessitar de apoios de Nível I (Tier I; apoio menos intensivo) ou de Nível II (Tier II; apoio mais intensivo e prolongado) e, com o mesmo instrumento, fazer o *follow-up* das intervenções desenhadas a partir da avaliação inicial (Berkeley et al., 2009).

Apesar dessas contribuições, há que se considerar algumas limitações deste estudo. Uma delas se refere à constituição da amostra de participantes, que se limitou a estudantes da cidade do Porto e arredores. Desta forma, a amostra não reflete toda a diversidade da população portuguesa e, consequentemente, a generalização dos resultados fica mais restrita aos estudantes com características mais próximas à dos participantes deste estudo. Esse dado aponta para a necessidade da realização de outras investigações, com amostras mais alargadas, ou especificas de outras regiões de Portugal, para verificação da reprodutibilidade dos resultados.

Outro aspecto a ser considerado é que a prova não reflete o desempenho de todo o conteúdo ministrado no currículo de matemática das escolas portuguesas. Um importante conteúdo, a saber, a geometria, foi deixada de fora, por envolver habilidades diferentes das contempladas neste estudo e pelo impacto que sua inclusão teria na extensão da prova e no tempo requerido para resolvê-la. Apesar disso, os autores reconhecem a importância da área e, por isso mesmo, a necessidade do desenvolvimento de outro instrumento específico para esta finalidade. Ainda assim, com todas as limitações que possa ter, este estudo contribui para a realização de avaliações objetivas, mas psicologicamente fundamentadas, do desempenho em matemática de crianças das séries iniciais do ensino básico português.

Referências

Almeida, L. S., Lemos, G., Guisande, M. A., & Primi, R. (2008). Inteligência, escolarização e idade: Normas por idade ou série escolar? *Avaliação Psicológica, 7*, 117-125.

Ashlock, R. B. (2005). *Error patterns in computation*. Columbus, OH: Pearson.

Berkeley, S., Bender, W. N., Gregg Peaster, L., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learn Disabilities, 42*(1), 85-95. doi:10.1177/0022219408326214

Brannon, E. M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition, 83*(3), 223-240.

Brissiaud, R., & Sander, E. (2010). Arithmetic word problem solving: A situation strategy first framework. *Developmental Science, 13*(1), 92-107.

Brown-Chidsey, R., Bronaugh, L., & McGraw, K. (2009). *RTI in the classroom: Guidelines and recipes for success*. New York: Guilford Press.

Burton, G. M. (1993). *Number sense and operations: Addenda series, grades K-6*. Reston, VA: National Council of Teachers of Mathematics.

Capovilla, A. G. S., Raad, A. J., Berberian, A. A., Dias, N. M., & Trevisan, B. T. (2007). Avaliação de aritmética em crianças de 1ª a 4ª série: Prova de Aritmética. In A. G. S. Capovilla & F. C. Capovilla (Eds.), *Teoria e Pesquisa Neuropsicológica* (pp. 45-53). São Paulo, SP: Memnon.

Citoler, S. D. (1996). *Las difcultades de aprendizaje: un enfoque cognitivo – lectura, escritura, matemáticas*. Málaga, España: Aljibe.

Clark, A. C., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309-319.

Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General, 137*(1), 22-38.

Correa, J., & Moura, M. L. S. (1997). A solução de problemas de adição e subtração por cálculo mental. *Psicologia: Reflexão e Crítica, 10*(1), 71-86.

Dehaene, S. (2001). Author's response: Is number sense a patchwork? *Mind and Language, 16*(1), 89-100.

Delazer, M., & Benke, T. (1997). Arithmetic facts without meaning. *Cortex, 33*(4), 697-710.

Dixon, J. A., Deets, J. K., & Bangert, A. (2001). The representations of the arithmetic operations include functional relationships. *Memory and Cognition, 29*(3), 462-477.

Fayol, M. (2008). Acquisition of arithmetic knowledge. *L'acquisition de L'arithmétique Élémentaire, 24*(1), 87-90.

Gurganus, S. (2004). Promote number sense. *Intervention in School and Clinic, 40*(1), 55-58.

Hardiman, P. T., Dufresne, R., & Mestre, J. P. (1989). The relation between problem categorization and problem solving among experts and novices. *Memory & Cognition, 17*, 627-638.

Haydu, V. B., Costa, L. P., & Pullin, E. M. M. P. (2006). Resolução de problemas aritméticos: Efeito de relações de equivalência entre três diferentes formas de apresentação dos problemas. *Psicologia: Reflexão e Crítica, 19*(1), 44-52.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrica, 30*, 179-185.

Howden, H. (1989). Teaching number sense. *Arithmetic Teacher, 36*(6), 6-11.

Jensen, A. R., & Whang, P. A. (1994). Speed of accessing arithmetic facts in long-term memory: A comparison of Chinese-American and Anglo-American children. *Contemporary Educational Psychology, 19*(1), 1-12.

Jordan, N. C. (2007). The need for number sense. *Educational Leadership, 65*(2), 63-66.

Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences, 20*(2), 82-88.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*, 878.

Lopes, J. A. (2010). *Conceptualização, avaliação e intervenção nas dificuldades de aprendizagem: A sofisticada arquitectura de um equívoco*. Braga, Portugal: Psiquilíbrios.

Miller, K., Perlmutter, M., & Keating, D. (1984). Cognitive arithmetic: Comparison of operations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(1), 46-60.

Moore, C. F., Dixon, J. A., & Haine, B. A. (1991). Components of understanding in proportional reasoning: A fuzzy set representation of developmental progressions. *Child Development, 62*(3), 441-459.

Moreau, S., & Coquin-Viennot, D. (2003). Comprehension of arithmetic word problems by fifth-grade pupils: Representations and selection of information. *British Journal of Educational Psychology, 73*, 109-121.

Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction, 9*(4), 329-389.

Newman, M. A. (1983). *Strategies for diagnosis and remediation*. Sidney, Australia: Harcourt, Brace Jovanovich.

Nys, J., & Content, A. (2010). Complex mental arithmetic: The contribution of the number sense. *Canadian Journal of Experimental Psychology, 64*(3), 215-220.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*(3), 396-402.

Prather, R. W., & Alibali, M. W. (2008). Understanding and using principles of arithmetic: Operations involving negative numbers. *Cognitive Science, 32*(2), 445-457.

Primi, R. A. L. (1998). Considerações sobre a análise factorial de itens com resposta dicotómica. *Psicologia, Teoria, Investigação e Prática, 3*, 225-234.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12*, 287-297.

Reys, B. J. (1991). *Developing number sense: Addenda series, grades 5-8*. Reston, VA: National Council of Teachers of Mathematics.

Semenza, C. (2002). Conceptual knowledge in arithmetic: The core of calculation skills. *Cortex, 38*(3), 285-288.

Singer-Dudek, J., & Greer, R. D. (2005). A long-term analysis of the relationship between fluency and the training and maintenance of complex math skills. *Psychological Record, 55*(3), 361-376.

Vieira, E. (2001). Representação mental: As dificuldades na atividade cognitiva e metacognitiva na resolução de problemas matemáticos. *Psicologia: Reflexão e Crítica, 14*(2), 439-448.

Wagner, R. K., & Compton, D. L. (2011). Dynamic assessment and its implications for RTI models. *Journal of Learning Disabilities, 44*(4), 311-312. doi:10.1177/0022219411407859

Wilson, D. T., Wood, R., & Gibbons, R. (1991). *Testfact: Test scoring, item statistics, and item factor analysis*. Chicago, IL: Scientific Software.

Zentall, S. S., Smith, Y. N., Lee, Y. B., & Wieczoreck, C. (1994). Mathematical outcomes of attention-deficit huperactivity disorder. *Journal of Learning Disabilities, 27*(8), 510-519.

Recebido: 1º/04/2013
Aceite final: 16/04/2013