



Centro de Educação  
Campus Universitário  
Cidade Universitária  
Recife-PE/BR CEP: 50.670-901  
Fone/Fax: (81) 2126-8952  
E. Mail: edumatec@ufpe.br  
[www.gente.eti.br/edumatec](http://www.gente.eti.br/edumatec)

**Érica Michelle Silva Cavalcanti**

**Para variar: Compreensões de estudantes dos anos iniciais  
diante de aspectos da variabilidade**

**Recife  
2011**

**Érica Michelle Silva Cavalcanti**

**Para variar: Compreensões de estudantes dos anos iniciais diante  
de aspectos da variabilidade**

Dissertação apresentada ao Programa de Pós-Graduação em Educação Matemática e Tecnológica, como requisito parcial para obtenção do título de Mestre em Educação Matemática e Tecnológica.

Orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Gilda Lisbôa  
Guimarães

Recife  
Fevereiro / 2011

Cavalcanti, Érica Michelle Silva

Para variar: compreensões de estudantes dos anos iniciais  
diante de aspectos da variabilidade / Érica Michelle Silva  
Cavalcanti. – Recife: O Autor, 2011.

125 f. : il.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Gilda Lisboa Guimarães

Dissertação (Mestrado) - Universidade Federal de  
Pernambuco, CE, Programa de Pós-Graduação em Educação  
Matemática e Tecnológica, 2011.

Inclui Bibliografia.

1. Ensino fundamental 2. Educação Estatística 3.  
Variabilidade - Estatística I. Guimarães, Gilda Lisboa  
(Orientadora) II. Título

CDD 372.21

UFPE (CE 2011-035)



ALUNA

ERICA MICHELLE SILVA CAVALCANTI

TÍTULO DA DISSERTAÇÃO

"Para variar: compreensões de estudantes dos anos iniciais diante de aspectos da variabilidade."

COMISSÃO EXAMINADORA:

---

Presidente e Orientador  
Prof<sup>ª</sup>. Dr<sup>ª</sup>. Gilda Lisboa Guimarães

---

Examinador Externo  
Prof<sup>ª</sup>. Dr<sup>ª</sup>. Verônica Yumi Kataoka

---

Examinador Interno  
Prof<sup>ª</sup>. Dr<sup>ª</sup>. Verônica Gitirana Gomes Ferreira

Recife, 21 de fevereiro de 2011.

*A minha mãe, pela confiança e apoio em todas as decisões tomadas por mim.*

*A Gilda, minha orientadora, pelo amor e dedicação praticados ao ensinar e fazer pesquisa, o que foi meu maior estímulo profissional.*

## AGRADECIMENTOS

Agradeço, primeiramente, à intervenção divina pelos momentos de calma e lucidez nas situações em que busquei apoio espiritual para não me desesperar (o que tive vontade de fazer muitas vezes, contudo, consegui me controlar).

À minha mãe e meu irmão Bruno pelo apoio incondicional, o que foi importantíssimo para mim, mesmo estando distantes.

À minha “grande família” formada por tios, primos e sobrinhos, especialmente Luíza e Bruninho cujos risos e brincadeiras me deram tantas alegrias.

Agradeço à minha tia Arleide, além de meus primos Amélia, Marcelo e Guigo, pela acolhida oferecida em minha trajetória profissional. Agradecimentos especiais a Badeco (*in memoriam*) que sempre me recebeu como a uma filha.

Agradeço à Gilda, minha orientadora, que como uma amiga me incentivou a refletir e procurar fazer sempre o melhor pela pesquisa e educação. Para isso, desequilibrou-me muitas vezes.

Às minhas queridas amigas Cybelle, Patrícia Batista, Vanessa, Lucília, Patrícia Luz, Iane, Rielda, Mabel e Tâmara com as quais compartilhei e compartilho momentos agradáveis e outros nem tanto, porém todos de crescimento.

À Márcia Barbosa, professora que muito me incentivou ao estudo acadêmico, tanto no ensino como na pesquisa.

Ao Grupo de Estudos em Educação Estatística no Ensino Fundamental (GrEF), que tanto ajudou na construção e (re)construção da pesquisa. Especialmente as prestativas Mabel, Milka e Edilza pelas contribuições desde o pré-projeto.

Aos pesquisadores Verônica Gitirana, Cláudia Borim, Antonio Roazzi e Verônica Yumi pela disponibilidade e interesse em contribuir com sugestões e outras análises na pesquisa.

Aos pesquisadores dos Processos de Ensino Aprendizagem da Educação Matemática e Científica, inclusive Pilar, que a partir da disciplina de Seminários também contribuíram com todos os processos dessa pesquisa.

Aos mestrandos da EDUMATEC, sempre dispostos a discutir nos momentos adequados e a relaxar nos momentos mais que oportunos, especialmente Dayse e Flávia.

Ao corpo de professores e secretaria da EDUMATEC, cujos ensinamentos e orientações foram importantes na conclusão de tal pesquisa. Especialmente Marlene, com toda sua “paciência”, que proporcionaram bons momentos de descontração.

Agradeço também à direção das escolas, ao corpo docente e principalmente aos estudantes que participaram desse estudo. A disponibilidade e vontade de ajudar dos mesmos viabilizaram tal pesquisa.

Finalmente, agradeço à FACEPE (Fundação do Amparo a Ciência e Tecnologia de Pernambuco) e ao Programa de bolsas REUNI - UFPE, pelo incentivo à pesquisa.

## RESUMO

A necessidade de conhecer e tomar decisões a partir de informações tratadas estatisticamente faz da variabilidade um conceito fundamental, uma vez que a Estatística existe porque os dados variam. Contudo, poucos estudos se preocuparam em pesquisar de modo sistemático compreensões de estudantes dos anos iniciais do Ensino Fundamental quanto à variabilidade, apesar de no Brasil o ensino da Estatística ser recomendado no currículo escolar para esses anos. Assim, o objetivo principal do presente estudo foi investigar as compreensões apresentadas por estudantes do 2º e 5º ano a respeito do conceito de variabilidade estatística. Para isso, utilizou-se cinco atividades abordando aspectos diferentes da variabilidade, que foram propostos a 48 (quarenta e oito) sujeitos do 2º e 5º ano, com os quais realizou-se entrevistas clínico-piagetianas, uma vez que as justificativas dadas pelos estudantes às questões propostas foi o interesse maior na pesquisa. Os aspectos da variabilidade explorados foram: explanação da variabilidade; identificação de ponto máximo / moda; predição a partir da moda, do ponto máximo e da tendência; quantificação de variação entre dois pontos; conservação de quantidade total; representação da variabilidade, de frequência nula e de acréscimos; comparação entre conjuntos de dados; além de identificação e proposição de ausência de variabilidade. Constatou-se que os estudantes do 5º ano apresentaram um desempenho significativamente maior do que aquele dos estudantes do 2º ano, em pelo menos metade das questões. O melhor desempenho do 5º ano ocorreu nos seguintes aspectos: explanação da variabilidade em dados qualitativos e dados numa série temporal; localização de ponto máximo; comparação entre pontos; representação da variabilidade; predição a partir do ponto máximo; predição a partir da tendência do gráfico e identificação de ausência de variabilidade. A explanação da variabilidade, quando solicitada aos estudantes após representarem dados, assim como a representação de frequência nula e a representação de acréscimos foram aspectos facilmente compreendidos pelos dois grupos de estudantes. Numa gradação dos aspectos da variabilidade que se mostraram mais complexos para os dois grupos pode-se destacar, numa ordem decrescente: a comparação entre conjuntos de dados; a predição a partir da moda; a comparação entre pontos com quantificação da variação; finalmente, a proposição de ausência de variabilidade. Foi realizada uma análise multidimensional (MDS), a qual evidenciou que o comportamento dos alunos do 2º ano foi diferente dos alunos do 5º ano. No gráfico do 2º ano ficou evidente que a possibilidade de respostas tendo como base a experiência de vida dos mesmos foi um fator importante. Já para os alunos do 5º ano, foram encontrados 4 (quatro) grupos que englobavam diferentes aspectos da variabilidade: representação da variabilidade e da frequência nula; explanação da variabilidade; localização e predição de ponto máximo; comparação entre conjuntos e quantificação da variação, os que se mostraram mais complexos. Assim, esse estudo evidenciou que estudantes desde o 2º ano de escolaridade são capazes de compreender aspectos da variabilidade, o que pode ser potencializado se os mesmos vivenciarem na escola situações de ensino que os desafiem a analisar e refletir sobre dados tratados estatisticamente.

**Palavras-chave:** Educação Estatística, variabilidade, anos iniciais do Ensino Fundamental.

## ABSTRACT

The necessity of knowing and making decisions from statistical information makes variability a fundamental concept, since it exists because of data variation. Nevertheless, few studies were concerned about doing a systematic research about students' – in the early years of basic education - comprehensions of variability, despite teaching statistics in Brazil is recommended for the school curriculum in such years. Thus, the present study's objective was of investigating the comprehensions presented by 48 students from 2nd to 5th grade, to whom clinic-piagetian interviews were held, once the students' answers to the questions proposed was of our greatest interest. The explored aspects of variability were: explanation of variability; identification of the maximum point \ mode; prediction from the mode, maximum point and the tendency; quantification of the variation between two points; conservation of the total quantity; representation of the variability, the null frequency and the increments; comparison between data groups; besides the identification and proposition of variability absence. It was verified that 5th graders had a more significant performance than the 2nd graders in at least half of the questions. The best performance of the 5th grade occurred on the following aspects: explanation of variability on situations of qualitative data and data in a temporal series; localization of the maximum point; comparison between points; representation of variability; prediction from the maximum point; prediction from the tendency of the graphic and identification of variability absence. The explanation of variability, when requested to the students after representing data, as well as the representation of the null frequency and the representation of increments were aspects easily comprehended by the two groups of students. In a gradation on the aspects of variability which were considered more complex for the two groups, we can highlight, in a decrescent order, the comparison between data groups; the prediction from the maximum point; the comparison between points with quantification of variation; finally, the proposition of variability absence. A multidimensional analysis (MDS) was held and it evinced that the 2nd graders' behavior was different from the 5th graders'. It was also possible to observe that on the 2nd grade's graphic, the possibility for the answers, which were based on the students' life experience, was, clearly, an important element. However, for the 5th graders, 4 groups which embodied different aspects of variability were found: representation of variability and the null frequency; explanation of variability; localization and prediction of the maximum point; comparison between data groups and quantification of variation, which were considered to be more complex. Therefore, this study showed that students, since the 2nd grade of schooling, are capable of understanding aspects of variability, which can be potentiated if students experience teaching situations which challenge them to analyse and reflect about statistical data at school.

**Keywords:** Statistical Education, variability, early years of Basic Education.

## SUMÁRIO

### RESUMO

### ABSTRACT

INTRODUÇÃO.....	10
-----------------	----

### **CAPÍTULO 1 – Revisão da literatura.....12**

1.1 A Análise Exploratória de Dados na Educação Básica.....	12
---	----

1.2 Variabilidade: Conceito fundamental da estatística.....	15
---	----

1.3 Componentes do raciocínio variacional.....	16
--	----

1.4 Variabilidade de dados.....	19
---------------------------------	----

1.5 Compreensões de estudantes quanto à variabilidade.....	23
--	----

1.6 Níveis de aprendizagem: O SOLO (Structure of Observed Learning Outcomes) como instrumento de análise.....	42
---	----

### **CAPÍTULO 2 – Percurso metodológico.....46**

2.1 Participantes.....	46
------------------------	----

2.2 Procedimentos.....	47
------------------------	----

2.2.1 Atividade 1.....	48
------------------------	----

2.2.2 Atividade 2.....	49
------------------------	----

2.2.3 Atividade 3.....	51
------------------------	----

2.2.4 Atividade 4.....	53
------------------------	----

2.2.5 Atividade 5.....	55
------------------------	----

2.3 Procedimentos de análise dos dados.....	58
---	----

### **CAPÍTULO 3 – Análise das atividades.....59**

3.1 Atividade 1.....	59
----------------------	----

3.1.1 Discussão da atividade 1.....	66
-------------------------------------	----

3.2 Atividade 2.....	68
----------------------	----

3.2.1 Discussão da atividade 2.....	78
-------------------------------------	----

3.3 Atividade 3.....	81
----------------------	----

3.3.1 Discussão da atividade 3.....	83
-------------------------------------	----

3.4 Atividade 4.....	84
----------------------	----

3.4.1 Discussão da atividade 4.....	88
-------------------------------------	----

3.5 Atividade 5.....	90
3.5.1 Discussão da atividade 5.....	96
3.6 Compreensão de aspectos da variabilidade.....	100
<b>CAPÍTULO 4 – Análise multidimensional.....</b>	<b>105</b>
<b>CAPÍTULO 5 – Considerações Finais.....</b>	<b>115</b>
<b>REFERÊNCIAS.....</b>	<b>123</b>

## INTRODUÇÃO

Conceitos estatísticos, assim como a linguagem própria a este tema, são utilizados constantemente em nosso cotidiano. Estes estão em diversos âmbitos como educação, saúde, economia, política e ciência. Afirmações e tomadas de decisões são feitas, baseadas em informações tratadas estatisticamente.

A mídia vem desempenhando um papel de propagadora de tais informações. Cada vez mais ela utiliza argumentos e/ou apresentações de informações através de análises e representações gráficas e de indicadores estatísticos como medidas que resumem certas distribuições de dados como, por exemplo, a média.

É importante ressaltar que, apesar da presença marcante da estatística no cotidiano das pessoas, muitos não a percebem. A razão da não percepção da estatística em diversas afirmações realizadas e amplamente veiculadas pode estar no que comumente se entende desse campo do saber. Desse modo, acredita-se que ainda sejam insuficientes as discussões atuais quanto à abordagem da Estatística e seus principais conceitos já no início da escolarização.

No Brasil, o ensino da Estatística é uma recomendação do currículo escolar da Educação Básica, desde os anos iniciais. Entretanto, a literatura vem afirmando que ainda é ausente um entendimento mais profundo de como estudantes desenvolvem a compreensão de conceitos fundamentais da Estatística, embora o número de pesquisas em tal campo tenha aumentado nos últimos anos.

A consideração da importância da exploração de conhecimentos estatísticos também na Educação Básica fez com que surgisse o Grupo de Estudos em Educação Estatística no Ensino Fundamental (GrEF), formado por pesquisadores (professores e estudantes de graduação e pós-graduação), no qual muitas das discussões tratadas no presente estudo foram aprofundadas, com a intenção de se obter maiores esclarecimentos quanto ao que se é possível ensinar de Estatística nos anos iniciais do Ensino Fundamental.

Uma das razões para a necessidade da aprendizagem de conceitos estatísticos já na Educação Básica é o grande desenvolvimento da informática. Afinal, com o advento dos computadores, a humanidade passou a ter a possibilidade de lidar com uma quantidade enorme de informações sobre o mundo nunca antes

possível. A quantidade de informações que vem sendo levantada sobre a realidade é produzida para que o homem tenha cada vez mais condições de compreender e dominar o mundo que o rodeia. Para tal, todo esse montante de dados precisa ser tratado e analisado. Analisar significa buscar regularidades, tendências, desvios, ou seja, investiga-se a característica (variabilidade) de um conjunto de dados.

A aprendizagem da Estatística envolve a compreensão de diversos conceitos, entretanto, nesse estudo será abordado, particularmente, o conceito de variabilidade, uma vez que o mesmo é um dos conceitos essenciais para o desenvolvimento do pensamento estatístico. Apesar de ser fundamental, ainda tem sido pouco explorada a compreensão dos estudantes em início de escolarização em relação à variabilidade.

Desse modo, o objetivo desse estudo foi investigar as compreensões apresentadas por estudantes de 2º e 5º anos do Ensino Fundamental em relação ao conceito de variabilidade estatística.

Assim, no Capítulo 1 a revisão da literatura é apresentada. Nesse, buscou-se ressaltar a importância da Análise Exploratória dos Dados (AED) para a melhor compreensão de conceitos estatísticos desde a Educação Básica.

No Capítulo 2 descreveu-se o percurso metodológico adotado na pesquisa, o que englobou sujeitos, procedimentos de coleta de dados e de análise dos mesmos.

Nos Capítulos 3 e 4 os resultados e discussões dos mesmos são apresentados. No Capítulo 3 com a análise de cada atividade, a partir de questões norteadoras. No Capítulo 4 faz-se uma análise multidimensional objetivando perceber o comportamento dos sujeitos nas questões de modo mais global.

Por último, no Capítulo 5, considerações a respeito de todo o exposto no estudo são realizadas, nas quais resalta-se os principais resultados encontrados e contribuições, tanto para o ensino, quanto para a pesquisa da temática em discussão.

## **CAPÍTULO 1 – REVISÃO DA LITERATURA**

### **1.1 A Análise Exploratória de Dados na Educação Básica**

Conforme afirmação de Batanero, Estepa e Godino (1991) hoje tem-se uma nova filosofia nos estudos estatísticos, introduzida por Tukey (1977), denominada Análise Exploratória de Dados (AED). Esses autores alertam que, anteriormente, a análise de dados se baseava fundamentalmente em procedimentos que conduziam a uma importância exclusiva dos cálculos estatísticos, em detrimento da importância visual da representação dos dados. Acrescentam que a análise se equiparava a um modelo confirmatório, isto é, de aceitação ou não de uma hipótese, observando-se, então, apenas regularidades. Diferente do modelo tradicional, a filosofia da AED, parte da exploração exaustiva dos dados para a geração de novas hipóteses. Assim, não são apenas as regularidades dos dados que importam a AED, são também os desvios.

Pelo potencial da AED em gerar novas hipóteses num estudo, Batanero et al (1991) salientam que a análise exploratória é especialmente útil em estudos iniciais os quais se tem pouca informação sobre os objetos de investigação. Esse tipo de análise pode ser utilizado em diversas áreas como Biologia, Ciências Humanas, Economia.

É pertinente esclarecer que o surgimento da AED e a distinção entre a mesma e o enfoque anterior foram possíveis, em grande parte, graças ao desenvolvimento tecnológico. Isso porque a exploração dos dados que permitem uma ampla variedade de gráficos e estudos mais avançados de relações entre variáveis é potencializada pela utilização de diversas ferramentas nos métodos de análise que simplificam cálculos estatísticos.

Considerando a conveniência da Estatística para a Educação Básica, a AED tem as seguintes características que a fazem apropriada ao ensino em etapas anteriores ao Ensino Superior, conforme Batanero et al (1991):

- Possibilidade de gerar situações de aprendizagem contextualizadas em temas que sejam de interesse para o estudante (a partir de dados gerados pelos próprios estudantes ou obtidos em publicações estatísticas);

- Forte apoio de representações gráficas (pela geração de múltiplas representações de um conjunto de dados e conversão de uma representação em outra);
- Emprego, preferencialmente, de estatísticas de ordem (que diminuem o efeito de valores atípicos e facilitam a atribuição de significado pelo estudante da escola básica);
- Não necessita de uma teoria matemática complexa (pois utiliza noções matemáticas mais elementares e procedimentos gráficos mais simples);
- Utiliza diferentes escalas (uma nova escala para uma variável permite outras formas de manipulação dos dados).

Percebe-se que a AED, na perspectiva dos autores citados, preocupa-se com um ensino adequado da Estatística de modo a criar situações de aprendizagens motivadoras e que tenham significado para o estudante. Evita-se, dessa forma, o uso de fórmulas e procedimentos complexos, pelo menos quando se trata do ensino voltado para estudantes da Educação Básica.

Das características elencadas nos pontos acima, aquele que corresponde à importância de representações gráficas na análise de um conjunto de dados é de particular interesse no presente estudo. Uma vez que buscamos analisar a compreensão dos estudantes em relação a esse tipo de representação a qual expressa a variabilidade de uma dada distribuição, mostrando o comportamento de um determinado fenômeno observado.

Os gráficos vêm sendo amplamente utilizados e bastante difundidos, principalmente pelos meios de informação, como a mídia impressa, indicando que a mesma já percebeu a relevância das representações de dados em gráficos como constataram Guimarães e Cavalcanti (2008).

Assim, quando pretende-se uma formação estatística adequada na Educação Básica, acreditamos ser importante considerar situações de aprendizagem contextualizadas que foquem a exploração dos dados e não apenas os procedimentos de cálculo.

É pertinente, a partir de então, uma discussão maior a respeito da formação estatística necessária aos estudantes nos anos iniciais de escolarização, conforme orientações curriculares nacionais voltadas para a etapa inicial da educação formal.

A inserção de conteúdos estatísticos no currículo escolar brasileiro responde à necessidade de incorporar as discussões científicas, desenvolvidas no campo da Estatística, ao ensino formal. Discussões que compartilham dos princípios da filosofia da AED e que acreditam que a ênfase do ensino da Estatística deve estar na análise de dados, como afirmam Ponte, Brocardo e Oliveira (2005). A análise dos dados permite que o indivíduo assuma uma perspectiva investigativa, formulando questões, coletando dados, organizando-os, fazendo sistematizações e representações, procedimentos que vão envolver interpretações e reflexões.

No Brasil, tais discussões surgem no currículo escolar dos anos iniciais do Ensino Fundamental, orientando o ensino para o desenvolvimento do pensamento estatístico em 1997, com os Parâmetros Curriculares Nacionais de Matemática (PCN). Na mesma perspectiva defendida por autores que têm contribuído com pesquisas na área da Educação Estatística, os PCN sugerem que o espírito de investigação e exploração deva orientar o desenvolvimento do estudo em Estatística. A ideia expressa no documento é que o estudante construa procedimentos para coletar, organizar, comunicar e interpretar dados, utilizando representações como tabelas, gráficos e outras de uso corrente.

Quanto aos professores, que são aqueles aos quais as orientações de ensino expressas em documentos como os PCN mais interessam, é exigido que os mesmos se adaptem a essa realidade e aprimorem seu conhecimento sobre esse conteúdo de modo a desenvolverem atividades didáticas motivadoras e significativas para os estudantes, conforme Silva (2007).

No entanto, não é apenas a formação inicial e continuada de professores que precisa de maiores subsídios, no que se refere a conhecimentos conceituais e didáticos de Estatística. Guimarães, Gitirana, Marques e Cavalcanti (2009) argumentam que poucos estudos têm sido publicados na área, tanto no que diz respeito à compreensão de conceitos, quanto a propostas didáticas que auxiliem a formação estatística dos professores. As autoras afirmam ainda, que os livros didáticos e manuais do professor focam bastante a interpretação de gráficos e

tabelas, carecendo de maior atenção às etapas inerentes de uma pesquisa estatística, assim como de conceitos estatísticos como população, amostra, frequências relativa e absoluta, média, dentre tantos outros. Desse modo, se recursos que poderiam ajudar aos professores em sua prática docente, a desenvolverem um ensino da Estatística mais eficaz, requerem maiores incentivos, por conseguinte, a formação dos professores fica comprometida.

Assim, a exigência de esforços maiores por parte daqueles que são responsáveis diretos pela instrução dos estudantes, que são os professores, cabe não apenas a esses como também a todos que de alguma forma estão preocupados com a formação estatística da população desde o início da escolarização, o que inclui pesquisadores e escritores de livros didáticos. A necessidade de tais esforços acontece principalmente pela concepção atual de que todo cidadão precisa saber quando um argumento estatístico está sendo empregado devidamente, como afirmam Ponte, Brocardo e Oliveira (2005). Assim, educar-se estatisticamente é essencial quando se almeja uma educação para a cidadania.

## **1.2 Variabilidade: Conceito fundamental da estatística**

O conceito de variabilidade é essencial à estatística. Alguns autores diriam mais, que é o coração da estatística, uma vez que não haveria necessidade desta se os dados não variassem, como argumentam Watson e Kelly (2002b).

O termo variabilidade muitas vezes é utilizado como sinônimo de variação, contudo, há uma distinção, em Estatística, no significado de tais palavras. Reading e Shaughnessy (2004), citados em Silva (2007), argumentam que a variabilidade refere-se à característica da entidade que é observada, propensa a variar ou mudar, enquanto a variação será a descrição ou medida desta característica. De acordo com Silva (2007), a preocupação em medir a variabilidade caracteriza um estudo de variação.

No presente estudo se reconhece a variabilidade e variação como termos de significados distintos. Assim, a utilização do termo variação se restringirá aos processos de medição (quantificação) da variabilidade de um conjunto de dados.

Contudo, cabe esclarecer que muitos estudos que abordam o conceito de variabilidade não estão preocupados com a distinção explicitada acima.

Estudos sobre conceitos de variação/variabilidade são bem recentes. Reading (2004) afirma que o estudo de Shaughnessy, Watson, Moritz e Reading (1999) foi um dos primeiros trabalhos que tentaram compreender, de modo sistemático, o que acontecia com o entendimento de variabilidade dos estudantes. Somente em 2003, no terceiro *Fórum de Pesquisa em Alfabetização, Pensamento e Raciocínio Estatístico* (SRTL-3), o conceito de variabilidade foi o tema central do encontro, assim como, das edições de novembro de 2004 e maio de 2005 do SERJ (*Statistics Education Research Journal*).

Como será possível verificar nos tópicos seguintes, o conceito de variabilidade envolve aspectos distintos, cujo desenvolvimento requer o conhecimento dos componentes fundamentais do raciocínio próprio a este conceito. Além do entendimento de seus componentes outros estudos mostrarão que mesmo ao explorar-se a variabilidade em uma situação específica, como a encontrada num conjunto de dados, o conceito de variabilidade pode mudar, dependendo do tipo de dados trabalhado. Outro ponto importante, que será aprofundado posteriormente, refere-se à análise de um conjunto de dados, na qual a variabilidade intrínseca ao conjunto pode ser compreendida a partir de uma visão local ou global dos dados.

### **1.3 Componentes do raciocínio variacional**

Garfield e Ben-Zvi (2005) concordando com a centralidade da variabilidade no estudo estatístico, uma vez que esta torna possível tomar decisões em momentos de incerteza, a partir de interpretação, modelação e predição de dados, delineiam sete componentes de um abrangente modelo epistemológico que, segundo os autores, são a base para um entendimento aprofundado (compreensão conceitual) da variabilidade. Os autores enfatizam ainda que desde os anos iniciais o conceito de variabilidade pode ser explorado, a partir de atividades e discussões formais e informais.

Ao sugerirem os componentes do raciocínio variacional, Garfield e Ben-Zvi (2005) apresentaram ideias-chave sobre os mesmos e como poderiam ser

desenvolvidos. A seguir, o Quadro 1, inspirado em Silva (2007), apresenta os itens dos sete componentes propostos pelos autores de forma sintetizada. Por ser um modelo bastante amplo vai englobar aspectos complexos da variabilidade. Complexo do ponto de vista em que normalmente são tratados no ensino conceitos estatísticos como desvio padrão, intervalo interquartil, distribuições especiais, modelos residuais, os quais não são o foco aqui, uma vez que são temas mais comuns a etapas mais avançadas da escolarização.

**Quadro 1: Componentes sintetizados do raciocínio variacional propostos por Garfield e Ben-Zvi (2005)**

<b>Componentes</b>	<b>Ideias-chave</b>	<b>Como avaliar</b>
1) Desenvolvimento de idéias intuitivas de variabilidade	A variabilidade está em toda parte. Variabilidade é uma característica global de um conjunto de dados, uma entidade.	A partir: da descrição da variabilidade ou formato de uma distribuição; de predições; da reflexão de formas de reduzir a variabilidade de um dado contexto; da comparação de gráficos pensando qual tem a medida de variação maior ou menor.
2) Descrição e representação da variabilidade	Gráficos de dados mostram diferenças e padrões que ajudam a incidir sobre características globais da distribuição; Diferentes gráficos revelam distintos aspectos variacionais de um conjunto de dados.	A partir: da interpretação de gráficos e descrição da variabilidade de cada variável, dadas medidas sumárias; da escolha de medidas de centro e medidas de variabilidade apropriadas a distribuições particulares; da construção de gráficos ao se fornecer centro e variação.
3) Uso da variabilidade em comparações	Comparações de dois ou mais conjuntos de dados representados em gráficos sob uma mesma escala permite verificação de variabilidade e especulações.	A partir: da comparação de gráficos de mesmas temáticas, para saber no que diferem; de respostas a questão qual gráfico tem menor ou maior variabilidade, quando se tem que coordenar, forma, centro e medidas diferentes de variação.
4) Reconhecimento da variabilidade em tipos especiais de distribuição	Em uma distribuição normal, a média e o desvio padrão fornecem informações úteis sobre a variabilidade.	A partir: da construção de gráficos apresentando a variação de dados, ao se fornecer a média e o desvio padrão de uma distribuição normal; determine se a variabilidade da variável $x$ pode ser explicada em função da variável $y$ , em conjuntos bivariados.

5) Identificação de padrões de variabilidade em montagem de modelos	Os dados podem, por vezes, ser reorganizados e transformados para melhor revelar padrões ou encaixar um modelo.	A partir da verificação se um conjunto de dados assemelha-se a uma distribuição normal ou se um gráfico de dados bivariados sugerem uma relação linear.
6) Uso da variabilidade para prever amostras ou resultados aleatórios	Existe variabilidade em resultados de eventos, que pode ser prevista; Amostras variam de modo previsível, com base no tamanho da amostra e população, assim, a variabilidade pode ser mais facilmente explicada e descrita.	A partir: de respostas à questão qual sequência estatística é mais plausível ao se fornecer opções de amostras para um tamanho de amostra dado; da proposição de uma ou mais amostras de dados de uma população determinada; de respostas a qual resultado é mais provável em experimentos aleatórios.
7) Consideração da variabilidade como parte do pensamento estatístico	A variabilidade no pensamento estatístico está em produção de dados e em análise de dados.	A partir: da investigação de um problema em um conjunto de dados, o que requer descrição e explanação da variabilidade na resolução de problemas; da realização dos passos de uma investigação estatística, revelando se/como os estudantes consideram a variabilidade dos dados.

Do modelo proposto descrito no quadro, o presente estudo deter-se-á aos aspectos, daqueles ressaltados por Garfield e Ben-Zvi (2005), relacionados à descrição da variabilidade, à predição de resultados, à representação da variabilidade (em situações de construção de gráfico), à comparação de conjuntos de dados distintos representados em gráficos e à reflexão de formas de se reduzir a variabilidade de um conjunto. Acredita-se que sejam enfoques da variabilidade pertinentes de serem investigados quando se tem interesse nas compreensões de estudantes dos anos iniciais de escolarização quanto ao conceito.

Tais aspectos da variabilidade estão localizados no modo de avaliação (terceira coluna do quadro) de alguns dos componentes explicitados. Assim, tomando-se como exemplo a habilidade de refletir sobre formas de redução da variabilidade de um conjunto, esse é um aspecto do componente 1 do modelo de Garfield e Ben-Zvi (2005).

## 1.4 Variabilidade de dados

Variabilidade existe em diversas situações, não apenas em conjuntos de dados coletados, como também em situações amostrais, probabilísticas e outras. Afinal, amostras de uma mesma população podem variar. Assim como também, apesar da probabilidade, há variabilidade nas chances de um evento ocorrer. Diferente de estudos como o de Watson e Kelly (2002a; 2002b), que integraram distintas situações como essas, em que existe variabilidade, no presente estudo o foco está na variabilidade de conjuntos de dados, apresentados em representações gráficas.

Aprofundando a análise de variabilidade de dados, ainda pode-se encontrar uma distinção entre variabilidade de dados categóricos e variabilidade de dados quantitativos (KADER & PERRY, 2007).

Com o objetivo de desenvolver o coeficiente de “*unlikeability*”, como medida de variabilidade para variáveis categóricas, Kader e Perry (2007) defendem a diferenciação entre os dois tipos de dados. Os autores afirmam que o conceito de *unlikeability* é mais natural que variação a partir da média (para dados quantitativos). *Unlikeability* foca a frequência com que uma observação difere da outra e não a quantificação dessa diferença. É em Loosen, Lion, Lacante (1985), citados em Kader e Perry (2007), que esses últimos vão se fundamentar para a proposta do conceito de *unlikeability* acima descrito.

Conforme Kader e Perry (2007), Loosen et al (1985) apresentaram a estudantes de graduação em Psicologia dois conjuntos de dados representados por blocos vermelhos (conjunto 1) e blocos amarelos (conjunto 2). Abaixo, segue a representação física de tais blocos, de acordo com os primeiros autores:



Assim, refletindo sobre a proposta de Kader e Perry (2007) pode-se inferir que os estudantes pensaram no conjunto 1 como tendo 6 categorias, pois não tinha nenhuma barra de igual altura e no conjunto 2 como tendo apenas 2 categorias, pois só tinham duas alturas diferentes, então, nesse raciocínio é lógico concluir que o primeiro conjunto variou mais.

Como Kader e Perry (2007) estavam preocupados em desenvolver o coeficiente de *unlikeability* com dados categóricos, é pertinente a exemplificação oferecida pelos autores desse tipo de dados, uma vez que o estudo descrito acima baseou-se em dados quantitativos. Assim, como exemplo de atividade em que se explora variabilidade em dados categóricos os autores apresentaram a seguinte situação:

Grupo 1: 7 respostas na Categoria A; 3 respostas na Categoria B Grupo 2: 5 respostas na Categoria A; 5 respostas na Categoria B Grupo 3: 1 resposta na Categoria A; 9 respostas na Categoria B
--

Acima, cada grupo contém dez respostas e duas categorias de respostas (A e B), o que variou foi a frequência das categorias. Os autores afirmam que a comparação dos diferentes grupos permite os seguintes questionamentos: Qual grupo variou mais? Qual variou menos? Para resolução de tais questões, os autores salientam que no grupo 2 apenas cinco respostas estão numa mesma categoria, variou mais, enquanto que no grupo 3 nove respostas estão num mesmo grupo, o que torna os valores mais parecidos. O grupo 3, então, apresenta-se mais homogêneo.

A partir da explicitação das duas situações em que a variabilidade de dados toma um significado diferente, em um caso mais intuitiva (variabilidade de dados categóricos) e no outro mais complexa (variabilidade de dados quantitativos envolvendo a compreensão de média), Perry e Kader (2005) vão alertar que os professores podem estar pensando num conceito de variabilidade e os estudantes em outro. Isso porque ao analisarem a variabilidade numa distribuição gráfica estudantes podem estar pensando com que frequência as observações diferem de uma para outra e não em o quanto diferem.

A discussão proposta por esses autores suscita questões importantes de serem consideradas no âmbito do presente estudo, uma vez que pode-se estar investigando a influência do tipo de dado representado em gráfico de barras, quando articulado com aspectos que promovem o desenvolvimento do raciocínio de variabilidade, como comparação de conjunto de dados e predições de resultados a partir da análise de informações representadas num gráfico.

Percebe-se que apesar de ser recente os estudos abordando o conceito de variabilidade estatística na perspectiva de uma Análise Exploratória dos Dados, tal qual proposta por Tukey (1977), já se pode encontrar na literatura discussões importantes a respeito do conceito. Discussões que apontam para situações diferentes em que a variabilidade pode estar sendo explorada por professores e compreendida por estudantes.

Situações diferenciadas são importantes para a compreensão de um conceito, pois, de acordo com Vergnaud (1996), no ensino e na aprendizagem um conceito não pode ser reduzido a sua definição, afinal, as situações e problemas a resolver é que permitem que o mesmo adquira sentido para o sujeito. Vergnaud (1996) argumenta que um conceito é formado por três conjuntos: *S, I, R*.

S: conjunto das situações que dão sentido ao conceito (referência);  
 I: conjunto das invariantes nas quais assenta a operacionalidade dos esquemas (o significado);  
 R: conjunto das formas pertencentes e não pertencentes à linguagem que permitem representar simbolicamente o conceito, as suas propriedades, as situações e os procedimentos de tratamento (o significante). (VERGNAUD, 1996, p.166).

O autor salienta que a investigação do desenvolvimento e funcionamento de um conceito no processo de aprendizagem ou em sua utilização considera, necessariamente, os três conjuntos.

A partir da proposta de Vergnaud (1996) conclui-se que o desenvolvimento da variabilidade enquanto conceito e a melhor compreensão do mesmo pelos estudantes requerem práticas de ensino que abranjam situações diversas, representações distintas e conhecimento dos invariantes (propriedades) da mesma. No campo da pesquisa, pressupõe-se também a necessidade de investigações futuras a respeito da tríade (*S, I, R*) que compõe o conceito de variabilidade

estatística, pois não se tem ainda na literatura discussões a esse respeito, o que é fundamental para o desenvolvimento do estudo de todo campo conceitual.

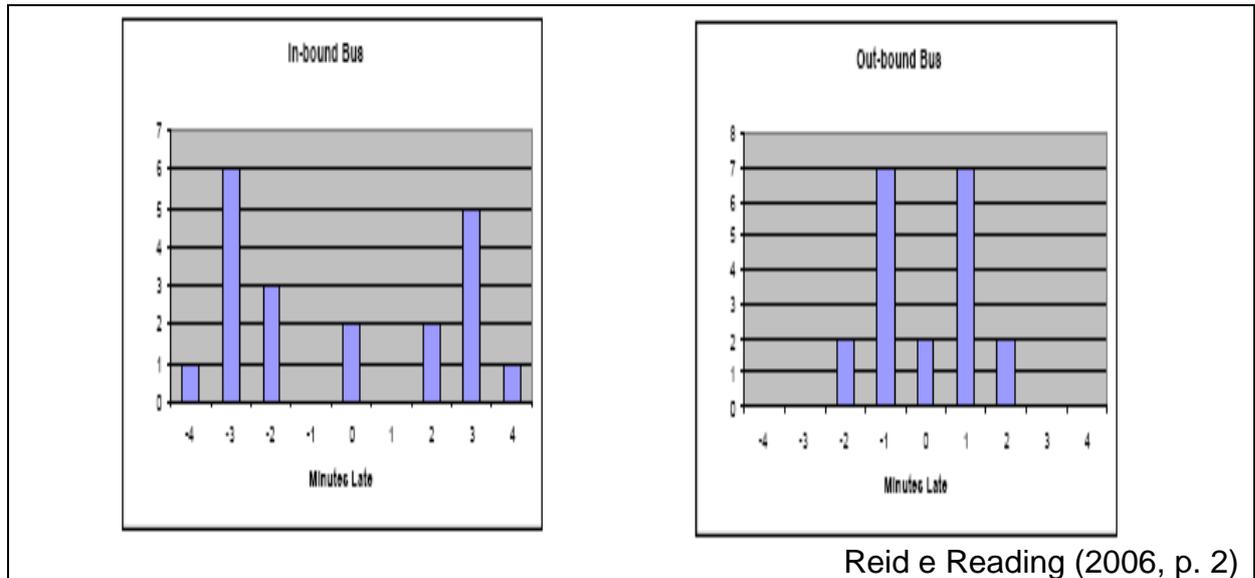
### **1.5 Compreensões de estudantes quanto à variabilidade**

Após situar a variabilidade estatística sob um enfoque mais teórico da construção conceitual, nesse momento é pertinente voltar-se para pesquisas que têm procurado investigar, experimentalmente, que compreensões os estudantes têm apresentado a respeito do conceito de variabilidade.

Reid e Reading (2006) e Cooper e Shore (2008) desenvolveram estudos sobre o conceito de variabilidade e variação com estudantes de graduação. Reid e Reading (2006) tiveram como objetivo desenvolver uma hierarquia para as compreensões ou considerações de estudantes de graduação sobre variação. Para tal, propuseram um pré-teste a 32 estudantes no início de um curso introdutório de estatística e um pós-teste a 23 estudantes no término do curso. Foram utilizadas quatro questões: definição de variabilidade (Q1); comparação de conjuntos de dados (Q2); probabilidade (Q3); amostragem e probabilidade (Q4).

Das questões exploradas pelos autores, as duas primeiras estão mais relacionadas ao foco do presente estudo e por isso cabe entender melhor sua elaboração e considerações dos estudantes a seu respeito.

A primeira questão 1 (Q1) perguntava o que a variabilidade significava para os estudantes e pedia uma explanação verbal e/ou exemplo. A questão 2 (Q2) solicitava a comparação de dados apresentados em dois gráficos sobre o tempo que os ônibus levavam em duas rotas. Os gráficos, a seguir, exemplificam a comparação, entre conjuntos de dados, requerida na atividade.



É importante esclarecer que o enunciado dessa questão explicitava que as duas rotas de ônibus foram monitoradas na saída e na chegada. O número zero minuto indica que o ônibus estava no tempo previsto, enquanto um número negativo indica que o ônibus estava adiantado. O Quadro 2, mostra a classificação realizada pelos autores das respostas dos estudantes tanto para a primeira questão, quanto para a segunda, considerando níveis hierárquicos de compreensão.

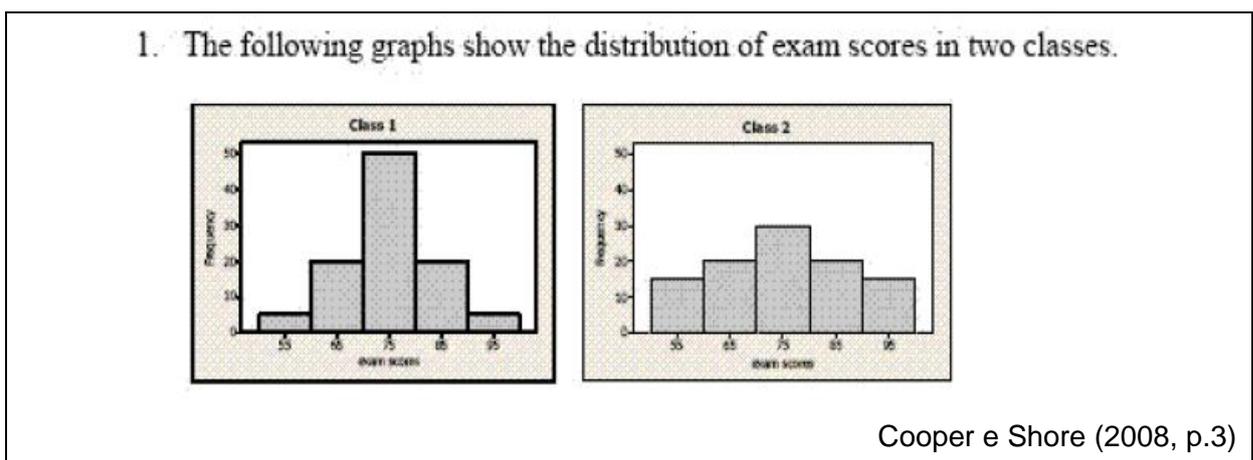
**Quadro 2: Classificação das respostas dos estudantes nas duas primeiras questões de acordo com hierarquia construída**

Níveis hierárquicos	Q1- Definição de variabilidade	Q2 – Comparação entre dois gráficos
Não consideração da variação	Não referência a nenhuma variação.	Referiam-se a medidas de centro, mas não a medidas de variação.
Fraca consideração	Considerava uma fonte de variação, mas tinham dificuldades de expressá-la.	Referiam-se a amplitude e/ou descrição básica do formato.
Consideração desenvolvida	Descreviam claramente uma fonte de variação.	Referiam-se a medidas de localização e descrição mais detalhada da distribuição.
Forte consideração	Descreviam claramente mais de uma fonte de variação.	Forneciam mais informações sobre a distribuição, tal como explicitações de proporção.

Os autores verificaram que o nível de compreensão ou consideração dos estudantes diferiu entre as questões, assim como variou do início para o fim do

curso. Houve uma redução na proporção de estudantes mostrando não consideração de variação para ambas as questões. Reid e Reading (2006) salientam ainda que no pós-teste as respostas foram mais sofisticadas no que se refere à terminologia utilizada pelos estudantes e à ênfase na medida e modelação da variação ao invés de simplesmente descrição da mesma. Quanto ao tipo de mudança percebida no desempenho dos estudantes, após o curso, para cada questão, observou-se que na Q1 mais de 40% dos estudantes apresentaram aumento em suas compreensões no final do curso, entretanto, na Q2 o mesmo não ocorreu, verificou-se uma redução no nível de compreensão em 26%. A partir dos dados encontrados os autores argumentaram que o contexto da atividade interferiu no desempenho dos estudantes e alertaram para a necessidade de uma atenção maior no currículo escolar.

Cooper e Shore (2008), também em cursos introdutórios de estatística, identificaram concepções equivocadas de estudantes a respeito dos conceitos de variabilidade de dados e medidas de centro, utilizando a representação em gráficos. A partir de entrevistas, o conceito de variabilidade foi apresentado, em histogramas, a 186 estudantes de graduação. A figura, abaixo, corresponde aos gráficos apresentados aos estudantes na pesquisa.



Na atividade, os sujeitos precisariam comparar os dois gráficos que mostravam a distribuição de notas de exame em duas classes. É importante esclarecer que a mesma escala foi utilizada (10, 20, 30, 40 e 50), assim como os eixos horizontais apresentaram os mesmos valores (55, 65, 75, 85 e 95). Os dois

conjuntos de dados apresentavam ainda a mesma média, mediana, amplitude e formato de distribuição. Conforme as autoras, o esperado seria os sujeitos considerarem que o segundo gráfico (classe 2) apresentou uma variabilidade maior. Entretanto, as autoras observaram que apenas 27% responderam que a classe 2 variou mais, 20% afirmaram que os dois gráficos tinham a mesma variabilidade e aproximadamente metade dos estudantes responderam que a classe 1 teve maior variabilidade. Conforme argumentam as autoras, os alunos que optaram pela classe 1 expressaram uma concepção equivocada uma vez que consideraram que o histograma com maior variabilidade nas barras indicava maior variabilidade no conjunto de dados. Já os estudantes que acreditaram que os dois gráficos tinham uma mesma variabilidade, consideraram que a amplitude era a mesma, corroborando da ideia de que existe uma tendência entre estudantes, quando lidam com histogramas, em focar a escala horizontal (eixo x), ignorando a significância das frequências.

Cooper e Shore (2008) salientam que o mais preocupante desses resultados é que metade dos estudantes julgou a variabilidade focando na altura das barras, variação da frequência, ao invés dos valores dos dados. As autoras sugerem que a fonte dessa confusão pode ser a não diferenciação, pelos estudantes, dos diversos tipos de representação gráfica (envolvendo barras) e seus diferentes métodos de análise da variabilidade. Assim, essas autoras acreditam que as dificuldades dos estudantes estão em compreender que a representação da variabilidade dos dados em histograma é diferente de compreender representações em outros gráficos de barras.

É importante ressaltar que Loosen, Lioen e Lacante (1985) não utilizaram histogramas, porém ao apresentarem a estudantes um conjunto de blocos com alturas distintas também constataram dificuldades semelhantes por parte dos estudantes, uma vez que os mesmos percebiam muito mais a diferença na altura das “barras” (blocos) do que o desvio padrão. Apesar das representações diferenciadas, ambos os estudos – Cooper e Shore (2008) e Loosen et al (1985) – atribuem as dificuldades dos estudantes na comparação da variabilidade de distribuições ao pouco conhecimento dos mesmos com os tipos de representações

utilizados em distribuições estatísticas e as peculiaridades de cada um no que se refere à análise da variabilidade.

A comparação entre os estudos de Loosen et al (1985) e Cooper e Shore (2008) ainda permite a afirmação de que apesar do segundo ser bem recente maiores avanços na compreensão de estudantes de cursos de graduação, no que se refere à comparação entre distribuições não foram constatados, o que só confirma a necessidade de outros estudos que busquem uma compreensão mais sistemática do entendimento do conceito aqui explorado, desde os anos iniciais.

As pesquisas até então descritas exigiam de estudantes, em níveis mais avançados de ensino, habilidades como definição da variabilidade e comparação de gráficos, as quais também serão investigadas em estudantes em anos iniciais de escolarização. Contudo, a grande diferença entre os estudos voltados para esses distintos públicos, que já se pode adiantar, é que existe uma preocupação maior com a compreensão dos estudantes, em níveis mais avançados, de conceitos relacionados à medida da variabilidade. Para isso, medidas como amplitude (mais simples) e desvio padrão (de maior complexidade) são exploradas na atividade de alguma forma ou esperadas nas respostas dos estudantes.

Em se tratando dos anos finais do Ensino Fundamental, Santos e Gitirana (1999) investigaram estudantes de 6ª série (7º ano – aproximadamente 12 anos de idade) em leitura de gráfico de barra com categoria ordinal, na qual o eixo horizontal apresentava uma série temporal. Uma sequência de atividades foi utilizada em torno do software educacional Tabletop (1994) para investigação das estratégias dos estudantes ao interpretar gráficos. Um pré-teste, seguido de 4 (quatro) encontros de intervenção, mais o pós-teste caracterizaram tal estudo. No pré-teste os autores perceberam que os sujeitos apresentaram dificuldades nas questões que envolviam leituras mais globais do gráfico (variações, taxa de variação, extrapolação). Com o processo de intervenção que envolvia análises de diferentes gráficos e de Diagrama de Venn, os estudantes passaram a considerar mais a variação, principalmente na questão de extrapolação, na qual os sujeitos conseguiram fazer considerações qualitativas e globais em suas interpretações. Contudo, a questão que pedia a taxa de maior crescimento no gráfico continuou sendo uma grande dificuldade (apenas

5,88% dos sujeitos acertaram) sendo a principal estratégia utilizada no pós-teste (62,5%) a recategorização do gráfico em barras altas e baixas.

Ben-Zvi (2002) também investigou compreensões de estudantes dos anos finais do Ensino Fundamental, na faixa etária de 13 anos de idade. Em duplas, 80 estudantes criaram questões de pesquisa, hipóteses, representaram dados em gráficos e fizeram explicações, a partir do tema “imigração em Israel”. Ben-Zvi constatou que os estudantes foram hábeis em: 1) Formular questões de pesquisa e hipóteses globais (67,5%); 2) Interpretar gráficos globalmente em uma variedade de formas (87,5%) percebendo períodos de imigração, existência de variabilidade, ciclos, tendências ou associações; 3) Explicar o que era “ondas de imigração” e sugerir uma representação gráfica para o fenômeno, numa perspectiva que indicava um entendimento global dos dados (45%). Além de interpretar gráficos os estudantes foram fluentes em escolher uma variedade de gráficos para expor formatos globais dos dados e em manipular representações.

Tanto o estudo de Santos e Gitirana (1999), quanto o estudo de Ben-Zvi (2002) exploraram dados que exprimiam uma continuidade (distribuição de uma série temporal) e focaram nas interpretações locais e globais que se pode fazer ao analisar um conjunto de dados representado num gráfico. Contudo, chegaram a resultados bem distintos no que se refere à análise (local/global) dos dados, mesmo tratando-se de estudantes de faixa etária aproximada.

Apesar dos resultados distintos a que chegaram tais estudos, é pertinente esclarecer que ambos compartilham da ideia de que na análise dos dados é possível visualizar os dados local e globalmente. Ben-Zvi (2002) distingue visão local / global dos dados da seguinte forma:

Entendimento local de dados envolve focar em valores individuais (ou em poucos deles) dentro de um grupo de dados (uma entrada particular em uma tabela, um ponto único em um gráfico). Entendimento global refere-se à habilidade para pesquisar, reconhecer, descrever e explicar padrões gerais em um conjunto de dados (ex: mudança no decorrer do tempo, tendências) (BEN-ZVI, 2002, p. 1).

Para a diferença nos resultados das pesquisas descritas acima, pode-se buscar explicações no fato de que esse estava preocupado com habilidades outras,

envolvendo a variabilidade de dados, a serem desenvolvidas junto aos estudantes. O estudo de Santos e Gitirana (1999) se propôs a investigar habilidades relacionadas, especificamente, a análise de conjuntos de dados direcionada por questões que exigiam leituras locais ou globais dos dados, representados em um tipo de gráfico, o de barras. Enquanto que o estudo de Ben-Zvi (2002) investigou aspectos diversos envolvendo a pesquisa, desde a elaboração de questões até a análise de dados, em que se pode estar tendo visões locais e globais dos dados.

Entretanto, Ben-Zvi (2002) mesmo salienta que sua pesquisa contrariou dados de pesquisas anteriores, o que provavelmente deveu-se às peculiaridades metodológicas e curriculares adotadas. Outra novidade que esse autor apresenta nesse estudo, a partir dos dados encontrados, é que visões locais e globais dos dados estão integradas dinamicamente e flexivelmente. Tal afirmação deve-se a constatação de que os estudantes apresentaram vários níveis de entendimento dentro de suas visões local e global, combinando ambas as visões de forma variável. Ben-Zvi (2002) sugere que o desenvolvimento dos estudantes não parte de uma visão local para uma visão global, sofisticações acontecem em ambas as visões. Para este desenvolvimento é possível duas trajetórias: vertical (sofisticação dentro de uma visão) ou horizontalmente (sofisticação nas visões local e global).

Os resultados apresentados por Ben-Zvi (2002), no estudo acima explicitado, realmente contraria estudos anteriores, devido aos fatores já levantados. Contudo, é importante esclarecer aqui que os sujeitos de sua pesquisa, no início do curso referido anteriormente, enfatizavam visões locais dos dados e no final do curso eram capazes de perceber os dados globalmente. Isso já indica que mesmo os estudantes mostrando-se competentes em combinar ambas as visões foi necessário existir um trabalho de intervenção que possibilitasse o desenvolvimento da visão global dos dados. Assim, já se é possível vislumbrar encaminhamentos, a partir de intervenções didáticas, a fim de melhor desenvolver outra visão dos dados, além da local ou pontual, pois se ambas as visões estão integradas é importante que o ensino contemple-as como um todo.

Antes de discutir a compreensão da variabilidade dos dados pelos estudantes dos anos iniciais de escolarização é pertinente estabelecer uma relação entre o que se entende por visão local / global dos dados e o conceito de variabilidade.

Ao se analisar um conjunto de dados se está percebendo a variabilidade do conjunto, ou seja, faz-se uma análise global. Quanto à análise pontual, os estudos aqui destacados (BEN-ZVI, 2002; GUIMARÃES, 2002; SANTOS e GITIRANA, 1999) concordam que em tal visão dos dados o foco da análise está em valores individuais de uma dada distribuição, como um ponto no gráfico.

Entretanto, questiona-se aqui se na identificação de um ponto específico no gráfico (visão local) também não se está considerando a variabilidade do conjunto. Isso porque ao se analisar um conjunto de dados, mesmo baseando-se em pontos individuais, foi necessário perceber aquele ponto dentro de um conjunto, ou seja, comparações precisaram ser realizadas. Afinal, como analisar um conjunto de dados sem fazer comparações?

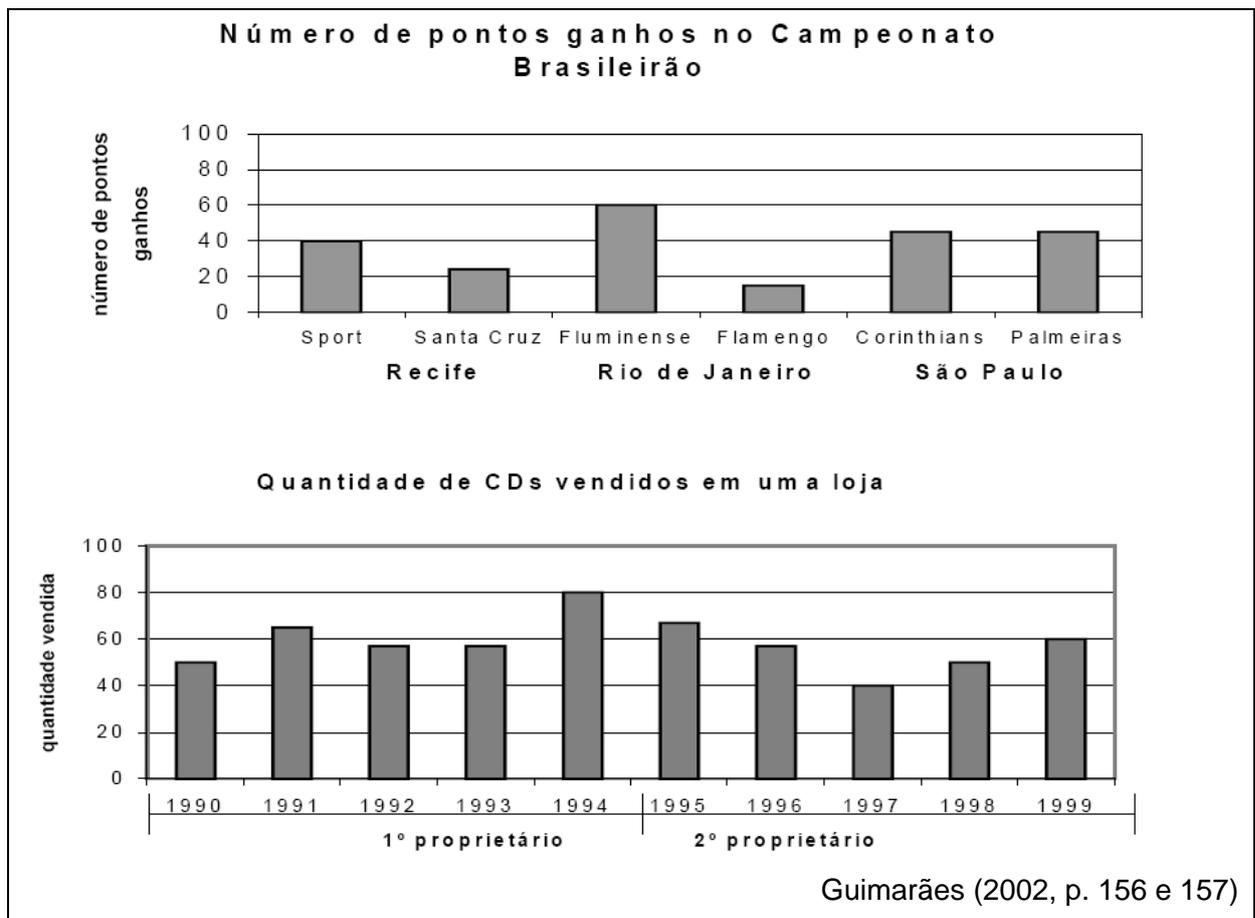
Conforme proposta de Ben-Zvi (2002), ambas as visões (global e local) estão integradas. Desse modo o questionamento anterior só salienta a necessidade de outros estudos, que indiquem como ocorre a relação entre as diferentes análises de um conjunto de dados.

É importante ressaltar que apesar de Ben-Zvi (2002) sugerir que as visões locais e globais estejam integradas o autor não discute nesse estudo a relação entre variabilidade e ambas as visões. As reflexões apresentadas aqui partem da necessidade de entender como a variabilidade, que vem sendo concebida de modo tão abrangente, ao caracterizar um conjunto de dados, pode ser compreendida e explorada dentro da discussão já realizada em muitos estudos, no Brasil inclusive, de análises ou interpretação dos dados baseadas numa visão mais local (também chamada de pontual) ou global.

A seguir são descritos os estudos interessados nas compreensões de crianças dos anos iniciais do Ensino Fundamental a respeito do conceito de variabilidade. Estudos nos quais algumas habilidades já investigadas junto aos estudantes dos anos finais do Ensino Fundamental reaparecem, assim como percebem-se pesquisas em que análises das compreensões dos estudantes focando em visões locais e globais apresentadas pelos mesmos também são utilizadas.

Guimarães (2002) investigou as compreensões de 107 estudantes brasileiros de 3ª série (4º ano), ao interpretarem e construírem gráficos de barra. Cinco tipos de

atividades foram elaborados para esse fim, nos quais os sujeitos interpretaram gráficos de barra de categorias nominais e ordinais (eixo horizontal), inclusive gráficos com valores múltiplos para um descritor (relação entre duas variáveis), respondendo a questões pontuais (ou locais) e globais. Quanto à construção de gráficos, a partir de um banco de dados, os estudantes construíram tanto gráficos de categorias nominais como ordinais. Abaixo, segue exemplificação de gráficos utilizados em tal estudo, o primeiro de categoria nominal e o segundo ordinal:



A autora percebeu que localizar pontos extremos (análise pontual dos dados), independente do tipo de categoria (nominal ou ordinal), foi uma tarefa fácil para aqueles sujeitos, entretanto, os mesmos apresentaram dificuldades em interpretar variação nos dois tipos, em questões que solicitava a comparação entre dois pontos e quantificação da variação. A mesma dificuldade com a compreensão da variação foi encontrada na construção de gráficos, quando os estudantes precisavam representar os dados com barras. Conforme exemplificação de Guimarães (2002), as estratégias mais utilizadas por esses estudantes ao representarem acréscimos de

peso de bebês, foi registrar apenas o peso do último mês ou somar todos os pesos, sem expressar o aumento, isso em dados ordinais. Os estudantes ao localizarem ausência de variação também tiveram facilidade, indicando que compreender ausência de variação é diferente de compreender variações, pelo menos quando se tratou de dados em que as categorias eram ordinais.

Outra situação apresentada pela autora refere-se à extrapolação de dados, na qual um percentual superior a 87% das crianças demonstrou compreensão da questão, quando essas foram solicitadas a ir além das informações apresentadas no gráfico. Como exemplo de uma questão de extrapolação utilizada, tem-se: “*Qual a quantidade de CDs que você acha que vão ser vendidos em 2000? Por quê?*”. Nessa questão se tinha informações no gráfico até 1999, desse modo, exigia-se uma leitura global dos dados para predição de resultados, habilidade que muitos sujeitos da pesquisa demonstraram domínio, uma vez que a análise global do gráfico foi uma estratégia amplamente utilizada. A autora, ainda, intrigada com as respostas incorretas dos alunos, resolveu ao final da situação experimental escolher uma dupla que havia a princípio apresentado dificuldades em compreender as questões sobre variação e dar continuidade realizando uma conversa sobre a situação. Para tal, apresentou e refletiu com a dupla uma situação mais familiar e com pequenas intervenções na interação as crianças demonstraram compreender o que estava sendo solicitado e transferiram o mesmo raciocínio da questão mais familiar para a questão anterior (de experimentação), que haviam respondido de forma inadequada. O trecho de falas, a seguir, esclarece esse ponto:

“J” e “T” resolveram uma tarefa similar onde o que aumentava era o peso de 3 peixes durante 3 meses. As duas também consideravam que a resposta correta implicava na soma de todos os valores.

Pq: Tenho duas amigas, uma pesava 45 e a outra 50 (escrevo em uma folha). Agora essa (apontando para 45) tá pesando 49 e essa 51. Quem foi que engordou mais?

As duas: ela (apontando para a que passou de 45 para 49)

Pq: Por quê?

T: Porque ela engordou 4Kg e aqui só 1.

Pq: Ah! E qual dos peixes engordou mais?

J: Peraí, deixa eu ver, 4 Engordou 4.

Pq: 4 onde?

J: aqui, porque 26, 27, 28, 29. Dá 4 (vão resolvendo todos os aumentos corretamente).

(GUIMARÃES, 2002, p. 178)

Assim, apesar das dificuldades da maioria dos estudantes com o aspecto da variabilidade abordado nesse estudo, Guimarães (2002) salienta que a análise variacional pode ser compreendida por crianças de aproximadamente 9 (nove) anos de idade, uma vez que respostas adequadas e avanços foram observados.

Arcoverde, Barbosa e Guimarães (2004) analisaram o conhecimento de estudantes de 3ª série (4º ano) do Ensino Fundamental interpretando gráficos de barras, num processo de ensino-aprendizagem a partir de sequências didáticas, desenvolvidas em 4 (quatro) aulas. Nas duas primeiras aulas foram propostas atividades de interpretação de gráficos com categorias nominais e ordinais, nas quais eram feitas perguntas que exigiam tanto análises pontuais como globais dos dados. Devido às dificuldades dos alunos em comparar os dados ao quantificá-los, sendo necessárias leituras mais globais, a terceira aula foi planejada a fim de explorar esse tipo de interpretação nos gráficos. Nessa aula, propôs-se que os estudantes trabalhassem em duplas, acreditando-se que avanços na compreensão dos estudantes seriam alcançados, contudo, tal procedimento não elevou o desempenho dos estudantes. Na última aula foi proposta pelas pesquisadoras a construção de gráfico com a intenção de verificar se a representação de dados pelos estudantes possibilitaria uma melhor compreensão da variação, atividade que gerou resultados positivos. Desse modo, apesar das dificuldades iniciais de compreensão da variabilidade realizando uma análise global dos dados, depois de uma curta intervenção, na qual estudantes analisaram gráficos coletivamente e construíram gráficos, uma evolução na compreensão da variação /variabilidade foi constatada.

Com estudantes de faixa etária aproximada, tanto com os sujeitos da pesquisa de Guimarães (2002), quanto com os do estudo de Arcoverde et al (2004), Watson e Kelly (2002b) num estudo que envolvia uma intervenção, investigaram o ensino da variabilidade em situações de acaso (na qual media-se a chance de um evento ocorrer) e análise de dados com crianças entre 8 e 9 anos de idade. Na investigação, os autores utilizaram pré e pós-testes com 72 crianças de três escolas, que tiveram dez aulas durante oito semanas com um especialista em Matemática para crianças. As aulas que envolveram a variabilidade de dados, nosso foco neste estudo, realizaram-se com situações em que:

- Os estudantes, em duplas, classificaram doces coloridos e construíram um gráfico de colunas, posteriormente, fizeram inferências quanto às possíveis razões para a variabilidade verificada nas cores dos doces;
- Os estudantes deviam representar em gráficos a quantidade de pessoas de suas famílias, coletivamente. Objetivou-se desenvolver ideias a respeito da definição de dados a serem coletados, diferentes representações dos dados e descrição do formato geral desses mesmos dados. Durante tal atividade, as crianças representaram os dados ao ficarem de pé, em colunas, converteram os dados em torres de blocos e construíram um gráfico de colunas;
- Os estudantes desenvolviam uma pesquisa, com a finalidade de gerar dados a partir da medição de quanto tempo os mesmos poderiam ficar em cada pé, com os olhos fechados. Depois foi construído um gráfico (dot plot empilhados) para comparação dos dois conjuntos de dados. As crianças foram solicitadas a analisar pontuações mais comuns, incomuns e a amplitude.

Os autores salientam que os estudantes se mostraram particularmente interessados em coletar dados sobre eles mesmos e afirmam que, apesar desses estudantes não alcançarem o mais alto nível de desempenho possível, a análise do pré e pós-teste demonstrou progressos estatisticamente significantes no entendimento dessas crianças da variabilidade nas situações observadas. Um exemplo oferecido foi a situação de predição de resultados a partir de dados representados em gráficos, na qual respostas dos estudantes baseadas em padrões foi mais comum no pós-teste. Watson e Kelly (2002b) afirmam que os resultados também são educacionalmente significativos uma vez que foi constatado ser possível a inclusão deste conceito no currículo deste ano de ensino.

Watson e Kelly (2002a) também investigaram sete crianças de seis anos de idade a respeito da emergência de ideias focando quatro temas que envolviam variabilidade de dados, além de variabilidade em situações de probabilidade (ou acaso). Os temas tratados pelos autores foram: 1) apreciação de incerteza e variação; 2) observação e criação de representações; 3) apreciação da necessidade de contar e conservar quantidades; 4) uso de dados para interpretação e predição. Para o desenvolvimento dos temas foram elaboradas quatro atividades, que

integravam dois ou três temas. O Quadro 3 apresenta a descrição e resultados do desempenho dos estudantes em três atividades das quatro que exploravam a variabilidade de dados (uma delas não será descrita aqui, por tratar-se da variabilidade em situações de probabilidade, o que não é o foco do presente estudo):

**Quadro 3: Descrição e resultados de atividades de Watson e Kelly 2002a**

Atividades	Descrição	Resultados
Criação, interpretação e predição de pictograma.	Cartões representativos de sete crianças e livros lidos por estas. Pedia-se que os estudantes representassem quanto cada criança tinha, adicionassem um livro da biblioteca para cada criança e respondessem qual o total de livros.	Todos os estudantes contaram os livros. A arrumação dos livros pelos estudantes foi variada: um fez de forma que não poderia ser contado; 3 colocaram os livros em torno das crianças, podendo contá-los; e 3 organizaram em colunas. Predizer quem gostaria de receber um livro de presente no Natal foi fácil, pois respondiam: <i>“Lisa porque leu mais”</i> . Quanto à predição de quantos livros uma nova criança da classe teria lido foi mais difícil e os estudantes utilizaram histórias imaginativas, sem uma análise mais cuidadosa dos dados.
Demonstração de habilidades para ler e interpretar gráficos de barras.	Um gráfico de barras apresentou o modo como 18 crianças chegaram à escola um dia, se de carro, de ônibus, de bicicleta ou andando. Foi perguntado quantos a mais chegaram de ônibus que de bicicleta e quantas crianças estavam na classe. Os estudantes, ainda, obtiveram ajuda na leitura da escala do gráfico.	Cinco dos sete estudantes da pesquisa solucionaram o problema implícito $7 - 3 = 4$ ou $3 + 4 = 7$ . Cinco contaram o número de crianças da classe, mas essa tarefa foi mais difícil que no pictograma. Para predizer como um novo aluno chegaria à escola, um estudante usou o gráfico e respondeu ônibus porque mais crianças chegaram de ônibus, os demais contaram histórias baseadas em experiências pessoais. Explorou-se ainda o entendimento dos estudantes da conservação da quantidade total de crianças ao mover as colunas do gráfico, para isso foi proposto: se o ônibus não viesse? Se chovesse? Todos os estudantes deram sugestões sensíveis de caminhos alternativos para a escola.
Explorar variação de dados no contexto da previsão do tempo.	A atividade baseou-se na história de alguns estudantes que gravaram a temperatura máxima diária em Hobart (Austrália) durante um ano e encontraram como média $17^{\circ}\text{C}$ . Os estudantes foram	Todos os estudantes demonstraram familiaridade com o contexto tempo e fizeram algum comentário de variação dentro disso. Ao sugerirem seis temperaturas máximas para o ano, apenas três respostas foram consideradas razoáveis (ex.: 29, 11, 20, 30, 24, 15). Para a pergunta: qual a temperatura mais alta do ano, dois desses três estudantes utilizaram

	perguntados o que isso dizia da temperatura em Hobart e se todos os dias tinham a temperatura máxima de 17°C.	temperaturas tão altas quanto às maiores sugeridas. Ao expressarem a temperatura máxima em um ano numa figura ou gráfico, quatro desenharam figuras, um mostrou a variação entre verão e inverno.
--	---	---

Para uma análise mais ampla, Watson e Kelly (2002a) utilizaram o *SOLO*<sup>1</sup> (Structure of Observed Learning Outcomes), modelo estrutural que será explicitado mais adiante. A intenção foi verificar níveis de aprendizagem nas respostas dos estudantes, as quais foram classificadas em: icônicas (I), baseadas principalmente em imaginação; uso de elementos únicos (S); uso de elementos múltiplos (M). As autoras observaram que houve predominância de respostas icônicas, apesar de constatar o reconhecimento intuitivo da variação, embora sem justificção, a partir do momento em que as crianças perceberam a variação em situações de incerteza. Os autores afirmam que isso reforça a ideia de que a discussão com crianças muito jovens é apropriada para construção da intuição destacando que os sujeitos da pesquisa não tiveram instruções específicas a respeito dos tópicos propostos.

Já Watson e Kelly (2005) objetivando identificar uma hierarquia de compreensões intuitivas de 73 estudantes sobre a variação em um contexto familiar (o tempo), propuseram aos participantes da pesquisa (sujeitos de escolas públicas da Austrália, entre cinco e quinze anos de idade) uma atividade a partir de entrevistas. A atividade exigia quatro habilidades:

- Explanação: explicar a variação existente por trás de uma temperatura média;
- Sugestão de dados: oferecer valores consistentes que expressassem a variação em torno de uma medida de posição (no caso da atividade a partir de uma média anual de temperatura 6 (seis) valores possíveis precisavam ser registrados);
- Gráficos: construir um gráfico (temperaturas de um ano inteiro) e selecionar o melhor gráfico (linha, colunas e histograma) para representar uma situação;
- Definição: descrever em palavras a variação de modo mais geral.

---

<sup>1</sup> Esse modelo está apresentado com mais detalhes no tópico seguinte.

Como resultados, Watson e Kelly (2005) constataram que habilidades relacionadas ao gráfico (construção e seleção), assim como a habilidade de definir tal conceito, apresentaram um desenvolvimento no decorrer da escolaridade, enquanto que a habilidade de explicar a variação foi a mais uniforme. De acordo com os autores, a falta de mudanças para a explanação da variação indica ausência de discussões a respeito dos conceitos de média e variação. Quanto à habilidade de sugerir dados houve melhora no desempenho dos estudantes no decorrer do ensino, entretanto, diferente das habilidades anteriores, nenhum sujeito atingiu ao nível 4 (nível máximo considerado). Associação entre habilidades também foi um aspecto observado na análise e percebeu-se que 85% dos estudantes apresentaram alta performance para sugerir valores que compõem uma média e baixa performance para explicar variação; associação moderada entre habilidades relacionadas a gráfico e a habilidade explanação; associação mais forte que as anteriores entre as habilidades voltadas para gráfico e relativas à sugestão de dados, essa associação foi similar ao par definição e explanação.

Watson (2009) retomando produções de estudantes de estudos anteriores (WATSON e KELLY, 2002a; WATSON e KELLY, 2005) e analisando-as a partir de outro enfoque, verificou a influência da variação e da “expectativa” no desenvolvimento de estudantes, quando solicitado que os mesmos representassem conjuntos de dados em gráficos. Watson (2009) concorda com a distinção sugerida por Reading e Shaughnessy (2004) para os termos variabilidade e variação (sendo a variabilidade entendida como a característica de uma entidade ou conjunto de dados e a variação como a exploração dessa variabilidade a partir de descrições ou medidas), desse modo esclarece que em tal estudo preocupou-se em investigar o ato de descrever ou representar a variabilidade nos gráficos criados. Quanto à “expectativa” aparece em contraste com o termo variação, significando o valor esperado (ex.: média) numa distribuição em que se é possível prever resultados. Contudo, a autora também ressalta que apesar do contraste entre variação e “expectativa”, elas estão relacionadas, pois a variação cria distribuições, uma vez que só faz sentido construir distribuições se houver variação, entretanto, a previsão de determinadas distribuições é possível. A autora argumenta ainda que o ideal é

que ambos os conceitos (variação e expectativa) sejam desenvolvidos de modo integrado no currículo escolar.

Partindo dos esclarecimentos conceituais acima, como método Watson (2009) entrevistou 109 estudantes, entre seis e quinze anos de idade, os quais responderam a três atividades apresentando situações diferenciadas: Livros (material manipulativo); tempo (conceito de média); doces (chances de resultados). Nesse estudo a autora analisa o conceito de variação e de expectativa. As duas primeiras atividades são as que interessam ao presente estudo, então, os resultados das mesmas terão uma maior atenção aqui. Ambas as atividades já foram referidas em estudos anteriores da autora, explicitadas acima: a dos livros, utilizada em Watson e Kelly (2002a), descrita na página 35; a do tempo, em Watson e Kelly (2002a; 2005), descrita na página 36.

As respostas dos estudantes foram analisadas dentro da estrutura cognitiva hierárquica *SOLO*, como em outros estudos. Respostas consideradas de nível mais elevado articulavam nas construções dos gráficos tanto o entendimento de variação, quanto de expectativa. A atividade dos “livros” foi mais fácil, pois os estudantes mais escolarizados conseguiram atingir o nível mais alto, o que não aconteceu com a que utilizou o “tempo” como contexto. Entretanto, na atividade do livro, houve uma maior variação de desempenho, pois alguns sujeitos mais escolarizados apresentaram o nível mais baixo (respostas idiossincráticas), enquanto na atividade do “tempo” nenhum desses sujeito apresentou esse nível de resposta.

Os diferentes níveis hierárquicos utilizados foram percebidos em todas as tarefas propostas aos estudantes, contudo, o reconhecimento da expectativa ocorreu apenas nos mais altos níveis. Conforme Watson (2009), o desempenho dos estudantes revelou que a tendência dos mesmos foi reconhecer primeiramente a variação e em seguida a expectativa. Isso porque crianças de 6 (seis) anos de idade já demonstraram um entendimento inicial desse conceito, enquanto que o reconhecimento da expectativa só foi evidenciado em crianças a partir de 11 (onze) anos. A autora alerta que isso ocorre contraditoriamente à ordem em que esses conceitos são iniciados no currículo escolar, pois formalmente o ensino de medidas de posição (ex.: média aritmética) e medidas de chance de eventos (ex.: contar casos favoráveis ou possíveis em resultados probabilísticos), que estão associados

ao conceito de expectativa, é realizado anteriormente ao ensino de medidas de variação (ex: desvio padrão), devido à complexidade de cálculo desse último. Entretanto, Watson (2009) afirma que crianças, desde cedo, são capazes de construir pictogramas para representar um determinado conjunto.

No estudo de Watson (2009) as compreensões dos estudantes, sujeitos da pesquisa foram analisadas a partir do desempenho dos mesmos nas distintas atividades propostas. Contudo, é importante salientar que apesar da habilidade exigida dos estudantes ser a mesma, construção de gráficos, as diferenças entre as atividades dos “livros” e do “tempo” vão além do contexto. Isso porque na dos “livros” um conjunto de dados é oferecido para a construção de um pictograma com cartões representativos, já na do “tempo”, uma média anual de temperatura máxima é dada, para a construção de um gráfico, nessa última os dados a serem representados teriam que ser criados. Percebe-se que outros fatores poderiam estar influenciando o desempenho dos alunos, que não apenas o contexto da atividade. Fatores que de fato influenciaram, uma vez que o desempenho dos estudantes em ambas as atividades não foi o mesmo. No entanto, a complexidade diferenciada entre tais atividades fica evidenciada nos resultados, porém, não é realizada uma discussão aprofundada a esse respeito no texto citado.

Além do contexto diferenciado, cada atividade continha questões envolvendo conceitos distintos, questões que não foram analisadas em separado no estudo de Watson (2009). A autora em tal estudo analisou cada atividade, no *SOLO*, como um todo, sem focar nas especificidades conceituais de suas questões. Entretanto, acredita-se que seria pertinente verificar e evidenciar que algumas questões exigem uma complexidade conceitual maior que outras, como na dos “livros”, em que responder adequadamente a questão “qual criança leu mais livros” pode ter sido mais fácil para os sujeitos da pesquisa do que responder a questão “qual seria a melhor estimativa de quantos livros um novo garoto poderia ter lido”.

Vale esclarecer ainda que Watson (2009) afirma ter analisado as representações de dados em gráficos pelos estudantes, contudo, seus resultados só foram possíveis porque a autora analisou não apenas os gráficos construídos pelos sujeitos da pesquisa como também a interpretação que os mesmos davam ao que construíram, conforme evidências mostradas no decorrer do texto da pesquisadora.

Assim, os distintos estudos aqui descritos evidenciaram que estudantes que nunca tiveram uma instrução sistemática sobre variabilidade/variação apresentaram noções intuitivas respondendo às situações propostas.

Dos estudos realizados no Brasil descritos anteriormente, a compreensão do conceito de variabilidade esteve mais relacionada à quantificação da variabilidade, focando então a variação de um conjunto de dados como medida, pois apesar de não exigir aquelas medidas de variabilidade já reconhecidas (por exemplo, o desvio padrão), no momento em que se pede a quantificação de uma comparação se está medindo. Tal habilidade de quantificar variações em um dado conjunto desenvolveu-se nos estudos realizados no Brasil a partir da interpretação de gráficos por estudantes, aos quais eram propostas questões focando análises pontuais (locais) ou globais. Desse modo, percebe-se que as habilidades requeridas dos estudantes ao lidarem com variabilidade nos estudos desenvolvidos no país (SANTOS e GITIRANA, 1999; GUIMARÃES, 2002; ARCOVERDE, BARBOSA e GUIMARÃES, 2004) eram diferentes daquelas habilidades exploradas pelas pesquisas realizadas na Austrália (WATSON e KELLY, 2002a; 2002b; 2005; WATSON, 2009), embora esses últimos também considerassem esse tipo de medição, como constatado na atividade 2 (dois) de Watson e Kelly (2002a), na qual os sujeitos eram perguntados quantos a mais chegaram de ônibus que de bicicleta, ao analisarem um gráfico.

Os autores aqui no Brasil concordam que quantificar a variabilidade, a partir de acréscimos e decréscimos, não é um conceito facilmente compreendido, como localização de pontos extremos, uma vez que tanto estudantes de 4º ano (entre 9-10 anos de idade) quanto estudantes de 7º ano (entre 12-13 anos de idade) apresentaram dificuldades. Contudo, percebe-se também que intervenções a fim de tornar o contexto mais familiar possibilitam avanços na construção do conceito de variabilidade.

O entendimento de estudantes a respeito de tal conceito é o foco nesse estudo, no entanto, é importante salientar que compreensões de professores atuantes na Educação Básica também vêm sendo investigadas e resultados, como o de Silva (2007), demonstram que docentes formados em Matemática, os quais tiveram uma instrução em Estatística ligada a conceitos como desvio padrão, apresentaram, inicialmente, ausência de significação dessa medida variacional ou

raciocínio idiossincrático. A autora afirma, ainda, que os poucos professores que já utilizavam a pesquisa como metodologia de ensino de Estatística o faziam de maneira restrita, empregando apenas a distribuição de frequência e sua representação gráfica. A falta de discussão nas aulas desenvolvidas por esses docentes, quanto a medidas de tendência central (ex.: a média) e medidas de dispersão (variação) pode ter decorrido do tipo de variável escolhida na pesquisa, a qualitativa.

Ao se explorar o tipo de variável qualitativa em pesquisas envolvendo estudantes na Educação Básica, são comuns temas relacionados à preferência alimentar ou musical de um grupo de sujeitos. A exploração de medidas de tendência central e medidas de variação, tais como a média e o desvio padrão, respectivamente, só faz sentido quando se trabalha com variáveis quantitativas, que Kader e Perry (2007) vão denominar dados quantitativos.

A partir do que Silva (2007) constatou quanto à metodologia adotada por professores ao ensinarem Estatística e refletindo a discussão proposta por Kader e Perry (2007), no tópico anterior (página 19), pode-se pensar na importância de se ter um desenvolvimento conceitual amplo da variabilidade na formação de estudantes da Educação Básica em Estatística. Já é possível perceber que a exploração de distintas habilidades ao lidar com conjuntos de dados não é o suficiente. É necessário reconhecer que os tipos de dados analisados também vão interferir na compreensão dos estudantes do conceito de variabilidade e que para a construção de conhecimentos mais elaborados ou profundos o que influencia essa construção precisa ser investigado.

Vê-se nos estudos apresentados que o interesse pela melhor compreensão e desenvolvimento do conceito de variabilidade é crescente, principalmente pelo entendimento que se tem atualmente de que variabilidade é um conceito central da estatística. Entretanto, apesar dessa centralidade, poucos são os grupos de pesquisadores preocupados com a construção do conceito por estudantes dos anos iniciais do Ensino Fundamental e muitos são os aspectos ou habilidades relacionados à variabilidade / variação que podem influenciar o entendimento e evolução da mesma, como constatado na revisão literária realizada. Ressalta-se, desse modo, a necessidade de mais estudos voltados para a compreensão do

desenvolvimento do conceito de variabilidade, quando se inicia o ensino da Estatística.

### 1.6 Níveis de aprendizagem: O SOLO (Structure of Observed Learning Outcomes) como instrumento de análise

Estudos como o de Watson e Kelly (2002a; 2002b; 2005) e Watson (2009), que desenvolveram pesquisas interessadas nas compreensões de estudantes da Educação Básica, quanto ao conceito de variabilidade, utilizaram um modelo denominado de SOLO (Structure of Observed Learning Outcomes), para níveis de aprendizagem nas tarefas explorando o conceito de variabilidade estatística. O modelo foi empregado para análise das respostas dos estudantes às atividades propostas e baseia-se em teorias neo-piagetianas para o pensamento concreto-simbólico. Tal modelo foi elaborado por Biggs e Collis (1982). Abaixo, segue o Quadro 4 que sumariza o SOLO, conforme Watson (2009).

**Quadro 4: Síntese do SOLO para níveis de aprendizagem**

Nome	Elementos nas respostas para a tarefa	Conflitos
<b>Nível pré-estrutural ou icônico (IK)</b>	Respostas não comprometidas com nenhum elemento da tarefa esperado para uma solução concreto-simbólica. São dadas respostas imaginativas ou baseadas em experiências pessoais.	Não reconhecimento de conflito /contradição.
<b>Nível 1 (Uniestrutural)</b>	Respostas empregam elementos relevantes únicos da tarefa.	Não reconhecimento de conflito / contradição.
<b>Nível 2 (Multiestrutural)</b>	Múltiplos elementos empregados nas respostas, geralmente de forma sequencial.	Reconhecimento de conflitos / contradição, mas inabilidade para resolvê-los.
<b>Nível 3 (Relacional)</b>	Respostas integram os múltiplos elementos da tarefa para a solução do problema.	Resolução de conflitos que surgem.
<b>Nível 4 (Abstração estendida)</b>	Resposta vai além do nível relacional ao expor outros elementos relevantes para extensão da tarefa.	Pode sugerir a possibilidade de novos conflitos e resolver ou dar alternativas a isso.

Sumário adaptado de Biggs & Collis (1982), Pegg (2002a) e Watson & Moritz (2000), conforme Watson (2009, p. 37).

Percebe-se que no modelo estrutural do SOLO os níveis evoluem da não consideração das respostas às atividades elaboradas sobre o conceito que se quer abordar até a capacidade de se fazer relações entre conceitos, para se chegar a uma solução adequada de um problema proposto.

Na tentativa de elucidar cada um dos níveis apresentados no quadro acima, cabe exemplificá-los a partir da classificação realizada por Watson (2009) de respostas de estudantes em seu estudo.

Na atividade dos “livros”, em que ofereceu-se fichas representando livros lidos e leitores para a construção de um pictográfico, respostas em um *nível pré-estrutural* ou *idiossincrático* foram aquelas nas quais os estudantes contaram quantos livros cada leitor leu, contudo não expuseram a variação na representação gráfica. No nível *uniestrutural* os estudantes representaram um único aspecto da variação, a quantidade de livros lidos por crianças. No nível *multiestrutural* a variação foi mostrada em formato retangular, porém, predizer resultados a partir dos dados representados foi mais difícil. No *nível relacional* as respostas indicavam uma intuição de “expectativa” na exposição da variação, como por exemplo, na organização do pictograma (valores ordenados de mínimo para máximo). Entretanto, esse entendimento de média não se estendeu na consideração de que em média as meninas leram mais que os meninos. Finalmente no *nível de abstração estendida* as respostas precisaram empregar variação e expectativa na questão de predição. Como respostas típicas, houve a representação da variação entre garotos e garotas e a separação da estimativa para ambos.

Já na atividade do “tempo” foi proposta uma média da temperatura máxima anual de uma cidade e os estudantes precisaram construir um gráfico que representasse a temperatura durante um ano dessa mesma cidade. No nível pré-estrutural houve representações estáticas na natureza, embora se expusesse algo sobre temperatura máxima. O nível *uniestrutural* caracterizou-se por representações indicando variação de estação, porém, não se apresentou dados ou frequência. No nível *multiestrutural* os gráficos construídos sugeriam frequências de temperatura aleatoriamente ou considerando o tempo durante o ano. No nível *relacional* houve respostas demonstrando “expectação” intuitiva no contexto do tempo, ao expor no gráfico estações do ano (mais quente, mais frio). No último nível, o de abstração

estendida, as respostas focaram variação e “expectativa” com discussão da variação diária.

É importante salientar que a classificação realizada por Watson (2009) das respostas de estudantes ao analisar suas produções buscou atender seu objetivo em tal estudo, que foi investigar a influência da variação e da expectativa no desenvolvimento das compreensões de estudantes. O que sugere que para a utilização do modelo estrutural *SOLO* torna-se necessário pensar o conceito que se quer explorar e a articulação do mesmo com outros conceitos relacionados.

Em resumo, a discussão desenvolvida durante o capítulo, permite a exposição dos seguintes pontos:

- A estatística desempenha um papel cada vez mais importante na vida cotidiana, pelas afirmações e decisões tomadas em função do tratamento estatístico dado as informações comunicadas;
- No âmbito educacional já se reconhece a necessidade do ensino de Estatística desde os anos iniciais do Ensino Fundamental;
- Dentro da Estatística o conceito de variabilidade é essencial e poucas pesquisas vêm abordando tal conceito nos anos iniciais de escolarização, de modo a investigar habilidades distintas, a partir de situações diferenciadas;
- Diversos são os aspectos importantes de serem investigados e relacionados quando se explora o conceito de variabilidade como: representação da variabilidade; descrição da mesma; predição de resultados a partir da análise de um dado conjunto; quantificação da variação ao se comparar pontos num gráfico; sugestão de valores para composição de um conjunto de dados e muitos outros aspectos;
- Estudos realizados revelam que crianças muito jovens já são capazes de perceber a variabilidade de dados representados, assim como também podem representar a variabilidade de um conjunto.

A partir do que foi exposto na revisão literária, pode-se pensar que apesar de já ser possível encontrar algumas respostas ao problema das compreensões de estudantes dos anos iniciais, quanto ao conceito de variabilidade, muitos são os aspectos ou habilidades importantes ao desenvolvimento do conceito

(representação da variabilidade em gráficos, descrição da variabilidade, a quantificação da mesma, a predição de resultados pela análise de um conjunto de dados, comparação entre conjuntos e tantas outras) e é necessário que se saiba das relações entre eles.

Desse modo, interroga-se: Existe uma gradação de compreensão por parte dos estudantes entre os aspectos? Estudantes de diferentes anos de escolaridade podem apresentar compreensões distintas do conceito de variabilidade? Essas questões mobilizam a presente pesquisa, motivadas por uma questão maior: Que compreensões do conceito de variabilidade estatística apresentam estudantes de 2º e 5º anos do Ensino Fundamental?

No capítulo seguinte, explicita-se os objetivos principal e específicos que surgem das questões acima descritas, além da metodologia adotada para contemplar os objetivos elencados.

## **CAPÍTULO 2 - PERCURSO METODOLÓGICO**

O objetivo principal do presente estudo foi investigar as compreensões apresentadas por estudantes de 2º e 5º anos do Ensino Fundamental a respeito do conceito de variabilidade. Mais especificamente pretendeu-se averiguar:

- ❖ A compreensão de estudantes de 2º e 5º anos de escolarização referente aos diferentes aspectos do conceito de variabilidade, tais como representação da variabilidade, descrição da mesma, comparação entre conjuntos de dados, predição de resultados, quantificação da variação;
- ❖ A existência de uma gradação na compreensão desses estudantes quanto aos diferentes aspectos do conceito de variabilidade.

Para o alcance dos objetivos acima elencados, realizou-se uma aproximação com o método clínico-piagetiano, o qual tem como uma de suas características principais a obtenção de justificativas para respostas oferecidas. Desse modo, o processo pelo qual o sujeito chega a sua resposta é o interesse maior numa pesquisa baseada na teoria de Piaget, conforme afirmação de Carraher (1998). Os participantes da pesquisa precisaram ser entrevistados individualmente e serem instigados a exporem suas compreensões a respeito do conceito explorado junto a eles. Assim, as ações adotadas na coleta de dados justificam a pequena quantidade de sujeitos e escolas investigadas, como explicitação a seguir.

### **2.1 Participantes**

A pesquisa foi composta por 48 (quarenta e oito) estudantes de 2 (duas) escolas públicas municipais do Recife, sendo 24 (vinte e quatro) estudantes do 2º ano (entre 7 e 8 anos de idade) do Ensino Fundamental e 24 (vinte e quatro) estudantes do 5º ano (entre 9 e 10 anos de idade). Na coleta de dados, procurou-se ainda variar as turmas de um mesmo ano de escolaridade numa mesma escola. Assim, para o 2º ano de uma das escolas, os dados apresentados no capítulo seguinte foram obtidos em ao menos duas turmas, o mesmo aconteceu com o 5º ano.

## 2.2 Procedimento

Os sujeitos, individualmente, responderam a cinco tarefas envolvendo aspectos diferenciados do conceito de variabilidade. Os estudantes receberam as atividades de forma impressa. Durante a coleta de dados a interação entre pesquisadora e sujeito aconteceu com a intencionalidade de garantir que as atividades planejadas fossem compreendidas pelos estudantes, conforme elaboração. Desse modo, a busca de justificativas para as respostas dos estudantes foi constante e a partir dessas foi possível analisar de modo mais profundo como os mesmos compreenderam as atividades e procederam em suas resoluções. Para isso, as falas dos sujeitos no decorrer das tarefas foram audio-gravadas.

As cinco atividades a serem descritas foram apresentadas, aos estudantes, uma de cada vez. Houve uma randomização em pelo menos três das cinco atividades, o que será melhor explicitado posteriormente.

A seguir, apresentam-se as cinco tarefas tal como foram propostas aos estudantes. Na apresentação das atividades procurou-se fazer uma descrição das mesmas, destacando que aspectos do conceito de variabilidade cada uma delas explora. Cabe salientar, nesse momento, que as atividades foram selecionadas baseando-se principalmente no modelo epistemológico, proposto por Garfield e Ben-zvi (2005), para o desenvolvimento de um profundo conhecimento do raciocínio de variabilidade.

Conforme descritos no capítulo anterior, os componentes do raciocínio do conceito de variabilidade (página 16) são bem abrangentes e dos sete componentes recomendados, elegeu-se os aspectos julgados importantes e possíveis de serem explorados com os estudantes a serem investigados na pesquisa. Desse modo, os aspectos explorados nas atividades descritas a seguir variam de uma atividade para outra e são os seguintes:

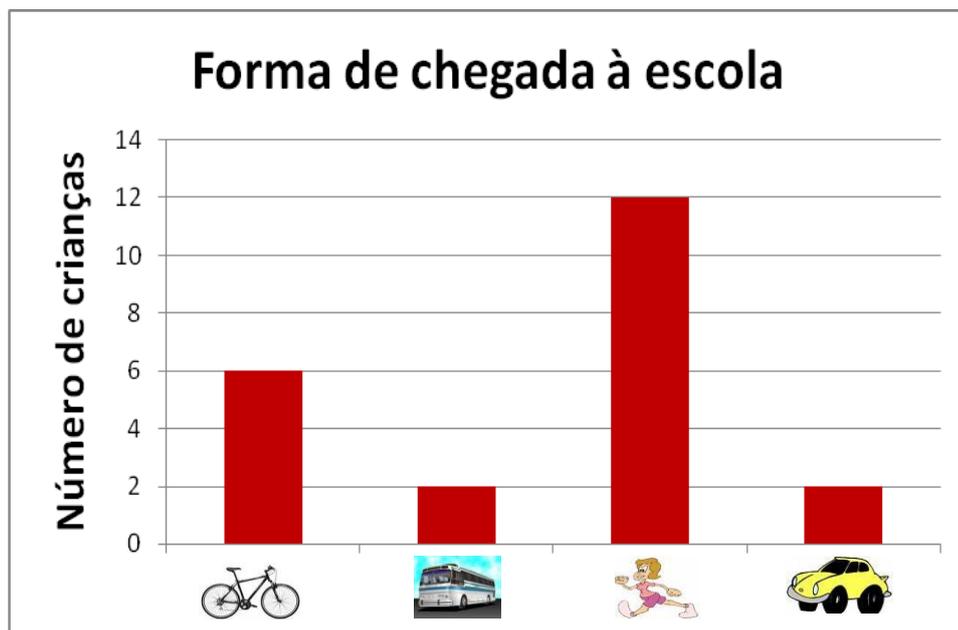
- Descrição ou explanação da variabilidade;
- Predição de resultados a partir da variabilidade dos dados;
- Comparação entre conjuntos de dados;
- Representação da variabilidade;

- Redução da variabilidade (nesse caso decidiu-se verificar o aspecto da redução de variabilidade também em situações em que se propõe a ausência da mesma).

### 2.2.1 Atividade 1

A atividade 1 foi adaptada da pesquisa realizada por Watson e Kelly (2002a). A mesma tem como objetivo explorar a interpretação de gráficos e a predição de resultados a partir do que o gráfico expõe. As informações comunicadas no gráfico são próximas ao que frequentemente se verifica nos gráficos apresentados pelos livros didáticos, aos quais os estudantes normalmente têm acesso.

O gráfico abaixo mostra os resultados da pesquisa que a professora Patrícia realizou quanto à forma que as crianças de sua turma chegou a escola em um dia.



- 1) Você pode dizer como essas crianças chegaram à escola nesse dia? A forma de chegada variou de uma criança para outra? (*explicação da variabilidade de dados*)
- 2) Qual o meio de transporte mais utilizado por essas crianças nesse dia? (*ponto máximo e moda*)
- 3) Considerando o gráfico, como você pensa que uma nova criança chegaria à escola? (*predição a partir da moda*)
- 4) Quantos a mais chegaram caminhando que de ônibus? (*comparação entre dois pontos*)

- 5) Se tivesse ocorrido, nesse dia, uma greve de ônibus, mas essas crianças tinham que ir à escola, não podiam faltar. O que precisaríamos mudar no gráfico (como ele ficaria)? (*manipulação das barras*)
- 

Ao adaptar-se a atividade utilizada pelos autores acima citados, precisou-se alterar informações do gráfico, para aproximarmos os dados à realidade foco de nosso estudo. Assim, a quantidade de crianças e o meio de transporte utilizado pelas mesmas foram modificados. Foi acrescentada a questão de número 1, que também se julga pertinente quando se interpreta a variabilidade de dados.

As questões 1 e 3, elaboradas a partir dos dados representados no gráfico, objetivam a exploração de alguns dos componentes do modelo epistemológico, proposto por Garfield e Ben-Zvi (2005), tais como descrição da variabilidade (questão 1) e predição de resultados (questão 3).

As questões 2 e 4 têm um caráter mais pontual que global (a 2 referente ao ponto máximo e a 4 referindo-se à comparação entre dois pontos). Porém, acredita-se ser importante, aqui, a apresentação dessas questões aos estudantes, uma vez que são exploradas pelo livro didático. Além do mais, compreende-se que ambas as questões, apesar de pontuais, para serem respondidas satisfatoriamente precisam considerar que os dados variaram, ou seja, a variabilidade precisou ser verificada.

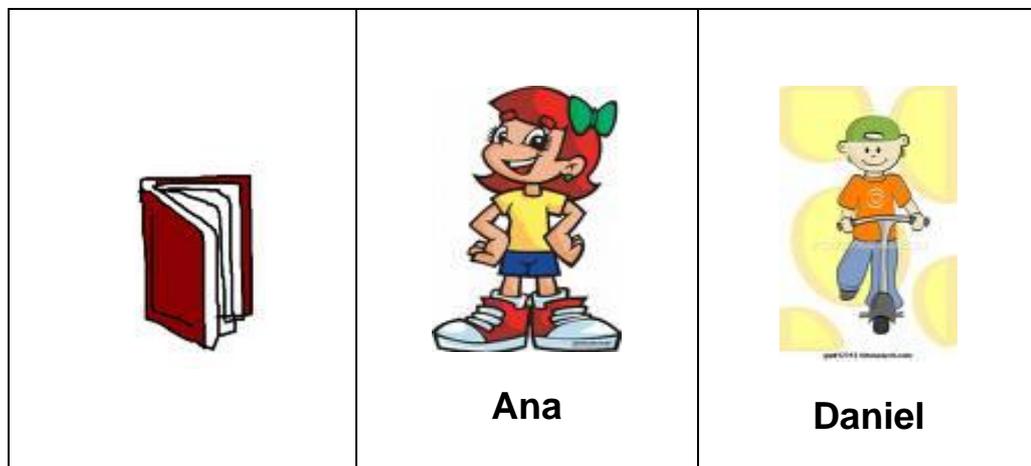
A questão 5, proposta por Watson e Kelly (2002a) como envolvendo a conservação da quantidade total de crianças após uma manipulação, pode ser considerada no modelo epistemológico sugerido por Garfield e Ben-Zvi (2005) como uma atividade em que os estudantes precisam pensar em formas de redução da variabilidade. Entende-se, nesse caso, que a redução da variabilidade acontece a partir do momento em que se diminui as opções de chegada das crianças à escola.

### **2.2.2 Atividade 2**

A atividade 2, adaptada de Watson (2009), ressalta a construção de gráfico como forma de representar a variabilidade de dados. Como na tarefa anterior, questões de interpretação também são exploradas, entretanto, na segunda atividade isso ocorre a partir das construções dos próprios alunos. Predições, à medida que se incorpora situações novas ao gráfico, são exigidas mais uma vez. Situando a

presente atividade na discussão proposta por Garfield e Ben-zvi (2005), quanto aos componentes do raciocínio variacional, a representação da variabilidade também é um componente importante a ser desenvolvido para compreensão desse conceito. Assim, entende-se que ao construirmos um gráfico estamos representando a variabilidade de um conjunto de dados.

Temos aqui vários cartões. Uns representam crianças e outros representam livros lidos por elas (comunicar as seguintes informações: Ana leu 4 livros; Daniel leu 1; Luíza leu 6; Bruno leu 3).



Três cartões representativos de crianças e livro

#### Representação (parte 1)

(Serão disponibilizados para os estudantes 18 cartões, 14 representando livros lidos e 4 representando as crianças leitoras)

- 1) Você pode usar esses cartões para mostrar essas informações? (*Representação da variabilidade*)
- 2) Se alguém entrar na sala e observar a figura que você fez o que (ela ou a criança) poderia dizer da leitura de livros por essas crianças? Como a pessoa poderia descrever a figura? (*Explicação da variabilidade*)
- 3) Você sabe o que é um gráfico? Você pode organizar um com esses cartões? (para as crianças que não fizeram um gráfico)

#### Representação (parte 2)

(Apresentar mais um cartão representando uma nova criança)

- 4) Suponha que Henrique não leu nenhum livro. Você pode mostrar na sua figura que ele não leu livros? (*Representação de ausência de variabilidade*)

(Apresentar mais cinco cartões representando livros lidos)

- 5) Agora, suponha que todos foram à biblioteca e leram mais um livro cada um. Você pode mudar sua figura para mostrar que todos eles leram mais um livro? (*Representação de acréscimos*)
  - 6) Quem você acha que provavelmente gostaria de receber um livro no Natal? Por que você pensa isso? (*Predição a partir do ponto máximo*)
- 

- **Construção:** Na parte 1 da representação, a forma de comunicar os dados terá que ser definida e explicitada pelo estudante. Já na parte 2 de representação, modificações na construção anteriormente produzida serão necessárias e também será interessante verificar como o estudante se depara com novas situações interferindo na produção que se pensou acabada (questões 4 e 5). A questão 3 foi especialmente construída para aqueles estudantes que não construíram o pictograma com os cartões, será uma forma de refletir as razões.
- **Interpretação dos dados:** Tanto no primeiro momento de representação (construção) como na segunda parte, pede-se uma exposição das características dos dados representados. Na questão 2 abordamos a explicitação pelos estudantes das características dos dados representados e a descrição da variabilidade do conjunto representado.
- **Predição:** A questão 6 requer uma leitura além dos dados, ou seja, as questões abrem espaço para extrapolações que podem ou não considerar os dados representados pelo estudante.

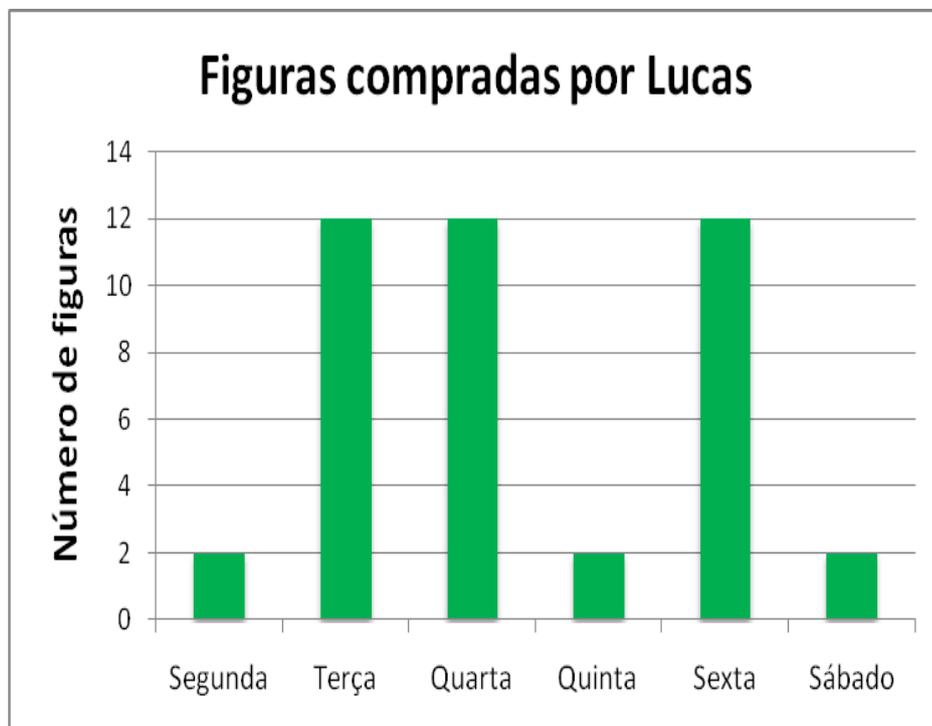
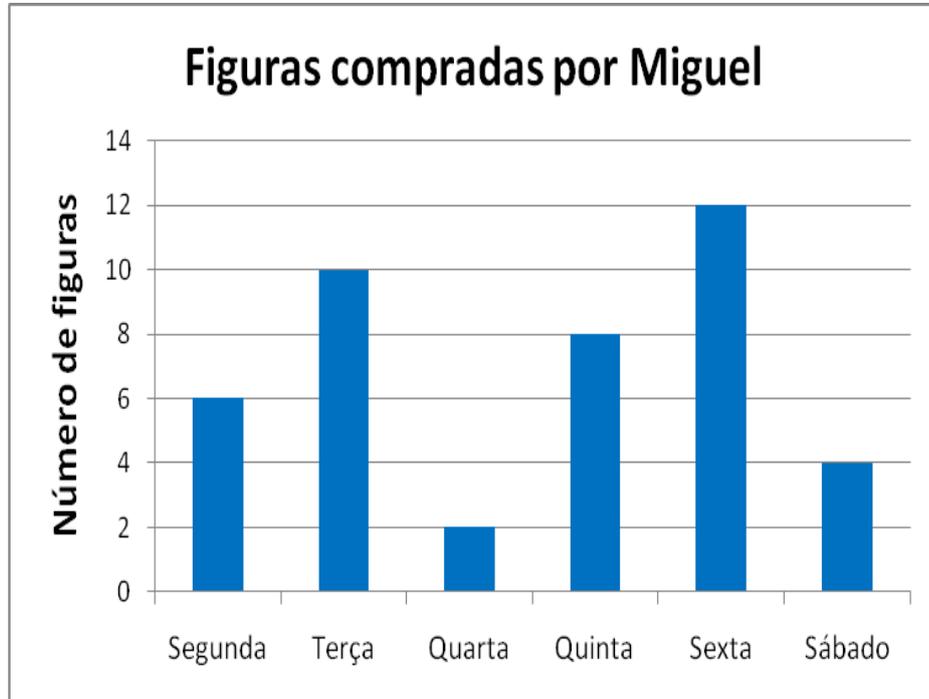
### 2.2.3 Atividade 3

A atividade 3 foi adaptada do estudo de Loosen, Lioen e Lacante (1985), a qual exige a comparação entre conjuntos de dados diferenciados. Tratam-se de dados quantitativos que vão expressar a variabilidade de modo diferente, de quando a variabilidade corresponde a dados categóricos, conforme Kader e Perry (2007).

Ao pensarmos no modelo epistemológico proposto por Garfield e Ben-Zvi (2005), a atividade corresponde a um de seus componentes, pois explora a comparação da variabilidade de conjuntos de dados representados em gráficos, sob uma mesma escala.

---

Miguel e Lucas compraram figurinhas para suas coleções. Observe esses 2 gráficos que mostram a quantidade comprada por cada um durante uma semana. Quem variou mais na quantidade de figurinhas compradas durante essa semana? Por quê?



Os estudantes poderiam pensar que o conjunto 1 variou mais porque não tem blocos de mesma altura, enquanto que o conjunto 2 tem blocos de apenas duas alturas diferentes, logo apresentaria menor variabilidade. Contudo, de acordo com o conceito de desvio padrão, medida de variabilidade que considera a distância dos dados em relação à média, o conjunto 2 variou mais. Kader e Perry (2007) ainda alertam que apesar desses conjuntos se referirem a dados quantitativos, intuitivamente, pode-se pensar formando categorias, que, em se tratando de variabilidade, enfatiza o quão frequente as observações mudam.

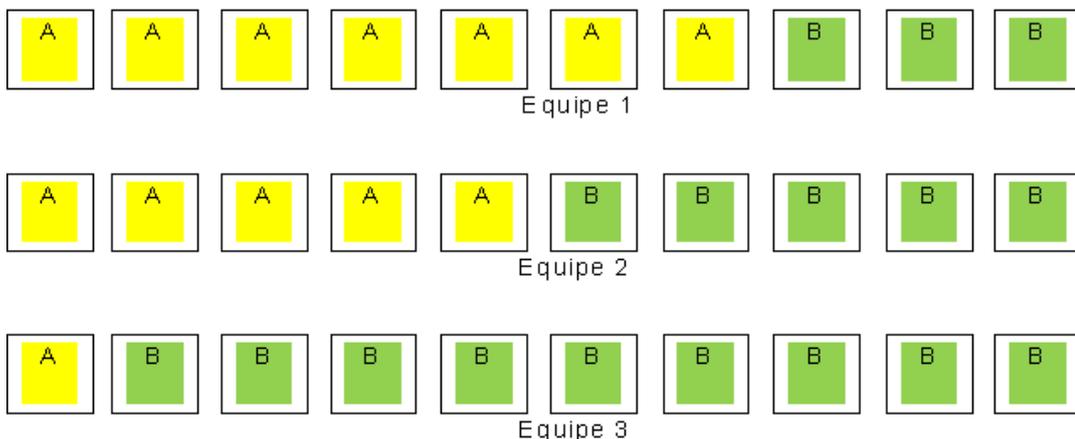
#### 2.2.4 Atividade 4

A atividade 4, baseada em Kader e Perry (2007), também aborda a comparação entre conjunto de dados, como na atividade anterior. Porém, a presente atividade refere-se a dados categóricos e o conceito intuitivo de variabilidade, nesse caso, torna-se mais evidente que quando analisamos a variação a partir de dados quantitativos, conforme os autores. Diferente das atividades anteriores, essa não explora a variabilidade em gráfico de coluna. Outro diferencial da atividade é que os conjuntos de dados foram apresentados de dois modos distintos, metade dos sujeitos recebeu a *representação 1* e a outra metade a *representação 2*.

---

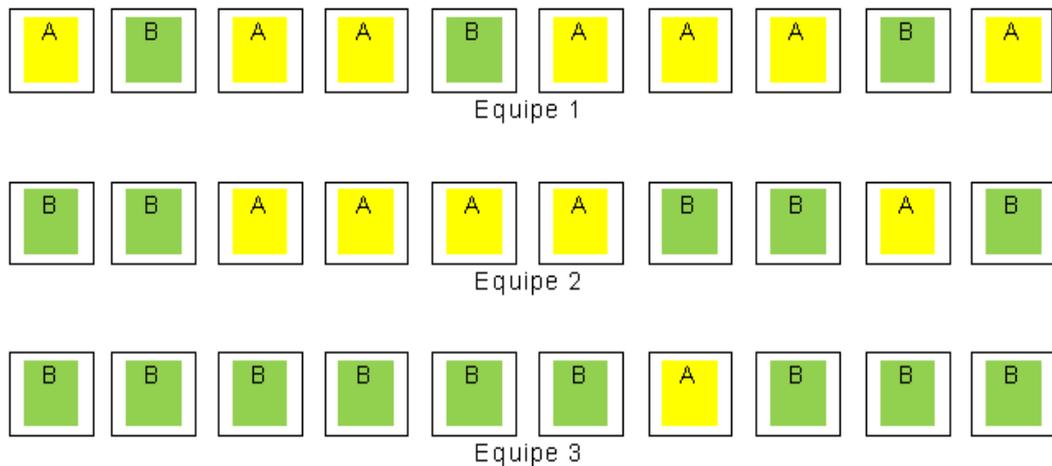
Na escola em que Pedro estuda foram formadas três equipes, com dez pessoas em cada grupo. Cada estudante de cada grupo para responder a uma questão escolhia a resposta A ou B. Assim ficaram as respostas das equipes:

##### Representação 1



- 1) Qual equipe teve uma maior variação das respostas?
- 2) Qual equipe variou menos?

Representação 2

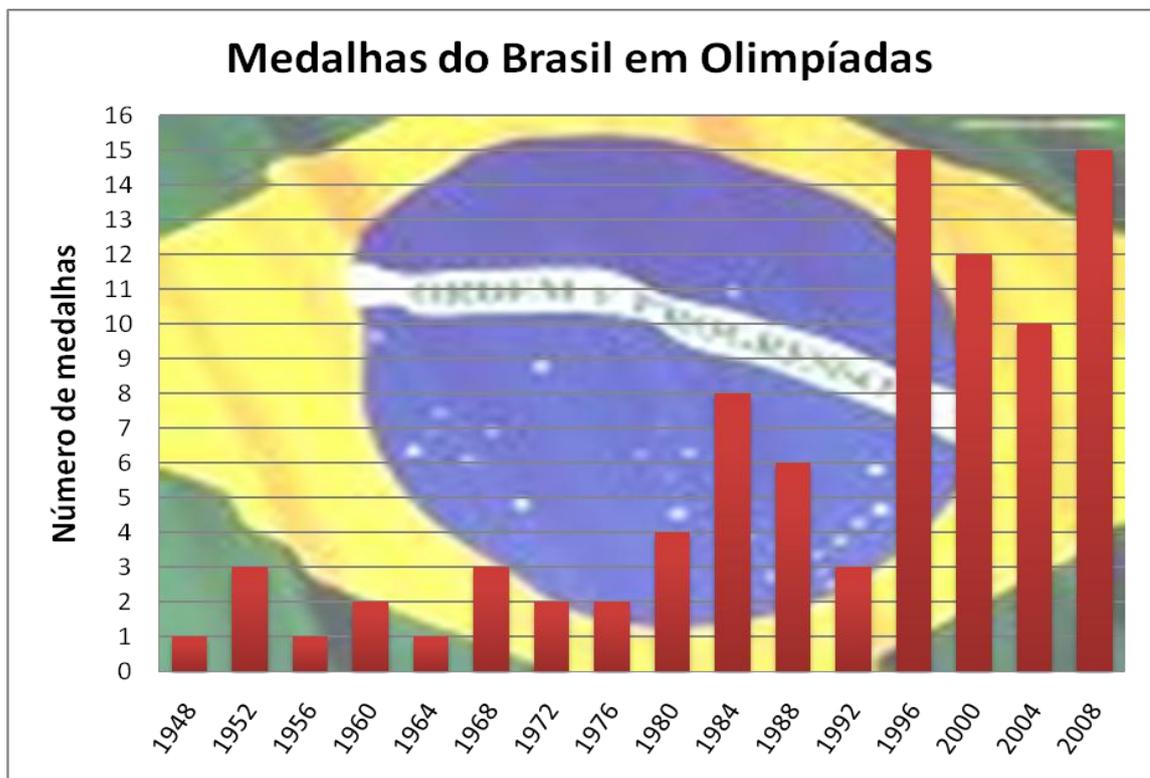


A mudança no formato da representação se justifica pela necessidade de investigar se a organização dos dados poderia estar interferindo nas compreensões dos estudantes. Assim, optou-se por duas formas de representação. Na primeira a frequência das categorias (resposta A ou B) está agrupada, de modo que o quantitativo das mesmas está bem evidenciado. Na segunda, a frequência de ambas as categorias está misturada, nesse caso os tipos de respostas foram representados de forma aleatória.

Na atividade 4, diferente da atividade 3, considerar a maior ou menor variabilidade requer pensarmos em quão diferente as respostas numa categoria (opção de resposta “A”) diferem das respostas na outra categoria (opção de resposta “B”). Em conjuntos de dados categóricos não faz sentido pensar no conceito de desvio padrão, medida de variabilidade comum quando se trata de dados quantitativos (KADER e PERRY, 2007).

### 2.2.5 Atividade 5

A atividade 5 explora a interpretação da variabilidade dos dados representados em um gráfico de colunas, assim como a atividade 1. Entretanto, o gráfico da presente atividade foi veiculado pela mídia, numa revista semanal de grande divulgação no Brasil. Além disso, aqui tem-se uma série temporal (como na atividade 3) e a partir do que o gráfico comunica podemos propor aos estudantes questões que exploram a tendência do gráfico ao longo de um período de tempo determinado.



- 1) Como o Brasil vem se saindo na conquista de medalhas? (*Explicação da variabilidade*)
- 2) Na próxima olimpíada, em 2012, como você imagina que o Brasil vai se sair? Quantas medalhas você acha que o Brasil vai conquistar? (*predição a partir da tendência*)
- 3) a) O que aconteceu com a quantidade de medalhas em 1972 e 1976? (*Identificação de ausência de variabilidade*)  
 b) O que aconteceu em 2000 e 2004? Mudou? Como teria que ser se quiséssemos que não mudasse? (*Proposição de ausência de variabilidade*)

Nas três primeiras questões enfatizou-se apenas o caráter global dos dados. Objetivou-se com as mesmas investigar que visões do gráfico se sobressaem nas leituras realizadas pelos estudantes. Ben-zvi (2002) salienta que visões locais e globais dos dados estão integradas, sendo possíveis vários níveis de entendimento dentro de ambas as visões. A terceira questão requer que se considere a possibilidade de ausência de variabilidade no gráfico. Essa questão se subdivide em duas. Na questão 3a buscou-se a identificação de ausência de variabilidade, uma vez que o gráfico já apresentou categorias (anos) com mesma frequência. Já a questão 3b requeria que o estudante propusesse tal ausência, para isso o sujeito precisaria perceber que os dados variaram.

Com as questões explorando a manipulação da variabilidade, tanto na atividade 1 (questão 5) quanto na atividade 5 (questão 3), relaciona-se tais atividades aos estudos de Kader e Perry (2007), que afirmam ser a variabilidade de dados categóricos diferente da variabilidade de dados quantitativos. Podemos perceber que existe uma relação direta entre as atividades 1 e 4, assim como entre as atividades 3 e 5, quando tentamos responder a pergunta referente à ausência de variabilidade. Ausência de variabilidade nas atividades 1 e 4 (dados categóricos) seria uma forma apenas de chegar à escola (atividade 1) e respostas de dez alunos de uma equipe numa única opção, “A” ou “B” (atividade 4). Já a ausência de variabilidade nas atividades 3 e 5 (dados quantitativos) seria mesma quantidade de figurinhas compradas por dia (atividade 3) e quantidade constante de medalhas conquistadas pelo Brasil em Olimpíadas (atividade 5).

Como já afirmou-se, as tarefas selecionadas aqui abordam diversos aspectos da variabilidade estatística, que vão desde a sua representação até a comparação entre diferentes conjuntos de dados. Os enfoques diferenciados do conceito de variabilidade geraram uma diversidade também de atividades e questões, como explicitado acima, cujas sequências foram alteradas, o que resultou em quatro sequências, conforme Tabela 1, a seguir:

**Tabela 1: Frequência de estudantes do 2º e 5º ano nas sequências de atividades**

Sequência	2º ano	5º ano
1; 2; 3; 4.a; 5	6	6
1; 2; 3; 4.b; 5	6	6
2; 1; 3; 4.a; 5	6	6
2; 1; 3; 4.b; 5	6	6

A Tabela 1 mostra a mudança na apresentação de pelo menos três das cinco atividades, resultando em quatro sequências, em que a ordem das atividades 1 e 2 foi alterada, assim como modificou-se a organização dos dados na atividade 4. Dos 24 (vinte e quatro) estudantes de um mesmo ano de escolarização, 12 (doze) responderam numa sequência diferente dos outros doze estudantes, isso tanto para o 2º quanto para o 5º ano. A mudança na sequência das duas atividades (a primeira de interpretação de gráfico e a segunda de construção de pictograma, a partir da representação de dados em cartões) justifica-se pela necessidade de verificar se a resolução de atividades de construção de gráfico interfere na atividade de interpretação.

Quanto à atividade 4, como já exposto na descrição da mesma, contou com organizações (representações) diferentes para os mesmos dados, pois pretendia-se verificar a influência da representação dos dados nas compreensões dos estudantes, se a organização (agrupamento) ou não da frequência das categorias interferiria nas respostas dos mesmos. Desse modo, metade dos estudantes tanto do 2º quanto do 5º ano recebeu a *representação 1*, (frequência da categoria agrupada) e a outra metade recebeu a *representação 2* (frequência da categoria aleatória). Tais variações geraram as quatro sequências explicitadas na Tabela 1, cada sequência contou com doze (12) sujeitos, que distinguindo-os por ano de escolarização, contou cada ano com seis sujeitos.

### **2.3 Procedimentos de análise dos dados**

Os dados coletados a partir das justificativas dos estudantes às atividades propostas aos mesmos foram categorizados e analisados com o auxílio do software SPSS (Statistical Package for the Social Sciences).

Testes de significância foram realizados, no qual comparou-se adequação ou não das respostas aos questionamentos direcionados aos sujeitos do 2º e 5º ano. O índice de significância considerado no presente estudo foi aquele em que o nível-p foi de 0,05.

## CAPÍTULO 3 – ANÁLISE DAS ATIVIDADES

No presente capítulo realiza-se descrição dos resultados das cinco atividades utilizadas como instrumento de pesquisa, assim como discussão das mesmas. As atividades são analisadas partindo das questões elaboradas, as quais exploravam aspectos distintos da variabilidade, ao término de cada atividade os resultados encontrados são confrontados com o que já argumentaram estudos anteriores a respeito de tais aspectos.

Por último, procura-se fazer um resgate do desempenho dos estudantes de ambos os grupos (2º e 5º ano) nas questões norteadoras de tal estudo de modo mais geral. Pretende-se elucidar, numa análise preliminar, os questionamentos delineados anteriormente a respeito da influência da escolaridade na compreensão do conceito de variabilidade, a partir dos aspectos investigados, destacando aqueles mais complexos.

### 3.1 Atividade 1

A Atividade 1 explora distintos aspectos da variabilidade de dados. Foi apresentado um gráfico com dados categóricos, a partir do qual realizou-se 5 (cinco) questões.

A primeira questão (Q1.1) requeria dos estudantes a *explicação da variabilidade* do conjunto de dados. Esperava-se que os estudantes explicitassem que havia uma variabilidade de meios de locomoção apresentados no gráfico, ou seja, a *variabilidade de categorias*. A Tabela 2 apresenta os tipos de respostas dadas pelos estudantes para a pergunta: “*Você pode dizer como essas crianças chegaram à escola nesse dia? A forma de chegada variou de uma criança para outra?*” (*Explicação da variabilidade de categorias*).

**Tabela 2: Percentual de respostas adequadas na Q1.1 por ano de escolaridade**

Tipos de respostas	2º	5º
1) Não percebe os dados, cria histórias	4,2	4,2
2) Refere-se apenas a um elemento	33,3	4,2
3) Expõe variação sem esgotar categorias	33,3	12,5
4) Expõe variação dos transportes	29,2	79,1

Respostas como “*uma de bicicleta, outro de ônibus, outra a pé e outra de carro*” (S521) demonstram que o estudante percebeu a variabilidade de categorias e por isso justificativas do tipo foram consideradas como adequadas. Além dessas, considerou-se como apropriadas respostas que contemplaram não apenas as categorias como também as frequências, por exemplo: “*6 vieram de bicicleta, 2 de ônibus, 12 caminhando e 2 de carro*” (S514). Poucos estudantes foram hábeis em oferecer respostas nesse nível, no 2º ano (8,3%) e no 5º ano (20,8%).

Conforme apresentado na Tabela 2, surgiram ainda outros tipos de respostas como não se referir aos elementos representados e apresentando argumentos como o sujeito (S220) que responde “*chegaram trabalhando, escrevendo...*”. Foi frequente entre os estudantes do 2º ano (33,3%) a referência a apenas um elemento do gráfico como, por exemplo, responder que todos vieram de ônibus “*porque acho que o colégio é longe*” (S211). É possível perceber nesse trecho de fala que a vivência do sujeito foi mais relevante que o próprio gráfico para a resposta. Ainda encontramos estudantes que escolhiam duas ou três categorias e afirmavam: “*vieram de ônibus e de carro*” (S205).

Observa-se que os estudantes do 2º ano não apresentaram uma preocupação em esgotar todas as categorias apresentadas. Quando comparamos esses percentuais, verificou-se uma diferença bastante significativa no desempenho dos dois grupos de sujeitos ( $t(46) = 3,93, p < 0,001$ ), uma vez que 79% dos estudantes do 5º ano e 29% dos estudantes do 2º ano responderam adequadamente a essa questão.

A segunda questão (Q1.2) buscava investigar a compreensão dos estudantes sobre moda e ponto máximo do conjunto de dados. Na Tabela 3, a seguir, apresentamos os tipos de respostas para a questão: “*Qual o meio de locomoção mais utilizado por essas crianças nesse dia?*” (Ponto máximo / moda).

**Tabela 3: Percentual de respostas adequadas na Q1.2 por ano de escolaridade**

Tipos de respostas	2º	5º
1) Escolhe ônibus	33,3	16,7
2) Escolhe caminhando	41,7	79,2
3) Escolhe bicicleta	8,3	---
4) Escolhe carro	16,7	---
5) Cita vários transportes	---	4,2

A forma de chegada *caminhando* seria aquela considerada adequada, de acordo com os dados apresentados aos estudantes no gráfico. Contudo, menos da metade (41,7%) dos sujeitos do 2º ano escolheram tal categoria. Essa constatação se agrava quando se verifica, a partir das justificativas, que desses 41,7% apenas 29,2% realmente compreenderam a questão. Isso porque os outros 12,5% escolheram *caminhando* argumentando “*Andando, porque muita gente não pega carro*” (S203), o que é um fato comum no cotidiano desses sujeitos, entretanto, as informações apresentadas no gráfico não foram levadas em consideração.

Entre os estudantes do 5º ano equívocos desse tipo também ocorreram. Porém, dos 79,2% que citaram a barra *caminhando* 66,7% justificaram adequadamente sua escolha, como o estudante que aponta para a barra *caminhando* e afirma: “*Porque passa de todo mundo*” (S221).

Comparando o desempenho dos dois grupos de estudantes, constatou-se que os sujeitos do 5º ano foram mais hábeis na leitura do ponto máximo no gráfico do que os sujeitos do 2º ano ( $t(46) = 2,75, p = 0,009$ ).

Verificou-se ainda outros tipos de respostas, como os sujeitos que optaram por qualquer outra barra ou um estudante que citou vários transportes. Dentre os que escolheram outro meio de locomoção que não *caminhando*, um número expressivo de estudantes, principalmente no 2º ano, citaram *ônibus*. A opção pelo *ônibus* como forma mais utilizada, contrariando as informações do gráfico, esteve relacionada com outros fatores que não o tamanho das barras no gráfico ou frequência expressada na escala, os que optaram pelo mesmo justificaram: “*ônibus porque cabe mais gente*” (S205).

A terceira questão (Q1.3) requeria dos estudantes a habilidade de *predição a partir da moda*. Apesar da diferença de desempenho entre os grupos de estudantes, conforme Tabela 4, o teste estatístico não identificou diferença significativa entre os grupos ( $t(46)=1,41, p= 0,16$ ). A seguinte questão foi solicitada: “*Pedro é uma dessas crianças. Como você imagina que ele chegou à escola?*” (*Predição a partir da moda*).

**Tabela 4: Percentual de respostas adequadas na Q1.3 por ano de escolaridade**

Tipos de respostas	2º	5º
1) Aponta qualquer barra justificando com situações imaginativas	70,8	62,5
2) Aponta qualquer barra (sem justificar)	20,8	4,2
3) Aponta barra caminhando justificando com situações imaginativas	4,2	16,7
4) Faz predição	4,2	16,7

Considerou-se como justificativas que indicaram compreensão da *moda*, enquanto medida que subsidiou a predição, respostas como: “*De pé, por que a maioria veio de pé*” (S502). Contudo, apenas 16,7% dos estudantes do 5º ano deram esse tipo de resposta e no 2º ano esse índice baixou para 4,2%. Desse modo, constatou-se que mesmo os estudantes percebendo a *moda*, eles não utilizaram essa informação para fazer as predições.

Outros estudantes (4,2% no 2º ano e 16,7% no 5º) também apontaram a barra *caminhando*. Entretanto, suas justificativas evidenciaram claramente respostas mais comprometidas com suas vivências ou imaginação do que preocupadas em considerar os dados representados no gráfico. A resposta “*acho que a pé porque muita gente não tem carro, nem bicicleta...*” (S508) é um exemplo de tal constatação. Porém, a maioria dos sujeitos de ambos os anos escolares (70,8% no 2º ano e 62,5% no 5º ano) escolheu qualquer outra barra justificando também com vivências ou imaginação, como o sujeito (S510) que afirma: “*Bicicleta porque é melhor, carro polui o ar*”. Ainda foi possível encontrar respostas em que se escolhia qualquer barra sem justificção alguma, o que aconteceu principalmente no 2º ano (20,8%).

Ao se comparar o desempenho dos estudantes nessa questão (Q1.3) com a questão anterior (Q1.2), percebeu-se que os sujeitos que ofereceram uma resposta adequada na Q1.3 (predição a partir da *moda*) fizeram o mesmo na Q1.2 (identificação da *moda* / ponto máximo). Isso significa que poucos estudantes foram hábeis não apenas em encontrar a *moda* como também em refletir a respeito da mesma. Por outro lado, pode-se concluir que nem sempre habilidade em perceber a *moda*, tarefa amplamente explorada pelos Livros Didáticos e professores dos anos iniciais quando se trabalha com leituras de gráficos, possibilita que os estudantes reflitam sobre os dados para fazer predições.

A quarta questão (Q1.4) propunha a *comparação entre dois pontos*, na qual os estudantes precisariam expressar quanto variou, ou seja, uma medida. Tal tarefa mostrou-se difícil, assim como a atividade anterior, principalmente para os estudantes do 2º ano, pois nesse último nenhum estudante acertou. Quanto aos estudantes do 5º ano, 37,5% dos mesmos acertaram a questão. A diferença de desempenho entre os grupos foi significativa ( $t(46) = 3,71, p = 0,001$ ).

Ao se comparar o desempenho dos estudantes do 2º e 5º ano infere-se que provavelmente os últimos vivenciaram situações de ensino que permitiram aos mesmos compreenderem melhor a tarefa. Assim, medir variação parece ser uma atividade trabalhada na escola, mesmo que de modo não sistemático, pois é possível encontrar questões como essa em Livros Didáticos dos anos iniciais, apesar de em um número bem reduzido. Outra possibilidade para a diferença de desempenho entre estudantes dos dois grupos é que sujeitos do 5º ano podem estar generalizando conhecimentos construídos em outros eixos da Matemática. Um exemplo pode ser o bloco Números e Operações, uma vez que questões como essa (quantos a mais / a menos) são exploradas no início da escolarização e fazem parte do campo conceitual das estruturas aditivas<sup>2</sup>.

A Tabela 5 revela os tipos de respostas para a questão: “*Quantos a mais chegaram caminhando que de ônibus?*” (*Comparação entre dois pontos*).

**Tabela 5: Percentual de respostas adequadas na Q1.4 por ano de escolaridade**

Tipos de respostas	2º	5º
1) Diz não saber	4,2	---
2) Cria situação imaginativa	16,7	4,2
3) Soma os dois valores	12,5	---
4) Diz um valor ou barra qualquer	16,7	20,8
5) Cita uma ou as duas frequências	50,0	37,5
6) Aponta a diferença e conta as linhas de grade (não sabe ler escala)	---	16,7
7) Responde corretamente	---	20,8

<sup>2</sup> Magina, Campos, Nunes e Gitirana (2001) argumentam, baseadas na Teoria dos Campos Conceituais de Gerard Vergnaud, que esse tipo de problema pode ser classificado como de comparação no campo conceitual aditivo (estrutura aditiva).

Observa-se que foram encontrados diferentes tipos de resposta, entretanto, apenas as duas últimas, em destaque na Tabela 5, podem ser considerados como adequados. Os estudantes que ofereceram a resposta classificada como 7 (20,8%) responderam corretamente expressando a diferença entre as barras e demonstraram compreender a escala, o que só ocorreu entre estudantes do 5º ano. Já os estudantes classificados na categoria 6 (16,7%), apontaram a diferença entre as barras, mas demonstraram não compreender a escala uma vez que suas respostas basearam-se na diferença das linhas de grade. Esse tipo de resposta também foi constatado apenas entre estudantes do 5º ano.

O estudo de Guimarães (2002) já apontou para a dificuldade dos alunos em compreender os intervalos das escalas, principalmente quando não unitárias. Desse modo, pelo conhecimento desse fator enquanto dificultador considerou-se como adequado também o tipo de resposta em que os estudantes contavam as linhas de grade existentes entre uma barra e outra para expressar a diferença entre as mesmas. Afinal, em ambos os tipos de respostas a proposta da questão foi compreendida, uma vez que um valor da medida de variação foi exposto.

Outros tipos de respostas surgiram, nos quais tentativas de resolução do problema foram realizadas. Contudo, para a maioria dos estudantes a questão gerou dificuldades de compreensão. O tipo de resposta ou estratégia mais frequente, tanto no 2º ano (50%) quanto no 5º (37,5%), caracterizou-se pela citação de uma das frequências ou das duas. Nesse caso, o que os estudantes realizavam era uma comparação entre as barras (*caminhando* e *ônibus*) sem medição, ou seja, percebiam que existia variação, porém, não compreenderam que precisavam expressá-la com um valor. A justificativa “*veio mais andando que de ônibus, porque de ônibus tá baixo e andando tá alto*” (S504) evidencia essa conclusão. Outros estudantes, 16,7% no 2º ano e 20,8% no 5º ano, indicaram um valor, como o estudante que respondeu “*vieram 4 porque a escola é perto*” (S205) ou referiram-se a qualquer outra barra (como *carro*). Esses, contudo, não ofereceram maiores explicações quanto a suas escolhas, que permitissem outras inferências.

Também nessa questão (Q1.4) foi possível encontrar respostas baseadas apenas em situações imaginativas, sem qualquer alusão a valores ou menção a outros elementos do gráfico, o que ocorreu pouco (16,7%, no 2º ano e apenas 4,2%

no 5º), porém nos dois grupos de estudantes. Para exemplificar, tem-se o estudante que responde: *“Porque o ônibus não deu para carregar mais”* (S217). Por fim, constatou-se tipos de respostas encontrados apenas no 2º ano. Nesse caso, os estudantes afirmavam não saber (4,2%), assim como somavam os dois valores das barras *caminhando* e *ônibus* (12,5%). É importante esclarecer que na soma dos valores os estudantes utilizaram frequências imaginárias, isto é, a escala não foi considerada, como mostra a resposta do sujeito (S204) ao afirmar: *“2 a mais (1 de ônibus mais 1 a pé)”*. Para esse estudante uma criança chegou à escola de ônibus e outra a pé.

A quinta questão (Q1.5) envolvia a habilidade de manipulação das barras, a partir do momento que se propunha a retirada de uma categoria, diminuindo a variabilidade no conjunto. Na Tabela 6, a seguir, apresentam-se os tipos de respostas encontrados para a questão: *“Se tivesse ocorrido nesse dia uma greve de ônibus, mas as crianças tinham que ir à escola, não podiam faltar. O que precisaríamos mudar no gráfico (como ele ficaria)?”* (Manipulação das barras).

**Tabela 6: Percentual de tipos de respostas na Q1.5 por ano de escolaridade**

Tipos de respostas	2º	5º
1) Resposta imaginativa ou confusa	12,5	---
2) Apenas caminhando	37,5	41,7
3) Cita mais de uma opção	33,3	50,0
4) Outra opção que não caminhando	16,7	8,3

Os dados da Tabela 6 revelam que a maioria dos estudantes, de ambos os grupos, aceitaram propor caminhos alternativos para o problema apresentado aos mesmos. Então, se o ônibus não podia ser utilizado naquele dia (uma barra foi eliminada), um ou mais caminhos alternativos foram citados (uma ou mais barras foram apontadas) pelos estudantes como solução. Justificativas como *“Iam ter que ir a pé...”* (S501) e ainda *“Vai ficar sem criança (aponta para a barra ônibus) e vinham de carro”* (S516) revelam a escolha de uma entre as formas dispostas no gráfico. O sujeito S501 optou pela barra *caminhando* e o sujeito S516 fez a opção por outra barra (carro).

A distinção realizada entre a escolha da forma *caminhando* e as demais foi necessária, uma vez que verificou-se um alto percentual de estudantes tanto no 2º

ano (37,5%), quanto no 5º ano (41,7%) citando *caminhando* como a forma que poderia estar sendo utilizada já que o ônibus estava em greve. Na verdade, acredita-se que esses estudantes ofereceram respostas coerentes com a realidade deles.

Ainda foi possível encontrar respostas que buscaram oferecer mais de uma opção para o problema proposto, como o sujeito (S509) que afirma: *“Podiam vir de bicicleta ou a pé”*. Finalmente, obtiveram-se respostas imaginativas ou confusas, nas quais não se explicitou elementos representados no gráfico, como o sujeito (S211) que justifica: *“Ele não ia poder entrar, mas ia providenciar pra entrar na escola”*. Respostas desse tipo foram poucas (12,5%), encontradas apenas entre os estudantes do 2º ano.

A partir do exposto na Tabela 5 e dos trechos de fala evidenciados acima, pode-se argumentar que dentre as respostas dos estudantes para a Q1.5 considerar adequações ou não é problemático, uma vez que não foi possível obter um mínimo de garantia de que os estudantes estivessem realizando algum tipo de leitura do gráfico ou simplesmente baseando suas escolhas em fatos vivenciados por eles no cotidiano.

Essa discussão será retomada no tópico a seguir, momento em que buscou-se relacionar os resultados encontrados no presente estudo com a literatura utilizada como fundamento para o mesmo.

### **3.1.1 Discussão da atividade 1**

Como já explicitado na metodologia, a atividade 1 foi elaborada por Watson e Kelly (2002a) e adaptada no presente estudo, o que torna possível a comparação dos resultados descritos acima com aqueles encontrados por tais autores. Entretanto, sabe-se que é importante respeitar as especificidades de cada estudo. A pesquisa realizada por Watson e Kelly (2002a), por exemplo, contou com apenas 7 (sete) sujeitos, os quais durante a atividade precisaram de ajuda na leitura da escala no gráfico utilizado. Aqui, entende-se que tal ajuda seria um direcionamento que geraria respostas diferentes daquelas que os estudantes poderiam estar oferecendo espontaneamente. Desse modo, optou-se por não fazê-lo. Essa opção, por outro

lado, acarreta menores possibilidades de comparação. Abaixo, discute-se as questões da atividade 1 possíveis de serem relacionadas com estudos anteriores.

Para a questão (Q1.2) que requeria a leitura da *moda*/ponto máximo no gráfico, os resultados revelam um baixo percentual (29,2%) de adequação da resposta entre os estudantes do 2º ano, estando os estudantes do 5º ano com um índice significativamente melhor. O baixo percentual no desempenho do 2º ano contraria o estudo de Watson e Kelly (2002a) com crianças de faixa etária aproximada (seis anos), no qual todos os estudantes localizaram o ponto máximo ou *moda*. Entretanto, salienta-se, mais uma vez, que esses sujeitos contaram com ajuda na leitura da escala do gráfico, o que se optou, aqui, por não fazer.

Já a questão de predição a partir da moda (Q1.3), mostrou-se uma questão de difícil compreensão tanto nos resultados do presente estudo quanto nos resultados dos autores citados. Em ambas as pesquisas, os estudantes que não utilizaram o gráfico para predizer contaram histórias imaginativas ou baseadas em suas vivências. Como em outras atividades a habilidade de predição (a partir do ponto máximo, assim como da tendência do gráfico) também foi testada, essa discussão será retomada mais adiante.

Quanto à questão de comparação entre dois pontos no gráfico (Q1.4), na qual se expressa com um valor a variação entre duas barras, mostrou-se difícil para os dois grupos de estudantes, como já discutido anteriormente. Os sujeitos do 5º ano obtiveram um desempenho significativamente melhor que os do 2º ano (no qual não houve acerto). Entretanto, os resultados de Watson e Kelly (2002a) são bem diferentes, uma vez que cinco dos sete estudantes (na faixa etária de seis anos) solucionaram o problema proposto. Os demais responderam “sete a mais”, o valor da maior frequência. Esse tipo de resposta, na qual se expõe o valor de uma das barras também surgiu no presente estudo. Os dados dos autores citados diferem não apenas aqui como também daqueles alcançados por Guimarães (2002).

Contudo, apesar das dificuldades constatadas nas pesquisas realizadas no Brasil a respeito dessa questão, tanto Guimarães (2002) como Arcoverde et al (2004) concordaram que pequenas intervenções foram suficientes para que estudantes de 4º ano compreendessem a tarefa. Ao se comparar os resultados encontrados na Q1.4 com o que a literatura vêm discutindo a respeito das

compreensões apresentadas por estudantes dos anos iniciais, percebe-se que de fato o estudo de Watson e Kelly (2002a) representou uma novidade, principalmente por se tratar de crianças tão pequenas. Como essas crianças obtiveram ajuda com a leitura da escala, uma explicação possível pode advir de tal orientação, o que merece maiores investigações.

Por último, tem-se a questão de *manipulação das barras* (Q1.5) na qual a proposta era reduzir a variabilidade de um conjunto a partir da retirada de uma categoria. A proposta era que os estudantes buscassem caminhos alternativos de modo que conservassem a quantidade total de crianças no gráfico. Verificou-se nos resultados da questão, que sugerir opções para aquelas crianças que não poderiam utilizar o ônibus foi uma tarefa fácil. Afinal, a maioria dos estudantes de ambos os grupos citaram uma ou mais alternativas para se chegar à escola. O mesmo aconteceu com os sujeitos da pesquisa realizada por Watson e Kelly (2002a).

Entretanto, cabe a reflexão a respeito da possibilidade de que os estudantes quando optaram por uma ou mais formas de se chegar à escola não estivessem fazendo uma leitura do gráfico, no sentido de distribuir a frequência de crianças que não poderiam utilizar o ônibus (duas no caso) nas demais barras. Afinal, os estudantes poderiam estar respondendo de acordo com suas experiências apenas. Acredita-se que na proposta da questão a leitura da frequência e distribuição da mesma pelos estudantes deveriam ter sido exploradas para que se pudesse avaliar com maiores subsídios a habilidade dos sujeitos em conservar a quantidade total no gráfico ao se reduzir categorias e conseqüentemente a variabilidade do conjunto.

### **3.2 Atividade 2**

A atividade 2 englobou seis questões. Diferente da atividade 1, as questões na atividade 2 emergiram da situação em que os estudantes representaram a variabilidade de um conjunto de dados, quando cartões mostrando livros e leitores foram oferecidos aos mesmos.

Assim, iniciou-se a primeira questão (Q2.1) solicitando que os estudantes representassem uma figura que mostrasse a quantidade de livros lidos por quatro

crianças. Procurou-se esclarecer que a figura construída por eles fosse auto-explicativa, ou seja, apenas olhando qualquer pessoa poderia saber a quantidade de livros que cada criança leu. Para isso, os estudantes foram informados, logo de início, que uma foto da representação criada seria mostrada a outra pessoa para ela saber as quantidades de livros lidos por cada criança. A Tabela 7, a seguir, apresenta os tipos de respostas para a questão: “*Você pode usar esses cartões para mostrar a quantidade de livros que essas crianças leram?*” (*Representação da variabilidade*).

**Tabela 7: Percentual de respostas adequadas na Q2.1 por ano de escolaridade**

Tipos de respostas	2º	5º
1) Expõe cartões misturados	25,0	8,3
2) Organiza em grupos (pilhas) que não mostram a variabilidade	41,7	20,8
3) Organiza em grupos podendo contar	8,3	20,8
4) Organiza em filas podendo contar e comparar	25,0	25,0
5) Faz pictograma	---	25,0

A Tabela 7 destaca três tipos de respostas (três últimas) consideradas adequadas, uma vez que as representações criadas em tais casos permitiram a expressão do critério mínimo proposto, que foi saber a quantidade de livros que cada criança leu sem a necessidade de manipulação dos cartões na contagem. Adequações das representações quanto à proposta da tarefa foram mais frequentes entre os sujeitos do 5º ano (70,8%) do que entre os do 2º ano (33,3%). Isso significa que para os estudantes do 5º ano essa foi uma atividade fácil, pois a maioria compreendeu. Constatou-se uma diferença significativa ( $t(46) = 2,75$ ,  $p = 0,009$ ) no desempenho entre os grupos de estudantes.

A análise mais detalhada dos tipos de respostas revela que surgiram tipos de respostas, nos quais não se atendeu a necessidade de mostrar a variabilidade de livros lidos por crianças, sem manipulação dos cartões. Houve o tipo de resposta menos elaborado, de acordo com a proposta da questão, que foi a disposição dos cartões sobre a mesa sem qualquer correspondência entre quantidade de livros lidos e crianças (Figura 1). A exposição dos cartões de modo misturado foi realizada por 25% dos estudantes do 2º ano e apenas 8,3% estudantes do 5º. Um percentual alto de estudantes do 2º ano (41,7%) organizou os cartões em grupos, cujos cartões

também não podiam ser contados sem manuseá-los, pois formaram pilhas (Figura 2), no 5º ano o percentual foi de 20,8%.



**Figura 1: Exemplo de exposição de cartões misturados (S217)**



**Figura 2: Exemplo de organização de cartões em pilhas (S210)**

Quanto aos tipos de respostas apropriados, apenas 8,3% dos estudantes do 2º ano e no 5º ano 20,8% organizaram os cartões em grupos que podiam ser contados, conforme Figura 3. Um tipo de resposta julgada mais elaborada que essa foi a organização dos cartões em filas (Figura 4), pois admite não apenas a contagem como também facilita comparações entre quantidades. A organização em filas apareceu na mesma frequência para os dois anos, 25%. Finalmente, houve aquele tipo de representação que além de permitir contagens e comparações em filas, utilizou a mesma base para as filas (Figura 5). Essa representação foi denominada pictograma, configuração que se aproxima do gráfico de barras e aconteceu apenas entre estudantes do 5º ano (25%).



**Figura 3: Exemplo de organização em grupos que podiam ser contados (S518)**



**Figura 4: Exemplo de organização dos cartões em filas (S220)**



**Figura 5: Exemplo de organização de pictograma (S512)**

A segunda questão (Q22) da atividade 2, como na atividade 1, também explorou a habilidade dos estudantes na *explicação da variabilidade*. Aqui,

entretanto, fatores diversos determinaram outros tipos de respostas e índices maiores no percentual de justificativas consideradas adequadas. Fatores que serão retomados mais adiante. Na Tabela 8, a seguir, apresentam-se os tipos de respostas encontradas para a pergunta: “*Se alguém entrar na sala o que poderia dizer olhando para essa figura?*” (*explicação da variabilidade*).

**Tabela 8: Percentual de respostas adequadas na Q2.2 por ano de escolaridade**

Tipos de respostas	2º	5º
1) Diz não saber	12,5	4,2
2) Opina sobre uma criança	8,3	---
3) Valoriza crianças serem leitoras	41,7	62,5
4) Foco no ponto máximo	8,3	4,2
5) Descreve as quantidades	29,2	29,2

A Tabela 8 destaca pelo menos três tipos de respostas classificadas como adequadas. Das respostas oferecidas pelos estudantes, no 2º ano 79,2% foram analisadas como apropriadas dentro das possibilidades da questão e no 5º ano esse número foi altíssimo, uma vez que 95,8% dos sujeitos justificaram partindo de uma percepção mínima dos dados representados. A diferença de desempenho entre os grupos não foi significativa ( $t(46) = 1,76$ ,  $p = 0,084$ ).

Um tipo de resposta, pouco frequente, porém presente nos dois grupos (12,5% no 2º ano e 4,2% no 5º ano) foi o estudante não responder à questão, justificando que não sabia. Observou-se que 8,3% dos estudantes do 2º ano expressaram uma opinião sobre uma das crianças, como o sujeito que aponta para o cartão representando *Daniel* e responde: “*Ele é comportado*” (S218).

O tipo de resposta predominante, no 2º ano (41,7%) e principalmente no 5º ano (62,5%) foi a valorização das crianças representadas nos cartões enquanto leitoras. A resposta do sujeito (S510) “*são legais, inteligentes e se interessam pelas coisas*”, como exemplo, evidencia que apesar da maioria dos estudantes (52,1%) não expressarem claramente em suas falas comparações (quantificando ou não) entre os dados representados, os mesmos foram coerentes com a questão, uma vez que a mesma requeria a opinião deles. De fato, ressaltar que existiam várias crianças leitoras é importante e desejado pelos educadores. Acredita-se que nessa questão para que o estudante realizasse comparações entre os dados, oferecendo

respostas quantitativas, seria necessário um outro tipo de questão na qual deixasse explícito essa necessidade. Aqui, optou-se por não explorar mais a tarefa nesse momento, pois outros direcionamentos poderiam comprometer as questões e atividades seguintes.

As respostas que demonstraram leitura do ponto máximo foram consideradas adequadas, pois essa era também uma possibilidade de resposta. De fato, essa questão foi um tanto vaga, então diferentes respostas podem ser consideradas corretas. O foco no ponto máximo foi pouco frequente entre os dois grupos de sujeitos, no 2º ano 8,3% e no 5º ano, 4,2%, como por exemplo: *“Essa daqui (aponta para Luiza). Ela leu mais! (S222).*

Finalmente, a descrição das quantidades foi realizada por 29,2% dos estudantes de ambos os grupos. Um exemplo desse tipo de resposta pode ser: *“Luiza pegou a maior quantidade de livros, Ana um pouco da metade, Bruno quase metade...” (S520).* Considerou-se a ação de descrever as quantidades como um tipo mais elaborado de resposta. Tal descrição revelou uma preocupação com o conjunto de dados como um todo durante a explanação, não apenas com o ponto (categoria) de maior frequência, mesmo que se concorde que para a leitura do ponto máximo foi necessário realizar comparações entre os dados.

Na terceira questão (Q2.3), o que se pretendia era a obtenção de maiores esclarecimentos a respeito da representação da quantidade de livros lidos no formato de pictograma. A dúvida era: Os estudantes que não construíram pictograma, não conheciam gráficos ou escolheram não utilizar esse tipo de representação? Para isso, questionou-se: *Você sabe o que é um gráfico?*

Constatou-se que três sujeitos do 5º ano refizeram suas figuras e construíram pictograma, totalizando 9 construções desse tipo. O percentual de estudantes que criaram pictograma subiu de 25% para 37,5%. Os demais sujeitos responderam que não sabiam o que era gráfico. Antes da questão (Q2.3), dos três que refizeram suas figuras, o primeiro dispôs os cartões misturados, o segundo organizou em filas e o terceiro organizou em grupos, cujos livros puderam ser contados.

A quarta questão (Q2.4) requeria dos estudantes a modificação de suas construções a partir da aceitação ou não da frequência nula na representação. Para atender ao objetivo da questão, a seguinte pergunta foi proposta aos estudantes:

“Henrique não leu nenhum livro. Você pode mostrar que ele não leu livro algum?”  
(Representação de ausência de frequência).

Nos dois grupos, a maioria dos estudantes aceitou incluir *Henrique* na figura, o que foi uma aceitação da possibilidade de se representar ausência de frequência. No 2º ano 66,7% concordaram com essa possibilidade e no 5º ano 70,8%. Não houve diferença significativa entre os grupos ( $t(46) = 0,30$ ,  $p = 0,76$ ). A seguir, a Figura 6 exemplifica representação de estudante (S219) que considerou a frequência nula.



**Figura 6: Exemplo de representação que considera frequência nula (S219)**

A quinta questão (Q2.5) também solicitava que os estudantes modificassem a representação construída. Porém, na Q2.5 a proposta foi que os estudantes acrescentassem um livro lido para cada criança representada. A Tabela 9, abaixo, apresenta os tipos de respostas encontradas para a questão: “*Todas essas crianças foram à biblioteca e leram mais um livro cada uma. Você pode mostrar na sua figura que todos leram mais um livro?*” (Representação de acréscimos).

**Tabela 9: Percentual de respostas adequadas na Q2.5 por ano de escolaridade**

Tipos de respostas	2º	5º
1) Diz não saber	4,2	---
2) Segura cartões e não distribui	16,7	---
3) Junta os cartões com demais espalhados	12,5	4,2
4) Não acrescenta para todos os leitores	4,2	12,5
5) Acrescenta um cartão para cada leitor	62,5	83,3

Dos tipos de respostas da Tabela 9, considerou-se como apropriado apenas o último, em destaque, no qual os estudantes acrescentaram um livro para cada leitor. Essa mostrou-se uma tarefa fácil, principalmente para os estudantes do 5º ano, cujo desempenho foi de 83,3% na adequação da resposta. Os estudantes do 2º ano conseguiram uma frequência de 62,5% para esse tipo de resposta, não sendo significativa a diferença ( $t(46) = 1,63, p = 0,109$ ) entre os dois grupos de sujeitos. A Figura 7, a seguir, evidencia a distribuição de cinco cartões (representando livros) entre cinco crianças, inclusive *Henrique*, o qual não tinha lido nenhum livro, como proposto na questão anterior.

Ainda foi possível encontrar aqueles estudantes que receberam os cartões representando livros, contudo, não concordaram em acrescentar para todos os leitores (Figura 8). Esse tipo de resposta foi pouco frequente, porém presente entre os dois grupos de sujeitos (4,2% no 2º ano e 12,5% no 5º ano).



**Figura 7: Exemplo de representação que considera frequência nula e faz acréscimo (S509)**



**Figura 8: Exemplo de representação que não considera frequência nula e não faz acréscimos para todos (S516)**

É interessante observar na Figura 8 que o sujeito (S516) deu todos os cartões para a criança com menos livros (*Daniel*), o que não foi comum a todos os sujeitos que ofereceram esse tipo de resposta. Entretanto, nenhum desses acrescentou livros à criança que leu mais. Tal fato pode indicar que esses sujeitos compreenderam a questão, contudo, não acreditaram ser “justo” dar mais um livro para quem já tinha lido tantos. Desse modo, manipularam os cartões de acordo com suas crenças, acarretando uma diminuição na variabilidade de livros lidos, uma vez que a distância entre as frequências diminuiu sensivelmente.

Ainda na Q2.5, observou-se outros tipos de respostas constatadas apenas entre estudantes do 2º ano. Foram os tipos em que os sujeitos seguraram os cartões e não fizeram a distribuição (16,7%) ou afirmaram não saber fazer (4,2%). Em ambos os casos, a questão não foi compreendida pelos sujeitos.

A sexta questão (Q2.6) requeria dos estudantes predição a partir do ponto máximo na figura representada pelos mesmos. A seguir, a Tabela 10 apresenta os tipos de respostas para a questão: *Quem você acha que provavelmente gostaria de receber um livro de presente no Natal? Por quê? (Predição a partir do ponto máximo).*

**Tabela 10: Percentual de respostas adequadas na Q2.6 por ano de escolaridade**

Tipos de respostas	2º	5º
1) Escolhe quem leu menos	29,2	37,5
2) Escolhe um leitor qualquer e justifica	50,0	12,5
3) Escolhe quem leu mais	20,8	50,0

Tem-se destacado na Tabela 10 o tipo de resposta considerado aqui como adequado. No 5º ano, 50% dos estudantes escolheram aquela criança que tinha lido mais. Contudo, no 2º ano esse percentual foi bem inferior (20,8%), o que ocasionou uma diferença significativa entre os dois grupos ( $t(46) = 2,17$ ,  $p = 0,035$ ). A justificativa *“Luiza porque gosta de ler. Leu 6”* (S216), é um exemplo de resposta apropriada para a Q2.6, que demonstra habilidade em prever resultados a partir dos dados representados.

Entretanto, encontramos respostas em que estudantes escolheram a criança que leu menos (29,2% no 2º ano e 37,5% no 5º ano). Como exemplo, expõe-se a resposta: *“Henrique porque só tem um”* (S221). Está claro que os estudantes que escolheram a criança que leu menos compreendeu os dados, pois foram capazes de fazer comparações entre os mesmos. Porém, tais respostas também indicam que para esses estudantes mais relevante em suas predições que compreender os dados e fazer predições a partir dos mesmos foi a oportunidade de transformá-los de acordo com suas crenças. Isto é, esses sujeitos, talvez por uma questão de “justiça”, concordaram que o estudante que leu menos é que gostaria de receber um livro de presente. Mais uma vez estudantes ao tomarem esse tipo de decisão tenderam a diminuir a variabilidade na quantidade de livros lidos, como na questão anterior.

Por último, houve o tipo de resposta em que os estudantes não se preocuparam em fazer uma leitura dos dados, no sentido de compará-los para realizar predições. Esse foi o tipo mais comum entre os estudantes do 2º ano, metade (50%) ofereceu respostas como *“Ana porque tá mais séria”* (S503). Percebe-se no trecho de fala que crenças e vivências parecem orientar as decisões de tais sujeitos, pois acreditam que uma pessoa séria é que provavelmente gostaria de receber um livro como presente. No 5º ano a escolha de um leitor qualquer como resposta também foi realizada, contudo, foi pouco frequente (12,5%).

### 3.2.1 Discussão da atividade 2

Os resultados alcançados na atividade 2 permitem a comparação de algumas dessas questões com estudos anteriores, como os de Watson (2009) e também de Watson e Kelly (2002a), a partir dos quais a atividade foi adaptada.

A Q2.1, na qual se requeria a representação da variabilidade de livros lidos, no presente estudos sujeitos do 5º ano obtiveram um desempenho significativamente melhor (70,8%) que os sujeitos do 2º ano (33,3%). Ao se relacionar esses dados com aqueles do estudo de Watson e Kelly (2002a), realizado com 7 (sete) crianças na faixa etária de 6 (seis) anos, constata-se que os tipos de respostas encontrados foram os mesmos.

No estudo dos autores citados, apenas um sujeito representou a quantidade de livros de modo que não poderia ser contado (em pilhas). Nós consideramos esse tipo de resposta inadequada uma vez que a questão solicitou a leitura de uma foto, conforme instrução oferecida e não a manipulação dos cartões. Aqui, esse tipo de resposta foi frequente (41,7%) entre os estudantes do 2º ano. No atual estudo optou-se por fazer uma distinção entre os estudantes que separaram os cartões por grupos de modo que fosse possível a contagem sem manipulação. No estudo citado três sujeitos expuseram cartões representando livros em torno do cartão representando uma criança. Da mesma forma, no 2º ano 8,3% formaram grupos e 25% formaram filas. No estudo de Watson e Kelly (2002a) os três últimos sujeitos organizaram os cartões em colunas (entendido aqui como pictograma), contudo, no presente estudo esse tipo de construção só ocorreu entre estudantes do 5º ano.

Uma explicação para esse fato pode ser o questionamento por parte dos autores citados a respeito da quantidade total de livros do conjunto, o que pode ter levado os mesmos a refletirem sobre a representação criada por eles. Acredita-se que quando os sujeitos na pesquisa são estimulados a refletirem a respeito de suas construções tem-se um processo de ensino, o que torna o estudo uma pesquisa de intervenção. Esse não foi o objetivo do atual estudo, caracterizado por um processo de diagnose. Assim, optou-se por evitar questões que pudessem interferir de modo direto nas respostas dos estudantes.

As demais questões da atividade 2 não permitem uma discussão mais profunda de seus resultados com a literatura utilizada como base, uma vez que nos estudos anteriores não foram analisadas individualmente, porém como um todo. Apenas a Q2.6 permite comparação, pois no estudo de Watson e Kelly (2002a) os autores afirmaram que prever quem gostaria de receber um livro de presente no Natal (*predição a partir do ponto máximo*) foi fácil, já que os sujeitos escolheram aquela criança que leu mais. Diferente dos resultados dos autores citados, no presente estudo outros tipos de respostas foram encontrados para a questão e a resposta considerada adequada foi constatada com metade dos alunos do 5º ano e com apenas 20,8% dos estudantes do 2º ano. Como já se afirmou anteriormente no estudo de Watson e Kelly (2002a) os autores se preocuparam em oferecer instruções a respeito da leitura de escala que certamente interferiu nas respostas dos estudantes.

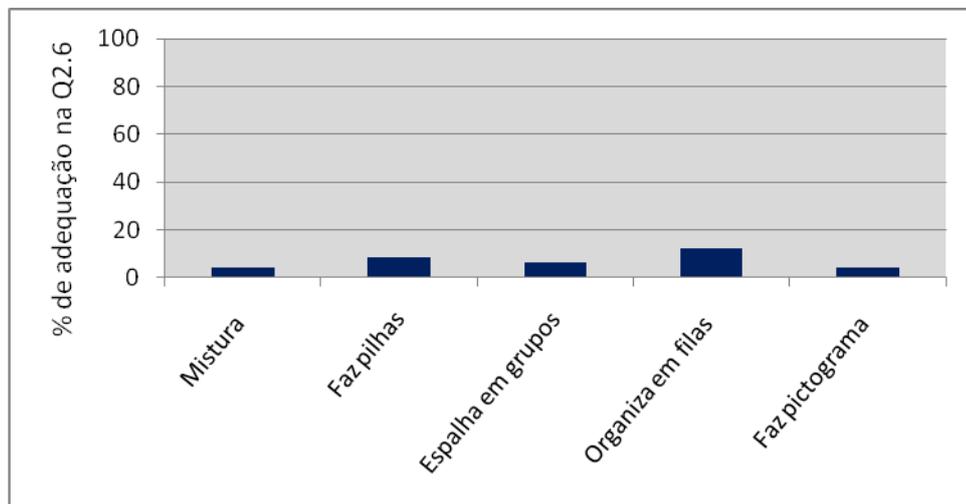
Watson (2009) realizou outro estudo o qual utilizou essa mesma atividade com estudantes entre 6 e 15 anos de idade. A autora juntou as respostas à atividade e classificou-as como um todo em níveis de acordo com o *SOLO* (explicitado no Capítulo 1). Assim ela encontrou respostas de nível menos elaborado (pré-estrutural), ou seja, aquelas em que os estudantes não expuseram a variação na representação gráfica de modo que se poderia contar. No nível mais elevado (abstração estendida) estavam as respostas em que os estudantes eram capazes de empregar variação e expectativa na questão de predição. Apenas estudantes na faixa etária dos treze e quinze anos de idade aproximadamente foram hábeis em reconhecer expectativa e integrá-la com o conceito de variação.

Uma vez que as questões dessa atividade exploraram aspectos diferenciados da variabilidade, optou-se por não utilizar o *SOLO* como classificação das respostas dos estudantes. Decidiu-se por discutir as questões em separado para compreender de forma mais detalhada como os sujeitos compreenderam cada um dos aspectos.

A partir dessa opção os resultados obtidos no presente estudo nos permitem alguns questionamentos. Será que ao se relacionar o desenvolvimento dos estudantes nas questões Q2.1 (representação) e Q2.6 (predição) os sujeitos que foram hábeis em prever a partir do *ponto máximo* também obtiveram um bom desempenho na representação? De acordo com os níveis estabelecidos por Watson

(2009) a habilidade em se relacionar representação da variabilidade e predição (que está articulado com o conceito de expectativa) está num nível mais elevado, sendo predição uma questão chave, presente apenas nos níveis mais avançados. Assim, seria de se esperar que aqueles estudantes que fizeram predição a partir do ponto máximo também criassem representações mais elaboradas.

Entretanto, a partir dos nossos dados pode-se afirmar que não houve uma relação direta entre representação e predição, afinal os 35,4% sujeitos que realizaram a predição esperada na Q2.6 se distribuíram entre todos os tipos de representação na Q2.1 como pode ser observado na Figura 9.



**Figura 9: Percentual de respostas adequadas da Q2.6 quando associada com tipos de representação (Q2.1)**

Assim, a relação entre representar e predizer parece não ser tão simples o que pressupõe a necessidade de outras investigações, a fim de maiores esclarecimentos.

Já ao se relacionar as questões de predição (Q2.6) e explanação da variabilidade (Q2.2), pode-se pensar que aqueles estudantes que foram hábeis em fazer predição a partir do ponto máximo (35,4%, no geral) poderiam ter oferecido as explicações mais elaboradas no que se refere à análise do conjunto de dados representado por eles. Entretanto, constatou-se que os que apresentaram respostas adequadas ao predizerem, quando explanaram ofereceram respostas de diversos tipos (adequados e não adequados) sem uma tendência que revelasse uma relação

direta entre as questões. A seguir, a Tabela 11 expõe os dados a respeito da relação entre ambas as questões.

**Tabela 11: Percentual de respostas na relação entre Q2.2 e Q2.6**

Tipos de respostas na Q2.2	Respostas inadequadas (Q2.6)	Respostas adequadas (Q2.6)
Diz não saber	6,3	2,1
Opina sobre uma criança	4,2	---
Valoriza crianças serem leitoras	31,2	20,8
Foco no ponto máximo	6,3	---
Descreve as quantidades	16,6	12,5
<b>Total</b>	<b>64,6</b>	<b>35,4</b>

Percebe-se que dos 35,4% que responderam adequadamente houve um pequeno percentual de estudantes (2,1%) afirmando não saber explicar a respeito da variabilidade. Por outro lado, foi possível encontrar sujeitos que na explanação mantiveram o foco no ponto máximo, porém, quando foram solicitados a fazer predição a partir do ponto máximo não o consideraram. Ainda ao se comparar os percentuais de respostas inadequadas e adequadas na questão de predição (Q2.6) não se percebe diferenças expressivas entre os mesmos.

### 3.3 Atividade 3

A atividade 3 solicitou a comparação entre dois conjuntos de dados, cujas variáveis eram quantitativas. Essa especificidade dos conjuntos requeria uma análise diferenciada da variabilidade dos dados representados e complexa, quando comparados com conjuntos cuja variável é categórica (qualitativa). A seguir, a Tabela 12 apresenta os tipos de respostas dos estudantes para a questão (Q3.1): *“Miguel e Lucas compraram figurinhas para suas coleções. Observe esses 2 gráficos que mostram a quantidade comprada por cada um durante uma semana. Quem variou mais na quantidade de figurinhas compradas durante essa semana? Por quê?”*.

**Tabela 12: Percentual de tipos respostas na atividade 3 por ano de escolaridade**

Tipos de respostas	2º	5º
1) Escolhe sem justificar	8,3	---
2) Resposta confusa /imaginativa	16,7	12,5
3) Justifica variou mais como comprou menos	8,3	4,2
4) Justifica variou mais como comprou mais	66,7	75,0
5) Justifica variou mais como frequências diferentes	---	8,3

Apesar da diversidade de tipos de respostas para a questão (Q3.1) nenhum dos estudantes conseguiu dar uma justificativa correta. Foi possível encontrar como tipo de respostas a escolha por parte dos estudantes de um dos conjuntos de dados, sem justificção alguma apenas no 2º ano (8,3%). Respostas confusas ou imaginativas foram oferecidas pelos dois grupos de sujeitos, no 2º ano 16,7% e no 5º ano 12,5%, que respondiam conforme o exemplo: *“Lucas é mais interessante. O outro não quer fazer tarefa (S215).*

Foi interessante perceber que as respostas da maioria dos sujeitos estavam vinculadas a concepção de “variou mais” como “comprou mais”. Isso tanto no 2º ano (66,7%) quanto no 5º (75%). A justificativa do estudante (S214) que afirma *“Lucas porque a quantidade tá mais alta” (S214)* evidencia tal concepção.

Acredita-se que esses estudantes não compreenderam a tarefa de verificar mudanças na distribuição e basearam suas respostas no termo “mais”, bem conhecido por eles, optando por aquela criança que teria comprado uma maior quantidade de figurinhas, na opinião dos mesmos. Outros estudantes, que foram poucos (8,3% no 2º ano e 4,2% no 5º ano) justificaram suas escolhas de modo inverso, pois optaram justamente por aquela criança que teria comprado menos figurinhas durante a semana. Como exemplo, tem-se o sujeito que responde: *“Miguel porque tem menos e ele (Lucas) tem mais” (S224).*

O tipo de resposta encontrada apenas entre estudantes do 5º ano (8,3%) também foi pouco frequente, porém já suscita discussões importantes a serem realizadas mais adiante, quando se compara tais resultados com estudos anteriores. Aqui, os sujeitos ao justificarem “variou mais” como frequências diferentes demonstraram compreender o termo “variou mais”, contudo, utilizá-lo de modo adequado ao tipo de variável representada não foi possível. Um exemplo desse tipo

de resposta pode ser: “*Miguel porque não tem nenhum igual e aqui (Lucas) tem iguais*” (S517).

### 3.3.1 Discussão da atividade 3

Os resultados encontrados para a atividade 3 permitem comparações importantes entre tal atividade e o estudo realizado por Loosen, Lioen e Lacante (1985), a partir do qual a atividade foi adaptada. É importante salientar que a pesquisa dos autores citados foi realizada com estudantes de graduação do curso de Psicologia, ou seja, com adultos. Nessa pesquisa, Loosen et al (1985) observaram que apenas 36% dos sujeitos acertaram a questão de comparação entre conjuntos de dados quantitativos. A metade dos sujeitos escolheu o conjunto em que todas as barras (frequências das categorias) eram diferentes, o que corresponde no nosso estudo ao tipo de resposta em que os estudantes justificaram *variou mais como frequências diferentes*. Apesar desse tipo de resposta ter ocorrido pouco (8,3%) e apenas com estudantes do 5º ano, cabe reafirmar que esse tipo de resposta demonstra compreensão do significado de *variar*.

Os resultados encontrados por Cooper e Shore (2008) ao investigarem estudantes de cursos introdutórios de estatística em atividades também de comparação entre conjuntos de dados confirmam a grande dificuldade em se realizar análise de comparação entre conjuntos, considerando os diferentes tipos de representação e diferentes métodos de análise da variabilidade.

Desse modo, pelo conhecimento das dificuldades de estudantes ao compararem conjuntos de dados, como já salientado na literatura, não seria prudente esperar que estudantes de faixa etária de 7 (sete) e 10 (dez) anos de idade demonstrassem compreensões elaboradas.

Entretanto, argumenta-se que os resultados do presente estudo são importantes, uma vez que comparação da variabilidade de conjuntos é um dos componentes para o desenvolvimento do raciocínio variacional (GARFIELD e BEN-ZVI, 2005) e precisa-se saber o que estudantes dos anos iniciais estão compreendendo a respeito de tal componente.

Kader e Perry (2007) ao discutirem os resultados de Loosen et al (1985) propuseram que entender *variou mais como frequências diferentes* é uma compreensão intuitiva de variabilidade, mais adequada quando os dados são categóricos (qualitativos). Entretanto, refletindo a respeito dos dados encontrados e considerando o argumento de Kader e Perry (2007), pode-se concluir que entender *variou mais como frequências diferentes* foi o único tipo de resposta para a questão que revelou alguma compreensão de variabilidade, mesmo que inadequada.

Na atividade seguinte tal discussão é retomada, a fim de se obter maiores esclarecimentos das compreensões apresentadas pelos estudantes ao compararem conjuntos de dados.

### 3.4 Atividade 4

A atividade 4 requeria dos estudantes a comparação entre conjuntos de dados, como na atividade anterior, contudo, as variáveis foram categóricas (qualitativas) e a representação dos conjuntos no formato de gráfico de barras não foi utilizada. A atividade contou com duas questões, as quais partiram da seguinte situação proposta aos estudantes: *“Em uma escola foram formadas três equipes, cada uma com dez pessoas. Todas as pessoas na equipe escolhiam uma resposta (A ou B), para uma questão”*. A primeira questão (Q4.1) foi apresentada aos estudantes com a pergunta: *“Qual equipe teve uma maior variação das respostas?”*. Após escolha e justificativa da mesma, preferiu-se a segunda questão (Q4.2): *“Qual equipe variou menos?”*.

A seguir, apresentam-se as Tabelas 13 e 14, nas quais buscou-se distinguir as escolhas pelas equipes realizadas pelos estudantes para ambas as questões. A Tabela 13, abaixo, revela as escolhas dos sujeitos na Q4.1.

**Tabela 13: Percentual de escolha de equipe na Q4.1 por ano de escolaridade**

Escolha da equipe	2º	5º
1) Não escolhe	4,2	---
2) Escolhe equipe 1	16,7	8,3
3) Escolhe equipe 2	4,2	16,7
4) Escolhe equipe 3	75,0	75,0

Tem-se destacado na Tabela 13 os percentuais de escolha da equipe 2, a opção correta. Constatou-se que poucos estudantes (4,2% no 2º ano e 16,7% no 5º ano) citaram tal equipe, justificando de modo apropriado. Um exemplo de resposta adequada para a Q4.1 pode ser: “A equipe 2 porque ficou equilibrado, nem muito A nem muito B” (S517).

A escolha mais frequente foi da equipe 3 para os dois grupos de sujeitos (75% para o 2º e 5º ano) que respondiam: “A equipe 3, porque tem mais verde” (S219). Entre os que escolheram a equipe 1 o percentual no 2º ano foi maior (16,7%) do que no 5º ano (8,3%). A resposta “equipe 1 porque tem mais A e pouco B, na segunda foi mesma coisa...” (S519) evidencia esse tipo de escolha.

É possível perceber que as escolhas inadequadas dos estudantes das equipes 1 e 3 obtiveram justificativas semelhantes no que se refere à compreensão do conceito de variabilidade, pois ao escolherem a equipe os estudantes focaram na categoria (resposta A, cor amarela; ou resposta B, cor verde) de maior frequência. Na equipe 1 a categoria mais frequente foi a resposta A e na equipe 3 foi a resposta B. O trecho de fala do sujeito (S519) acima indica ainda a dificuldade em aceitar a equipe 2 como a que *variou mais*, uma vez que a frequência das categorias foi a mesma, cinco para cada uma, enquanto que outras equipes dispôs de categorias com maior frequência que as da equipe 2. Por último, houve um percentual baixo de sujeitos (4,2%) que não escolheu equipe alguma, o que só ocorreu no 2º ano.

Abaixo, a Tabela 14 apresenta as escolhas de equipes realizadas pelos estudantes na Q4.2.

**Tabela 14: Percentual de escolha de equipe na Q4.2 por ano de escolaridade**

Escolha da equipe	2º	5º
1) Não escolhe	4,2	4,2
2) Escolhe equipe 1	45,8	45,8
3) Escolhe equipe 2	25,0	25,0
4) Escolhe equipe 3	25,0	25,0

A Tabela 14 destaca a escolha adequada para a equipe que *variou menos*, que foi a 3. Tanto no 2º quanto no 5º ano 25% dos estudantes optaram corretamente. Entretanto, no 2º ano apenas 16,7% ofereceram justificativas que evidenciaram uma real compreensão da questão. Como exemplo de justificativa

adequada à questão tem-se: “*Equipe 3, quase tudinho é B... Só um A*” (S502). O trecho de fala evidencia que o estudante comparou as frequências entre as categorias e escolheu apropriadamente a equipe em que as respostas (A ou B) na equipe estiveram mais próximas do que seria a ausência de variabilidade no conjunto, ou seja, mesma resposta (A ou B) para todos na equipe.

Outras escolhas foram realizadas e a mais frequente (45,8% tanto no 2º como no 5º ano) foi a opção pela equipe 1. Para exemplificar esse tipo de escolha, tem-se a justificativa: “*Equipe 1, só botou 3 B*” (S509). Quanto aos estudantes que escolheram a equipe 2 como a que *variou menos*, o percentual foi de 25% para os dois grupos de sujeitos. Como exemplo: “*Equipe 2 porque tem pouco amarelo e pouco verde*” (S219). Finalmente, houve aqueles estudantes que não escolheram, o que foi pouco frequente (4,2% para ambos os grupos de sujeitos).

As escolhas equivocadas na Q4.2 mostraram-se mais difíceis de serem compreendidas, uma vez que quando se analisa as justificativas dos sujeitos percebe-se que, a princípio, parece existir certa incoerência. Afinal, o fato de escolher a equipe 1 porque uma das categorias tem uma frequência baixa é exatamente o que poderia justificar a escolha adequada da equipe 3, pois uma das categorias dessa equipe tem a menor frequência dentre todas as equipes. Porém, quando se faz uma associação (Tabela 15, abaixo) entre as escolhas dos estudantes em ambas as questões, percebe-se que:

**Tabela 15: Associação entre escolhas de equipe na Q4.1 e Q 4.2**

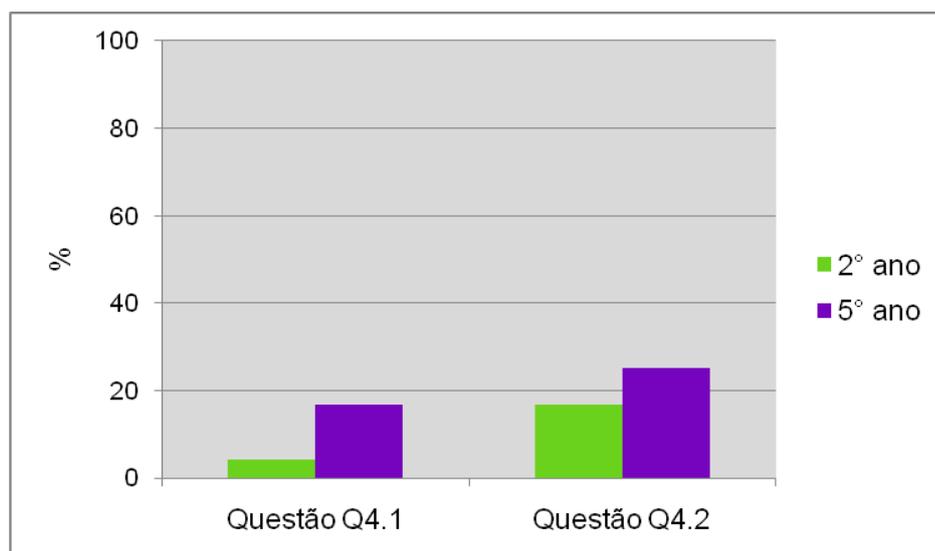
Associação entre escolhas na Q4.1 e Q4.2	%
Não escolhe / oferece respostas confusas	6,3
Q4.1 equipe 3 → Q4.2 equipe 2	25,0
Q4.1 equipe 3 → Q4.2 equipe 1	45,8
Q4.1 equipe 1 → Q4.2 equipe 3 (acerto parcial)	12,5
Q4.1 equipe 2 → Q4.2 equipe 3 (acerto)	10,4

A Tabela 15 evidencia que na verdade apenas 6,3% dos estudantes ofereceram respostas incoerentes, pois ofereceram respostas confusas como a escolha da mesma equipe para ambas as questões. Um percentual de 10,4% dos sujeitos escolheu adequadamente as equipes 2 e 3, respectivamente, respondendo corretamente as duas questões. Daqueles que acertaram apenas a Q4.2 (12,5%)

todos escolheram a equipe 1 como a que mais variou (Q4.1). Quanto aos 70,8% que indicaram com justificativas coerentes a equipe 3 na Q4.1, a maioria (45,8%) escolheu a equipe 1 na Q4.2 e 25% escolheram a equipe 2 para essa segunda questão (Q4.2).

Ao se refletir a respeito do alto percentual de estudantes que escolheu a equipe 3 na Q4.1 e equipe 1 na Q4.2 (45,8%) infere-se que provavelmente esses sujeitos pensaram nas equipes 1 e 3 como opostas, uma vez que as frequências de uma mesma categoria (resposta A ou B) mudaram bastante de uma equipe para a outra. Por outro lado, a equipe 2 por contar com a mesma frequência em ambas as categorias pode ter motivado estudantes a pensarem que a mesma não variou. A resposta do sujeito para a Q4.2 *“Equipe 2 porque teve 5 A e 5 B, não mudou nada”* (S213), pode ser um indicador de tal concepção.

Após a descrição e análise das escolhas dos estudantes como respostas para as duas questões da atividade 4, cabe nesse momento uma comparação do desenvolvimento dos mesmos em ambas, partindo das escolhas apropriadas realizadas. Abaixo, a Figura 10 expõe o percentual de acerto tanto na Q4.1 quanto na Q4.2, considerando os dois grupos de estudantes.



**Figura 10: Percentual de acertos na atividade 4 por ano de escolaridade**

A atividade 4, como a atividade anterior, mostrou-se complexa para esses estudantes. Contudo, os sujeitos dos dois grupos obtiveram um desempenho melhor

aqui. Constatou-se que tanto na Q4.1 como na Q4.2 os estudantes do 5º ano ofereceram respostas mais apropriadas para as questões do que os estudantes do 2º ano. Na Q4.1 o percentual de acerto no 2º ano foi de 4,2% e no 5º ano 16,7%. Já na Q4.2 esse número subiu para 16,7% entre os estudantes do 2º ano e no 5º ano alcançou 25%. Apesar da diferença de percentual, ao se comparar o desempenho de ambos os grupos de estudantes nas duas questões, estatisticamente, a mesma não foi significativa, nem na Q4.1 ( $t(46) = 1,42$ ,  $p = 0,163$ ), tampouco na Q4.2 ( $t(46) = 0,70$ ,  $p = 0,49$ ).

Verificou-se ainda que a segunda questão (Q4.2), a qual solicitou a escolha da *equipe que variou menos* apresentou um desempenho um pouco melhor comparada com a primeira questão (Q4.1), que requisitou a escolha da *equipe que variou mais*. Ao se relacionar as duas questões tem-se que poucos sujeitos acertaram as duas questões (10,4%, no geral). Entretanto, quem acertou a Q4.1 não errou a Q4.2. Isso confirma a dificuldade maior da Q4.1 e também indica que, na situação proposta, compreender *variou mais* foi primordial para a compreensão de *variou menos*.

### 3.4.1 Discussão da atividade 4

A atividade 4 foi elaborada por Kader e Perry (2007), contudo os autores não possuem dados empíricos a respeito da mesma, uma vez que essa não chegou a ser utilizada como instrumento de pesquisa. Desse modo, as comparações entre os estudos são impossíveis. Apesar disso, é possível e necessário comparar o desenvolvimento dos estudantes nas atividades 3 e 4, que se constituem tarefas de comparação entre conjuntos de dados, a partir do que os autores citados argumentaram a respeito das mesmas.

Kader e Perry (2007) afirmam que intuitivamente estudantes quando comparam conjuntos de dados pensam na variabilidade formando categorias, de acordo com as diferentes frequências. Assim, na atividade 3, os estudantes do 5º ano que responderam justificando *variou mais como frequências diferentes*, na verdade, pensaram que o conjunto de figurinhas compradas por Miguel foi aquele que mais variou porque todas as barras (frequências) eram diferentes, como se

formassem seis categorias, o que não aconteceu com o conjunto de Lucas, que formariam apenas duas categorias.

Na atividade 4, por outro lado, comparar apenas a quantidade de categorias não seria suficiente para resolução das questões, uma vez que os três conjuntos (equipes) continham a mesma quantidade: Duas categorias (resposta A ou B), entretanto, com frequências variadas. Tal distinção entre ambas as atividades já indica que o procedimento dos estudantes ao raciocinar nas duas atividades é diferente. Afinal, como salientam Kader e Perry (2007) na atividade 3 os estudantes que pensam na variabilidade intuitivamente transformam frequências em categorias, enquanto que na atividade 4, de acordo com os resultados aqui encontrados, os estudantes perceberam as categorias e compararam as frequências.

Os estudantes compararam as frequências na atividade 4 quando responderam equivocadamente e também nas respostas apropriadas, pois nas duas questões (*variou mais* e *variou menos*) ao fazerem inadequadamente contraposição entre as equipes 1 e 3 mantiveram o foco no fato de que nessas equipes a frequência de uma mesma categoria mudou bastante. Do mesmo modo, compararam frequências quando afirmaram que a equipe 2 não variou porque as frequências eram iguais, o que não convêm. Por outro lado, nas respostas corretas igualmente houve comparação das frequências das categorias para se concluir que a equipe 2 variou mais porque entre as categorias a menor frequência foi 5 (cinco) e a equipe 3 variou menos porque a menor frequência foi 1 (um).

Ao se relacionar a discussão de ambas as atividades realizada até aqui com a proposta de Kader e Perry (2007), explicitada anteriormente, pode-se refletir que mesmo ao se concordar com o argumento dos autores de que estudantes tendem a transformar dados quantitativos (atividade 3) em categorias, reafirma-se que o procedimento dos sujeitos ao raciocinarem a respeito da variabilidade nas duas atividades foi diferente, mesmo que intuitivamente comparassem a variabilidade entre os conjuntos pensando nos dados como qualitativos. Isso porque na atividade 3 alguns estudantes (8,3%) transformaram frequências em categorias e na atividade 4 percebeu-se as categorias, porém comparou-se as frequências.

### 3.5 Atividade 5

Finalmente a atividade 5 contou com três questões, entretanto, a última se subdivide em duas, pois entende-se que apesar de relacionadas têm objetivos diferenciados. A primeira questão (Q5.1) requeria dos estudantes a *explanação da variabilidade* de um conjunto, cujos dados quantitativos foram apresentados num gráfico de barras em uma série temporal. A seguir, a Tabela 16 expõe os tipos de respostas encontradas para a questão: “*Como o Brasil vem se saindo na conquista de medalhas?*” (*Explanação da variabilidade*).

**Tabela 16: Percentual de respostas adequadas na Q5.1 por ano de escolaridade**

Tipos de respostas	2º	5º
1) Diz não saber/ não responde	16,7	8,3
2) Explanação confusa /imaginativa	33,3	16,7
3) Explanação superficial	20,8	12,5
4) Relata pontos extremos	16,7	20,8
5) Descreve alguns anos	4,2	12,5
6) Descreve crescimento geral	8,3	29,2

Dos tipos de respostas elencados acima, considerou-se como apropriadas as três últimas (em destaque), nas quais os sujeitos relataram pontos extremos do gráfico (16,7% no 2º ano e 20,8% no 5º ano), descreveram a quantidade de medalhas que o Brasil ganhou em todos ou alguns anos (4,2% no 2º ano e 12,5% no 5º) e descreveram o crescimento geral da quantidade de medalhas no decorrer dos anos (8,3% no 2º ano e 29,2% no 5º). Como exemplificação de tais tipos de respostas, tem-se: “*Bem porque 1996 e 2008 foi 15 medalhas*” (S514), o estudante nesse caso explicita sua análise a partir do ponto máximo (maior número de medalhas conquistadas). Quando outro sujeito afirma: “*Em 1948 as olimpíadas tá fraca, em 52 tá mais ou menos bom...*” (S501), ele descreveu os anos, comparando as barras. Quanto à descrição do crescimento geral no gráfico, fica evidente na justificação: “*Muito bem porque a partir de 1996 eles foram se saindo bem até 2008*” (S513), aqui, o estudante indica que separou a série temporal em dois grandes períodos, o primeiro de poucas medalhas e o segundo de muitas.

Nos três tipos de respostas descritos anteriormente, os estudantes ao explanarem demonstraram comparar os dados representados. Desse modo, os sujeitos no 2º ano obtiveram um percentual de 29,2% na adequação das respostas, enquanto que no 5º ano esse percentual subiu bastante (62,5%), o que gerou uma diferença significativa no desempenho dos mesmos ( $t(46) = 2,41, p = 0,020$ ).

Entretanto, outros tipos de respostas surgiram, nos quais não se identificou nas afirmações comparação dos dados no gráfico. Poucos estudantes, principalmente no 5º ano afirmavam não saber ou não responderam (16,7% no 2º ano e 8,3% no 5º). A explanação confusa/imaginativa foi o tipo de resposta mais comum entre os estudantes do 2º ano (33,3%), já no 5º ano o percentual foi de apenas 16,7%. Como exemplo, tem-se: *“Ganhou mais medalhas que outros times”* (S202). O trecho de fala demonstra o equívoco do estudante, uma vez que o gráfico apresenta a variabilidade na quantidade de medalhas de um país, não de países ou “times”. Outro tipo de resposta inadequada foi a explanação superficial (20,8% no 2º ano e 12,5% no 5º), na qual não se expressou a variabilidade dos dados representados. A resposta do sujeito *“Porque ganhou muitas medalhas [de 1948 até 2008?] sim”* (S221) evidencia a superficialidade da mesma, pois a variação na quantidade de medalhas conquistadas no decorrer dos anos foi ignorada.

A segunda questão (Q5.2) solicitou por parte dos estudantes a predição a partir da tendência dos dados no gráfico, já que indicava uma série temporal. A Tabela 17, abaixo, apresenta os tipos de respostas para a questão: *“Na próxima olimpíada, em 2012, como você imagina que o Brasil vai se sair? Quantas medalhas você acha que o Brasil vai conquistar?” (Predição a partir da tendência).*

**Tabela 17: Percentual de respostas adequadas na Q5.2 por ano de escolaridade**

Tipos de respostas	2º	5º
1) Diz não saber	4,2	---
2) Diz não ser possível saber	4,2	---
3) Resposta incoerente /confusa	12,5	4,2
4) Expõe valor absurdamente alto	16,7	8,3
5) Igual a 2008	20,8	4,2
6) Superior a 2008	25	70,8
7) Inferior a 2008	16,7	12,5

Na Tabela 17, destacou-se os tipos de respostas nos quais se realizaram predições consistentes com os dados representados no gráfico. Os estudantes que afirmaram que a quantidade de medalhas seria igual a 2008 (20,8% no 2º ano e apenas 4,2% no 5º), superior a esse ano (25% no 2º ano e 70,8% no 5º) e aqueles que alegaram que seria inferior ao mesmo (16,7% no 2º ano e 12,5% no 5º) ofereceram respostas consideradas como apropriadas, pois eram acontecimentos possíveis. Para exemplificar o tipo de resposta mais frequente nos dois grupos de sujeitos (em que se oferecia um valor superior), tem-se o estudante que alegou: *“Maior que 2008! (Aponta para o 16 na escala)”* (S501). Quanto àqueles que ofereceram como resposta um valor inferior, um exemplo pode ser: *“Acho que vai ser bem...13 ou 14”* (S513).

Os tipos de respostas descritos totalizaram um percentual de adequação da resposta à questão (Q5.2) de 62,5% no 2º ano e 87,5% no 5º, o que acarretou uma diferença significativa no desempenho de ambos os grupos de estudantes ( $t(46) = 2,04$ ,  $p = 0,047$ ).

Quanto aos tipos de respostas consideradas como inadequadas, 4,2% de estudantes do 2º ano afirmou não saber e 4,2% não ser possível saber. Nos dois grupos de sujeitos houve aqueles que ofereceram respostas incoerentes/imaginativas, como *“mil, tem muita gente na olimpíada”* (S202) e ainda respostas cujos valores expressados foram absurdamente altos, por exemplo: *“200 medalhas, porque depende dos jogadores se fizerem bem”* (S213).

É importante esclarecer que maiores justificativas não foram oferecidas nessa questão (Q5.2). Entretanto, os sujeitos que ofereceram respostas apropriadas (62,5% no 2º ano e 87,5% no 5º) demonstraram habilidade em comparar os dados representados e em expressar um valor ou apontar para um local no gráfico correspondente ao tamanho da barra que acreditavam adequado, pois foi comum a referência a escala, principalmente no tipo de resposta mais frequente (maior que 2008). Os poucos estudantes que justificaram respondiam: *“Vai ser 16 medalhas, porque aqui tem 16 (escala)”* (S503).

Do trecho de fala acima, percebe-se que o valor máximo da escala do gráfico (dezesesseis), para o qual não houve frequência, pode ter induzido os estudantes a indicarem tal valor como predição para 2012, o que também explicaria o alto

percentual do tipo de resposta *maior que 2008*. Isso pode fazer com que se conclua que os estudantes que indicaram o valor 16 não fizeram predição.

Pode-se imaginar ainda, que o aumento da quantidade de medalhas é o desejo de todos e, por isso, encontramos respostas como *“Vai ser maior do que tudinho aqui (que todas as barras)”* (S509).

A terceira questão requeria a consideração da ausência de variabilidade, como isso envolve habilidades distintas a mesma foi subdividida em Q5.3a e Q5.3b. No primeiro caso (Q5.3a), solicitou-se a identificação de ausência de variabilidade ao se comparar duas barras de igual frequência no gráfico e posteriormente a proposição da ausência de variabilidade quando se comparou duas barras de frequências diferentes. A Tabela 18, abaixo, apresenta os tipos de respostas para a questão: *“O que aconteceu com a quantidade de medalhas em 1972 e 1976?”* (*identificação de ausência de variabilidade*).

**Tabela 18: Percentual de respostas adequadas na Q5.3a por ano de escolaridade**

Tipos de respostas	2º	5º
1) Diz não saber	20,8	4,2
2) Não compreende necessidade de comparar	4,2	---
3) Soma a frequência das barras	4,2	---
4) Resposta imaginativa /confusa	16,7	---
5) Diz que as barras eram menores	33,3	37,5
6) Diz que as barras estavam iguais	20,8	29,2
7) Diz a frequência das barras	---	29,1

Em destaque, na Tabela 18, encontram-se os dois tipos de respostas compreendidos aqui como apropriados, uma vez que estudantes argumentaram que as barras estavam iguais (20,8% no 2º ano e 29,2% no 5º) e outros expressaram a frequência das barras demonstrando compreensão de leitura da escala, o que aconteceu apenas com estudantes do 5º ano (29,1%). Ambos os tipos totalizaram um percentual de 20,8% de adequação das respostas no 2º ano e 58,3% no 5º ano. A diferença no desempenho dos grupos também foi significativa ( $t(46) = 2,82$ ,  $p = 0,007$ ).

Outros tipos de respostas foram oferecidos, nos quais não se verificou compreensão da questão, o que aconteceu principalmente no 2º ano que apresentou

uma diversidade maior de respostas inadequadas. Os tipos encontrados apenas no 2º ano foram formados por respostas em que não se compreendeu a necessidade de comparar as duas barras (4,2%), como o sujeito que afirma: “*A quantidade é 16, essas são 2*” (S214). Assim como o tipo em que o estudante somou as frequências de ambas as barras (4,2%), por exemplo: “*Ganhou só 2...4! Ganhou 4!*” (S215). Além do tipo em que se ofereceu uma resposta imaginativa / confusa (16,7%), como: “*Teve briga!*” (S223).

Quanto aos tipos de respostas presentes nos dois anos, que demonstraram pouco comprometimento com a proposta da questão, pôde-se constatar pelo menos dois tipos. Afirmações nas quais estudantes alegavam não saber (20,8% no 2º ano e apenas 4,2% no 5º) constituíram um desses tipos. O outro, foi aquele em que os estudantes alegaram que as barras estavam menores, nesse caso houve comparação do conjunto de dados como um todo, contudo, entre as barras indicadas não houve a comparação necessária. Esse último tipo de resposta foi o mais frequente para os dois grupos de sujeitos (33,3% no 2º ano e 37,5% no 5º). Um exemplo de resposta pode ser: “*Eu acho que essas medalhas tavam muito poucas*” (S211).

Já a Q5.3b, desenvolveu-se a partir da Q5.3a, com o objetivo de relacionar os resultados das mesmas, na busca de diferenças de compreensão entre elas. A última questão requeria que os estudantes comparassem duas barras de frequências distintas refletindo a respeito do que seria ausência de variabilidade entre as barras. A seguir, a Tabela 19 expõe os tipos de respostas encontrados para a questão: “*O que aconteceu em 2000 e 2004? Mudou? Como teria que ser se quiséssemos que não mudasse?*” (*Proposição de ausência de variabilidade*).

**Tabela 19: Percentual de respostas adequadas na Q5.3b por ano de escolaridade**

Tipos de respostas	2º	5º
1) Diz não saber/ não responde	12,5	4,2
2) Resposta confusa	8,3	8,3
3) Diz que não houve mudança	8,3	---
4) Inversão das frequências	---	4,2
5) 2004 maior que 2000	---	4,2
6) 2000 e 2004 maiores	16,7	12,5
7) 2000 e 2004 menores	20,8	12,5
8) Permanece como está	---	12,5
9) Frequência tinha que ser igual	33,3	41,7

Na Tabela 19, percebe-se uma ampla variedade de tipos de respostas, principalmente entre os estudantes do 5º ano. Em destaque, o tipo apropriado foi o mais frequente, entretanto, não se constituiu maioria para nenhum dos grupos de sujeitos. Nesse caso, os estudantes afirmaram que a frequência tinha que ser igual, o que aconteceu com 33,3% dos sujeitos no 2º ano e 41,7% no 5º. A diferença no desempenho dos grupos, constatada na porcentagem, não se revelou significativa ( $t(46) = 0,58$ ,  $p = 0,561$ ).

Dos tipos de respostas considerados inadequados tem-se aqueles presentes nos dois grupos de sujeitos e aqueles constatados apenas em um dos grupos. Alegar não saber ou não responder foi um tipo verificado em ambos os grupos (12,5% no 2º ano e apenas 4,2% no 5º). Assim como as afirmações confusas (8,3% para o 2º e 5º ano), em que situações imaginativas são criadas e os dados representados no gráfico não foram relevantes na articulação da resposta, por exemplo: *“Tinha que ficar sem fazer nenhuma natação. [Qual?] (aponta para 2004)”* (S512). Os dois últimos tipos de respostas encontrados em ambos os grupos são de certo modo contraditórios e pouco diferiram nas frequências. Afinal, argumentar que a não variação (mudança) na quantidade de medalhas dos anos indicados seria 2000 e 2004 maiores (16,7% no 2º ano e 12,5% no 5º) é o contrário de afirmar que 2000 e 2004 tinham que ser menores (20,8% no 2º ano e 12,5% no 5º).

Os demais tipos de respostas apresentados foram averiguados somente em um dos grupos. Como o tipo em que sujeitos argumentaram que não houve mudança (8,3% apenas no 2º ano), nesse caso, a partir do momento que estudantes

alegaram que a quantidade de medalhas não mudou nos dois anos não foram questionados a respeito de como teriam que ser as barras para expressar a não mudança.

No 5º ano um percentual baixo de sujeitos demonstrou compreender a ausência de variação como inversão da frequência das barras (4,2%), por exemplo, tem-se: “2000 ganhou 15 medalhas e 2004 - 2, tinha que ser 2000 - 2 e 2004 - 15” (S504). Com o trecho de fala percebe-se que o estudante foi hábil em ler a escala das barras, contudo, compreender o que seria ausência de variabilidade na situação proposta foi difícil.

Um percentual também baixo (4,2%) foi encontrado para o tipo de resposta em que se alegou que (para não mudar) a barra representando 2004 tinha que ser maior que 2000, o que não é correto.

Por fim, houve o tipo de resposta, apenas no 5º ano, no qual argumentou-se que para não mudar as barras tinham que permanecer como estavam (12,5%), como exemplo, tem-se o sujeito ao afirmar que as barras ficariam “*desse mesmo jeito*” (S510). Apesar da aparente coerência da resposta, a mesma não faz muito sentido uma vez que os mesmos sujeitos declararam anteriormente que a quantidade de medalhas em 2000 e 2004 mudou.

### **3.5.1 Discussão da atividade 5**

A atividade 5 foi proposta, a partir da reflexão de que as especificidades do tipo de variável representada no gráfico (quantitativa e temporal) gerariam possibilidades diferenciadas de análise da variabilidade do conjunto. Desse modo, escolheu-se um gráfico cujos dados apesar de mostrar crescimento, também apresentava oscilações. A opção por esse tipo de distribuição encontrou respaldo no estudo desenvolvido por Ben-Zvi (2002), no qual um gráfico com características semelhantes no que se refere à temporalidade e oscilações foi utilizado como instrumento de pesquisa.

Em tal estudo o autor objetivou investigar visões locais e globais realizadas por estudantes ao analisarem gráficos. O mesmo concluiu que ambas as visões estão integradas, sendo possível encontrar níveis diversos de entendimento dentro

das mesmas. Mesmo que o autor em tal estudo não tenha definido os diversos níveis a que faz referência, ao se considerar a proposta de Ben-Zvi (2002) e relacioná-la com os resultados encontrados no presente estudo, tornou-se possível discutir de modo mais profundo a última atividade realizada com os estudantes.

As duas primeiras questões possuíam um caráter mais global uma vez que não se questionava pontos no gráfico. Então, poder-se-ia esperar que os estudantes apresentassem respostas também de caráter mais global. Entretanto, na Q5.1 (explicação da variabilidade) estudantes demonstraram maior familiaridade com visões locais (pontuais), pois preferiram explicitar suas análises relatando pontos extremos do gráfico e descrevendo alguns anos pontualmente. Esses tipos de respostas, apesar do predomínio da visão local, foram considerados adequados, uma vez que expressaram percepção dos dados e comparação entre os mesmos.

Quando os estudantes descreveram o crescimento geral do gráfico como, por exemplo, *“primeiro ganhou pouquinho depois foi aumentando”* (S502), foi que evidenciou-se uma análise mais próxima da percepção dos dados como um todo, ou seja, a característica principal do gráfico que foi a temporalidade tornou-se mais explícita. Isso porque os estudantes demonstraram nessas respostas pelo menos a leitura de dois grandes períodos na distribuição, o primeiro de conquista de poucas medalhas e o segundo de muitas medalhas. Apesar de oscilações, também presentes no gráfico, não ter sido citadas.

Por outro lado, na Q5.2 (predição a partir da tendência) pode-se inferir que houve uma preocupação em se predizer resultados para a quantidade de medalhas da próxima olimpíada partindo da percepção de que as quantidades oscilaram no decorrer dos anos, ou seja, o crescimento não foi constante. O tipo de resposta que contemplou tal leitura foi aquela em que estudantes afirmaram que a quantidade de medalhas em 2012 seria inferior a 2008. Os estudantes apresentaram dificuldades em justificar suas afirmações, o que sujeita tais conclusões a maiores investigações.

Entretanto, ao se analisar esse tipo de resposta como um todo, sem distinguir ano de escolaridade, tem-se que dentre os 14,6% dos estudantes que predisseram o ano de 2012 como inferior a 2008 na quantidade de medalhas, 8,3% foram hábeis em descrever o crescimento geral do gráfico na questão de explicação da variabilidade (Q5.1), o tipo de resposta aqui considerado mais adequado. Essa

relação entre os tipos de repostas para ambas as questões pode indicar que maior habilidade em ler os dados do gráfico possibilitou a percepção de oscilações e predições em função da mesma. A partir de tais conjecturas, reafirma-se a necessidade de outras investigações a fim de se obter mais esclarecimentos a respeito das compreensões dos estudantes a questão.

Ainda discutindo a questão de predição, é possível comparar os resultados aqui constatados com os do estudo de Guimarães (2002). Tanto no presente estudo quanto no da autora os estudantes demonstraram habilidade em fazer *predição a partir da tendência*, Guimarães (2002) denominou extrapolação, pois em ambos os estudos a maioria dos sujeitos conseguiu oferecer respostas consistentes.

Ao se comparar a questão de predição da atividade 5 (a partir da tendência do gráfico) com as demais (Q1.3 e Q2.6) que envolveram predições a partir da moda. Dos estudantes, 75% responderam adequadamente a Q5.2, 10,4% acertaram a Q1.3 e 35,4% acertaram a Q2.6. Predizer a partir da tendência do gráfico foi mais fácil do que predizer a partir da moda. Entretanto, foi curioso observar que todos os estudantes que acertaram a Q1.3 obtiveram um bom desempenho na Q5.2, mas nem todos os que acertaram a Q2.6 acertaram a Q5.2.

Comparando o desempenho dos estudantes na Q1.3 e Q 2.6, constatou-se que também que houve estudantes que apresentaram respostas apropriadas em uma questão e na outra não. Tais dados indicam que a Q2.6 foi uma questão que suscitou reflexões divergentes, uma vez que acredita-se que esses estudantes sentiram maior liberdade na Q2.6 para fazer predições de acordo com suas crenças. Afinal, como já ressaltado na discussão da atividade 2, para estudantes dos dois grupos de escolaridade mais relevante em suas predições que compreender os dados e fazer predições a partir dos mesmos foi a oportunidade de transformá-los de acordo com suas crenças. Isso ocorreu com aqueles estudantes que justificaram que quem leu menos gostaria de receber um livro de presente.

A terceira questão da atividade 5, que foi subdividida em duas, diferente das demais explicitava dois pontos para serem comparados. A identificação de ausência de variabilidade pode ser diretamente comparada com o estudo de Guimarães (2002). A autora percebeu que estudantes (na faixa etária de nove anos de idade) não demonstraram dificuldades em tal questão. Aqui, para os estudantes do 2º ano

essa foi uma tarefa complexa, entretanto, no 5º ano a maioria (58,3%) ofereceu respostas adequadas.

Quanto à proposição de ausência de variabilidade, não se obteve na literatura utilizada um parâmetro, pois apesar da questão Q1.5 (atividade 1) envolver ausência de variabilidade enquanto aspecto, os estudantes não foram solicitados a propô-la. Na Q1.5 a redução já foi estipulada na questão e os sujeitos precisariam aceitar tal mudança e distribuir a frequência da barra que ficaria nula entre as demais.

Ao se associar as questões Q5.3a Q5.3b observou-se uma relação direta entre as mesmas. Isto é, foram poucos os sujeitos que responderam adequadamente a uma e a outra não, pois 47,9% dos estudantes se equivocaram em ambas e 25% responderam corretamente as duas. Esses dados não foram ocasionais, uma vez que se verificou uma diferença significativa na comparação do desempenho dos estudantes nas questões ( $\chi^2 (1) = 8,83; p = 0,003$ ).

A partir do que foi discutido na atividade 5 e considerando o argumento de Ben-Zvi (2002) a respeito da integração entre as visões locais e globais ao se analisar um conjunto de dados, concorda-se que ambas as visões realmente estão integradas, uma vez que ao se propor uma questão de caráter global os estudantes podem refletir sobre os dados partindo de uma visão local e ainda assim apresentarem respostas adequadas, como ocorreu na Q5.1 (explicação da variabilidade), conforme discussão acima.

Ao se finalizar a exposição e discussão dos resultados de cada atividade, ressalta-se que mudanças provocadas na sequência das mesmas (como explicitado no capítulo anterior), a fim de se investigar a influência da ordem das atividades 1 e 2, além da organização dos dados na atividade 4, não acarretaram diferenças expressivas de desempenho entre os sujeitos, ou seja, essas não foram significativas. Isso significa que iniciar o teste com tarefas de interpretação (atividade 1) ou de representação (atividade 2) não influenciou a compreensão dos estudantes a respeito dos aspectos da variabilidade. Assim como não influenciou na compreensão a mudança na organização dos dados realizada na atividade 4.

### 3.6 Compreensão de aspectos da variabilidade

Nesse momento, cabe uma apresentação sintetizada dos resultados e discussões realizados ao longo do capítulo, a fim de se obter uma visão mais ampla das compreensões demonstradas pelos estudantes a respeito do conceito de variabilidade estatística, explorado no presente estudo enquanto aspectos.

É importante tal síntese para que se possa realizar uma análise preliminar na tentativa de se obter respostas para os questionamentos realizados no Capítulo 1, a respeito da influência da escolaridade na compreensão de aspectos da variabilidade e da possibilidade de existir uma gradação entre os aspectos da variabilidade. A seguir, a Tabela 20 expõe o desempenho dos estudantes do 2º e 5º ano em dezesseis (16) questões:

**Tabela 20: Percentual de respostas adequadas para todas as questões por ano**

Questões	2º	5º	T-test (sig.)
Q1.1 ( <i>Explicação</i> )	29,2	79,1	X
Q1.2 ( <i>Localização de ponto máximo</i> )	29,2	66,7	X
Q1.3 ( <i>Predição: moda</i> )	4,2	16,7	---
Q1.4 ( <i>Comparação entre pontos</i> )	---	37,5	X
Q2.1 ( <i>Representação de dados</i> )	33,3	70,8	X
Q2.2 ( <i>Explicação</i> )	79,2	95,8	---
Q2.4 ( <i>Representação de frequência nula</i> )	66,7	70,8	---
Q2.5 ( <i>Representação de acréscimos</i> )	62,5	83,3	---
Q2.6 ( <i>Predição: ponto máximo</i> )	20,8	50,0	X
Q3.1 ( <i>Comparação entre conjuntos</i> )	---	---	---
Q4.1 ( <i>Comparação entre conjuntos</i> )	4,2	16,7	---
Q4.2 ( <i>Comparação entre conjuntos</i> )	16,7	25,0	---
Q5.1 ( <i>Explicação</i> )	29,2	62,5	X
Q5.2 ( <i>Predição: tendência</i> )	62,5	87,5	X
Q5.3a ( <i>Identificação de ausência</i> )	20,8	58,3	X
Q5.3b ( <i>Proposição de ausência</i> )	33,3	41,7	---

A respeito da influência da escolaridade, é possível perceber na Tabela 20 que os estudantes do 5º ano apresentaram um avanço maior nas compreensões da

variabilidade de dados do que os estudantes do 2º ano em todas as questões, sendo significativa a diferença no desempenho em pelo menos metade das questões.

Um olhar mais atento às questões mostra que os estudantes do 5º ano, apresentaram maiores dificuldades (desempenho inferior a 50%) nas questões que requeriam comparação entre conjuntos (Q3.1, Q4.1 e Q4.2) e nas questões de predição a partir da moda e comparação entre pontos (ambas da atividade 1) e a questão de proposição de ausência de variabilidade, na atividade 5.

Quanto aos estudantes do 2º ano, as únicas questões que eles obtiveram mais de 50% de acerto foram as de explanação (Q2.2), representação da frequência nula (Q2.4), Representação de acréscimos (Q2.5) e Predição a partir da tendência (Q5.2).

Assim, para estudantes do 2º ano tarefas simples para pelo menos metade dos estudantes do 5º ano revelaram-se complexas.

A diferença significativa de desempenho entre sujeitos do 2º e 5º ano em oito das dezesseis questões da Tabela 20 permite a reflexão de que os estudantes estão presenciando, ao longo dos primeiros anos de escolarização, situações de ensino que tem possibilitado avanços na compreensão dos mesmos. Entretanto, é importante salientar que os distintos aspectos da variabilidade não vêm sendo ensinados de modo intencional. Afinal, na revisão da literatura não se encontrou evidências desse fato e a discussão a respeito do conceito de variabilidade é nova, principalmente nos anos iniciais de ensino. Além disso, o Livro Didático apesar de trabalhar a interpretação de gráficos o faz de maneira a explorar a análise pontual, principalmente de pontos extremos, conforme constatação de Guimarães, Gitirana, Cavalcanti e Marques (2006).

O conhecimento de que o Livro Didático dos anos iniciais explora mais o tipo de análise pontual nas atividades com gráficos, pode fazer com que se suponha que estudantes, pelo menos do 5º ano, obteriam um desempenho melhor na questão de localização de ponto máximo, porém o mesmo não ultrapassou os 67%. O quadro se agrava quando se relaciona tal questão com as de predição a partir do ponto máximo e/ou moda, nas quais o desempenho não superou 16,7% (Q1.3) e 50% (Q2.6). Isso faz com que se acredite que a maioria dos estudantes (do 5º ano) tem

presenciado situações de ensino com tarefas de localização do ponto máximo, entretanto, não estão sendo estimulados a refletirem a respeito do mesmo.

Quanto às questões de predição citadas, apesar do baixo percentual de acerto, verificou-se que estudantes do 2º ano foram capazes de solucioná-las de modo adequado, sem que tenham vivenciado na escola muitas situações de ensino com gráficos. Acredita-se que as dificuldades podem estar relacionadas também à ausência de discussões diferenciadas na escola, que englobe tipos diferentes de análise de gráficos.

Os dados representados na Tabela 20, também suscitam algumas reflexões a respeito de uma possível gradação entre os distintos aspectos da variabilidade, explorados nas questões. Aqui, a gradação dos aspectos se deterá aos quatro aspectos mais complexos para os estudantes.

Na tentativa de fazer tal gradação tomou-se por base o desempenho dos estudantes principalmente do 5º ano. Isso porque esses sujeitos apesar de contarem com uma relativa experiência escolar, comparados com os do 2º ano, ainda assim apresentaram dificuldades maiores com pelo menos quatro aspectos distintos, que numa ordem decrescente de complexidade foram:

- 1º. Comparação entre conjuntos;
- 2º. Predição a partir da moda;
- 3º. Comparação entre pontos com quantificação da variação;
- 4º. Proposição de ausência de variabilidade.

É importante salientar que tais aspectos podem ser considerados complexos de acordo com a situação em que foram apresentados, pois um aspecto pode se mostrar de fácil compreensão numa atividade e na outra não. Um exemplo pode ser a explanação da variabilidade que foi explorada em três questões (Q1.1, Q2.2 e Q5.1). Ao se analisar a Tabela 20 constata-se que a maioria dos estudantes do 5º respondeu adequadamente nos três casos, ou seja, percebeu-se uma tendência. Entretanto, o mesmo não ocorreu com estudantes do 2º ano, pois a maioria obteve um bom desempenho em uma apenas (Q2.2), nas demais a frequência não chegou a 30%.

Ainda é possível argumentar que a “complexidade” desses quatro aspectos acontece provavelmente pela ausência de discussões no ensino a respeito de diferentes situações em que há variabilidade dos dados. Afinal, esses aspectos não podem ser considerados complexos do ponto de vista cognitivo, uma vez que estudantes do 2º ano ofereceram respostas adequadas a todos eles. A exceção é o aspecto da comparação entre pontos com quantificação da variação, o qual não foi compreendido por esse grupo de estudantes, porém, estudos anteriores (WATSON e KELLY, 2002; GUIMARÃES, 2002; ARCOVERDE et al, 2004) já evidenciaram que curtas intervenções permitiram avanços na compreensão do mesmo

Diante dos resultados encontrados e discussões suscitadas a partir dos mesmos, nesse momento cabe algumas considerações a respeito do *SOLO*, enquanto instrumento de análise, o qual foi utilizado em diversos estudos (WATSON, 2009; WATSON e KELLY, 2002a e outros), como já explicitado no Capítulo 1.

Tal qual se apresentou nos estudos citados, o *SOLO* classificou as respostas dos estudantes de acordo com a atividade (WATSON, 2009) ou englobando aspectos distintos como uma mesma habilidade (WATSON e KELLY, 2002a). Nesse último estudo os níveis para as respostas dos estudantes foram criados a partir de temas como: 1) *Observação e criação de representações*; 2) *uso de dados para interpretação e predição*. Entretanto, consideramos que dentro de um mesmo tema existem aspectos diferentes da variabilidade.

Aqui, acredita-se que uma análise mais profunda das compreensões do conceito de variabilidade pelos estudantes requer a distinção entre os diversos aspectos do conceito e análise sistemática dos mesmos. Desse modo, optou-se por não utilizar o *SOLO*, uma vez que tal instrumento não se mostrou adequado para os objetivos delineados no presente estudo. Afinal, pretendia-se compreender com profundidade o desenvolvimento de diferentes aspectos da variabilidade e não reduzi-los em função de temas que os englobassem.

Assim, a análise realizada foi pensada partindo das questões e posterior relação entre as mesmas. O que se pode concluir a partir dos resultados encontrados e discussões realizadas é que a diversidade de aspectos não foi o único fator que acarretou compreensões distintas. O tipo de atividade também

influenciou nas respostas dos estudantes, como pode ser constatado ao se analisar na Tabela 20, acima, a questão de explanação da variabilidade, na qual a maioria dos estudantes do 2º ano ofereceu respostas adequadas na Q2.1, contudo nas demais isso não ocorreu. Nas questões de predição também houve grande variação no desempenho dos estudantes, de acordo com a situação em que foi proposta.

Tais conclusões permitem que se reflita a respeito da importância que a proposta de Vergnaud (1996) referente aos três conjuntos que formam um conceito assume no presente estudo. Isso porque confirmou-se o que já era esperado, isto é, que a compreensão de um aspecto da variabilidade não significou a compreensão dos demais. Assim, tem-se que dentro da diversidade de aspectos explorados no atual estudo, os estudantes compreenderam melhor alguns e outros não.

Ao se relacionar os resultados aqui encontrados com a discussão estabelecida por Vergnaud (1996), pode-se afirmar que o conceito de variabilidade mantém uma estreita ligação com diversos outros e que compreender a variabilidade numa determinada situação não garante que a mesma seja entendida numa outra. Por exemplo, tem-se o aspecto da comparação entre conjuntos de dados, quando na atividade 3 não houve acerto, já na atividade 4 houve, apesar do baixo percentual. Como já explicitado antes, os dados na atividade 3 eram quantitativos e na atividade 4 qualitativos. A solução da tarefa na atividade 3 requeria a compreensão de conceitos como média e desvio padrão, esse último bastante complexo para os estudantes da faixa etária considerada. Além disso, essa tarefa mostrou-se difícil até mesmo para estudantes de graduação (LOOSEN et al, 1985).

Conclui-se que a resolução de problemas envolvendo o conceito de variabilidade estatística requer a compreensão de uma diversidade de situações em que a mesma pode ser encontrada, assim como o conhecimento de formas de representá-la e invariantes (propriedades) que a caracteriza.

No capítulo seguinte os diferentes aspectos da variabilidade são retomados com a finalidade de se obter maiores esclarecimentos a respeito das compreensões apresentadas pelos estudantes quanto ao conceito.

## CAPÍTULO 4 – ANÁLISE MULTIDIMENSIONAL

Nesse capítulo, as questões anteriormente discutidas são retomadas a fim de se investigar possíveis correlações entre os diferentes aspectos da variabilidade investigados. Para isso, realizou-se uma análise multidimensional envolvendo 15 (quinze) questões e o nível de escolaridade. Essas 15 questões foram discutidas separadamente no Capítulo 3 e agora serão analisadas conjuntamente. Assim buscou-se analisar correlações entre as questões que envolviam: localização de ponto máximo; predição a partir da moda, do ponto máximo e da tendência; comparação entre pontos; explanação da variabilidade; representação de variabilidade, representação de frequência nula e de acréscimos; além de identificação e proposição de ausência de variabilidade.

Para a análise multidimensional pretendida, utilizou-se o escalonamento multidimensional (Multidimensional Scaling - MDS).

De acordo com Oliver (1998) o objetivo principal do MDS é a construção de um espaço métrico com o menor número de dimensões possíveis, que permita representar as proximidades entre os elementos com o maior grau de fidelidade. A autora afirma que para compreender essa técnica é necessário o conhecimento de conceitos como proximidade (valor que assume a similaridade ou distância que existe entre dois objetos) e dimensionalidade (número de dimensões necessárias para representar um conjunto de objetos a partir dos índices de proximidade obtidos).

Aqui, utilizou-se o modelo euclidiano, definido por Oliver (1998) como a raiz quadrada da soma das diferenças entre os elementos ao quadrado. Para isso, o algoritmo ALSCAL (Algorithmic Scaling) foi empregado e incorpora os índices do “ajuste”. Com o ALSCAL, índices como Stress e RSQ (correlação múltipla ao quadrado) são fornecidos. O primeiro indica a qualidade com que a configuração reproduz a informação original. O zero é considerado um “ajuste” perfeito, enquanto que valores superiores a 0,2 correspondem a maus “ajustes” (PORCAR e ESCALANTE, 2009). Já o RSQ é um índice de porcentagem da variância explicada para a configuração obtida. Ele corresponde ao quadrado da correlação entre os dados e as distâncias. Um bom ajuste implica valores de RSQ próximo a 1 (um), isto

é, quanto mais próximo de 1 (um) o valor do RSQ mais ajustada está a configuração.

Aplicada a técnica do escalonamento multidimensional com as questões para os sujeitos do 2º ano, em duas dimensões, verificou-se que o índice de Stress (0,15240) revelou qualidade no ajuste. Quanto ao índice de RSQ (0,90309), o valor em termos percentuais foi de 90,309% (também utilizado), o que numa escala, pode ser compreendido como um valor bastante alto.

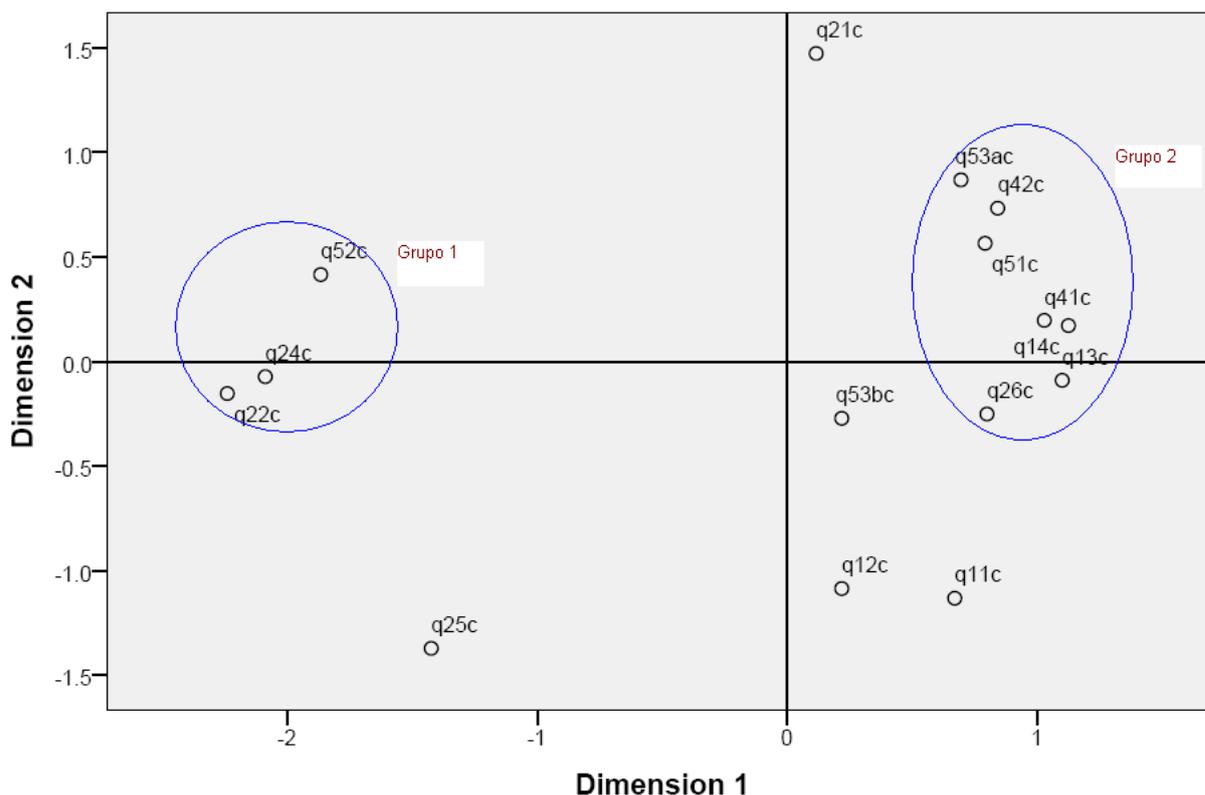
Cabe nesse momento um detalhamento maior dos valores encontrados para as dimensões 1 e 2, no 2º ano, em função das questões. A seguir, o Quadro 5 apresenta tais valores.

**Quadro 5: Estímulos (questões) por dimensões 1 e 2 no 2º ano**

Questões	Dimensão 1	Dimensão 2
q11c - Q1.1 ( <i>Explicação</i> )	0,6715	-1,1318
q12c - Q1.2 ( <i>Localização de ponto máximo</i> )	0,2193	-1,0855
q13c - Q1.3 ( <i>Predição: moda</i> )	1,1018	-0,0898
q14c - Q1.4 ( <i>Comparação entre pontos</i> )	1,1267	0,1720
q21c - Q2.1 ( <i>Representação de dados</i> )	0,1170	1,4732
q22c - Q2.2 ( <i>Explicação</i> )	-2,2404	-0,1529
q24c - Q2.4 ( <i>Representação de frequência nula</i> )	-2,0882	-0,0722
q25c - Q2.5 ( <i>Representação de acréscimos</i> )	-1,4239	-1,3714
q26c - Q2.6 ( <i>Predição: ponto máximo</i> )	0,8006	-0,2509
q41c - Q4.1 ( <i>Comparação entre conjuntos</i> )	1,0301	0,1970
q42c - Q4.2 ( <i>Comparação entre conjuntos</i> )	0,8434	0,7333
q51c - Q5.1 ( <i>Explicação</i> )	0,7930	0,5655
q52c - Q5.2 ( <i>Predição: tendência</i> )	-1,8664	0,4160
q53ac - Q5.3a ( <i>Identificação de ausência</i> )	0,6965	0,8688
q53bc - Q5.3b ( <i>Proposição de ausência</i> )	0,2192	-0,2714

A partir das informações do Quadro 5, acima, o gráfico apresentado na Figura 11 (a seguir) é gerado. O mesmo expõe a distribuição espacial das 15 (quinze) questões em duas dimensões, no modelo euclidiano, entre os sujeitos do 2º ano. Vale salientar que a proximidade entre as questões expressa o comportamento dos sujeitos para aquela questão, isso significa que há uma comparação do comportamento de todos os sujeitos do 2º ano para uma determinada questão em

relação às demais. Assim, a proximidade entre pontos no gráfico revela que os mesmos sujeitos que responderam adequadamente ou não a uma questão, fizeram o mesmo com aquelas questões correlacionadas.



**Figura 11: Distâncias euclidianas bidimensional para as questões no 2º ano**

A Figura 11 apresenta a dimensão 1 no eixo horizontal, que distingue as questões em que a maioria dos estudantes do 2º ano obteve um bom desempenho (à esquerda) e as questões em que a maioria ofereceu respostas inadequadas (à direita).

Apesar do gráfico evidenciar tal distinção do desempenho dos estudantes, é importante esclarecer que a correlação entre as questões não será definida simplesmente pelo percentual de adequação na resposta para a questão, pois se assim fosse as questões Q2.5 e Q5.2 (pontos q25c e q52c na Figura 11, respectivamente) estariam bem próximas, uma vez que apresentaram o mesmo percentual (62,5%). Entretanto, não é o que ocorre. Percebe-se que o ponto q25c está isolado e a distância entre os pontos q25c e q52c deve-se mais ao fato de que aqueles estudantes que responderam adequadamente a Q2.5 não foram os mesmos que responderam de modo adequado a Q5.2. Por outro lado, os pontos q52c, q22c,

q24c estão mais próximos (Grupo 1), o que revela maior correlação entre essas questões, no sentido de que os mesmos estudantes que compreenderam uma dessas questões, compreenderam as demais associadas. Já do lado direito encontram-se aquelas questões em que os estudantes do 2º ano obtiveram um desempenho ruim e em destaque percebe-se um agrupamento de questões (Grupo 2) que geraram grandes dificuldades para esses estudantes. Mais uma vez salienta-se que a maior proximidade entre as questões (q41c e q 14c, por exemplo), significa que os mesmos estudantes apresentaram aquele desempenho.

Ressalta-se que o agrupamento das questões não significa habilidades parecidas para resolução das questões, pois se assim fosse os pontos q41c e q42c estariam bem próximos uma vez que correspondem às questões de comparação entre conjuntos de dados categóricos. Do mesmo modo que as questões envolvendo explanação da variabilidade (q11c, q22c e q51c) também não estão próximas.

Na dimensão 2 (eixo vertical) requer uma subjetividade maior na análise. Com a configuração das questões em tal eixo infere-se que na parte inferior aparecem aquelas questões em que os estudantes apresentaram respostas cujas justificativas se relacionaram bastante com suas vivências ou imaginação, como foi o caso das questões Q1.1 e Q1.2 (pontos q11c e q12c), que inclusive estão próximos. No capítulo anterior, em que se descreveu os resultados encontrados para as questões, discute-se a tendência entre os estudantes do 2º ano em criar histórias imaginativas em ambas as questões, nas quais a análise dos dados representados no gráfico não foi predominante nas respostas oferecidas (páginas 66). Quanto ao ponto q25c, que está na posição inferior do eixo vertical e apresentou um valor próximo (para a dimensão 2, conforme Quadro 5) aos dos pontos q11c e q 12c, pode-se afirmar que de fato na questão Q2.5 (ponto q25c), que requeria o acréscimo de um livro para cada criança representada na figura criada pelos estudantes, para solucionar a questão de modo adequado a distribuição dos cartões entre as crianças representadas resolvia o problema proposto, o que significa que o estudante poderia responder adequadamente a questão sem que estivesse necessariamente envolvido com os dados. Assim, pode-se concluir que na configuração da Figura 11 a

dimensão 2 expressa uma relação das respostas com o maior ou menor envolvimento dos estudantes com os dados representados.

Se na parte inferior se localizaram as questões que evidenciaram menor envolvimento dos estudantes com os dados representados, seja porque não foi exigido para solução da questão, seja porque os estudantes não foram hábeis em refletir a respeito dos dados representados, na parte superior do eixo vertical encontra-se isolada a questão Q2.1 (ponto q21c), que foi a questão que mais demandou conhecimento dos dados apresentados para que os mesmos fossem representados. Nessa questão, independente de adequação na figura criada para representar a quantidade de livros lidos por crianças, os estudantes precisaram refletir a respeito dos dados. Desse modo, os sujeitos apresentaram uma compreensão diferenciada para essa questão, o que pode ser evidenciado pela ausência de respostas imaginativas ou baseadas em vivências para essa questão.

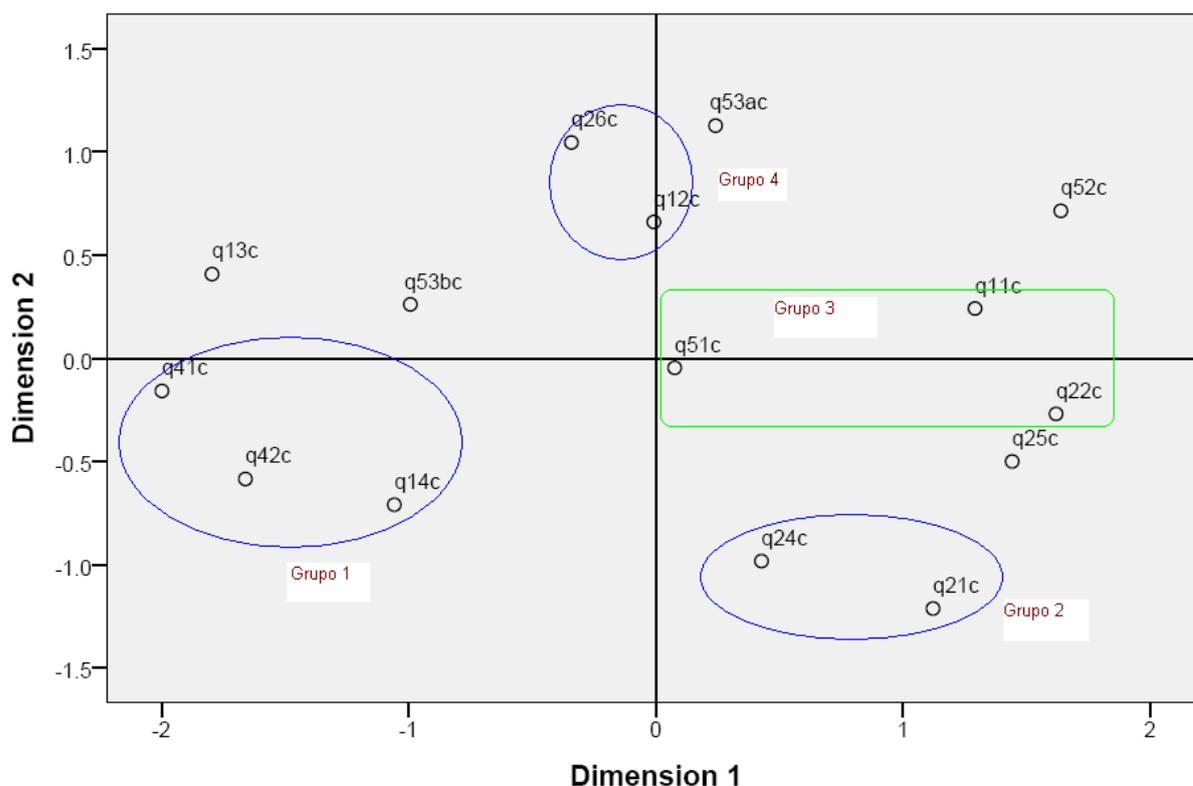
Assim como no 2º ano, no 5º também buscou-se analisar as 15 (quinze) questões já destacadas no modelo euclidiano para duas dimensões. Constatou-se que o índice de Stress encontrado ( $\text{Stress}=0,13387$ ) foi melhor que o do 2º ano, uma vez que esteve mais próximo de zero, “ajuste perfeito”. Entretanto, o índice de RSQ (0,89664) foi um pouco menor que o do 2º ano, em termos percentuais foi 89,664%, o que ainda assim é um valor alto.

Quanto aos valores para as dimensões 1 e 2 das questões, obteve-se os números apresentados no Quadro 6, que segue.

**Quadro 6: Estímulos (questões) por dimensões 1 e 2 no 5º ano**

Questões	Dimensão 1	Dimensão 2
q11c - Q1.1 ( <i>Explicação</i> )	1,2916	0,2416
q12c - Q1.2 ( <i>Localização de ponto máximo</i> )	-0,0081	0,6608
q13c - Q1.3 ( <i>Predição: moda</i> )	-1,7963	0,4082
q14c - Q1.4 ( <i>Comparação entre pontos</i> )	-1,0585	-0,7090
q21c - Q2.1 ( <i>Representação de dados</i> )	1,1215	-1,2121
q22c - Q2.2 ( <i>Explicação</i> )	1,6196	-0,2689
q24c - Q2.4 ( <i>Representação de frequência nula</i> )	0,4273	-0,9821
q25c - Q2.5 ( <i>Representação de acréscimos</i> )	1,4421	-0,4996
q26c - Q2.6 ( <i>Predição: ponto máximo</i> )	-0,3418	1,0456
q41c - Q4.1 ( <i>Comparação entre conjuntos</i> )	-1,9991	-0,1580
q42c - Q4.2 ( <i>Comparação entre conjuntos</i> )	-1,6612	-0,5844
q51c - Q5.1 ( <i>Explicação</i> )	0,0760	-0,0457
q52c - Q5.2 ( <i>Predição: tendência</i> )	1,6388	0,7147
q53ac - Q5.3a ( <i>Identificação de ausência</i> )	0,2417	1,1274
q53bc - Q5.3b ( <i>Proposição de ausência</i> )	-0,9937	0,2615

Para os valores expressos no Quadro 6, o seguinte gráfico (Figura 12) foi configurado, o qual mostra a qualidade da relação entre as questões.



**Figura 12: Distâncias euclidianas bidimensional para as questões no 5º ano**

Primeiramente, percebe-se que a relação entre as questões foi bem diferente, ao se comparar o desempenho dos estudantes do 2º e 5º ano. No 5º ano as questões estão mais dispersas no espaço do que no 2º ano. Isso já revela que no 5º ano, os estudantes compreenderam as questões de um modo distinto. É provável que as características peculiares de cada questão tenham acarretado compreensões tão diferentes entre os estudantes, que colocaram uma distância maior entre as mesmas. Isso indica também que foi mais difícil no 5º ano encontrar os mesmos estudantes respondendo adequadamente ou não a um grupo de questões, que estariam mais correlacionadas se fosse o caso. Afinal, os aspectos da variabilidade por serem bastante distintos suscitaram compreensões também diferenciadas.

Diferente do gráfico para os estudantes do 2º ano (Figura 11), no 5º ano a configuração, apesar de ainda distinguir as questões em que a maioria apresentou um bom desempenho a ordem foi invertida. Isso significa que do lado esquerdo concentraram-se as questões mais complexas para esses estudantes e do lado direito as questões em que houve uma compreensão maior.

Quanto à dimensão 2 a lógica não foi a mesma, o que é esperado, uma vez que respostas mais vinculadas à imaginação e dados do cotidiano foram mais frequentes entre os estudantes do 2º ano. A análise na perspectiva da dimensão 2 revela, na verdade, que as questões apesar da dispersão podem ser agrupadas conforme os grupos numerados na Figura 12.

Os grupos delimitados para a análise das questões com os sujeitos do 5º ano, faz com que se reflita que a dimensão 2 (eixo vertical) procurou distinguir as questões a partir de características comuns às mesmas. Assim, de acordo com a Figura 12, o Grupo 1 engloba questões bastante complexas para esses estudantes, que estavam relacionadas com: comparação entre conjuntos de dados e quantificação da variação na comparação entre dois pontos. A questão Q1.3 (ponto q13c), apesar de mostrar-se de difícil compreensão também não foi posta em associação no Grupo 1, pois diferente das demais questões desse grupo, requeria a identificação do ponto máximo e moda para sua resolução. O reconhecimento desse fator, enquanto diferenciador da questão Q1.3, é essencial para que se compreenda a aproximação de tal questão com a Q1.2 (de identificação do ponto máximo), pertencente ao Grupo 4. Ambas as questões na perspectiva da dimensão 2 apresentam valores parecidos, como pode ser evidenciado no Quadro 6, exposto anteriormente.

Quanto a questão Q5.3b (ponto q53b) que está localizada próxima às questões Q1.2 e Q1.3, pode-se afirmar que essa foi uma questão que gerou dificuldades para ambos os grupos e apesar de explorar ausência de variabilidade, a Q5.3b de fato compara dois pontos no gráfico, ou seja, estava mais comprometida com a visão pontual do que global dos dados representados assim como a questão Q1.2.

O Grupo 2, também na parte inferior da dimensão 2, encontram-se questões relacionadas com a *representação da variabilidade* (Q2.1) e da frequência nula (Q2.4). A questão Q2.5, apesar de também envolver representação encontra-se mais próxima da Q2.2, que é uma questão de explanação da variabilidade. A correlação entre ambas as questões indica que os mesmos estudantes que responderam adequadamente uma questão fez o mesmo com a outra.

No Grupo 3 estão as questões de *explanação da variabilidade*. A questão Q5.1 mesmo estando mais distante das outras (na perspectiva da dimensão 1), quando se observa a dimensão 2 ela esta muito próxima.

Já o Grupo 4 localiza-se na parte superior da configuração e diferente dos demais grupos requer o *conhecimento do ponto máximo* para solução da questão, seja para a habilidade de identificação, seja para o tipo de predição solicitado. Percebe-se que a questão Q5.3a (ponto q53ac) encontra-se próximo ao Grupo 4, indicando que os estudantes se comportaram de modo semelhantes nas três questões. Infere-se que a aproximação da questão de identificação de ausência de variabilidade com aquelas que partiram do ponto máximo justifica-se pelo caráter pontual das três questões. Afinal, na questão Q5.3a compara-se dois pontos (de ausência de variação na frequência).

Percebe-se que na análise do comportamento dos estudantes do 5º ano nas questões investigadas destacou-se as visões locais e globais que os estudantes podem estar realizando dos dados apresentados aos mesmos.

É importante esclarecer que dentre as questões há aquelas cujas formulações requerem uma visão mais pontual (por exemplo, localização de ponto máximo, comparação entre dois pontos) e outras que requerem visão mais global dos dados (como explanação da variabilidade, predição a partir da tendência). Contudo, concorda-se com a proposta de Ben-Zvi (2002), o qual afirma que ambas as visões estão integradas, sendo possível encontrar níveis de entendimento dentro das duas visões, conforme foi explicitado no Capítulo 1.

Assim, verificou-se que mesmo que a questão direcionasse a leitura do estudante para um tipo de visão, as visões por estarem integradas e possuírem níveis de entendimento provocaram aproximações entre questões que solicitavam visões diferentes, como foi o caso da questão Q1.4 (de comparação entre pontos com quantificação da variação) que requeria visão mais local, porém está no mesmo grupo (Grupo 1) das questões Q4.1 e Q4.2 (de comparação entre conjuntos), que requeriam visão mais global. As questões Q1.2 e Q5.2 ainda, apesar da diferença de desempenho dos estudantes para tais questões, possuem valores aproximados no que se refere à dimensão 2, porém exploram visões distintas. A primeira de

localização de ponto máximo, visão local e a segunda de predição a partir da tendência, visão global.

A partir da discussão realizada no presente capítulo, a respeito da análise multidimensional das questões, conclui-se primeiramente que o comportamento dos estudantes do 2º e 5º ano na resolução das questões foi bem distinto.

Além da adequação ou não da resposta oferecida, o que diferenciou a compreensão dos estudantes do 2º ano foi a possibilidade de representar dados oferecidos ou de imaginar dados vivenciados em outros momentos. Quando os estudantes não se envolveram com os dados, seja por falta de habilidade com análise dos dados, seja porque a questão não exigiu uma exploração maior dos dados, os estudantes justificaram com imaginação ou vivências.

Quanto aos estudantes do 5º ano, verificou-se uma maior dispersão entre as questões, evidenciando que o fato de um estudante responder adequadamente a uma questão não significou necessariamente a resolução adequada de uma outra. Com isso, percebe-se que a distinção entre os aspectos da variabilidade foi ressaltada entre os estudantes do 5º ano, o que é compreensível uma vez que a grande concentração de pontos entre os estudantes do 2º ano, no Grupo 2 da Figura 11, corresponde, na verdade, a concentração de estudantes expressando ausência de compreensão daquelas questões.

Assim, o escalonamento multidimensional realizado no presente capítulo confirma a maior compreensão dos estudantes do 5º ano dos aspectos da variabilidade investigados, o que se deve provavelmente não apenas a escolaridade maior desses sujeitos, como também a maturidade dos mesmos. Afinal, respostas imaginativas foram mais comuns entre os estudantes do 2º ano. Tal análise confirmou ainda que os aspectos da variabilidade podem ser agrupados por habilidades envolvidas para resolução das questões, o que evidenciou-se entre os estudantes do 5º ano, pois apesar das distinções entre os aspectos, fatores como tipo de visão dos dados (local ou global), ou ainda situação envolvida (como representação ou explanação) influenciaram a configuração multidimensional reproduzida.

## CAPÍTULO 5 – CONSIDERAÇÕES FINAIS

Na Estatística, o conceito de variabilidade é essencial, uma vez que não haveria necessidade de se fazer estatística se os dados não variassem. Apesar da centralidade de tal conceito, poucos estudos se preocuparam em investigar de modo sistemático as compreensões apresentadas por estudantes a respeito do mesmo, conforme ressaltado no Capítulo 1. Esse quadro se agrava quando o foco são os estudantes dos anos iniciais do Ensino Fundamental. Conhecer o que estudantes dos anos iniciais compreendem a respeito da variabilidade é fundamental, uma vez o ensino da Estatística é uma recomendação do currículo escolar brasileiro já nesses anos. Um entendimento profundo do conceito de variabilidade requer a exploração de seus componentes desde os anos iniciais (GARFIELD e BEN-ZVI, 2005).

A partir dos argumentos encontrados na literatura atual delimitou-se os objetivos geral e específicos do presente estudo. Assim, o objetivo principal foi investigar as compreensões apresentadas por estudantes de 2º e 5º anos do Ensino Fundamental a respeito do conceito de variabilidade, mais especificamente:

- A compreensão de estudantes de 2º e 5º anos de escolarização referente aos diferentes aspectos do conceito de variabilidade, tais como: representação da variabilidade, explanação da mesma, comparação entre conjuntos de dados, predição de resultados, quantificação da variação;
- A existência de uma gradação na compreensão desses estudantes quanto aos diferentes aspectos do conceito de variabilidade.

Para isso, propôs-se cinco atividades explorando diferentes aspectos da variabilidade a 48 (quarenta e oito) sujeitos de ambos os anos, com os quais foi realizada uma entrevista clínico-piagetiana, pois as justificativas dadas pelos estudantes às questões propostas foi o interesse maior na pesquisa.

Uma vez que foram identificados na literatura diferentes aspectos do conceito de variabilidade, atividades explorando tais aspectos foram adaptadas em questões que investigaram: explanação da variabilidade; localização do ponto máximo/moda; predição a partir da moda, a partir do ponto máximo e a partir da tendência de um gráfico; representação da variabilidade, de frequência nula e representação de

acrécimos; conservação de quantidade total de um gráfico; comparação entre pontos com quantificação da variação; comparação entre conjuntos de dados; identificação de ausência de variabilidade e proposição de ausência de variabilidade. Esses aspectos foram contemplados em 17 (dezesete) questões.

Para analisar as compreensões apresentadas pelos sujeitos, os tipos de respostas oferecidos pelos estudantes do 2º e 5º ano para cada questão foram comparados. Nesse processo elegeu-se o(s) tipo(s) mais adequado(s), com base em estudos anteriores. Além de tal análise, realizou-se também o escalonamento multidimensional (Multidimensional Scaling - MDS), com o algoritmo ALSCAL (Algorithmic Scaling), para verificar o comportamento dos estudantes nos anos investigados, em função das questões, as quais exploraram diferentes aspectos da variabilidade. Essas análises serão discutidas a seguir.

Da comparação entre estudantes do 2º e 5º ano a respeito da adequação da resposta, tem-se 16 (dezesesseis) questões em que foi possível verificar o desempenho dos sujeitos. Os estudantes do 5º ano apresentaram percentuais de adequação na resposta significativamente maior do que aqueles dos estudantes do 2º ano, em pelo menos metade das questões. Isso ocorreu nas questões que abordaram os seguintes aspectos: explanação da variabilidade de categorias; localização de ponto máximo; comparação entre pontos com quantificação da variação; representação da variabilidade; predição a partir do ponto máximo; explanação da variabilidade numa série temporal; predição a partir da tendência do gráfico e identificação de ausência de variabilidade. Desses aspectos, apenas naquele que requeria comparação entre dos pontos a maioria dos estudantes do 5º ano apresentou respostas inadequadas. Apesar disso, pode-se inferir que as experiências vivenciadas pelos sujeitos, inclusive com a escolarização oportunizaram os avanços que diferenciaram de modo significativo o desempenho de ambos os grupos de sujeitos.

Outros aspectos foram facilmente compreendidos pelos dois grupos de estudantes, uma vez que a maioria apresentou respostas adequadas aos mesmos. Esses foram: explanação da variabilidade quando solicitada aos estudantes após tarefa de representação dos dados; representação de frequência nula e representação de acréscimos.

Entretanto, houve aqueles aspectos que representaram grandes dificuldades para ambos os grupos de estudantes. Em tais aspectos a diferença no desempenho dos grupos não foi expressiva, os mesmos foram: predição a partir da moda; comparação entre conjuntos e proposição de ausência de variabilidade. Salienta-se que na comparação entre conjuntos de dados quantitativos não houve resposta adequada em nenhum dos grupos de estudantes.

Ao se buscar uma gradação dos aspectos da variabilidade que suscitaram grandes dificuldades para esses estudantes, pode-se destacar ao menos 4 (quatro), que numa ordem decrescente de complexidade foram:

- Comparação entre conjuntos de dados;
- Predição a partir da moda;
- Comparação entre pontos com quantificação da variação;
- Proposição de ausência de variabilidade.

A respeito da comparação entre conjuntos, Loosen, Lioen e Lacante (1985) constataram grandes dificuldades de estudantes de graduação com esse tipo de comparação com dados quantitativos. Desse modo não se podia esperar melhores resultados com crianças. Em se tratando de dados qualitativos, Kader e Perry (2007) argumentaram que a comparação dos conjuntos seria mais intuitiva. Entretanto, verificou-se no presente estudo grandes dificuldades dos estudantes também com a comparação nesse tipo de dado. Assim, a complexidade parece estar mais vinculada com a habilidade envolvida, do que com o tipo de dado representado.

A localização do ponto máximo e moda foi uma tarefa melhor compreendida, porém apenas pelos estudantes do 5º ano. Guimarães (2002), com estudantes de faixa etária aproximada (nove anos de idade), assim como Watson e Kelly (2002a), chegaram a resultados parecidos. Entretanto, apesar dessa facilidade dos estudantes do 5º ano em localizar ponto máximo e moda, os mesmos apresentaram dificuldades em realizar uma predição a partir da moda. A predição a partir da moda também foi uma tarefa difícil para os sujeitos da pesquisa de Watson e Kelly (2002a) na qual os estudantes justificavam suas respostas baseados em vivências e imaginação, o que demonstra pouco comprometimento com os dados representados.

Acredita-se que o bom desempenho dos estudantes na questão de localização de ponto máximo/moda e o baixo desempenho na de predição a partir da moda provavelmente deve-se a ausência de discussão na escola em relação à função dos dados identificados nas representações gráficas. Queiroz (2007) salienta que o trabalho com Estatística proposto nos anos iniciais do Ensino Fundamental está muito mais direcionado a apropriação da representação em si do que a sua função. Assim, os estudantes estão aprendendo a identificar o ponto máximo/moda, mas não sabem a utilidade da mesma para predizer situações futuras.

Apropriar-se dos conhecimentos estatísticos para fins inferenciais e não apenas descritivos é fundamental. Afinal, saber interpretar um conjunto de dados é importante, contudo, utilizar tais interpretações para tomadas de decisão é o interesse maior ao se fazer estatística, pois argumentos estatísticos são oferecidos frequentemente no cotidiano e saber avaliar-los devidamente é imprescindível. Entretanto, só é possível avaliar argumentos baseados em inferências estatísticas, mesmo que de maneira informal, com situações de ensino apropriadas, como por exemplo, a exploração de predições de resultados.

Quantificar a variação entre dois pontos foi uma atividade difícil, principalmente para estudantes do 2º ano, grupo no qual não houve resposta adequada. Isso confirma os resultados de Guimarães (2002) com crianças na faixa etária aproximada e contrariam os de Watson e Kelly (2002a), que obtiveram resultados positivos para esse aspecto da variabilidade. A diferença nos resultados evidencia o papel da intervenção. No estudo de Watson e Kelly (2002a), as autoras durante a entrevista realizaram pequenas intervenções em função das respostas dos participantes. Da mesma forma, Guimarães (2002) e Arcoverde et al (2004) também argumentaram que curtas intervenções possibilitaram avanços importantes na compreensão da quantificação da variação por estudantes dos anos iniciais.

É importante esclarecer que tais aspectos mostraram-se difíceis, provavelmente, porque os estudantes não estão vivenciando na escola situações de ensino que possibilitem aos mesmos compreenderem diferentes aspectos da variabilidade. Isso porque com curtas intervenções estudantes foram capazes de quantificar a variação na comparação entre dois pontos, como ressaltaram Guimarães (2002) e Arcoverde et al (2004), o que pode ser estendido aos demais

aspectos, uma vez que estudantes desde o 2º ano de escolaridade são capazes de compreender aspectos da variabilidade, o que pode ser potencializado se os mesmos vivenciarem na escola situações de ensino que os desafiem a analisar e refletir sobre dados tratados estatisticamente.

Quanto à solicitação de que o estudante modificasse os dados para que não houvesse variação entre dois pontos (proposição de ausência de variabilidade) não foram encontrados na literatura um parâmetro para esse aspecto da variabilidade. Porém, ao se relacionar apenas as questões que requeriam identificação de ausência de variabilidade e proposição da mesma constatou-se que poucos sujeitos ofereceram respostas adequadas a uma e outra não, isso revela correlação entre aspectos.

A análise multidimensional MDS e algoritmo ALSCAL foi realizada com as questões explorando os aspectos já descritos. Essa análise apresenta índices de ajustes para as proximidades das questões representadas em um modelo bidimensional. As questões foram representadas por ano de escolaridade.

Primeiramente, afirma-se que o comportamento das questões entre os dois grupos de estudantes mudou bastante. Das duas dimensões, para ambos os grupos na dimensão 1 distinguiu-se na configuração criada de um lado as questões em que os estudantes apresentaram melhor desempenho e do outro lado questões com desempenho ruim.

Já a dimensão 2 ressaltou no 2º ano que o diferenciou a compreensão desses estudantes foi principalmente o envolvimento ou não dos mesmos com os dados representados. Quando os estudantes não se envolveram com os dados, seja por falta de habilidade com análise dos dados, seja porque a questão não exigiu uma exploração maior dos dados, os estudantes justificaram com imaginação ou vivências. Quanto aos estudantes do 5º ano, verificou-se no eixo da dimensão 2 que as questões estavam mais dispersas, evidenciando que o fato de um estudante responder adequadamente a uma questão não significou necessariamente a resolução adequada de uma outra. Com isso, percebeu-se que a distinção entre os aspectos da variabilidade foi ressaltada entre os estudantes do 5º ano.

Assim, para os estudantes do 5º ano foi possível encontrar 4 (quatro) grupos principais: o primeiro grupo localizou-se na parte inferior da configuração e englobou

aspectos bastante complexos para os estudantes; o Grupo 2 (também na parte inferior da dimensão 2) mostrou correlação entre os aspectos envolvendo representação da variabilidade e da frequência nula; o Grupo 3 englobou as questões de explanação da variabilidade; quanto ao Grupo 4, localizou-se na parte superior da configuração e diferente dos demais grupos caracterizou-se pela necessidade de conhecimento do ponto máximo para solução da questão, seja para a habilidade de identificação, seja para o tipo de predição solicitado. O aspecto de identificação de ausência de variabilidade está próximo ao Grupo 4 e justifica-se pelo caráter mais pontual do mesmo, uma vez que se compara dois pontos.

Percebe-se que na análise do comportamento dos estudantes do 5º ano nas questões investigadas destacou-se as visões locais e globais que os estudantes podem estar realizando dos dados apresentados aos mesmos. É importante esclarecer que dentre as questões há aquelas cujas formulações requerem uma visão mais pontual (por exemplo, localização de ponto máximo, comparação entre dois pontos) e outras que requerem visão mais global dos dados (como explanação da variabilidade, predição a partir da tendência). Contudo, concorda-se com a proposta de Ben-Zvi (2002), o qual afirma que ambas as visões estão integradas, sendo possível encontrar níveis de entendimento dentro das duas visões.

Assim, verificou-se que mesmo a questão direcionando a leitura do estudante para um tipo de visão, as visões por estarem integradas e possuírem níveis de entendimento provocaram aproximações entre aspectos da variabilidade que solicitavam visões diferentes, como foi o caso do aspecto em que se requeria a comparação entre pontos com quantificação da variação, de visão mais local, porém está no mesmo grupo (Grupo 1) das questões envolvendo o aspecto da comparação entre conjuntos, de visão mais global. Confirma-se que as visões que podem ser realizadas na análise de dados (pontual ou global) estão integradas dinamicamente e flexivelmente, tal como propôs Ben-Zvi (2002).

Este estudo visou investigar de modo sistematizado as compreensões de estudantes dos anos iniciais do Ensino Fundamental, a respeito da variabilidade. A partir dele pode-se afirmar que uma de suas contribuições foi integrar em um mesmo estudo diferentes aspectos relacionados à variabilidade que podem ser explorados junto a estudantes dos anos iniciais. Ressalta-se que esses aspectos podem ser

trabalhados com estudantes dos anos iniciais uma vez que os resultados encontrados mostram que sujeitos do 2º ano conseguiram oferecer respostas adequadas para todos os aspectos investigados, exceto na quantificação da variação entre dois pontos. Isso, sem que tenham vivenciado situações formais de ensino a respeito dos mesmos na escola.

A respeito da comparação entre conjuntos, pode-se afirmar a partir do presente estudo que a comparação entre dados qualitativos não é simples e intuitiva como propuseram Kader e Perry (2007), o que requer muito trabalho dos professores e de estudos futuros que nos auxiliem a um ensino mais eficiente.

Outra contribuição desse estudo é a constatação de que os estudantes são capazes de avançar em seus conhecimentos, mesmo sem instruções formais a respeito de alguns aspectos da variabilidade. Um exemplo é a predição a partir do ponto máximo, tarefa que não é explorada com estudantes dos anos iniciais, contudo, no 5º ano houve um avanço significativo na compreensão da questão. Se eles avançaram sem instrução formal, com essa instrução todos poderão ter condições de compreender diferentes aspectos da variabilidade.

Os resultados encontrados permitiram ainda a delimitação de pelo menos quatro aspectos bastante complexos para os estudantes: comparação entre conjuntos de dados; predição a partir da moda; a comparação entre pontos com quantificação da variação; além da proposição de ausência de variabilidade. Entretanto, cabe salientar que tais aspectos provavelmente podem se tornar compreensíveis com curtas intervenções didáticas, o que é uma investigação importante a ser realizadas em estudos futuros.

Os diversos aspectos da variabilidade investigados e as diferentes compreensões encontradas para os mesmos suscitam a reflexão de que a variabilidade estatística enquanto conceito precisa ser melhor compreendida considerando o tripé da definição conceitual de Vergnaud (1996).

Acredita-se que quando discutiu-se aspectos da variabilidade, estavam sendo investigados diferentes situações em que esse conceito pode estar sendo explorado. Assim, como o desenvolvimento conceitual requer o reconhecimento do conjunto de situações, conjunto de representações e conjunto de invariantes que formam o conceito (VERGNAUD, 1996), não poderia ser diferente com a variabilidade. Afinal,

a resolução de problemas envolvendo o conceito de variabilidade estatística requer a compreensão de uma diversidade de situações em que a mesma pode ser encontrada, assim como o conhecimento de formas de representá-la e invariantes (propriedades) que a caracteriza.

Conhecer com maior profundidade a variabilidade é essencial, pois a mesma compõe o pensamento estatístico. Desse modo, compreender sua formação enquanto conceito, assim como acontece a ligação da mesma com outros conceitos relacionados, contribuirá não apenas para os avanços na pesquisa e no ensino da variabilidade, como também para o desenvolvimento da própria Estatística.

A partir do conhecimento da importância do conceito aqui explorado e da necessidade do mesmo ser trabalhado desde os anos iniciais do Ensino Fundamental, é possível apontar algumas implicações desse estudo para o ensino da variabilidade. Afinal, ao se reconhecer nos estudantes potencialidades para avanços no conhecimento, é imprescindível repensar as práticas de ensino para que a mesma possibilite evolução na aprendizagem sobre variabilidade.

Primeiramente, precisa-se repensar a prática de explorar, quando se trabalha com análise de conjuntos de dados, unicamente a localização de pontos extremos. A exploração maior desse aspecto da variabilidade como tem sido realizada, não garante, como mostraram os resultados, a reflexão dos dados do conjunto a partir do ponto máximo e moda, pois o aspecto da predição de resultados que requer a localização de um ponto extremo, mostrou-se bastante complexo.

Com os resultados aqui encontrados, cada vez mais se acredita que uma prática educativa eficaz, no que se refere ao desenvolvimento do conceito de variabilidade, precisa abordar os diferentes aspectos que compõem a variabilidade, evidenciando as propriedades do conceito, as formas diferenciadas que podem estar sendo utilizadas para representá-los e as situações propostas.

## REFERÊNCIAS

ARCOVERDE, D.; BARBOSA, R. H.; GUIMARÃES, G. **Uma proposta de seqüências didáticas sobre interpretação de gráficos em turmas de 3ª série.** Trabalho de Conclusão de Curso (Graduação em Pedagogia), UFPE, Recife, 2004.

BATANERO, C.; ESTEPA, A.; GODINO, J. D. Análisis Exploratorio de Datos: SUS posibilidades em la enseñanza secundaria. **Suma**, n. 9, p. 25-31, 1991.

Disponível em:

[http://170.210.81.106/faea/academica/departamentos/estadistica/catedras/Matematica/unidad\\_2\\_analisis\\_exploratorio\\_SUMA\\_91.pdf](http://170.210.81.106/faea/academica/departamentos/estadistica/catedras/Matematica/unidad_2_analisis_exploratorio_SUMA_91.pdf) Acesso: 16 de dez. de 2009.

BEN-ZVI, Dani. Seventh grade students' sense making of data and data representations. ICOTS 6. **Proceedings of the Sixth International Conference on Teachings Statistics**, 2002. Disponível em: <http://www.stat.auckland.ac.nz/> Acesso em: 15 de out. de 2008.

BRASIL, Ministério da Educação e do desporto. Secretaria de Educação Fundamental. **Parâmetros Curriculares Nacional: Matemática (1ª a 4ª)** Brasília, DF, 1997.

CARRAHER, Terezinha N. **O método clínico usando os exames de Piaget.** 5ª ed. São Paulo: Cortês, 1998.

COOPER, L.; SHORE, F. Students' misconceptions in interpreting center and variability of data represented via histograms and stem-and-leaf plots. **Journal of Statistics Education**, v. 16, n. 2, 2008. Disponível em: <http://www.amstat.org/publications/jse/v16n2/cooperpdf.html> Acesso em: 06 de maio de 2009.

GARFIELD, J.; BEN-ZVI, D. A framework for teaching and assessing reasoning about variability. **Statistics Education Research Journal**, v. 4, n. 1, p. 92-99, 2005. Disponível em: <http://www.stat.auckland.ac.nz/serj> Acesso em: 15 de out. de 2008.

GUIMARÃES, Gilda L. **Interpretando e construindo gráficos de barra.** Tese (Doutorado em Psicologia Cognitiva), UFPE, Recife, 2002.

GUIMARÃES, G.; CAVALCANTI, M. R. Gráficos apresentados pela mídia impressa. Anais do **II Simpósio Internacional de Pesquisa em Educação Matemática**, Recife, 2008.

GUIMARÃES, G.; GITIRANA, V.; MARQUES, M.; CAVALCANTI, M. Educação Estatística nos anos iniciais de escolarização. **Seminário Internacional de Pesquisa em Educação Matemática**, Brasília, 2009. Disponível em CD-ROM.

GUIMARÃES, G.; GITIRANA, V.; CAVALCANTI, M.; MARQUES, M. Atividades que exploram gráficos e tabelas em livros didáticos de matemática nas séries iniciais.

**Anais do III Seminário Internacional de Pesquisa em Educação Matemática – SIPEM**, Águas de Lindóia, outubro de 2006.

KADER, G.; PERRY, M. Variability for categorical variable. **Journal of Statistics Education**, v. 15, n. 2, 2007. Disponível em: <http://www.amstat.org/publications/jse/> Acesso em: 06 de maio de 2009.

LOOSEN, F.; LIOEN, M.; LACANTE, M. The standard deviation: some drawbacks to an intuitive approach. **Teaching Statistics**, v. 7, n. 1, p. 2-5, 1985.

MAGINA, S.; CAMPOS, T. M. M.; NUNES, T.; GITIRANA, V. **Repensando adição e subtração**: Contribuições da Teoria dos Campos Conceituais. 2ª ed. São Paulo: PROEM, 2001.

OLIVER, Maria del Carmen. **La atención a la diversidad desde los agrupamientos flexibles de alumnos**. 1998. 669 f. Tese (Doutorado em La Innovació Educativa i la Formació del Professorat). Facultat de Ciències de L'Educació. Universitat de Barcelona. Barcelona, 1998. (Capítol 7). Disponível em: <http://ldei.ugr.es/cddi/uploads/tesis/OliverVera1998.pdf> Acesso em: 22 de jan. de 2011.

PERRY, M.; KADER, G. Variation as unalikeability. **Teaching Statistics**, v. 27, n. 2, p. 58-60, 2005. Disponível em: <http://eric.ed.gov/ERICWebPortal/> Acesso em: 03 de jul. de 2009.

PONTE, J. P.; BROCARD, J.; OLIVEIRA, H. **Investigações Matemáticas em sala de aula**. Belo Horizonte: Autêntica, 2005.

PORCAR, M. L.; ESCALANTE, E. E. Límites de la explicación interna (norma de la internalidad): análisis de escalamiento multidimensional. **Revista Electrónica de Investigación y Docencia (REID)**, n. 2, Junho, p.59-77, 2009. Disponível em: <http://www.revistareid.net/revista/n2/REID2art4.pdf> Acesso em: 22 de jan. de 2011.

QUEIROZ, Esmeralda. **O uso do livro didático de Matemática por professores do Ensino Fundamental**. 2007. Dissertação (Mestrado em Educação). Centro de Educação. Universidade Federal de Pernambuco. Recife/PE, 2007.

READING, Chris. Student description of variation while working with weather data. **Statistics Education Research Journal**, v. 3, n. 2, p. 84-105, 2004. Disponível em: <http://www.stat.auckland.ac.nz/serj> Acesso em: 15 de out. de 2008.

READING, C; SHAUGHNESSY, M. Reasoning about variation. In: BEN-ZVI, D.; GARFIELD, J. **The challenge of developing Statistical Literacy, Reasoning and Thinking**. Netherlands: Kluwer, 2004, p. 201-226.

REID, J.; READING, C. A hierarchy of tertiary students' consideration of variation. **Proceedings of the Seventh International Conference on Teachings Statistics**.

África do Sul, 2006. Disponível em <http://www.stat.auckland.ac.nz/> Acesso em 15 de out. de 2008.

SANTOS, M. S.; GITIRANA, V. A interpretação de gráficos de barra, com variáveis numéricas, em um ambiente computacional de manipulação de dados. **Anais do XIV Encontro de Pesquisa Educacional do Nordeste (EPEN)**, Salvador, 1999. Disponível em CD-ROM.

SHAUGHNESSY, M.; WATSON, J.; MORITZ, J.; READING, C. School mathematics students' acknowledgements of statistical variation. Artigo apresentado na pré-sessão do simpósio There's More to Life than Centers, do **77º Annual National Council of Teachers of Mathematics Conference**, São Francisco, CA, 1999.

SILVA, Cláudia Borim da. **Pensamento Estatístico e Raciocínio sobre variação**: um estudo com professores de Matemática. Tese (Doutorado em educação Matemática), PUC-SP, São Paulo, 2007.

VERGNAUD, Gerard. A Teoria dos Campos Conceituais. In: BRUM, Jean (org.). **Didáticas das Matemáticas**. Tradução de Maria José Figueiredo. Lisboa: Instituto Piaget, 1996. Cap. 3, p. 155-191. (Coleção Horizontes Pedagógicos).

WATSON, J.; KELLY, B. A. Emerging concepts in chance e data. **Australian Journal of Early Childhood**, 27(4), 2002a, p. 24-28. Disponível em: <http://www.highbeam.com/doc/1G1-95915291.html> Acesso em: 09 de jun. de 2009.

WATSON, J.; KELLY, B. A. Can grade 3 students learn about variation? **Proceedings of the Sixth International Conference on Teachings Statistics**. África do Sul, 2002b. Disponível em: <http://www.stat.auckland.ac.nz/> Acesso em: 15 de out. de 2008.

WATSON, J.; KELLY, B. A. The winds are variable: student intuitions about variation. **School Science and Mathematics**, v. 105, n.5, p. 252-269, 2005. Disponível em: <http://www.eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/> Acesso em: 06 de set. de 2009.

WATSON, Jane. The influence of variation and expectation on the developing awareness of distribution. **Statistics Education Research Journal**, 8(1), pp 32-61, 2009. Disponível em: <http://www.stat.auckland.ac.nz/~iase/serj/> Acesso em: 15 de jun. de 2009.